



Thermal biology of *Tuta absoluta*: demographic parameters and facultative diapause

Mateus Ribeiro de Campos¹ · Philippe Béarez¹ · Edwige Amiens-Desneux¹ · Luigi Ponti^{2,3} · Andrew Paul Gutierrez^{3,4} · Antonio Biondi⁵ · Abhijin Adiga⁶ · Nicolas Desneux¹

Received: 9 June 2020 / Revised: 4 September 2020 / Accepted: 21 September 2020 / Published online: 6 October 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

The South American tomato pinworm, *Tuta absoluta*, (SATP) is now a devastating pest worldwide of crops in the family Solanaceae. Most prior studies of SATP's thermal biology were based on populations from tropical regions, and proved unsuitable for explaining its invasion of large areas of the Palearctic. A more holistic approach to the analysis of its thermal biology is essential background for developing models to assess its invasive potential. Our studies found that SATP has lower and upper thermal thresholds ($\theta_L = 5.37^\circ\text{C}$ and $\theta_U = 35.69^\circ\text{C}$, respectively) than South American populations used in prior studies ($\theta_L = 7.38^\circ\text{C}$ and $\theta_U = 33.82^\circ\text{C}$). Age-specific life tables were used to estimate the effects of temperature on its demographic parameters. Diapause in SATP had not been characterized prior to our study. We found facultative diapause in pupae developing from larvae exposed to relatively low temperatures (i.e., 2 and 5 °C) and short-day length for different exposure periods. The strength of diapause was measured as an increase in post-treatment developmental times of pupae (i.e., degree days) that on average were 2.45–3-fold greater than of pupae reared at favorable temperatures. A lower developmental threshold and a facultative diapause increase the invasive potential of SATP in temperate areas. Knowledge of this thermal biology is essential for predicting the potential geographic spread of this pest and to develop management and control strategies.

Keywords Invasive pest · Thermal biology · Temperature thresholds · Diapause · Phenotypic plasticity · Geographic variation

Abbreviations

SATP	South American tomato pinworm
θ_L	Lower temperature threshold
θ_U	Upper temperature threshold
R_0	Net reproductive rate
λ	Finite rate of increase
r_m	Intrinsic rate of increase
τ	Generation time
K	Degree-day
T	Temperature
dl	Minutes day length
E	Exposure periods

Key messages

- Demographic parameters estimated at different temperatures show that *T. absoluta* has a rapid developmental time and can reproduce at temperatures close to the development threshold.
- T. absoluta* from a temperate region (France) has a lower temperature threshold for development than strains from tropical regions in South America, but this awaits independent confirmation as it has important ecological implications for its invasive capacities.
- Facultative diapause in *T. absoluta* was confirmed experimentally in response to low non freezing temperatures, day length and exposure periods.

Communicated by M. Traugott.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10340-020-01286-8>) contains supplementary material, which is available to authorized users.

✉ Mateus Ribeiro de Campos
mateusrcampos@gmail.com

Extended author information available on the last page of the article

Introduction

Exotic invasive species threaten native species and ecosystems globally (Asplen et al. 2015; Lee et al. 2020; Requier et al. 2019; Roques et al. 2016). Among the most important is the highly invasive South American tomato pinworm, *Tuta absoluta* Meyrick (Lepidoptera: Gelechiidae) (SATP), a destructive invasive pest of tomato and other crops in the family Solanaceae worldwide (Biondi et al. 2018; Desneux et al. 2010; Guedes and Picanço 2012). From its native range in the Peruvian central highlands in South America, it reached eastern Spain in 2006, and spread rapidly to Central Europe and Mediterranean countries of Europe and Africa (Mansour et al. 2018; Sylla et al. 2017), and is expanding its range into Asia (Han et al. 2018, 2019). Its current range includes most Central and Southwest Asian countries neighboring China (Biondi et al. 2018; Desneux et al. 2011; McNitt et al. 2019), and threatens to invade important tomato producer-exporter countries such as China and United State (Campos et al. 2017; Han et al. 2019; Verheggen and Fontus 2019).

Sound data on the thermal biology of invasive species such as SATP are essential to developing models to predict the prospective geographic distribution and relative abundance of invasive species under extant and climate change weather, and for the development of integrated pest management (IPM) strategies (Gutierrez 1996; Kang et al. 2009). Temperature is a key factor in the survival, reproduction and abundance of poikilotherms species (Briere et al. 1999; Flitters and Messenger 1965; Gilbert and Gutierrez 1973; Lactin et al. 1995; Logan et al. 1976), but the effects are often not considered (Cherif et al. 2019; Iltis et al. 2020; Martins et al. 2016; Rank et al. 2020).

In this paper, we take a holistic approach to estimate the effects of temperature on the demographic parameters of SATP using age-specific life tables as a basis (Barrientos et al. 1998; Cherif et al. 2019; Krechmer and Foerster 2015; Martins et al. 2016). Life table parameters provide information about the limiting effects of temperature on the vital rates of species (Desneux et al. 2010; Ponti et al. 2015a, b). Ecological niche models based on occurrence data and aggregate weather and biological factors in the species' native range have been used to estimate their potential geographic range (Lawler et al. 2006; Tonnang et al. 2015). Mechanistic physiologically based population dynamics models based on non-linear biodynamic functions of the ecological and biological processes affecting the population dynamics of insects, including diapause (Gutierrez 1996; Gutierrez et al. 2015), can be used to predict and explain the geographic distribution and the relative abundance of species.

Our paper seeks to fill the gaps in the thermal biology of SATP across its wider range of temperatures (Ponti et al. 2015a) and to investigate the occurrence of diapause. Changes

in climatic factors may promote adaptive behavioral and physiological changes in insects such as diapause (Leather et al. 1993) that may be critical components in their life history strategy, enabling survival in unfavorable environmental conditions (Danks 2007; Musolin et al. 2019; Tauber et al. 1986). Diapause may be quiescence, facultative or obligate, and may be induced by environmental stimuli such as temperature, photoperiod and other factors (Denlinger 2002, 2008). Diapause in SATP had not been previously documented (Kahrer et al. 2019), though indications of diapause were preliminarily reported by Sannino and Espinosa (2010) in the last summer generation near Salerno, Italy, where 13 of 33 pupae became adults in 17.9 ± 0.41 days, while 20 pupae required 65.1 ± 1.16 days (see Ponti et al. 2015b).

Materials and methods

Growth chamber and greenhouse production of host plants

Pesticide-free tomato plants cv. Marmande grown from seed germinated in a climatic chamber at 24 ± 2 °C (T), $65 \pm 5\%$ relative humidity (RH), and 16:08 h (Light:Dark, L:D) photoperiod were used as hosts for SATP. After germination, the tomato seedlings were planted in 320-mL plastic pots in a commercial substrate (Tournesol®, Nice, France), and transferred to a greenhouse kept at 20 ± 2 °C, $65 \pm 5\%$ RH, and photoperiod of 10:14 h L:D. The mean daytime global irradiance inside the greenhouse was maintained at 449.5 W m^{-2} using shading as required. The plants were later replanted in 2.8-L pot using inert perlite as a substrate (Perlite Italiana srl, Corsico, Italy), and fertilized with a nutrient mineral solution developed by the French National Institute for Agricultural Research & Environment (INRAe) Sophia Antipolis, France ($\text{NO}_3^- = 1 \text{ mol m}^{-3}$; $\text{H}_2\text{PO}_4^- = 0.21 \text{ mol m}^{-3}$; $\text{SO}_4^{2-} = 0.055 \text{ mol m}^{-3}$; $\text{K}^+ = 0.641 \text{ mol m}^{-3}$, $\text{Ca}^{2+} = 0.215 \text{ mol m}^{-3}$; $\text{Mg}^{2+} = 0.114 \text{ mol m}^{-3}$ adjusted to pH 6.0 ± 0.2 using 0.2 M aqueous solution of H_2SO_4) (Han et al. 2014).

Tuta absoluta colonies

Colonies of SATP were established from field adults collected in July 2009 on the INRAe campus, Alénya, France. The colonies were maintained in eight cages ($55 \times 75 \times 80 \text{ cm}$) as mixed stages (egg, larval, pupal, and adult) on leaves from potted greenhouse tomato plants. The cages were housed in controlled environmental growth chambers maintained at 24 ± 2 °C, $65 \pm 5\%$ RH, and 16:08 h L:D. Large numbers of equal age individuals required for the experiments were obtained by providing adult females with preferred young leaves (Anastasaki et al. 2018) in a cage

for 48 h, and subsets of leaves with eggs were transferred to smaller cages for hatching. Newly hatched larvae were supplied daily with new leaves.

Laboratory cages for estimating vital rates

Small plastic cages each composed of two plastic cups (Biondi et al. 2012) were used in all laboratory experiments to assess the temperature-dependent vital rates, and to assess the effects of temperature (T , °C) \times day length (dl , minutes) \times exposure periods (E , days) on diapause induction. The top cup was 15 cm high with a volume of 700 mL with a one-cm diameter hole drilled in the bottom to allow the insertion of a leaf stem allowing access to water in the second 11 cm high, 350 mL cup. Leaves ~50 days of age (15 cm long with seven leaflets) collected from the mid- and upper levels of greenhouse plants were used as hosts for the larvae (Torres et al. 2001). An organdy cloth screen was affixed to the top of the first plastic cup with a rubber band to allow ventilation, and access to determine the development of the experimental insects.

Thermal biology experiments

Forty replicates at each of eight constant temperatures (6, 10, 15, 20, 25, 30, 33 and 36 °C \pm 1, 60 \pm 5% RH, and 16:08 h L:D) were used to estimate developmental times and thermal thresholds of each immature life stage. Each replicate consisted of three first-instar larvae per cage. All experiments were conducted in incubators (model TC445 S, Aqualytic® LIEBHERR, Tintometer GmbH, Austria) using a TFH 620 ebro® Electronic GmbH & Co. KG, Germany sensor to monitor temperatures. The time interval between observations decreased with increasing temperature. At 6, 10, 15, and 20 °C, observations were made every two days, while at 25, 30, 33, and 36 °C observations occurred daily. The temperatures used mirrored the range in previous studies (Barrientos et al. 1998; Krechmer and Foerster 2015; Martins et al. 2016; Ponti et al. 2015a) (Supplementary material, Table S1).

Lifespan and reproduction under winter conditions

On February 21, 2018, large numbers of eggs were collected on new leaves in the laboratory as described above and then transferred to field conditions in 55 \times 75 \times 80 cm screen cages where they remained until adult emergence. Field temperatures recorded by the meteorological station at INRAe were used to assess the effects of temperature on field survival. After emergence in spring, the adults were taken to the laboratory where 40 male–female pair replicates in small cages were maintained at 24 \pm 2 °C, 65 \pm 5%

RH and 16:08 h L:D. Daily oviposition and mortality of the females were recorded.

Life table analysis

The life table data were analyzed using the age-stage, two-sex life table methods (Chi 1988; Chi and Liu 1985). The net reproductive rate (R_0), mean generation time (τ), intrinsic rate of increase (r_m), and finite rate of increase (λ) were calculated using the TWOSEX-MSChart procedure (Chi 2019, Chi et al. 2020). The mean, variance, and standard error of the life table parameters were estimated using 100,000 bootstrap samples (Efron and Tibshirani 1986) for comparison across temperature treatments. Basic age-specific life table data consist of the average age of the insects from egg eclosion (x); the number alive at the beginning of each age (a_x); the proportion surviving to each life stage x (l_x) obtained by dividing the number of individuals living at the beginning of each age (a_x) by the initial number of initial eggs; the number of females produced per female at age x (m_x); the total number of females produced at age x ($l_x m_x$). Generation time (τ) was estimated as $\sum x l_x m_x / l_x m_x$. The intrinsic rate of natural increase (r_m) was estimated by solving the equation $\sum_{x=0}^{\infty} e^{-r_m(x+1)} l_x m_x = 1$ (Goodman 1982). The sex ratio (1:1) was estimated at the pupal stage (Bacci et al. 2019).

Effects of temperature on demographic functions

Rate of development

The developmental rate for each life stage at temperature T is the reciprocal of developmental time in days ($1/d(T = \text{°C})$). The developmental rate data were plotted on temperature and a linear regression fit (Eq. 1).

$$\frac{1}{d} = a + bT \quad (1)$$

The lower thermal threshold for development (θ_L) was computed at $1/d = 0$ as $-\frac{a}{b}$. Using θ_L , the thermal constant (K , degree days (dd)) was computed as $K = d(T - \theta_L)$ in the linear range of favorable temperatures (Campbell et al. 1974).

Effects of temperature on R_0 , r_m and fecundity

Our experimental data for R_0 , r_m and total fecundity and similar data from the literature were plotted on T . Non-linear concave polynomial functions were fit to both data sets. These functions provide estimates of the minimum, maximum and optimum temperatures for R_0 , r_m and total fecundity.

Diapause experiments

Preliminary studies on SATP larvae (Tarusikirwa et al. 2020) and pupae (Kahrer et al. 2019) demonstrated high survival at low temperatures, but the mechanism was not elucidated. In insects, cold hardening and diapause are common mechanisms to survive harsh periods and may be induced in any stage. In our study, we exposed 30 replicates of five 3rd and 4th instar larvae from the rearing colony maintained at 25 °C to one of 96 combinations of temperature × day length × exposure periods: T (2, 5, 10, and 15 ± 1 °C), dl (960, 900, 840, 600, 540, and 480 min) and exposure periods E (7, 14, 28, and 42 days). Test larvae were transitioned to the experimental temperatures using the scheme illustrated in Fig. 1. At the end of each exposure period, the surviving individuals (larvae and pupae) were transitioned to non-diapause laboratory conditions of 24 ± 2 °C, 65 ± 5% RH and 16:08 h L:D where they were observed at 2-day interval until adult emergence. The difference between observed developmental times in dd of treated pupae ($K(T,dl,E)$) to the average time under non-diapause conditions (i.e., the thermal constant K) was used as a metric of diapause.

Statistical analyses

IBM® SPSS® v21 software (2012) was used in all analyses. After confirming normality and homoscedasticity assumptions by running Kolmogorov–Smirnov test, one-way ANOVA was used to evaluate differences between the thermal requirement based on literature data and our experiments.

Results

Effects of temperature

The data on the effects of temperature on the rate of development in our experiments and those from the literature (Barrantes et al. 1998; Krechmer and Foerster 2015; Martins et al. 2016) were analyzed separately and in combination (Fig. S1 and Table 1). Fits to the combined data were made for heuristic purposes (see discussion). The results at 6 °C had high variability and mortality, and only data in the linear range of development (10–30 °C) were used in the linear regression analyses. A data point (egg stage; 0.4, 30 °C) from Krechmer and Foerster (2015) was deleted because it is an obvious outlier (Table S1).

The slopes (b) and intercepts (a) of the two sets of data for the egg stage were not statistically different ($F_{1,16}=2.126$, $p=0.195$ and $F_{1,16}=1.64$, $p=0.241$, respectively) (Fig. S1A). The lower thermal threshold for development ($\theta_{L,egg}$) from our data is 5.70 °C, and that for the literature data is 4.28 °C (Fig. S1A and Table 1). The combined data yielded $\theta_{L,egg}=4.94$ °C (Fig. S1B and Table 1).

For the larval data, the slopes and intercepts of the developmental rate data for the two data sets were not statistically different ($F_{1,19}=1.242$, $p=0.297$ and $F_{1,19}=6.907$, $p=0.427$, respectively) (Fig. S1C). $\theta_{L,larvae}$ from our data is 5.13 °C and 8.64 °C for the literature data (Fig. S1D and Table 1).

For the pupal data, the slopes and intercepts of the developmental rate for the two data sets were not different ($F_{1,19}=1.849$, $p=0.269$ and $F_{1,19}=1.387$, $p=0.211$, respectively) (Fig. S1E). $\theta_{L,pupae}$ was 5.31 °C for our data and 8.79 °C for the literature data (Fig. S1E and Table 1), while $\theta_{L,pupae}$ for the combined data was 7.59 °C (Fig. S1F and Table 1).

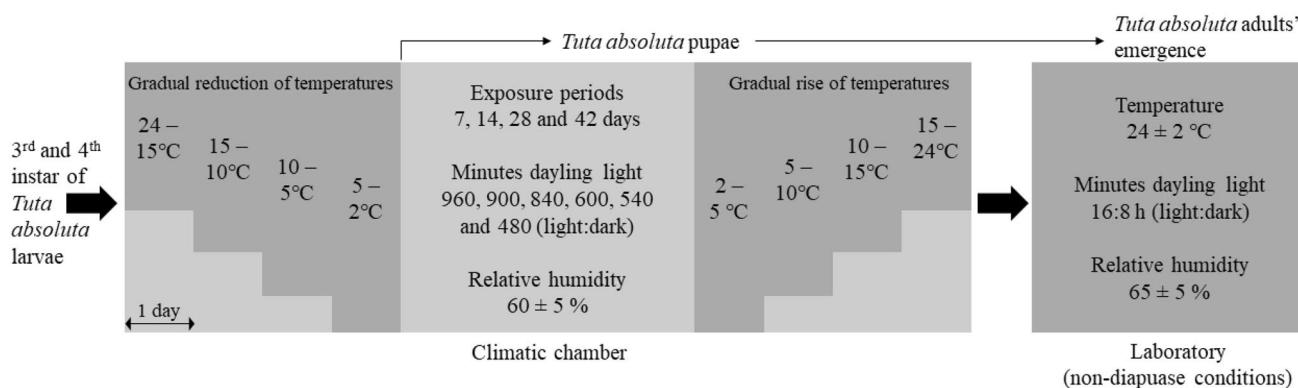


Fig. 1 Schematic diagram for the stepwise transition of temperatures in the diapause experiment for *Tuta absoluta*

The slopes and intercepts from egg eclosion to adult emergence (subscript E-A) were not statistically different ($F_{1,19}=1.084, p=0.328$ and $F_{1,19}=0.630, p=0.448$, respectively) (Fig. 2). $\theta_{L,E-A}$ from our data was $5.80\text{ }^{\circ}\text{C}$ ($R^2=0.992$), while the literature data yielded $8.28\text{ }^{\circ}\text{C}$ ($R^2=0.986$ [Fig. 2 and Table 1]). The combined data gave $\theta_{L,E-A}=7.38\text{ }^{\circ}\text{C}$ (Fig. 2 and Table 1).

The developmental times for all life stages on temperature are summarized in Table 2. No eggs hatched at $6\text{ }^{\circ}\text{C}$, and no development of any stage occurred at $36\text{ }^{\circ}\text{C}$. Developmental times in days for all life stages decreased with increasing temperature, but began to increase above $33\text{ }^{\circ}\text{C}$. Similarly, the longevity of adults in days decreased with increasing temperatures, with a sharp decline at $33\text{ }^{\circ}\text{C}$, with no survival at $36\text{ }^{\circ}\text{C}$ (Table S2).

The effects of temperature on life table parameters

The life table statistics are summarized in Table 3. For comparative purposes, the vital rates from the literature and our studies were also combined.

The combined data on net reproduction (R_0) data were fit with 3rd-order polynomial ($R^2=0.795, p=0.01$) (Fig. 3A), and our experimental data were fit with a 6th-order polynomial ($R^2=0.963$ and $p=0.003$; Fig. 3B). The combined data for r_m are right skewed and were fit with 4th-order polynomial yielding lower and upper temperature thresholds for r_m of $\theta_{L,rm}=6.65\text{ }^{\circ}\text{C}$ and $\theta_{U,rm}=36.16\text{ }^{\circ}\text{C}$, with a maximum at $\sim 22.5\text{ }^{\circ}\text{C}$ ($R^2=0.764$; Fig. 3C). A 3rd-order polynomial fit to our data yielded $\theta_{L,rm}$ and $\theta_{U,rm}$ estimates of $6.49\text{ }^{\circ}\text{C}$

and $33.82\text{ }^{\circ}\text{C}$, respectively, with a maximum at $\sim 25\text{ }^{\circ}\text{C}$ ($R^2=0.990$; Fig. 3D).

As expected, generation time (τ) decreased with increasing temperatures (Table 3), with the longest being 103.04 days at $10\text{ }^{\circ}\text{C}$.

Total fecundity per female using the combined data was fit with 6th-order polynomial ($R^2=0.812$ and $p=0.001$; Fig. 3E), yielding $\theta_{L,fec}$ and $\theta_{U,fec}$ values for oviposition of $6.10\text{ }^{\circ}\text{C}$ and $34.16\text{ }^{\circ}\text{C}$, respectively. Our data were fit with a 2nd-order concave polynomial ($R^2=0.916$ and $p=0.011$) with $\theta_{L,fec}=5.37\text{ }^{\circ}\text{C}$ and $\theta_{U,fec}=35.69\text{ }^{\circ}\text{C}$ (Fig. 3F).

Temperature affected the age-stage-specific survival (S_{xj}) or the probability that a newly hatched larva survives to age x and stage j . Figure S2A-F illustrates the transition between life stages showing the dispersion of developmental times of cohort members (Fig. S2A-F). The age-specific survival (I_x) curves at the six constant temperatures are illustrated in Fig. S3.

Reproductive lifespan

Data on the effects of age on daily oviposition at six constant temperatures are illustrated in Fig. 4A, B. Peak oviposition occurred on the second day with the oviposition rate declining to very low levels at 15 days. The age-specific oviposition surface as affected by temperature was fit by combining two functions (Eq. 2): an exponentially decreasing function for the effect of age at optimal $T=20.53\text{ }^{\circ}\text{C}$ and a normalized concave scalar function ($0 \leq \phi(T) \leq 1$) in favorable range ($T_{min}=5.37\text{ }^{\circ}\text{C}, T_{max}=35.69\text{ }^{\circ}\text{C}$) of the data (Fig. 3F):

Fig. 2 Linear regression of the development rate of *Tuta absoluta* egg to adult period on temperatures. Our experimental data (thin solid line), literature data (dash line) and the combined data (thick solid line) were fit with linear regressions

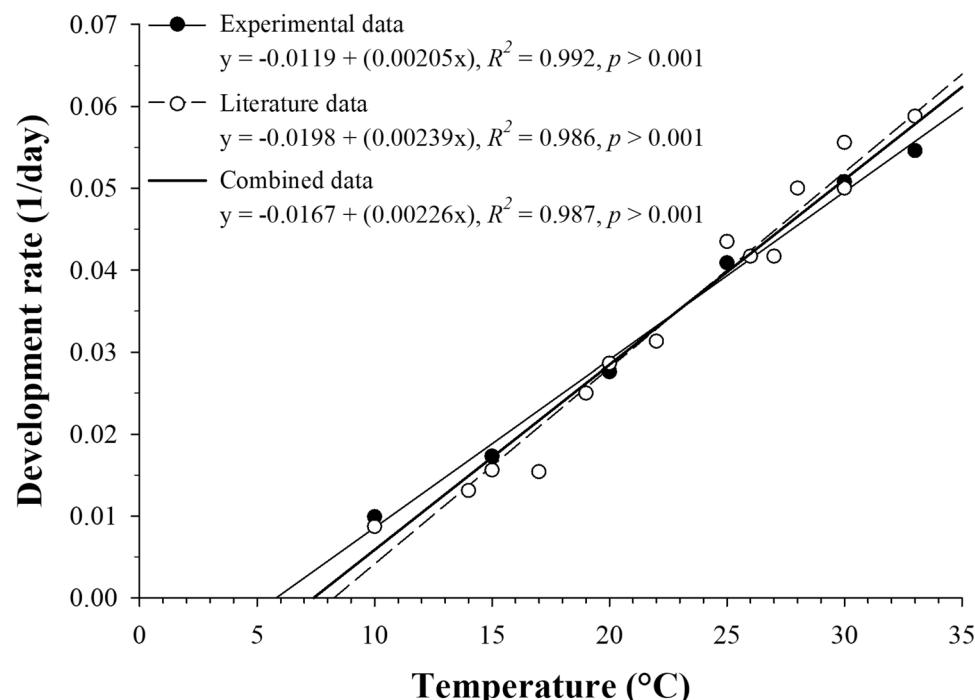


Table 1 The lower developmental threshold temperature (θ_L) and thermal constant in degree-days (K) estimated for the immature stages of *Tuta absoluta* based on our experiment data, literature data, and the combined data using the linear model

Model	Parameters	Linear regression	θ_L	K	R^2	<i>p</i> -model
		<i>a</i>	<i>b</i>			
Life stages of <i>Tuta absoluta</i>						
eggs	Experimental data	-0.0633	0.0111	5.7027	90.0900	0.9980
	Literature data	-0.0347	0.0081	4.2839	123.4567	0.8580
	Combined data	-0.0445	0.0090	4.9444	111.1111	0.8530
	<i>F</i> and <i>p</i> -value	$F_{1,16}=1.638,$ $p=0.2414$	$F_{1,16}=2.120,$ $p=0.1951$			
Larvae	Experimental data	-0.0190	0.0037	5.1351	270.2703	0.9780
	Literature data	-0.0441	0.0051	8.6470	196.0784	0.9270
	Combined data	-0.0340	0.0046	7.3913	217.3913	0.9640
	<i>F</i> and <i>p</i> -value	$F_{1,19}=6.9070,$ $p=0.4270$	$F_{1,19}=1.2420,$ $p=0.2970$			
Pupae	Experimental data	-0.0356	0.0067	5.3134	149.2537	0.9970
	Literature data	-0.0730	0.0083	8.7951	120.4819	0.9450
	Combined data	-0.0585	0.0077	7.5974	129.8701	0.9770
	<i>F</i> and <i>p</i> -value	$F_{1,19}=1.3870,$ $p=0.2110$	$F_{1,19}=1.8490,$ $p=0.2690$			
Eggs-adults	Experimental data	-0.0119	0.0020	5.8048	487.8048	0.9920
	Literature data	-0.0198	0.0024	8.2845	418.4100	0.9860
	Combined data	-0.0167	0.0022	7.3893	442.4778	0.9870
	<i>F</i> and <i>p</i> -value	$F_{1,19}=0.6300,$ $p=0.4480$	$F_{1,19}=1.0840$ $p=0.3280$			

Table 2 Development time (days) (mean \pm SE) of the different life stages of *Tuta absoluta* recorded at eight constant temperatures

Temperatures (°C)	Immature stages (days)			Egg to adults
	Eggs	Larvae	Pupae	
6	—	79.42 \pm 3.46 a	44.80 \pm 1.02 a	124.22 \pm 3.31 a*
10	22.77 \pm 0.62 a	48.30 \pm 0.68 b	30.44 \pm 0.93 b	119.56 \pm 2.33 b
15	9.32 \pm 0.27 b	32.16 \pm 0.65 c	16.46 \pm 0.68 c	75.18 \pm 0.93 c
20	6.27 \pm 0.20 c	19.42 \pm 0.52 d	10.78 \pm 0.49 d	48.87 \pm 1.47 d
25	4.80 \pm 0.11 d	12.30 \pm 0.46 d	7.37 \pm 0.13 d	24.47 \pm 1.10 d
30	3.70 \pm 0.15 d	12.46 \pm 0.13 d	5.55 \pm 0.12 e	21.71 \pm 1.21 de
33	3.73 \pm 0.08 d	10.10 \pm 0.32 d	5.71 \pm 0.22 e	21.43 \pm 1.48 e
36	—	—	—	—

*Egg hatch to adult emergence

Gutierrez et al. 2018). Equation 2 captures reproduction across the full range of age and temperature mapping the oviposition data in Fig. 3A.

$$\begin{aligned} \text{eggs/day/female} &= \phi(T)(ae^{(-bx)}) \\ &= (-78.66 + 16.84T - 0.41T^2)(54.753e^{(0.245 \text{ age})}) \\ \phi(T) &= 1 - \left[\frac{(T - T_{\min} - T_{mid})}{T_{mid}} \right]^2, \\ \text{with } T_{mid} &- \frac{(T_{\max} - T_{\min})}{2} = 20.53^\circ\text{C} \end{aligned} \quad (2)$$

Diapause

To determine if diapause occurs, we compared the post-treatment developmental times of individual subjected to assumed diapause inducing conditions (i.e., $K[T, dl, E]$) to the average developmental time constant ($K = 136.83dd$) of individuals under non-diapause conditions. We used the conservative $\theta_L = 7.38^\circ\text{C}$ for the egg to adult period computed from the combined data (Fig. 2 and Table 1) to compute all post-treatment developmental times of individuals in these experiments. The multiple linear regression analysis of the data for all individuals across all treatment is Eq. 3i.

$$\begin{aligned} K(T, dl, E) &= 105.4 - 5.601T + 0.0377dl + 9.619E \\ &\quad - 0.478T \times E - 0.0051dl \times E + 0.00018T \times dl \times E \end{aligned} \quad (3i)$$

$$R^2 = 0.578, df = 6013, F = 1374.17$$

All of the independent variable and interaction terms, except $T \times dl$, were significant ($|5.01| < t_{\text{value}} < 23.31$, $p < 0.05$). Note that the effects of temperature are negative and that the

Table 3 Means (\pm SE) of net reproductive rate (R_0), finite rate of increase (λ), intrinsic rate of increase (r_m) and mean generation time (τ) of *Tuta absoluta* under six constant temperatures estimated by bootstrap sampling (10,000 replicates)

Temperatures (°C)	Life table parameters of <i>Tuta absoluta</i>			
	R_0	λ	r_m	τ
10	14.10 \pm 4.12 bc	1.02 \pm 3.38 e	2.56 $10^{-2} \pm 3.30 \cdot 10^{-3}$ e	103.04 \pm 1.77 a
15	25.17 \pm 5.56 ab	1.05 \pm 4.32 d	5.33 $10^{-2} \pm 4.10 \cdot 10^{-3}$ d	60.57 \pm 1.27 b
20	32.72 \pm 6.49 a	1.08 \pm 4.96 c	8.16 $10^{-2} \pm 4.58 \cdot 10^{-3}$ c	27.49 \pm 1.11 c
25	36.53 \pm 9.64 a	1.15 \pm 1.34 a	14.32 $10^{-2} \pm 1.17 \cdot 10^{-2}$ a	25.12 \pm 0.60 cd
30	11.33 \pm 2.74 c	1.10 \pm 1.24 b	9.79 $10^{-2} \pm 1.12 \cdot 10^{-2}$ b	24.80 \pm 0.90 cd
33	3.13 \pm 1.50 d	1.05 \pm 2.79 cde	5.21 $10^{-2} \pm 2.69 \cdot 10^{-2}$ cde	21.91 \pm 1.89 d

Means followed by different letters in columns indicate significant differences among treatments (paired bootstrap test, $P < 0.05$)

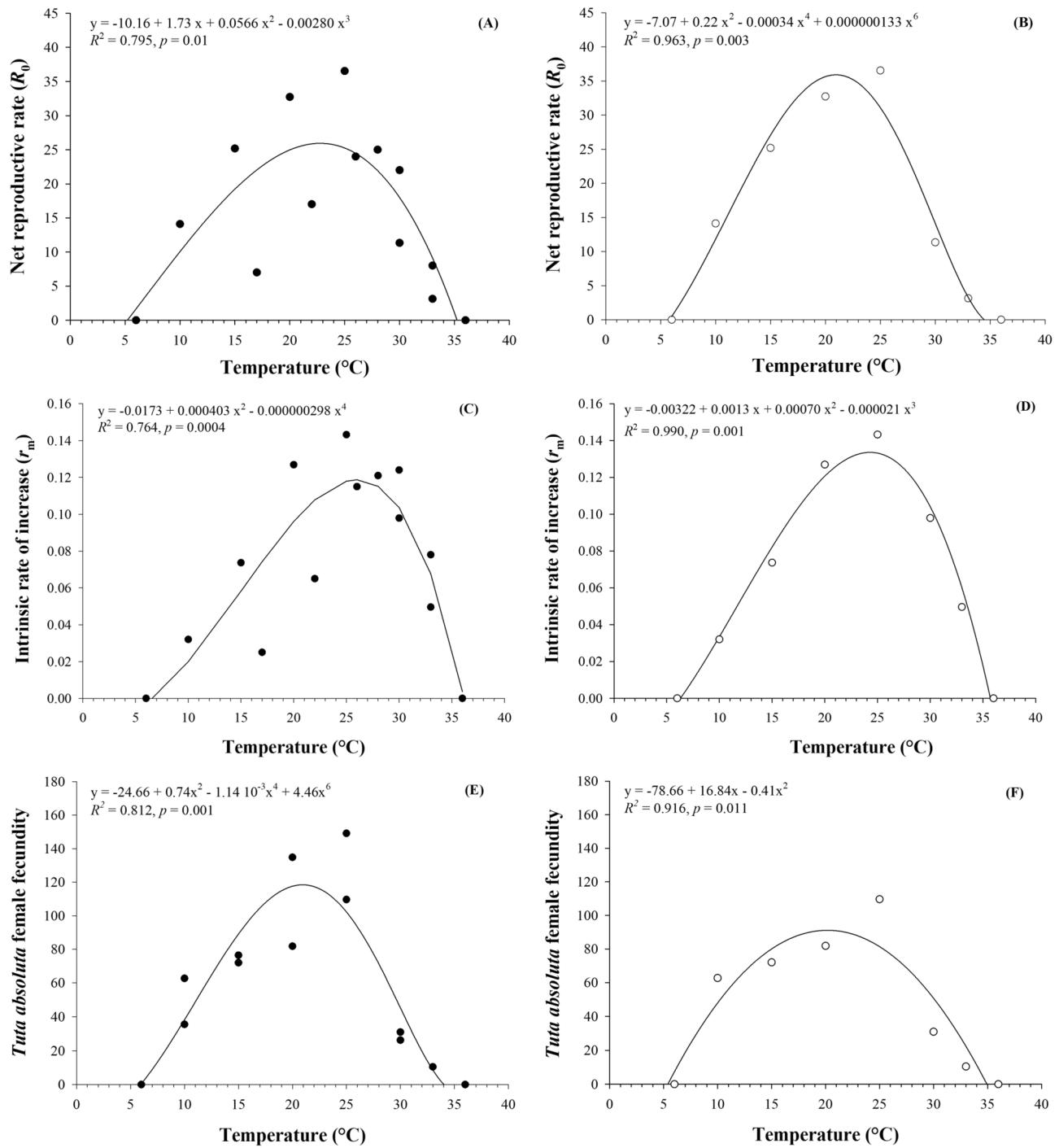


Fig. 3 Net reproductive rate (R_0), intrinsic rate of increase (r_m) and fecundity per female of *Tuta absoluta* at constant temperatures. Data from the literature and our experiments are combined in A, C, and E,

while only the experimental data are plotted in B, D, and F and are used to estimate θ_L and θ_U for each *T. absoluta* life table parameter

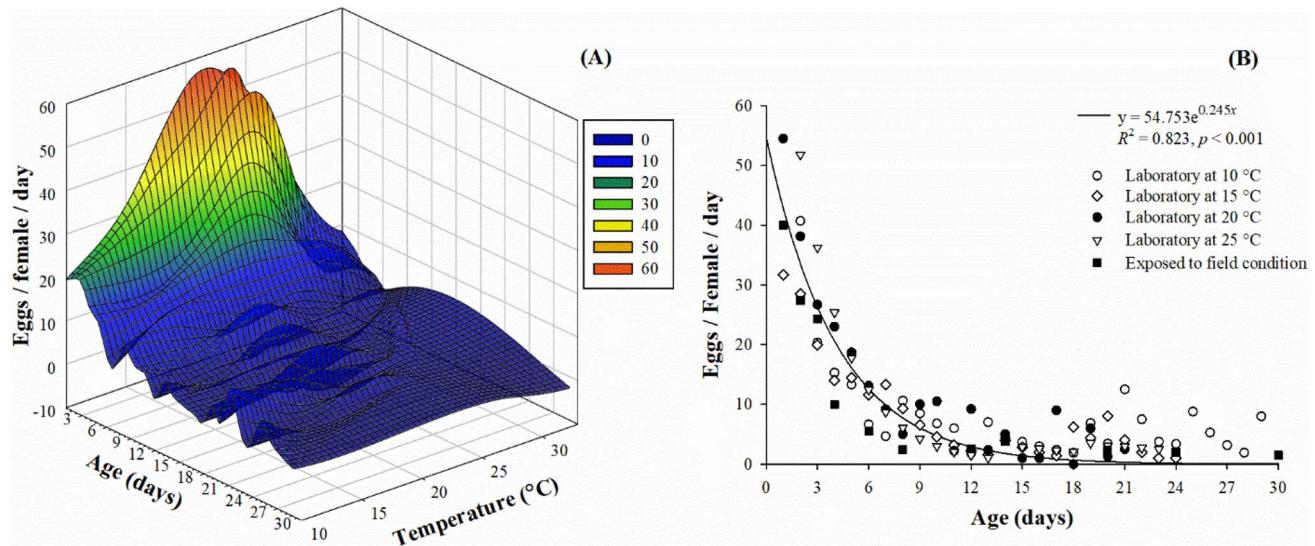


Fig. 4 Age-specific oviposition of *Tuta absoluta* females at six constant temperatures (10, 15, 20, 25, 30 and 33 °C). The data for egg/female/day are plotted as a 3-D figure on age (days) and temperature (a), while a statistical fit to all of the data eggs/female/day on age

(days) across temperatures is illustrated in (b). All females used in the experiments were derived from populations exposed to winter field condition (see text)

contribution of dl ($t_{value}=5.01$) to R^2 is about ~0.2% estimated by removing dl from the regression. The negative effect of T in Eq. 3i is due to the fact that treatment temperatures 10 and 15 °C are above the θ_L allowing development to continue during the treatment (see above).

Removing the 10 and 15 °C treatment data, the resulting regression model is Eq. 3ii. All independent variables and interaction terms except $T \times dl \times E$ are significant ($p < 0.05$; $|4.51| < t_{value} < 23.31$), with the contribution of dl to R^2 increasing to 8%.

$$K(T \leq 5^\circ\text{C}, dl, E) = -81.4 - 29.59T + 0.148dl + 15.20E - 0.0165T \times E - 1.132T \times E + 0.0074dl \times E \quad (3ii)$$

$$R^2 = 0.684, df = 2263, F = 814.43$$

Simplifying the notation, we define $\kappa = K(T, dl, E)$. To estimate the effect of each independent variable (T , dl , E) on post-treatment developmental times (i.e., the duration of diapause), we take the partial derivative with respect to each independent variable, using the average values of the other (i.e., $T [= 3.74^\circ\text{C}]$, $dl [= 730 \text{ min}]$ and $E [= 17.69 \text{ days}]$). For example, $\partial\kappa/\partial T = 29.59 - 0.165dl - 1.132 = -110.88dd$ estimates the decrease in the duration of diapause per degree increase in temperature. Similarly, $\partial\kappa/\partial dl = -0.60 dd$ is the decrease per minute increase in dl , and $\partial\kappa/\partial E = 5.563 dd$ is the increase per day of exposure time. By far, the most consistent response is to E ($t_{value}=34.13$), then T ($t_{value}=10.28$),

and dl ($t_{value}=8.48$). The data for $K(T, E)$ are mapped on E and $T \times dl$ in Fig. 5A, B.

The mean and standard deviation of developmental times (dd) of pupae that survived to adult emergence ($K(T, dl, E)$) are summarized in Table 4 with the relevant comparisons being to the developmental times of non-treatment pupae K (i.e., $136.83dd$). The post-treatment developmental times for 7 day exposure at 2 or 5 °C are the same as K . However, the developmental time of treated pupae increases with exposure time E , with pupae from the 42 day treatments

at 2 and 5 °C requiring an average of $407.65dd \pm 125.13$ and $334.59dd \pm 82.6$, respectively, to complete development (Table 4). The results shows a clear facultative diapause in $K(T, dl, E)$ pupae with average developmental times being threefold $> K$ at 2 °C and 2.45-fold at 5 °C. The proportions of pupae satisfying the inequality $K(T, dl, E) > K$ ranged from 60 to 80% at 2 and 5 °C and exposure periods 14, 28 and 42 days (Fig. 5C). The percentage diapause in the 7 day treatments ranged from 6 to 15% at 2 and 5 °C (Fig. 5C).

The mortality of pupae at 2 °C was 90% after 42 days, but decreased with shorter exposure periods and increased temperature (Fig. 5D). The same pattern with slightly lower mortality occurred at 5 °C.

Discussion

Temperature is an important abiotic factor exerting direct and indirect effects on SATP population development, dynamics, and invasive success (Kahrer et al. 2019; Machekano et al. 2018; Santana et al. 2019). Knowledge of the effects of temperature on demographic parameters can help predict the establishment and relative abundance of the pest, and could underpin the development of monitoring and control strategies for integrated pest management. However, the temperature range must be fully explored in such studies and must include the entire life cycle of the insect. During the early stages of its invasion in Europe, the effects of extreme temperatures on SATP were neglected, likely because the

Table 4 Mean and standard deviation of development times of *Tuta absoluta* in *K(TxP)* experiments averaged across day length treatments

Temperatures (°C)	Exposure periods (days)	Observed <i>K</i> (<i>T</i> , <i>dl</i> , <i>E</i>) Mean \pm SD
2	7	135.207 \pm 35.579
	14	145.569 \pm 26.486
	28	256.748 \pm 39.599
	42	407.653 \pm 125.130
5	7	134.718 \pm 25.777
	14	170.848 \pm 29.310
	28	198.986 \pm 40.256
	42	334.593 \pm 82.613

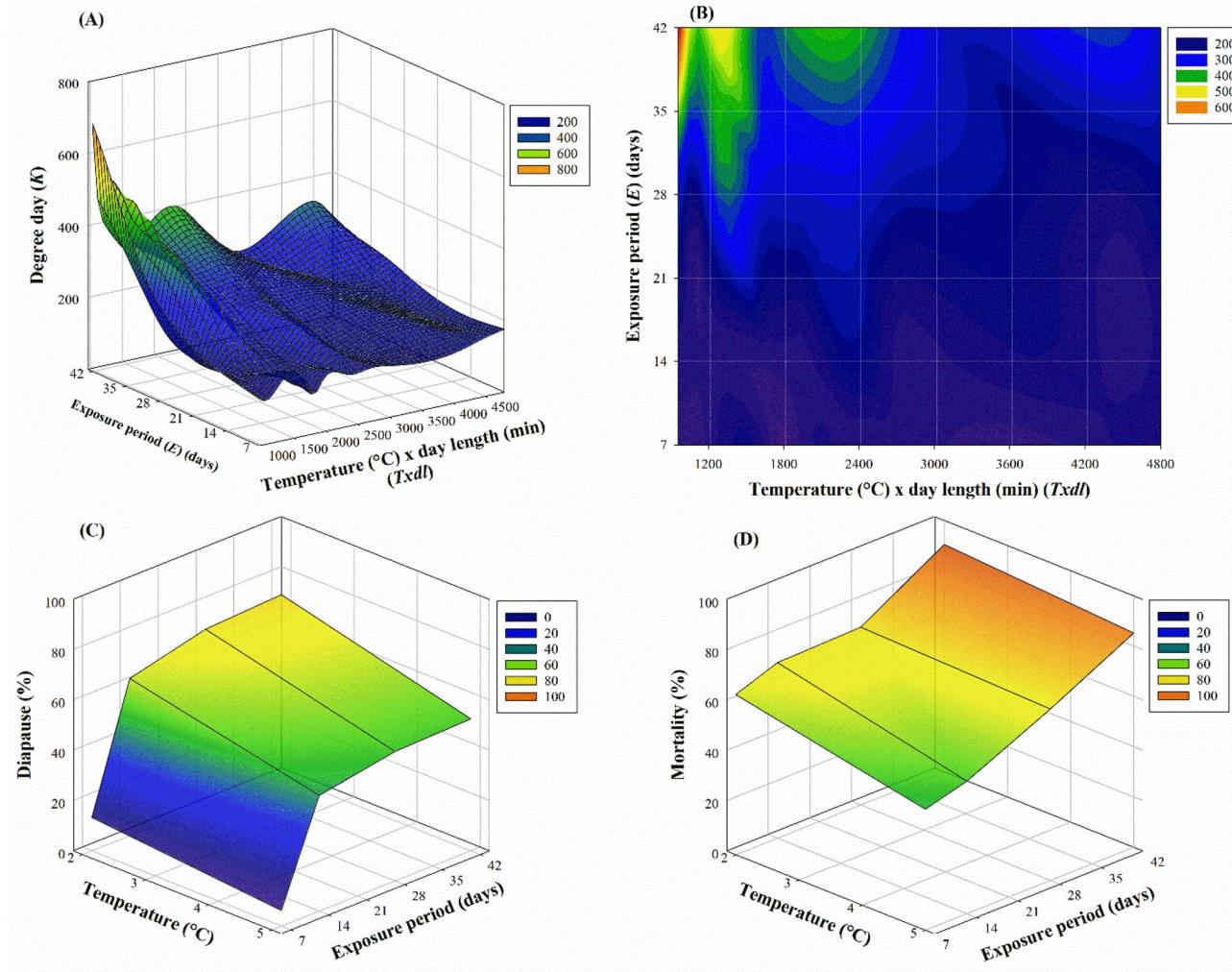


Fig. 5 Three-dimensional graphic summaries of the diapause experiment data: (a) a smoothed model fit of pupal developmental times (*K*, degree days) on the interaction of temperature (2 and 5 °C) x minutes day length (*Txdl*) (480, 540, 600, 840, 900, 960) and on exposure

time (*E*, 7, 14, 28 and 42 days), (b) a contour map of *K* (sub figure a) across the interaction (*Txdl*) and exposure time *E*, (c) the percentage diapause on temperatures (2 and 5 °C), and exposure times, and (d) the observed percentage mortality in the experimental data from C

tropical populations studied were not considered high risk for north temperate climates (Desneux et al. 2010). Early studies using South American populations are those of Barrientos et al. (1998), Krechmer and Foerster (2015), and Martins et al. 2016. The SATP's rapid spread in the colder climes of the Palearctic region was unexpected and stimulated studies to explore the thermal biology at extreme temperatures (Cuthbertson et al. 2013; Kahrer et al. 2019; Machekano et al. 2018; Van Damme et al. 2015). Our paper attempts a holistic analysis of the thermal biology of *T. absoluta* to explore some of these issues.

In the long term, natural selection on the genetic constitution of a population across generations occurs, but is slower (Nyamukondiwa et al. 2010, 2013; Sgrò et al. 2016) than required for invasion success. For successful establishment, exotic pests may have to overcome developmental obstacles in the new environment (Renault et al. 2018). Phenological variations (Briscoe et al. 2012; Chuine 2010) and phenotypic plasticity are important attributes enabling persistence in and adaptation to new, possibly harsher conditions (Manenti et al. 2017; Sgrò et al. 2016). Native environmental heterogeneity may contribute to species invasive success (Renault et al. 2018), and greater phenotypic plasticity in species coming from more heterogeneous environments may enable them to better adapt to novel environments than species from more homogenous environments (Manenti et al. 2017). The “latitudinal hypothesis” posits that developmental plasticity should increase from low to higher latitude in response to increasing thermal seasonality (Bozinovic et al. 2011; Ghalambor et al. 2007; Manenti et al. 2017).

Although SATP is thought to be from the Huancayo region in the central Andean highlands of Peru (Povolný 1994) with low variation in temperature throughout the year, the population that arrived in Europe was from central Chile, a region with broad variations in temperatures (Guillemaud et al. 2015). Our studies were conducted using a SATP population from France (45°N latitude), while the data in the literature are from Brazil (20°S and 25°S latitudes) (Krechmer and Foerster 2015; Martins et al. 2016), and from Chile (33°S) (Barrientos et al. 1998) (see weather summary in Table S3).

Thermal responses

Differences in the thermal requirement of the same species in different geographic regions are reported in the literature (e.g., *Plutella xylostella*; Lepidoptera: Yponomeutidae) (Marchioro and Foerster 2011; Shirai 2000). Linear regression analyses of the stage-specific developmental rate data from our experiments and the combined data reported in the literature from South America (Barrientos et al. 1998; Krechmer and Foerster 2015; Martins et al. 2016) show that the within stage slopes and intercepts were not significantly

different, but the estimated lower thermal thresholds for the larval and pupal stages were lower in our studies (Table 1). The differences in θ_L could be due to experimental methods (observation interval, materials), accuracy of the temperatures in the experiments, the population of SATP used, and/or to rapid adaptation after invasion in Europe (Biondi et al. 2018). For these reasons, we await confirming studies of our results on temperate populations of SATP. Hence, we used the combined data on egg to adult development from all sources to estimate the conservative lower and upper temperature thresholds of 7.38 °C and 33.82 °C, respectively.

In the range of temperatures tested (6–36 °C), only eggs did not survive at 6 °C, while no stage survived at 36 °C. Females were able to oviposit at temperatures above 6 °C, but not above 33 °C. Reproduction was highest at 20.53 °C and declined above and below this temperature. The longevity of SATP of adults was 88 days (538.6dd) in the open field at mean daily temperature of 13.5 °C, and 136 and 81 days under laboratory conditions of 10 (356.2dd) and 15 °C (617.2dd), respectively (Fig. S2A and B). Additional insights into the thermal limits of SATP were gained by fitting a nonlinear concave regression model to demographic parameters R_0 , r_m , total fecundity, age-specific fertility, and survivorship on temperature (Gutierrez 1996). The lower and upper thresholds for the demographic statistics were similar to those reported in the literature.

Diapause

Prior studies suggested that SATP has cold tolerance to otherwise lethal low temperatures (Kahrer et al. 2019; Machekano et al. 2018; Van Damme et al. 2015). We confirmed this in a field study where immature stages survived mean winter temperatures of 5.77 ± 4.29 °C, including a period when temperatures under snow dropped below 0 °C. Two types of physiological mechanisms enable insect survival during winter in the temperate zone: cold hardiness and diapause, though the relationship between them may not be separable (Denlinger 1986, 1991; Tougeron 2019). Diapause occurs during a specific stage of development, whereas cold hardiness may be expressed across a broad range of developmental stages (Denlinger 1991; Lee 1991). Diapause in insects has evolved as an adaptive strategy to survive harsh seasonal factors (temperature, moisture, food, etc.), an adaptation that is subject to both environmental influences and genetics (Tauber and Tauber 1981). Induction and maintenance of diapause under laboratory conditions have been demonstrated to vary considerably among species, geographic strains, and individuals of the same species (Fu et al. 2015; Masaki 1961). Diapause may be facultative or obligate. Facultative diapause in multivoltine species can be mediated by environmental cues inducing delayed physiological development as observed for SATP pupae.

In contrast, insects with obligate diapause enter diapause regardless of environmental cues (Danks 1987; Lees 1956; Tauber et al. 1986).

While the occurrence of diapause in SATP had not been reported for the South American strain(s), the observations of Sannino and Espinosa (2010) in Salerno, Italy, suggested the possibility of diapause in SATP. We found strong evidence for facultative diapause in SATP pupae in our experiments wherein 3rd and 4th instar larvae were exposed to 2 and 5 °C and short-day lengths for different periods. Facultative diapause in pupae was determined in treated individuals as a large increase in post-treatment developmental time of pupae after being returned to a favorable temperature. Compared to normal development, 2.45–3-fold increases in pupal developmental times occurred with low temperature and exposure time being the major contributing factors. The incidence of diapause was > 80% at 2 °C and > 60% at 5 °C.

Over the 14 years SATP has been present in Western Europe, it has gone through the process of establishment and expansion (Campos et al. 2017; Renault et al. 2018). The current level of facultative diapause in our French strain of SATP could be part of its innate plasticity, or it could have been selected during invasion of the Palearctic. In either case, it could be the springboard for greater selection for adaptation to harsher climates, enabling it to expand its range further making it even more difficult to control.

Overall, the data show that the moth has a rapid developmental time, a very short pre-oviposition period, and that it can reproduce at temperatures close to the development threshold enabling it to produce 6–9 generations per year in warmer temperate areas of the Mediterranean Basin. Furthermore, the moth is cold hardy (Kahrer et al. 2019) and, as shown here, can enter a facultative diapause as temperatures cool. These are characteristics that could enable SATP to invade large areas of the Palearctic. These insights accrued from a holistic analysis of its thermal biology.

Authors' contributions

MRC, LP, APG, AB and ND designed the research. MRC, PB and EAD led the trials. ND, AB and AA provided technical and material supports. MRC and APG analyzed the data. MRC, APG and LP wrote the manuscript. All authors read, edited and approved the manuscript.

Acknowledgements The authors wish to thank Lionel Salvy, Valérie Frandon, Anne-Violette Lavoie, Christiane Metay-Merrien, Richard Brun, Roger Boll, Sylvain Nuée, Christian Wdziekonski, Sylvain Benhamou, Anouck Lasserre, Lucie S. Monticelli, Yusha Wang, Yanyan Qu, Marianne Araújo Soares, Christine Becker, Peng Han, Timothée Fichant, Tara Malanga, Mathilda Idier, Ha Le-Thu Nguyen and Eva Thomine from INRAE for technical assistance and sharing their expertise during these studies.

Funding Project ASCII (FP7 IRSES No. 318246) for funding to ND and MRC, the University of Catania (Project 2016-18 "Emergent Pests and Pathogens and Relative Sustainable Strategies - 5A722192113) for financing to AB, the IPM Innovation Lab (USAID Cooperative Agreement No. AID-OAA-L-15-00001) for funding to ND and MRC, and the EUCLID project (H2020-SFS-2014, No. 633999) for funding to PB, EAD and ND. In kind funding accrued from the Center for Sustainable Agricultural Systems, Kensington, CA, USA. LP was supported by the MED-GOLD project that has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 776467.

Availability of data and materials The datasets used and analyzed during the current study are available from the corresponding authors on reasonable request.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

Consent for publication Not applicable.

Ethical approval and consent to participate Not applicable.

References

- Anastasaki E, Drizou F, Milonas PG (2018) Electrophysiological and Oviposition Responses of *Tuta absoluta* Females to Herbivore-Induced Volatiles in Tomato Plants. *J Chem Ecol* 44 (3):288–298
- Asplen MK, Anfora G, Biondi A, Choi D-S et al (2015) Invasion biology of spotted wing Drosophila (*Drosophila suzukii*): a global perspective and future priorities. *J Pest Sci* 88:469–494
- Bacci L, Silva EM, Martins JC, Soares MA et al (2019) Seasonal variation in natural mortality factors of *Tuta absoluta* (Lepidoptera: Gelechiidae) in open-field tomato cultivation. *J Appl Entomol* 143:21–33
- Barrientos ZR, Apablaza HJ, Norero SA, Estay PP (1998) English: threshold temperature and thermal constant for development of the South American tomato moth *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae). *Cienc Investig Agrar* 25:133–137
- Biondi A, Desneux N, Siscaro G, Zappalà L (2012) Using organic-certified rather than synthetic pesticides may not be safer for biological control agents: selectivity and side effects of 14 pesticides on the predator *Orius laevigatus*. *Chemosphere* 87:803–812
- Biondi A, Guedes RNC, Wan F-H, Desneux N (2018) Ecology, Worldwide Spread, And Management Of The Invasive South American Tomato Pinworm, *Tuta absoluta*: past, present, and future. *Annu Rev Entomol* 63:239–258
- Bozinovic F, Bastías DA, Boher F, Clavijo-Baquet S et al (2011) The mean and variance of environmental temperature interact to determine physiological tolerance and fitness. *Physiol Biochem Zool* 84:543–552
- Briere J-F, Pracros P, Le Roux A-Y, Pierre J-S (1999) A novel rate model of temperature-dependent development for arthropods. *Environ Entomol* 28:22–29
- Briscoe NJ, Porter WP, Sunnucks P, Kearney MR (2012) Stage-dependent physiological responses in a butterfly cause non-additive effects on phenology. *Oikos* 121:1464–1472
- Campbell A, Frazer BD, Gilbert N, Gutierrez AP et al (1974) Temperature requirements of some aphids and their parasites. *J Appl Ecol* 11:431–438

- Campos MR, Biondi A, Adiga A, Guedes RNC et al (2017) From the Western Palaearctic region to beyond: *Tuta absoluta* 10 years after invading Europe. *J Pest Sci* 90:787–796
- Cherif A, Attia-Barhoumi S, Mansour R, Zappalà L et al (2019) Elucidating key biological parameters of *Tuta absoluta* on different host plants and under various temperature and relative humidity regimes. *Entomol Gen* 39:1–7
- Chi H (1988) Life-table analysis incorporating both sexes and variable development rates among individuals. *Environ Entomol* 17:26–34
- Chi H (2019) TWOSEX-MSChart: a computer program for the age stage, twosex life table analysis. <https://140.120.197.173/ecology/Download/Twosex-MSChart.zip>.
- Chi H, Liu H (1985) Two new methods for the study of insect population ecology. *Acad Sin Inst Zool Monogr Ser Bull Inst Zool* 24:225–240
- Chi H, You M, Athihan R, Smith CL et al (2020) Age-stage, two-sex life table: an introduction to theory, data analysis, and application. *Entomol Gen* 40:102–123
- Chuine I (2010) Why does phenology drive species distribution? *Philos T R Soc B* 365:3149–3160
- Cuthbertson GSA, Mathers JJ, Blackburn FL, Korycinska A et al (2013) Population Development of *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) under simulated UK glasshouse conditions. *Insects* 4:185–197
- Danks HV (1987) Insect dormancy: an ecological perspective. Biological survey of Canada monograph series, Biological Survey of Canada (Terrestrial Arthropods), Ottawa, Canada
- Danks HV (2007) The elements of seasonal adaptations in insects. *Can Entomol* 139:1–44
- Denlinger DL (1986) Dormancy in tropical insects. *Annu Rev Entomol* 31:239–264
- Denlinger DL (1991) Relationship between cold hardiness and diapause. In: Lee RE, Denlinger DL (eds) Insects at low temperature. Springer, Boston, pp 174–198. https://doi.org/10.1007/978-1-4757-0190-6_8
- Denlinger DL (2002) Regulation of diapause. *Annu Rev Entomol* 47:93–122
- Denlinger DL (2008) Why study diapause? *Entomol Res* 38:1–9
- Desneux N, Luna MG, Guillemaud T, Urbaneja A (2011) The invasive South American tomato pinworm, *Tuta absoluta*, continues to spread in Afro-Eurasia and beyond: the new threat to tomato world production. *J Pest Sci* 84:403–408
- Desneux N, Wajnberg E, Wyckhuys KAG, Burgio G et al (2010) Biological invasion of European tomato crops by *Tuta absoluta*: ecology, geographic expansion and prospects for biological control. *J Pest Sci* 83:197–215
- Efron B, Tibshirani R (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat Sci* 1:54–75
- Flitters NE, Messenger PS (1965) Effect of temperature and humidity on development and potential distribution of the Mexican fruit fly in the United States. Technical bulletin / United States Department of Agriculture. U.S. Dept. of Agriculture, Washington
- Fu S, Chen C, Xiao L, He H et al (2015) Inheritance of diapause in crosses between the northernmost and the southernmost strains of the asian corn borer *Ostrinia furnacalis*. *PlosONE* 10:e0118186
- Ghalambor CK, McKay JK, Carroll SP, Reznick DN (2007) Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Funct Ecol* 21:394–407
- Gilbert N, Gutierrez AP (1973) A plant-aphid-parasite relationship. *J Anim Ecol* 42:323–340
- Goodman D (1982) Optimal life histories, optimal notation, and the value of reproductive value. *Am Nat* 119:803–823
- Guedes RNC, Picanço MC (2012) The tomato borer *Tuta absoluta* in South America: pest status, management and insecticide resistance. *EPPO Bull* 42:211–216
- Guillemaud T, Blin A, Le Goff I, Desneux N et al (2015) The tomato borer, *Tuta absoluta*, invading the Mediterranean Basin, originates from a single introduction from Central Chile. *Sci Rep* 5:8371
- Gutierrez AP (1996) Applied population ecology: a supply-demand approach. Wiley, New York
- Gutierrez AP, Ponti L, Giloli G, Baumgärtner J (2018) Climate warming effects on grape and grapevine moth (*Lobesia botrana*) in the Palaearctic region. *Agric For Entomol* 20:255–271
- Gutierrez AP, Ponti L, Herren HR, Baumgärtner J et al (2015) Deconstructing Indian cotton: weather, yields, and suicides. *Environ Sci Eur* 27:12
- Han P, Bayram Y, Shaltiel-Harpaz L, Sohrabi F et al (2019) *Tuta absoluta* continues to disperse in Asia: damage, ongoing management and future challenges. *J Pest Sci* 92:1317–1327
- Han P, Lavois A-V, Le Bot J, Amiens-Desneux E et al (2014) Nitrogen and water availability to tomato plants triggers bottom-up effects on the leafminer *Tuta absoluta*. *Sci Rep* 4:4455
- Han P, Zhang Y-n, Lu Z-z, Wang S et al (2018) Are we ready for the invasion of *Tuta absoluta*? Unanswered key questions for elaborating an Integrated Pest Management package in Xinjiang, China. *Entomol Gen* 38:113–125
- Iltis C, Moreau J, Pecharová K, Thiéry D et al (2020) Reproductive performance of the European grapevine moth *Lobesia botrana* (Tortricidae) is adversely affected by warming scenario. *J Pest Sci* 93:679–689
- Kahrer A, Moyses A, Hochfellner L, Tiefenbrunner W et al (2019) Modelling time-varying low-temperature-induced mortality rates for pupae of *Tuta absoluta* (Gelechiidae, Lepidoptera). *J Appl Entomol* 143:1143–1153
- Kang L, Chen B, Wei J-N, Liu T-X (2009) Roles of thermal adaptation and chemical ecology in *liriomyza* distribution and control. *Annu Rev Entomol* 54:127–145
- Krechmer SF, Foerster AL (2015) *Tuta absoluta* (Lepidoptera: Gelechiidae): thermal requirements and effect of temperature on development, survival, reproduction and longevity. *Eur J Entomol* 112:658–663
- Lactin DJ, Holliday NJ, Johnson DL, Craigen R (1995) Improved rate model of temperature-dependent development by arthropods. *Environ Entomol* 24:68–75
- Lawler JJ, White D, Neilson RP, Blaustein AR (2006) Predicting climate-induced range shifts: model differences and model reliability. *Glob Change Biol* 12:1568–1584
- Leather SR, Walters KFA, Bale JS (1993) The ecology of insect overwintering. Cambridge University Press, Cambridge
- Lee RE (1991) Principles of insect low temperature tolerance. In: Lee RE, Denlinger DL (eds) Insects at low temperature. Springer, Boston, pp 17–46. https://doi.org/10.1007/978-1-4757-0190-6_2
- Lee S, Lee Y, Lee S (2020) Population genetic structure of *Anoplophora glabripennis* in South Korea: invasive populations in the native range? *J Pest Sci* 93:1181–1196
- Lees AD (1956) The physiology and biochemistry of diapause. *Annu Rev Entomol* 1:1–16
- Logan JA, Wollkind DJ, Hoyt SC, Tanigoshi LK (1976) An analytic model for description of temperature dependent rate phenomena in arthropods 1. *Environ Entomol* 5:1133–1140
- Machekano H, Mutamiswa R, Nyamukondiwa C (2018) Evidence of rapid spread and establishment of *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) in semi-arid Botswana. *Agric Food Secur* 7:48
- Manenti T, Sørensen JG, Loeschke V (2017) Environmental heterogeneity does not affect levels of phenotypic plasticity in natural populations of three *Drosophila* species. *Ecol Evol* 7:2716–2724

- Mansour R, Brévault T, Chailleux A, Cherif A et al (2018) Occurrence, biology, natural enemies and management of *Tuta absoluta* in Africa. Entomol Gen 38:83–112
- Marchioro C, Foerster L (2011) Development and survival of the diamondback moth, *Plutella xylostella* (L.) (*Lepidoptera: Yponomeutidae*) as a function of temperature: effect on the number of generations in tropical and subtropical regions. Neotrop Entomol 40:533–541
- Martins JC, Picanço MC, Bacci L, Guedes RNC et al (2016) Life table determination of thermal requirements of the tomato borer *Tuta absoluta*. J Pest Sci 89:897–908
- Masaki S (1961) Geographic variation of diapause in insects. Faculty of Agriculture, Hirosaki University: 98
- McNitt J, Chungbaek YY, Mortveit H, Marathe M et al (2019) Assessing the multi-pathway threat from an invasive agricultural pest: *Tuta absoluta* in Asia. P R Soc B-Biol Sci 286:1–9
- Musolin DL, Dolgovskaya MY, Protsenko VY, Karpun NN et al (2019) Photoperiodic and temperature control of nymphal growth and adult diapause induction in the invasive Caucasian population of the brown marmorated stink bug, *Halyomorpha halys*. J Pest Sci 92:621–631
- Nyamukondiwa C, Kleynhans E, Terblanche JS (2010) Phenotypic plasticity of thermal tolerance contributes to the invasion potential of Mediterranean fruit flies (*Ceratitis capitata*). Ecol Entomol 35:565–575
- Nyamukondiwa C, Weldon CW, Chown SL, le Roux PC et al (2013) Thermal biology, population fluctuations and implications of temperature extremes for the management of two globally significant insect pests. J Insect Physiol 59:1199–1211
- Ponti L, Gilioli G, Biondi A, Desneux N et al (2015a) Physiologically based demographic models streamline identification and collection of data in evidence-based pest risk assessment. EPPO Bull 45:317–322
- Ponti L, Gutierrez AP, Altieri MA (2015b) Holistic approach in invasive species research: the case of the tomato leaf miner in the Mediterranean Basin. Agroecol Sustain Food Syst 39:436–468
- Povolný D (1994) Gnornoschemini of southern South America VI: identification keys, checklist of Neotropical taxa and general considerations (*Insecta, Lepidoptera, Gelechiidae*). Steenstrupia 20:1–42
- Rank A, Ramos RS, da Silva RS, Soares JRS et al (2020) Risk of the introduction of *Lobesia botrana* in suitable areas for *Vitis vinifera*. J Pest Sci 93:1167–1179
- Renault D, Laparie M, McCauley SJ, Bonte D (2018) Environmental adaptations, ecological filtering, and dispersal central to insect invasions. Annu Rev Entomol 63:345–368
- Requier F, Rome Q, Chiron G, Decante D et al (2019) Predation of the invasive Asian hornet affects foraging activity and survival probability of honey bees in Western Europe. J Pest Sci 92:567–578
- Roques A, Auger-Rozenberg M-A, Blackburn TM, Garnas J et al (2016) Temporal and interspecific variation in rates of spread for insect species invading Europe during the last 200 years. Biol Invasions 18:907–920
- Sannino L, Espinosa B (2010) *Tuta absoluta*: guida alla conoscenza e recenti acquisizioni per una corretta difesa. Edizioni L'Informatore Agrario, Verona, Italy
- Santana PA, Kumar L, Da Silva RS, Picanço MC (2019) Global geographic distribution of *Tuta absoluta* as affected by climate change. J Pest Sci 92:1373–1385
- Sgrò CM, Terblanche JS, Hoffmann AA (2016) What can plasticity contribute to insect responses to climate change? Annu Rev Entomol 61:433–451
- Shirai Y (2000) Temperature tolerance of the diamondback moth, *Plutella xylostella* (*Lepidoptera: Yponomeutidae*) in tropical and temperate regions of Asia. Bull Entomol Res 90:357–364
- Sylla S, Brévault T, Bal AB, Chailleux A et al (2017) Rapid spread of the tomato leafminer, *Tuta absoluta* (*Lepidoptera: Gelechiidae*), an invasive pest in Sub-Saharan Africa. Entomol Gen 36:269–283
- Tarusikirwa VL, Mutamiswa R, English S, Chidawanyika F et al (2020) Thermal plasticity in the invasive south American tomato pinworm *Tuta absoluta* (Meyrick) (*Lepidoptera: Gelechiidae*). J Therm Biol 90:102598
- Tauber CA, Tauber MJ (1981) Insect seasonal cycles: genetics and evolution. Annu Rev Ecol Syst 12:281–308
- Tauber MJ, Tauber CA, Masaki S (1986) Seasonal adaptations of insects. Oxford University Press, New York, NY, USA
- Tonnang HEZ, Mohamed SF, Khamis F, Ekesi S (2015) Identification and risk assessment for worldwide invasion and spread of *Tuta absoluta* with a focus on Sub-Saharan Africa: implications for phytosanitary measures and management. PlosONE 10:e0135283
- Torres JB, Faria CA, Evangelista WS, Pratissoli D (2001) Within-plant distribution of the leaf miner *Tuta absoluta* (Meyrick) immatures in processing tomatoes, with notes on plant phenology. Int J Pest Manag 47:173–178
- Tougeron K (2019) Diapause research in insects: historical review and recent work perspectives. Entomol Exp Appl 167:27–36
- Van Damme V, Berkvens N, Moerkens R, Berckmoes E et al (2015) Overwintering potential of the invasive leafminer *Tuta absoluta* (Meyrick) (*Lepidoptera: Gelechiidae*) as a pest in greenhouse tomato production in Western Europe. J Pest Sci 88:533–541
- Verheggen F, Fontus RB (2019) First record of *Tuta absoluta* in Haiti. Entomol Gen 38:349–353

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Mateus Ribeiro de Campos¹ · Philippe Béarez¹ · Edwige Amiens-Desneux¹ · Luigi Ponti^{2,3} · Andrew Paul Gutierrez^{3,4} · Antonio Biondi⁵ · Abhijin Adiga⁶ · Nicolas Desneux¹

Antonio Biondi
antonio.biondi@unict.it

Nicolas Desneux
nicolas.desneux@univ-cotedazur.fr

¹ INRAE, CNRS, UMR ISA, University Côte d'Azur, 06000 Nice, France

² Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile (ENEA), Centro Ricerche Casaccia, Via Anguillarese 301, 00123 Roma, Italy

³ Center for the Analysis of Sustainable Agricultural Systems Global (CASAS Global), Kensington, CA 94707-1035, USA

⁴ College of Natural Resources, University of California, Berkeley, CA 94720, USA

⁵ Department of Agriculture, Food and Environment, University of Catania, Via Santa Sofia 100, Catania, Italy

⁶ University of Virginia, Charlottesville, VA 22904, USA

Prioritizing allocation of COVID-19 vaccines based on social contacts increases vaccination effectiveness

Jiangzhuo Chen¹, Stefan Hoops¹, Achla Marathe^{1,4}, Henning Mortveit^{1,3}, Bryan Lewis¹, Srinivasan Venkatramanan¹, Arash Haddadan¹, Parantapa Bhattacharya¹, Abhijin Adiga¹, Anil Vullikanti^{1,2}, Aravind Srinivasan⁶, Mandy L Wilson¹, Gal Ehrlich⁵, Maier Fenster⁵, Stephen Eubank^{1,4}, Christopher Barrett^{1,2}, and Madhav Marathe^{1,2,*}

¹Network Systems Science and Advanced Computing Division, Biocomplexity Institute, University of Virginia, Charlottesville, VA 22904, USA

²Department of Computer Science, University of Virginia

³Department of Engineering Systems and Environment, University of Virginia

⁴Department of Public Health Sciences, University of Virginia

⁵Ehrlich & Fenster of the Ehrlich Group, Ramat-Gan 5268104, Israel

⁶Department of Computer Science, University of Maryland

*Contact author: marathe@virginia.edu

February 15, 2021

Abstract

We study allocation of COVID-19 vaccines to individuals based on the structural properties of their underlying social contact network. Even optimistic estimates suggest that most countries will likely take 6 to 24 months to vaccinate their citizens. These time estimates and the emergence of new viral strains urge us to find quick and effective ways to allocate the vaccines and contain the pandemic. While current approaches use combinations of age-based and occupation-based prioritizations, our strategy marks a departure from such largely aggregate vaccine allocation strategies. We propose a novel agent-based modeling approach motivated by recent advances in (*i*) science of real-world networks that point to efficacy of certain vaccination strategies and (*ii*) digital technologies that improve our ability to estimate some of these structural properties. Using a realistic representation of a social contact network for the Commonwealth of Virginia, combined with accurate surveillance data on spatio-temporal cases and currently accepted models of within- and between-host disease dynamics, we study how a limited number of vaccine doses can be strategically distributed to individuals to reduce the overall burden of the pandemic. We show that allocation of vaccines based on individuals' degree (number of social contacts) and total social proximity time is *significantly more effective* than the currently used age-based allocation strategy in terms of number of infections, hospitalizations and deaths. Our results suggest that in just two months, by March 31, 2021, compared to age-based allocation, the proposed degree-based strategy can result in *reducing an additional 56–110k infections, 3.2–5.4k hospitalizations, and 700–900 deaths just in the Commonwealth of Virginia. Extrapolating these results for the entire US, this strategy can lead to 3–6 million fewer infections, 181–306k fewer hospitalizations, and 51–62k fewer deaths compared to age-based allocation.* The overall strategy is robust even: (*i*) if the social contacts are not estimated correctly;

(ii) if the vaccine efficacy is lower than expected or only a single dose is given; (iii) if there is a delay in vaccine production and deployment; and (iv) whether or not non-pharmaceutical interventions continue as vaccines are deployed. For reasons of implementability, we have used degree, which is a simple structural measure and can be easily estimated using several methods, including the digital technology available today. These results are significant, especially for resource-poor countries, where vaccines are less available, have lower efficacy, and are more slowly distributed.

1 Introduction

New vaccines typically take a decade to develop and distribute, but vaccines for COVID-19, the disease caused by the novel coronavirus SARS-CoV-2, have been developed in record time to help mitigate the raging pandemic. As of February 13, 2021, the reported number of confirmed cases and deaths in the US stand at 27M and 467K; the reported number of confirmed cases and deaths worldwide stand at 103M and 2.3M respectively.¹. These numbers are likely to go up substantially in the coming months. Vaccines offer a safe and effective way to contain the pandemic quickly. However, the supply of COVID-19 vaccines is limited, so the challenge now is the distribution of these vaccines in a timely manner to bring the pandemic under control. If we have a sufficient number of vaccines to immunize 70-90% of the people in the United States (US), protection can be offered to both individuals who are immunized and those who are unimmunized through herd immunity.

In the next 3 months, the US is expected to have a total of only 100 million vaccines, which is sufficient to immunize only 30% of the population (15% if we account for two doses) and thus cannot provide herd immunity. Lacking that, the current focus of the vaccination is to protect individuals who are at a high risk of infection and mortality, as well as critical workers.

Vaccination priority is complex and intertwined with age, race, occupation, health equity, geography, and politics. Data shows that COVID-19 disproportionately affects older adults, Blacks, Hispanics, American Indians, gig and wage workers, and individuals with comorbidities. Many of these attributes are also correlated with low socio-economic status (SES), and high social vulnerability. There can be many criteria for prioritization, for example: (i) risk of infection; (ii) risk of death; (iii) risk of transmission if infected; and (iv) occupation, such as healthcare workers, teachers, cashiers, etc. Estimating the consequences of different prioritization strategies is further complicated by production limitations, requiring a vaccine schedule to be specified for each. Additionally, vaccine distribution requires complex logistical support, such as cold-chain storage, transportation, qualified personnel, and scheduling etc. for any prioritization scheme to achieve its results in an effective and equitable manner. See [4] for a comprehensive discussion on this topic.

The US Centers for Disease Control and Prevention (CDC) has announced a prioritization order based on the Advisory Committee on Immunization Practices (ACIP). It recommends healthcare personnel and long-term facility care residents be vaccinated first; followed by frontline essential workers, and those aged 75 years and older because they are at a higher risk of hospitalization, illness and death; followed by those aged 65-74 years; followed by those aged 16-64 years with underlying medical conditions and other essential workers. Although most states in the US follow a similar phased approach, there are subtle, but important, variations on who gets vaccinated first.

Our contributions. The current rate of vaccination in the US and other countries suggests

¹See <https://nssac.bii.virginia.edu/covid-19/dashboard/> for latest figures

that it can take between 6-24 months to complete vaccination campaigns for much of the world. At the same time, the discovery of multiple variant strains implies a rapid acceleration of the pandemic in several parts of the world. The number of new strains are likely to increase with increased prevalence. Thus a natural question to study is the following: *can we prioritize vaccine distribution so as to significantly reduce the overall burden of COVID-19 quickly?*

We propose prioritization schemes based on properties of individuals within social contact networks with the goal of bending the pandemic curve and improving overall pandemic outcome. We synthesize a digital twin of Virginia, which is a detailed social contact network model for the Commonwealth of Virginia (8 million individuals), and use an agent-based model (ABM) to study the effectiveness of various prioritization schemes. In contrast to other such networks, our networks incorporate detailed information about the population, their activities and the built infrastructure. Further information on how such a digital twin is constructed and its structural properties can be found in Section A.1. The ABM simulates disease propagation and a complex set of interventions, including various non-pharmaceutical interventions and vaccine allocation schemes.

Our prioritization schemes based on simple, individual-based yet computable, structural properties of the underlying social contact network are motivated by: (i) recent advances in network science that have studied such schemes in more abstract settings; (ii) our ability to construct detailed, realistic social contact networks at scale; (iii) our ability to simulate and assess such strategies even for complex disease transmission models and public health control measures; and (iv) recent progress in development of digital apps that can be used for measuring structural properties in large populations relatively accurately, rendering such schemes potentially operationalizable.

Our prioritization schemes can be stated simply as follows: *vaccinate individuals who typically exhibit high social contact (degree or total contact time in the social contact network)*. Some key points to note: (i) we focus on simple network structural properties that can be estimated in a privacy-preserving way, (ii) we do not insist on strict ordering of individuals nor an exact estimation of their social contacts, and (iii) while our *analysis* uses a realistic representation of the social contact networks, implementation of the policy does not require one to *synthesize* the social network.

There is folklore that degree based heuristics to allocate vaccines often work well. The folklore is based on mathematical results for highly structured random networks or on computational experiments based on relatively simple class of social contact networks [10,21,46,56]. But the folklore has never been tested in time-varying realistic social contact networks such as the one constructed here and intended to capture the network evolution due to adaptive NPIs and vaccine allocation that is undertaken in a time varying manner. Our results show for the first time that degree based heuristics are likely to work even for such time-varying social contact networks; see Sections H and E for further discussions on this topic.

Our results suggest that in just two months (i.e. by the end of March 2021), compared to age-based allocation, the proposed degree-based strategy can result in averting an *additional 56–110k infections (8–16%), 3.2–5.4k hospitalizations (8–13%), and 700–900 deaths (6–8%) just in the state of Virginia. Extrapolating these results per capita for the entire US, we estimate this strategy will lead to 3–6 million fewer infections, 181–306k fewer hospitalizations, and 51–62k fewer deaths compared to the age-based allocation*. The results continue to hold qualitatively and show that we can avert many more infections, hospitalizations, and deaths even if the current social distancing measures are relaxed. Furthermore, similar results hold even for vaccines with 50% efficacy; this is important, as most resource-poor countries do not have access to high efficacy vaccines at this point in time. The basic intuition behind our results is that vaccinating individuals with high

degree not only protects them but also confers significant protection to individuals who come in close proximity in their contact network.

A natural question is: *how can such individuals be identified?* This might be done by objectively determining who they are and seeking them out. Alternatively, at a time of interview, a person could be designated as “high degree” through identifying data or proxy characteristics to necessary statistical precision that show the individual belongs to such a critical group identified by the model. We discuss how currently deployed digital contact tracing apps can be modified in a very simple manner to achieve the goal of identifying high degree individuals (Section E). Such individuals can also be identified by observing that certain occupations naturally lead to a high level of social interactions. Our methods are robust to partial mis-estimation of these social contacts and their implementation does *not* require access to the social contact network.

2 Experiment Settings and Design

For the experiments, we use an agent-based simulation model, EpiHiper, which is described in Appendix F and has been used in previous studies [18]. The simulation’s input parameters specify the population demographics and contact network, COVID-19 disease model, initial configuration S_0 , non-pharmaceutical interventions (NPIs), and vaccination schedule. The simulation output is a dendrogram: a directed graph that tells us who infects whom and on what day. From the output data, we can compute many epidemiological measures such as daily new infections, cumulative infections, prevalence in each age group, total hospitalizations, and deaths, as well as many other measures.

2.1 Simulation parameterization

These studies use a synthetic population and contact network for Virginia, which is described in Appendix A.1. The initial conditions are calibrated to the conditions in Virginia as of January 1, 2021. Every simulation is run for 90 days, until March 31, 2021. Since the simulations are stochastic, each simulation is repeated for 30 replicates, and distributions of the measures are computed. The boxplots and curves in figures presented in Section 3 (Figure 1 through 11) are all based on data from 30 replicates. The curves show an uncertainty of one standard deviation above and below the mean.

Disease model. The disease model is the *best guess version* of “COVID-19 Pandemic Planning Scenarios” prepared by the US Centers for Disease Control and Prevention (CDC) SARS-CoV-2 Modeling Team [16] and has been used by multiple researchers in their papers. It is an SEIR model where state transitions follow the parameters as defined in the document. The disease states and transition paths are shown in Figure 24. Individuals of different age groups have different infectivity and susceptibility; dwell time distributions and state transition probability distributions are stratified by the following age groups: preschool (0-4 years), students (5-17), adults (18-49), older adults (50-64) and seniors (65+). Furthermore, individuals that are vaccinated have different disease parameter values than those that are not vaccinated. Detailed parameterization for unvaccinated individuals is summarized in Appendix G.

Initializations. The simulations are initialized at the county level by age group using the detailed data of confirmed cases from [60]. The initialization specifies the health state of each individual. Based on county-level cumulative confirmed cases through December 19, 2020, we derive the number

of prior infections in each county by scaling the cumulative number by a case ascertainment ratio of 3 (i.e., only one third of all infections are reported), then computing the number of prior infections in each age group of this county using the age distribution in cases. We randomly choose individuals in each age group in each county and set their health states to `recovered` to reflect that they have already been infected. Based on county-level daily confirmed cases from December 20, 2020 to January 5, 2021, we derive the number of individuals that are infected each day by the same scaling, and seed the simulation by setting randomly chosen individuals to `exposed` by day in each age group of each county.

Non-pharmaceutical interventions. We consider four NPIs: (i) *Infectivity reduction (IR)*. Infectivity is universally reduced (by 60%) through preventive behavior, e.g., mask wearing and hand washing. (ii) *Generic social distancing (GSD)*. A fraction (25%) of the population chooses to reduce non-essential (shopping, religion, and other) activities. (iii) *Virtual learning (VL)*. A fraction (50%) of K-12 students choose virtual learning. (iv) *Voluntary home isolation of symptomatic cases (VHI)*. With probability 75%, a symptomatic person chooses to stay home for 14 days, reducing the weights on household contacts by 50%. For this person, all outside contacts are disabled and at-home contacts are reduced by 50% temporarily during these 14 days.

Scenarios based on relaxing social distancing measures. We assume that these NPIs are in place when a simulation starts, but adherence may change during the simulation. We consider three scenarios for adherence to the NPIs:

- **As-is.** NPI parameters remain the same for the duration of the simulation.
- **Slow relaxation.** NPI parameters change every 30 days from January 30, 2021, so that in 7 months, infectivity reduction decreases from 60% to 10%, generic social distancing decreases from 25% to 10%, and virtual learning decreases from 50% to 25%. Note that this is used to specify the speed of relaxation. Nevertheless the results are only reported for the period until end of March.
- **Fast relaxation.** NPI parameters change every 30 days from January 30, 2021, so that in 5 months, they reach the same levels as in the *slow relaxation* scenario.

2.2 Vaccination: supply, schedule and priority groups

Vaccine schedule. As of this writing, we expect 400 million doses to be delivered by Pfizer-BioNTech and Moderna to the US by the end of July 2021, enough to vaccinate 200 million people. By assuming that 25 million people can be vaccinated per month, starting from late December 2020 until late July 2021, and that vaccines are allocated to all states proportional to population size, we consider a vaccination schedule as shown in Table 1, where 650K people are vaccinated per month in Virginia, and a schedule where they are vaccinated at half this rate. Therefore we consider three vaccination schedules: *none* (no vaccination), *fast* (vaccinating 650K people per month), and *slow* (vaccinating 325K people per month). The later schedule is intended to capture the current challenges faced in distributing the vaccines to individuals. For simplicity, all individuals vaccinated during each month are assumed to be vaccinated on the first day of that month; spreading the vaccines over the month does not change the overall results by much.

Vaccine efficacy. Overall vaccine efficacy is characterized by three numbers: (i) e_I , efficacy against infection; (ii) e_D , efficacy against severe illness (requiring hospitalization or leading to death) given

Table 1: Cumulative number of individuals vaccinated in each month of 2021. Note that in our experiment, where simulations run until the end of March, we consider vaccinations up to March only.

vaccination up to	US (million)	Virginia (thousand)
Jan	25	650
Feb	50	1300
Mar	75	1950
Apr	100	2600
May	125	3250
Jun	150	3900
Jul	175	4550
Aug	200	5200

infection; and (iii) e_T , efficacy against onward transmission given infection. We assume that $e_I = 90\%$ and $e_D = 50\%$ starting only 21 days after vaccination. In our sensitivity analyses, we also consider $e_I = 50\%$. In all cases, we ignore e_T .

Vaccination prioritization. The Pfizer-BioNTech vaccine and the Moderna vaccine are recommended for people aged at least 16 years and at least 18 years, respectively. In the experiments, we only allocate vaccines to people who are at least 18 years old. Among those people, we consider the following prioritization strategies.

- *No priority.* Everyone 18+ years old is vaccinated with the same probability. This is our baseline strategy.
- *Essential workers.* This strategy targets those who work for medical, care facilitation, retail, education, military, and government.
- *Older people.* This strategy prioritizes those who are at least 50 years old.
- *High degree.* Degree of an individual is the number of contacts per day. This strategy targets those in the top quartile among all 18+ years old in terms of degree.
- *Long total contact (also denoted as weighted degree).* Weighted degree of an individual is the total contact time this individual has with other people in a day. This strategy targets those in the top quartile among all 18+ years old in terms of weighted degree.

Most vaccines are allocated to the targeted groups, but we allow some to be given to other groups. This accounts for potential inaccuracy and precision in identifying and locating the targeted people. For example, since we do not know people's daily number of contacts, which may vary, we can only estimate it using proxy attributes, such as age, household size and occupation, or from data collected through digital devices. We consider the following rates of enforcement: 100%, 80%, and 60%.

2.3 Experimental design

The design consists of 4 factors: (i) 3 adherence scenarios (as-is, slow relaxation, fast relaxation); (ii) 3 vaccination schedules (none, fast, slow); (iii) 5 prioritization targets (no priority, essential

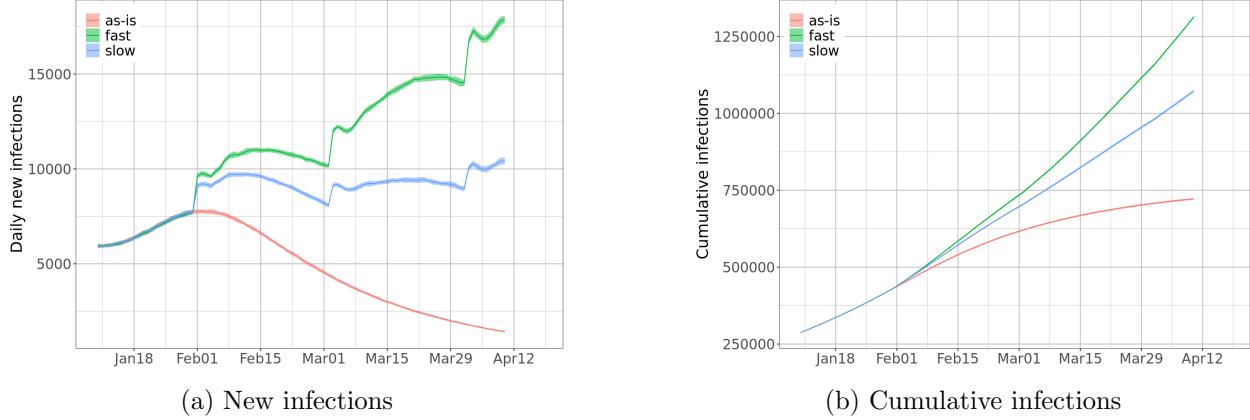


Figure 1: Incidence without vaccination under different NPI relaxation scenarios: (a) daily number of new infections; (b) cumulative number of infections. The sudden surge in new infections in the beginning of each month is caused by NPI relaxation. Without NPI relaxation, the incidence reaches peak in early February then starts decreasing. With slow relaxation, daily incidence increases then fluctuates. With fast relaxation, the incidence keeps rising.

workers, older people, high degree, high weighted degree); and (iv) 3 levels of priority enforcement (100%, 80%, 60%). Combining (iii) and (iv) we have the baseline (no-priority) plus 12 prioritized strategies named according to the target group (essential, old-age, high degree, high weighted degree) and the fraction of vaccine given to the target group (100%, 80%, 60%), e.g., “essential 100%” or “high degree 60%”. We also consider vaccines with a 50% efficacy against infection (e_I) and compare the effectiveness of degree-based vaccinations under this assumption against that under 90% efficacy.

3 Results and Analysis

In Figure 1, we show daily new infections under three scenarios (*as-is*, *slow relaxation* and *fast relaxation*) *without vaccination*. If NPI adherence can be maintained, then we expect infections to decrease after January. With slow relaxation, the infections will fluctuate around a level that will be a little higher than the current level. With fast relaxation, the infections show a steady increase in the next three months. The sharp increase every 30 days is caused by the implementation of the relaxation of NPIs and does not have any influence on the results presented.

3.1 Effectiveness of degree- and weighted degree-based strategies

Prioritizing vaccinations based on individual degree and weighted degree are extremely effective in controlling the pandemic. In particular, depending on the scenario, the reductions in the number of infections and hospitalizations by these schemes are over 50% more than the reductions from the age-based prioritization schemes. For example, assuming that the current non-pharmaceutical interventions remain at the same level over the next few months, our experiment shows that by the end of March 2021, degree-based schemes can result in 56–110k fewer infections, 3.2–5.4k fewer hospitalizations, and 700–900 fewer deaths in the state of Virginia, compared to age-based schemes. Note that the ranges come from different levels of priority enforcement (three levels for both age-

based and degree-based schemes). Figure 2 shows the estimated reductions by one of the age-based schemes and the further reductions by one of the degree-based schemes. Extrapolating these results for the entire US, we estimate that degree-based schemes will lead to 3–6 million fewer infections, 181–306k fewer hospitalizations, and 51–62k fewer deaths by the end of March, compared to age-based schemes. If the NPIs are relaxed, the reductions in infections, hospitalizations, and mortality are even more substantial. This implies that when conditions worsen, the marginal gains from a more effective strategy are even higher.

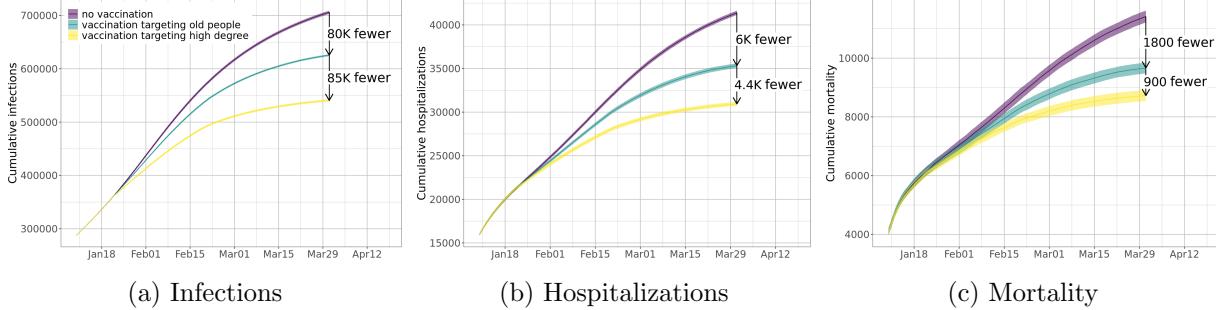


Figure 2: Vaccination targeting old people can reduce (a) total infections, (b) total hospitalizations, and (c) total mortality significantly, assuming current non-pharmaceutical interventions remain at the same level. Vaccination targeting high degree people can further reduce total infections, hospitalizations, and mortality. Numbers in the plots show total reductions up to the end of March 2021. Note that the “no vaccination” curves in (a) is the same as the “as-is” curve in Figure 1b.

Figure 3 compares incidence reduction up to March 31, 2021, under different prioritization strategies for the vaccine distribution schedule given in Table 1, also known as the *fast* schedule. We find that all strategies targeting either essential workers or high degree people outperform the no-priority distribution. The degree-based strategies reduce incidence more than any other strategy. For example, with no NPI relaxation (as-is), all degree-based strategies can reduce infections by over 20% while all other strategies can reduce infections by at most 20%. Strategies targeting older people perform worse than the no-priority distribution in terms of reducing incidence. Similar results are obtained for the *slow* vaccine distribution schedule, as shown in Figure 4. One reason to consider weighted degree-based heuristics is that they are potentially easier to implement in the current digital apps, we will discuss this further in later sections. **All degree-based strategies outperform the other strategies.**

Targeting high degree people is also the most effective strategy for reducing mortality. Prioritization of older people is effective in reducing mortality compared to other strategies, but not when compared to a high degree strategy. This is shown in Figures 5 and 6. Figure 7 shows that prioritizing people with high weighted degree (total contact durations) is even more effective than prioritizing those with high degree. For example, Figure 7a shows that, with no relaxation of NPIs, targeting people of high weighted degree can reduce infections by about 23–30%, compared to targeting high degree people, which can reduce infections by about 21–26%. In the case where NPIs are relaxed, the strategy prioritizing high weighted degree can cause over 40% reduction in infections if it can be implemented with high precision.

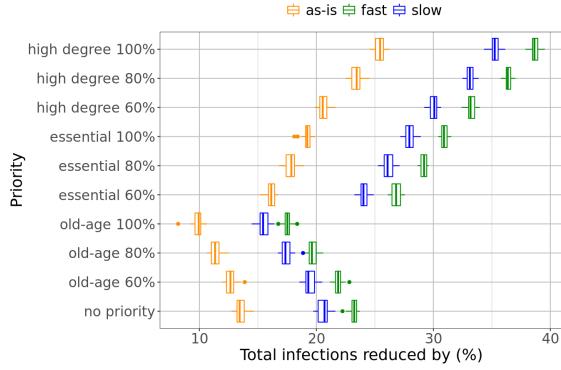


Figure 3: Total reduction in incidence under the fast vaccine distribution schedule. Degree-based strategies outperform all other ones, while age-based strategies are outperformed by all other ones.

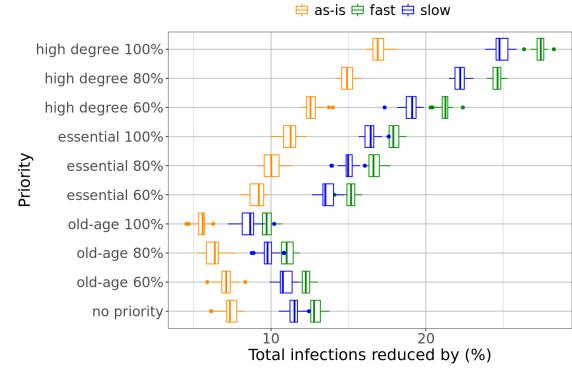


Figure 4: Total reduction in incidence under the slow vaccine distribution schedule. Degree-based strategies still reduce more infections than other strategies and are more effective if accuracy is higher.

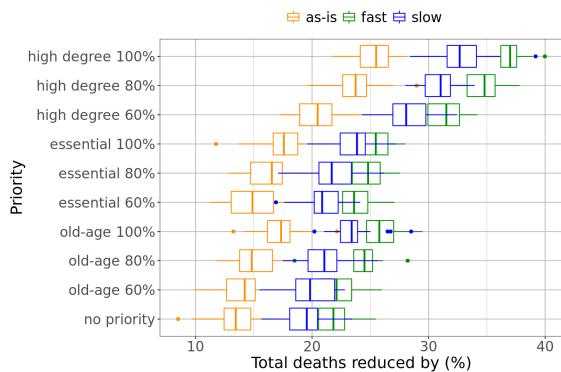


Figure 5: Total reduction in mortality under the fast vaccine distribution schedule. While degree-based strategies continue to perform better than other ones in reducing mortality, age-based strategies seem to be more effective than the baseline.

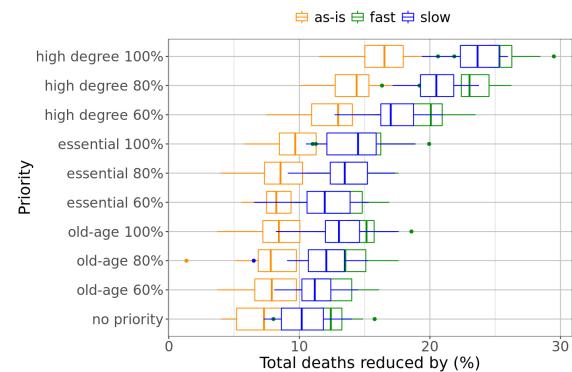


Figure 6: Total reduction in mortality under the slow vaccine distribution schedule. Observations in the fast vaccine distribution case remain true, except that the reductions now have smaller sizes.

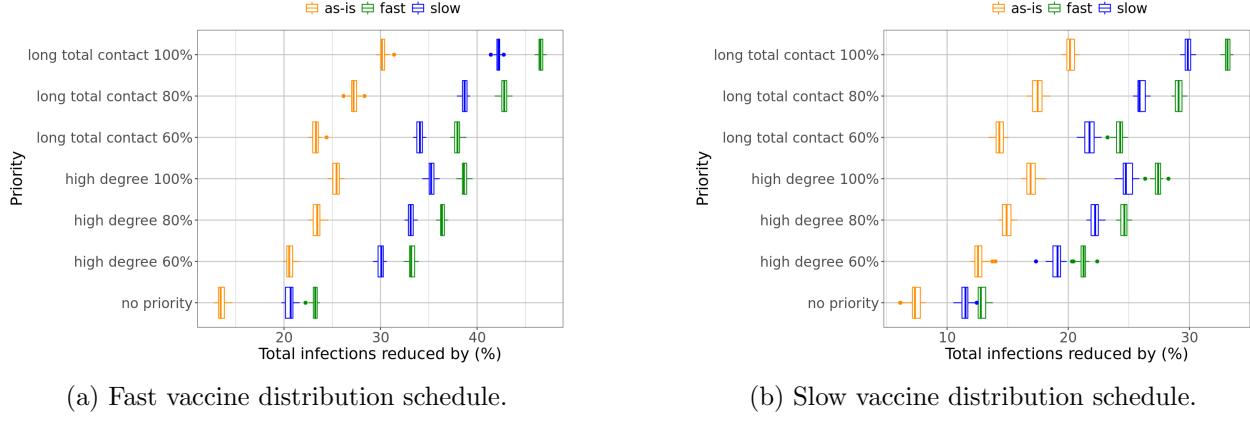


Figure 7: Comparison of degree and weighted degree-based strategies under (a) the fast vaccine distribution schedule; (b) the slow vaccine distribution schedule. Both can reduce infections much more than the baseline strategy. The weighted degree-based strategy outperforms the degree-based one at any prioritization level.

3.2 The high degree prioritization schemes are effective even when we cannot accurately estimate the degree of a node

Our results show that prioritization schemes based on degree and weighted degree (total contact time) work even when they are not accurately estimated. Specifically, even when we can only estimate the degree for 60% of the nodes (as being in the first quartile or not), we notice significant improvement in the overall control of the pandemic. This is highlighted in Figure 8, where we compare degree-based schemes of various accuracies with the age-based scheme and show improvement even at lower levels of accuracy. For example, consider infection reduction in Figure 8a: targeting high degree people with only 60% accuracy improves the reduction from 10% by the age-based strategy to 20% (with no relaxation), from 15% to 30% (with slow relaxation), or from 17.5% to 33% (with fast relaxation). Consider mortality reduction in Figure 8b: degree-based strategy with 60% accuracy improves the reduction from 17.5% to 20% (with no relaxation), from 23% to 27.5% (with slow relaxation), or from 26% to 32% (with fast relaxation). In fact, these strategies require neither knowledge of the exact degree of each person, nor that of the complete ranking of people by degree. They only depend on knowing which nodes have high degrees (are in the top quartile); they are tolerant to a certain amount of inaccuracy.

3.3 Effectiveness when social distancing measures are relaxed

The effectiveness of degree-based strategies holds in three hypothetical scenarios for social distancing: one in which there is no relaxation, and the other two wherein social distancing is progressively relaxed 5 or 7 months from now. Our results show that the value of these prioritization schemes is even higher when social distancing measures are relaxed quickly. Recall in Figure 2 we observe that, with no relaxation, the degree-based strategy results in another reduction of 85K infections and an additional reduction of 900 mortality, compared to the age-based strategy. In Figure 9, we find that with relaxation of NPIs, the degree-based strategy can reduce even more infections (152K with slow relaxation and 192K with fast relaxation) and more mortality (1.3K with slow relaxation and 1.5K with fast relaxation). These observations highlight the importance of vaccination prioritization if the current NPIs are relaxed, which will likely happen as vaccines get distributed.

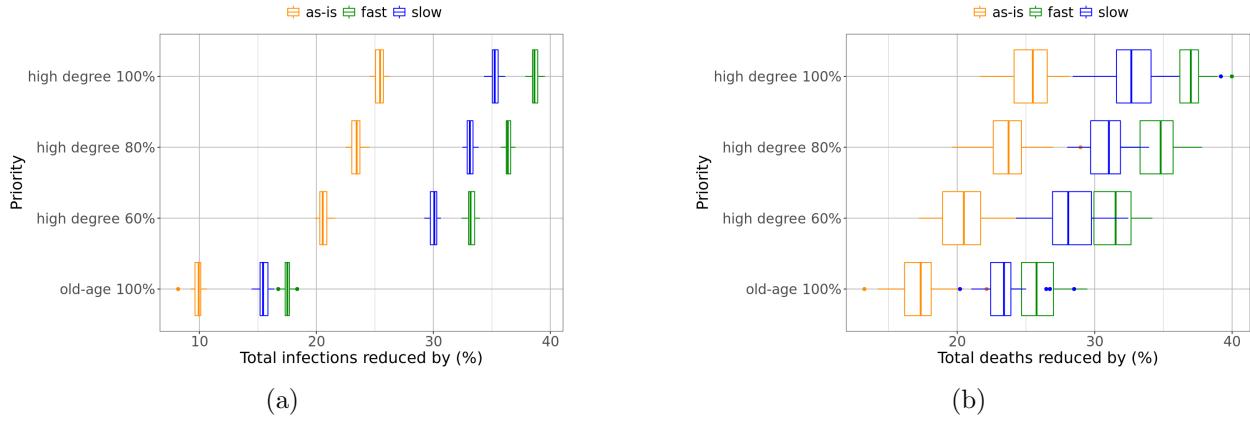


Figure 8: Even with lower (80% or 60%) accuracy in identifying and vaccinating high degree people, this strategy is still much more effective than the age-based strategy in (a) reducing infections, as well as (b) reducing mortality.

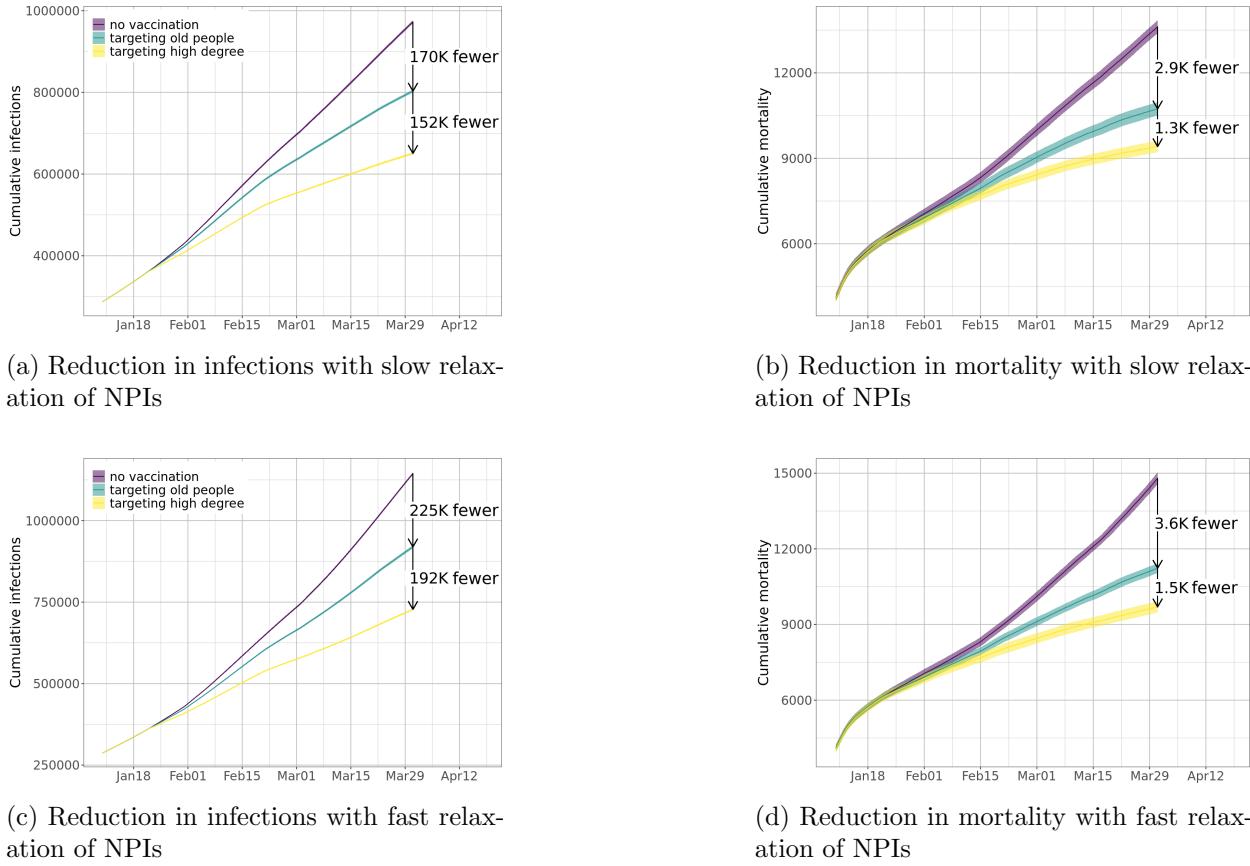


Figure 9: Reductions in infections and mortality from degree-based allocation strategies are even larger when NPIs are relaxed when compared to the age-based schemes. This figure shows reductions in (a) infections and (b) mortality with slow relaxation of NPIs; and reductions in (c) infections and (d) mortality with fast relaxation of NPIs.

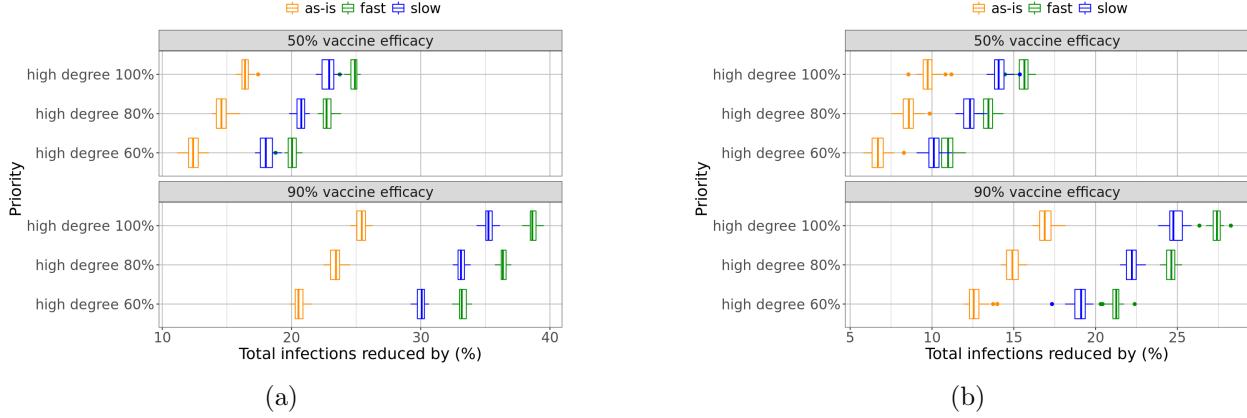


Figure 10: Comparison of effectiveness of degree-based strategies when vaccine efficacy is high (90%) and when it is low (50%), assuming (a) a fast distribution of vaccine schedule, or (b) a slow distribution of vaccine schedule. In both cases the effectiveness of vaccination becomes smaller with a lower vaccine efficacy, but the degree-based vaccination can still reduce infections significantly.

3.4 Effectiveness with low efficacy vaccines

We have assumed that vaccines have 90% efficacy regarding protection against infection (e_I). Our results also hold when the vaccine efficacy is lower than that of the current Pfizer and Moderna vaccines. We study this for two reasons: (i) there is an ongoing discussion about giving just one dose of these vaccines which may result in lower efficacy (about 50%) or approving a low efficacy vaccine², and (ii) most other vaccines under development are traditional vaccines and may also have a lower efficacy.

To this end we study the degree-based strategies assuming 50% vaccine efficacy. In Figure 10, we show that while the reduction in infections decreases with low efficacy vaccines, it is still significant. For example, under no NPI relaxation and with the fast vaccine distribution schedule, a degree-based strategy with 60% accuracy can reduce infections by about 12.5% with $e_I = 50\%$, compared to 20% with $e_I = 90\%$. Under slow relaxation and with the slow vaccine distribution schedule, the reduction in infections is about 10% with $e_I = 50\%$, compared to 18% with $e_I = 90\%$. In Figure 11, we use epidemic curves to show reductions from the no vaccination scenario by vaccinating high degree people with 80% accuracy (*high degree 80%* in Figure 10) with both high vaccine efficacy and low efficacy, assuming slow relaxation of NPIs. We find that even with 50% efficacy, the degree-based strategy can reduce infections by 202K, hospitalizations by 13.7K, and mortality by 3.4K, by the end of March 2021, just in Virginia.

4 Discussion

These results are obtained using a realistic, data-driven and highly resolved agent-based model and individual-based social contact network of the Commonwealth of Virginia. The agent-based model represents individual-level activities that are spatially explicit. The model represents the Commonwealth-built infrastructure in great detail and uses this to develop a realistic social contact

²<https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-takes-action-help-facilitate-timely-development-safe-effective-covid>

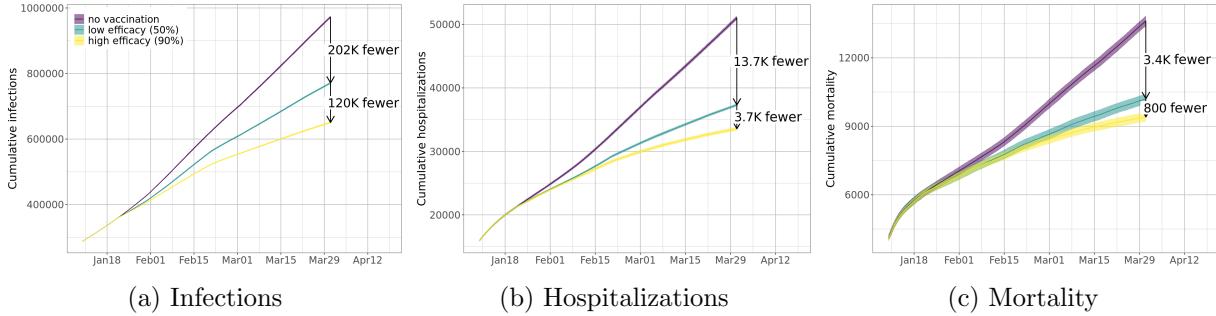


Figure 11: Even with a lower vaccine efficacy, a degree-based scheme can already significantly reduce (a) total infections, (b) total hospitalizations, and (c) total mortality. Numbers in the plots show total reductions up to the end of March 2021.

network. This allows us to: (i) capture details of within-host disease progression, as well as between-host transmission, including the impact of vaccines, (ii) model the complicated set of interventions that are currently in play, (iii) represent network-based vaccine prioritization schemes, (iv) represent the expected vaccine deployment schedule, including the expected mix of vaccine efficacy against infection, severe illness, and onward transmission estimates, (v) incorporate current surveillance data, and (vi) study counter-factual and hypothetical scenarios, such as a steady relaxation of social distancing measures. *This is the first study we know of that accounts for all of these components, not just for COVID-19, but for any infectious disease outbreak.*

The efficacy of the proposed policy is based on two important assumptions: (i) the synthetic contact network is a realistic representative of the real-world social contact world, and (ii) NPI-induced contact ‘thinning’ is applied homogeneously across the population. While the structural metrics may vary over time, we show the results are fairly robust to mis-identification of high degree individuals. We believe both these assumptions hold and discuss this in more detail below. Further discussion on this topic can be found in the Appendix, where we describe how our networks are synthesized, their structural properties, and the way the pandemic is simulated.

The potential efficacy of degree-based heuristics has been discussed in several earlier papers—this includes both provable analyses on different random graph models (under mean field assumptions in some cases), e.g., [3, 10, 51], and empirical analysis in various real world networks, e.g., [3, 22, 73]; a notion of weighted degree is also considered in [22]. However, it is important to note that these results are *not directly applicable* in our context for the following reasons: (i) many of the theoretical results show the efficacy of these methods for simple power law-type models—the networks we generate are similar to power law networks, but with a very different exponent; additionally the network exhibits other features of social networks (local clustering, low diameter, and relatively high expansion) and (ii) many of the results are shown when vaccines are applied at the start of the epidemic process, and the results do not say anything of what happens when the vaccine is applied temporally—this is important, because the temporal epidemic process infects individuals, thereby changing the network structure substantially, including the application of NPIs.

Nevertheless, the intuition behind the efficacy of such methods is simply stated as follows: vaccinating high degree nodes not only protects them, but also confers a higher level of indirect protection on their neighbors as they interact with many individuals who might themselves be conferred similar protection. Our data-driven approach shows, in fact, that real-world social net-

works have sufficient nodes of high degree to ensure that such heuristics are effective. Note that, by virtue of degree bias in social networks, even traditional approaches such as contact tracing will lead us to high degree individuals. The proposed approach makes identifying these individuals as a proactive, rather than reactive, step in infection control. It is important to note, however, that just the presence of high degree nodes does not guarantee that degree-based heuristics would work. See Section E for further discussion.

As discussed earlier, identification of nodes with high degrees can be done in multiple ways, including using digital apps that have been deployed for contact tracing, interviewing individuals, and identifying typical job categories or other demographic attributes that entail higher social interactions. Further, even when other prioritization schemes are considered, one can use high social contact to further prioritize the distribution. For example, when distributing vaccines based on age, one can further subselect individuals with higher social contact in the case of limited supply.

Our results suggest that degree-based prioritization should be considered by larger and resource-poor countries to quickly bend the epidemic curve and reopen the economy. The benefits of the proposed degree-based prioritization are so significant that even a partially successful campaign will likely have a large impact.

5 Conclusions and Limitations

We present an analysis of various vaccine prioritization strategies based on demographic attributes, occupation, and structural attributes of social contact networks. Our results show that vaccine prioritization schemes based on network degrees and total contact time can provide significant reductions in incidence, mortality, and hospitalizations. The results hold even for low efficacy vaccines and even when degrees and contact networks are estimated only approximately. Network-based prioritization is often more than twice as effective as other strategies. The results suggest that such methods should be considered when vaccines are available in limited supply; the benefits are likely to be greater in resource-poor and highly populated regions of the world. While individualized policies aimed at minimizing mortality do exist (e.g., comorbidities) and are part of the phased approach, lack of technology thus far had made it difficult to ‘individualize’ policies targeted on minimizing transmission. The advantage of our approach is in leveraging the mechanistic and network-based understanding of disease spread, and creating priority categories that cut across age, risk, and other demographic characteristics.

The study has a number of limitations, stated below. First, our network has been developed with a large number of data sources, and a number of assumptions have been made in constructing the networks, including travel patterns, distance traveled, etc. These modeling assumptions might affect the efficacy of the network-based strategies. To mitigate this, we have carried out extensive validation and assessed the impact of the uncertainty in some of the modeling parameters on the network structure. Our results indicate that the network structure is fairly robust. Second, the nodes initially infected were based on spatio-temporal and age distributions in publicly available data, but not on any network properties. If the majority of high degree nodes have already been infected and recovered, the effectiveness of targeting them for vaccination would be reduced. However, since the current vaccination policies are not based on serostatus, preferentially vaccinating higher degree individuals will still be beneficial. Furthermore, even after accounting for testing rates, a number of resource poor countries with large populations have so far had a relatively small outbreak. This means without vaccinations, these countries are likely to see a surge when normal

worldwide travel and economic activity is resumed. The strategy is also potentially advantageous in the presence of novel variants, which may escape natural immunity. See [4,55] for further discussion on this issue. In particular in [55], the authors point out that high degree nodes could have been infected early on in the pandemic but can pose challenges if they re-enter the pool due to waning immunity or lower immunity to new strains. This makes identifying and vaccinating high degree nodes important, even if they have been infected earlier.

Third, our base scenario has made assumptions regarding the background interventions in place. These are best estimates. Fourth, assuming that a vaccinated node gets infected, we assume that they can transmit like any other node (of course, they have a very small chance of being infected). Fifth, our results depend on estimating the degrees and weighted degrees of nodes. While we have shown that the results are robust to mis-estimation, the overall efficacy of the scheme does depend on the ability to infer these degrees.

Increasing compliance among some high degree individuals may be difficult; nevertheless, the results under such conditions will be more similar to one where the vaccine has lower efficacy and/or under degree mis-estimation. Further, when such high degree individuals are identified and vaccinated, they may themselves turn into influencers in their local community, much like the phenomenon observed on online social networks. This is a topic for immediately subsequent work. Ultimately, we believe that one can develop more comprehensive prioritization strategies that combine proposed metrics with serostatus, hesitancy surveys, and other static demographic variables to optimally reduce disease incidence and mortality.

Acknowledgments

The authors would like to thank members of the Biocomplexity COVID-19 Response Team and the Network Systems Science and Advanced Computing (NSSAC) Division for their thoughtful comments and suggestions related to epidemic modeling and response support. We thank members of the Biocomplexity Institute and Initiative, University of Virginia, Paul Eastham, Jeff Shaman, Chadi Saad-Roy, Theo Gibbs and Adam Sadilek for useful discussion and suggestions. This work was partially supported by National Institutes of Health (NIH) Grant 1R01GM109718, NSF BIG DATA Grant IIS-1633028, NSF Grant No.: OAC-1916805, NSF Expeditions in Computing Grant CCF-1918656, CCF-1917819, NSF RAPID CNS-2028004, NSF RAPID OAC-2027541, US Centers for Disease Control and Prevention 75D30119C05935, University of Virginia Strategic Investment Fund award number SIF160, Google Grant, and Defense Threat Reduction Agency (DTRA) under Contract No. HDTRA1-19-D-0007. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

References

- [1] B. Abbasi, M. Fadaki, O. Kokshagina, N. Saeed, and P. Chhetri. Modeling vaccine allocations in the covid-19 pandemic: A case study in australia. *Available at SSRN 3744520*, 2020.
- [2] N. Ahmed, R. A. Michelin, W. Xue, S. Ruj, R. Malaney, S. S. Kanhere, A. Seneviratne, W. Hu, H. Janicke, and S. K. Jha. A survey of covid-19 contact tracing apps. *IEEE Access*, 8:134577–134601, 2020.
- [3] R. Albert and A.-L. Barabasi. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.
- [4] R. M. Anderson, C. Vegvari, J. Truscott, and B. S. Collyer. Challenges in creating herd immunity to sars-cov-2 infection by mass vaccination. *The Lancet*, 396(10263):1614–1616, 2020.
- [5] A. Anglemyer, T. H. Moore, L. Parker, T. Chambers, A. Grady, K. Chiu, M. Parry, M. Wilczynska, E. Flemyng, and L. Bero. Digital contact tracing technologies in epidemics: a rapid review. *Cochrane Database of Systematic Reviews*, 8, 2020.
- [6] J. Aspnes, K. Chang, and A. Yampolskiy. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. *J. Comput. Syst. Sci.*, 2006.
- [7] C. L. Barrett, R. J. Beckman, M. Khan, V. A. Kumar, M. V. Marathe, P. E. Stretz, T. Dutta, and B. Lewis. Generation and analysis of large synthetic social contact networks. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pages 1003–1014. IEEE, 2009.
- [8] R. J. Beckman, K. A. Baggerly, and M. D. McKay. Creating synthetic baseline populations. *Transportation Research Part A: Policy and Practice*, 30(6):415–429, 1996.
- [9] D. Bertsimas, J. K. Ivanhoe, A. Jacquillat, M. L. Li, A. Previero, O. S. Lami, and H. T. Bouardi. Optimizing vaccine allocation to combat the covid-19 pandemic. *medRxiv*, 2020.
- [10] B. Bollobás and O. Riordan. Robustness and vulnerability of scale-free random graphs. *Internet Mathematics*, 2004.
- [11] L. Breiman. *Classification and regression trees*. Wadsworth statistics/probability series. Wadsworth International Group, 1984.
- [12] K. M. Bubar, K. Reinholt, S. M. Kissler, M. Lipsitch, S. Cobey, Y. H. Grad, and D. B. Larremore. Model-informed covid-19 vaccine prioritization strategies by age and serostatus. *Science*, 2021.
- [13] J. H. Buckner, G. Chowell, and M. R. Springborn. Optimal dynamic prioritization of scarce covid-19 vaccines. *medRxiv*, 2020.
- [14] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani. Dynamics of person-to-person interactions from distributed rfid sensor networks. *PloS one*, 5(7):e11596, 2010.

- [15] CDC. Planning parameters for COVID-19 outbreak scenarios. Circulated in COVID-19 Modeling working groups, March 31, 2020.
- [16] Centers for Disease Control and Prevention. Covid-19 pandemic planning scenarios. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios-h.pdf>, 2020. [Online, accessed January 15, 2021].
- [17] J. Chan, S. Gollakota, E. Horvitz, J. Jaeger, S. Kakade, T. Kohno, J. Langford, J. Larson, S. Singanamalla, J. Sunshine, et al. Pact: Privacy sensitive protocols and mechanisms for mobile contact tracing. *arXiv preprint arXiv:2004.03544*, 2020.
- [18] J. Chen, A. Vullikanti, S. Hoops, H. Mortveit, B. Lewis, S. Venkatramanan, W. You, S. Eubank, M. Marathe, C. Barrett, and A. Marathe. Medical costs of keeping the us economy open during covid-19. *Scientific reports*, 10(1):1–10, 2020.
- [19] R. Cohen, S. Havlin, and D. Ben-Avraham. Efficient immunization strategies for computer networks and populations. *Physical review letters*, 91(24):247901, 2003.
- [20] W. E. Deming and F. F. Stephan. On a least squares adjustment of a sampled frequency table when the expected marginal tables are known. *Annals Math. Stats*, 11(4):427–444, 1940.
- [21] M. Draief, A. Ganesh, and L. Massoulié. Thresholds for virus spread on networks. In *Proceedings of the 1st international conference on Performance evaluation methodologies and tools*, pages 51–es, 2006.
- [22] K. T. Eames, J. M. Read, and W. J. Edmunds. Epidemic prediction and control in weighted networks. *Epidemics*, 1(1):70 – 76, 2009.
- [23] S. Eubank, H. Guclu, V. S. Anil Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429:180–184, 2004.
- [24] S. Eubank, V. S. Anil Kumar, M. V. Marathe, A. Srinivasan, and N. Wang. Structure of Social Contact Networks and Their Impact on Epidemics. In *Discrete Methods in Epidemiology*, volume 70, pages 179–200. American Math. Soc., Providence, RI, 2006.
- [25] L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bon-sall, and C. Fraser. Quantifying sars-cov-2 transmission suggests epidemic control with digital contact tracing. *Science*, 2020.
- [26] B. H. Foy, B. Wahl, K. Mehta, A. Shet, G. I. Menon, and C. Britto. Comparing covid-19 vaccine allocation strategies in india: a mathematical modelling study. *International Journal of Infectious Diseases*, 2020.
- [27] A. Ganesh, L. Massoulié, and D. Towsley. The effect of network topology on the spread of epidemics. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, volume 2, pages 1455–1466. IEEE, 2005.
- [28] M. Génoin and A. Barrat. Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Science*, 7(1):11, 2018.

- [29] T. Germann, K. Kadau, I. Longini Jr, and C. Macken. Mitigation strategies for pandemic influenza in the united states. *Proceedings of the National Academy of Sciences*, 103(15):5935–5940, 2006.
- [30] D. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, 22:403–434, 1976.
- [31] M. E. Halloran, N. M. Ferguson, S. Eubank, I. M. Longini, D. A. T. Cummings, B. Lewis, S. Xu, C. Fraser, A. Vullikanti, T. C. Germann, D. Wagener, R. Beckman, K. Kadau, C. Barrett, C. A. Macken, D. S. Burke, and P. Cooley. Modeling targeted layered containment of an influenza pandemic in the United States. In *Proceedings of the National Academy of Sciences (PNAS)*, pages 4639–4644, March 10 2008.
- [32] A. Hayrapetyan, D. Kempe, M. Pál, and Z. Svitkina. Unbalanced graph cuts. In *ESA*, pages 191–202, 2005.
- [33] HERE, 2020. <http://www.here.com>, Accessed April 2020.
- [34] A. Hogan, P. Winskill, O. Watson, P. Walker, C. Whittaker, M. Baguelin, D. Haw, A. Lochen, K. Gaythorpe, K. Ainslie, et al. Report 33: Modelling the allocation and impact of a covid-19 vaccine. 2020.
- [35] M. E. Kretzschmar, G. Rozhnova, M. C. J. Bootsma, M. van Boven, J. H. H. M. van de Wijgert, and M. J. M. Bonten. Impact of delays on effectiveness of contact tracing strategies for covid-19: a modelling study. *The Lancet Public Health*, 5(8):e452 – e459, 2020.
- [36] V. S. A. Kumar, R. Rajaraman, Z. Sun, and R. Sundaram. Existence theorems and approximation algorithms for generalized network security games. In *Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference on*, pages 348–357. IEEE, 2010.
- [37] M. Lipsitch and N. E. Dean. Understanding covid-19 vaccine efficacy. *Science*, 370(6518):763–765, 2020.
- [38] E. Lofgren, M. E. Halloran, C. M. Rivers, J. M. Drake, T. C. Porco, B. Lewis, W. Yang, A. Vespignani, J. Shaman, J. N. S. Eisenberg, M. C. Eisenberg, M. Marathe, S. V. Scarpino, K. A. Alexander, R. Meza, M. J. Ferrari, J. M. Hyman, L. A. Meyers, and S. Eubank. Opinion: Mathematical models: A key tool for outbreak response. *PNAS*, pages 18095–18096, 2014.
- [39] K. Lum, Y. Chungbaek, S. Eubank, and M. Marathe. A two-stage, fitted values approach to activity matching. *International Journal of Transportation*, 4:41–56, 2016.
- [40] M. Marathe and A. Vullikanti. Computational epidemiology. *Communications of the ACM*, 56(7):88–96, 2013.
- [41] L. Matrajt, J. Eaton, T. Leung, and E. R. Brown. Vaccine optimization for covid-19: who to vaccinate first? *medRxiv*, 2020.
- [42] R. M. May and R. M. Anderson. Spatial heterogeneity and the design of immunization programs. *Mathematical Biosciences*, 72(1):83–111, 1984.

- [43] J. Medlock and A. P. Galvani. Optimizing influenza vaccine distribution. *Science*, 325(5948):1705–1708, 2009.
- [44] Microsoft. U.S. building footprints. <https://github.com/Microsoft/USBuildingFootprints>, 2020.
- [45] P. V. Mieghem, J. S. Omic, and R. E. Kooij. Virus spread in networks. *IEEE/ACM Transaction on Networking*, 2009.
- [46] P. V. Mieghem, D. Stevanovic, F. F. Kuipers, C. Li, R. van de Bovenkamp, D. Liu, and H. Wang. Decreasing the spectral radius of a graph by link removals. *IEEE Transactions on Networking*, 2011.
- [47] H. S. Mortveit, A. Adiga, C. L. Barrett, J. Chen, Y. Chungbaek, S. Eubank, C. J. Kuhlman, B. Lewis, S. Swarup, S. Venkatramanan, A. Wilson, D. Xie, and M. V. Marathe. Synthetic populations and interaction networks for the U.S. Technical report, NSSAC, University of Virginia, 2020. NSSAC Technical Report: #2019-025.
- [48] M. National Academies of Sciences, Engineering et al. Framework for equitable allocation of covid-19 vaccine. 2020.
- [49] T. National Center for Education Statistics (NCES). Last accessed: February 2020.
- [50] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015.
- [51] R. Pastor-Satorras and A. Vespignani. Immunization of complex networks. *Physical Review E*, 65(036104), 2002.
- [52] B. A. Prakash, D. Chakrabarti, N. C. Valler, M. Faloutsos, and C. Faloutsos. Threshold conditions for arbitrary cascade models on arbitrary networks. *Knowledge and information systems*, 33(3):549–575, 2012.
- [53] V. M. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, and G. J. Pappas. Optimal vaccine allocation to control epidemic outbreaks in arbitrary networks. In *IEEE Conference on Decision and Control*. IEEE, 2013.
- [54] V. M. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, and G. J. Pappas. Optimal resource allocation for network protection against spreading processes. *IEEE Transactions on Control of Network Systems*, 1(1):99–108, 2014.
- [55] C. M. Saad-Roy, C. E. Wagner, R. E. Baker, S. E. Morris, J. Farrar, A. L. Graham, S. A. Levin, M. J. Mina, C. J. E. Metcalf, and B. T. Grenfell. Immune life history, vaccination, and the dynamics of sars-cov-2 over the next 5 years. *Science*, 370(6518):811–818, 2020.
- [56] S. Saha, A. Adiga, B. A. Prakash, and A. K. S. Vullikanti. Approximation algorithms for reducing the spectral radius to control epidemic spread. In *Siam Data Mining (SDM)*, 2015.
- [57] S. Saha, A. Adiga, B. A. Prakash, and A. K. S. Vullikanti. Approximation algorithms for reducing the spectral radius to control epidemic spread. In *Proceedings of the 2015 SIAM International Conference on Data Mining*, pages 568–576. SIAM, 2015.

- [58] P. Sambaturu, B. Adhikari, B. A. Prakash, S. Venkatramanan, and A. Vullikanti. Designing effective and practical interventions to contain epidemics. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1187–1195, 2020.
- [59] G. Tennenholz, C. Caramanis, and S. Mannor. Sequential vaccination for containing epidemics. *medRxiv*, 2020.
- [60] The New York Times. Coronavirus (covid-19) data in the United States. <https://github.com/nytimes/covid-19-data>, last accessed on January 7, 2021, 2020.
- [61] The University of Oxford. The Multinational Time Use Study (MTUS). Last accessed: February 2020.
- [62] H. Tong, B. A. Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos. Gelling, and melting, large graphs by edge manipulation. In *Proc. of CIKM*, 2012.
- [63] United States Censuc Bureau. 2011-2015 5-year ACS commuting flows. Last accessed: April 2020.
- [64] United States Census Bureau. American Community Survey 2013-2017 5-year estimates. Last accessed: February 2020.
- [65] United States Department of Labor, Bureau of Labor Statistics. The American Time Use Survey (ATUS). Last accessed: February 2020.
- [66] U.S. Department of Transportation, Federal Highway Administration. The National Household Travel Survey (NHTS). Last accessed: February 2020.
- [67] P. Van Mieghem, D. Stevanović, F. Kuipers, C. Li, R. Van De Bovenkamp, D. Liu, and H. Wang. Decreasing the spectral radius of a graph by link removals. *Physical Review E*, 84(1):016101, 2011.
- [68] S. Venkatramanan, J. Chen, A. Fadikar, S. Gupta, D. Higdon, B. Lewis, M. Marathe, H. Mortveit, and A. Vullikanti. Optimizing spatial allocation of seasonal influenza vaccine under temporal constraints. *PLoS computational biology*, 15(9):e1007111, 2019.
- [69] S. Venkatramanan, J. Chen, S. Gupta, B. Lewis, M. Marathe, H. Mortveit, and A. Vullikanti. Spatio-temporal optimization of seasonal vaccination using a metapopulation model of influenza. In *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 134–143, Aug 2017.
- [70] Wikipedia. Clustering coefficient, July 2020.
- [71] Wikipedia. Degeneracy, July 2020.
- [72] B. Wilder, S.-C. Suen, and M. Tambe. Preventing infectious disease in dynamic populations under uncertainty. In *AAAI Conference on Artificial Intelligence*, 2018.
- [73] Y. Yang, A. McKhann, S. Chen, G. Harling, and J.-P. Onnela. Efficient vaccination strategies for epidemic control using network information. *Epidemics*, 27:115 – 122, 2019.

- [74] Y. Zhang, A. Adiga, S. Saha, A. Vullikanti, and B. A. Prakash. Near-optimal algorithms for controlling propagation at group scale on networks. *IEEE Transactions on Knowledge and Data Engineering*, 28(12):3339–3352, 2016.

A Models and Methods

We study vaccine allocation strategies using agent-based simulations, which compute COVID-19 disease spread in a population (e.g. Virginia) through a social contact network. In this section, we describe a synthetic contact network of Virginia and the agent-based simulation model of COVID-19. Then we describe vaccine allocations based on different prioritizations, especially strategies targeting high degree people in the population. The overall framework is described in Figure 13.

A.1 Generating synthetic populations and networks

A synthetic population of a region may be regarded as a digital twin of the real population of that region. Here we provide a compact summary of the model and the methodology behind constructing synthetic populations and their contact networks in the case of the US; see [47] for details. Our work builds on earlier techniques for a first principles approach for constructing synthetic populations [7, 23, 24]. These populations and networks are central to the EpiHiper simulation model.

To construct a population for a *geographic region R* (e.g., Virginia), we first choose a collection of *person attributes* from a set \mathcal{D} (e.g., age, gender, and employment status) and a set \mathcal{T}_A of *activity types* (e.g., Home, Work, Shopping, Other, and School). The precise choices of \mathcal{D} and \mathcal{T}_A are guided by the particular scenarios or analyses the population will serve. Described at a high level, we (i) construct people and places, (ii) assign activity sequences to people, (iii) map each activity for each person to a location (including the time of the visit), and (iv) from this, we derive a contact network using co-occupancy to infer edges. The construction is broken down in a sequence of steps outlined as follows.

Using *iterative proportional fitting* (IPF) [8, 20] the **base population** model constructs a set of individuals \mathcal{P} where each person has assigned demographic attributes from \mathcal{D} . By design, this ensures that \mathcal{P} matches the actual distributions and Public Use Microdata Sample (PUMS) data from the US Census [64], which is the input data for the model. Additionally, this model partitions \mathcal{P} into a set \mathcal{H} of *households*, where the notion of household encompasses the traditional notion of “family”, but also any other subset of individuals residing in the same *dwelling unit* (e.g., dormitories, army barracks, or prisons).

After household assignment, each individual $p \in \mathcal{P}$ is assigned a week-long activity sequence $\alpha(p) = (a_{i,p})_i$ where each *activity* $a_{i,p}$ has a *start time*, a *duration*, and an *activity type* from \mathcal{A} . Data sources used for this step include National Household Travel Survey (NHTS) [66], American Time Use Survey (ATUS) [65] and Multinational Time Use Study (MTUS) [61]; these sources are fused to form consistent, week-long activity sequences. We write $\alpha: \mathcal{P} \rightarrow \mathcal{A}$ for the mapping assigned to each person. For this construction, we use Fitted Values Matching (FVM) for adults [39], and Classification And Regression Tree (CART) for children (see, e.g., [11]).

The **location model** constructs a set of spatially embedded locations \mathcal{L} consisting of *residence locations* where households live, and activity locations where people conduct their non-Home activities. This construction is highly granular and is rooted in data such as the MS Building data [44], HERE/NAVTEQ data [33] for points-of-interest (POIs) and land-use classifications, National Center for Education Statistics (NCES) [49] data for public schools, as well as LandScan³,

³<https://landscan.ornl.gov/>.

OpenStreetMap⁴, and Gridded Populations of the World (GPW) v4⁵. A point plot of the locations of Virginia is shown in Figure 12.

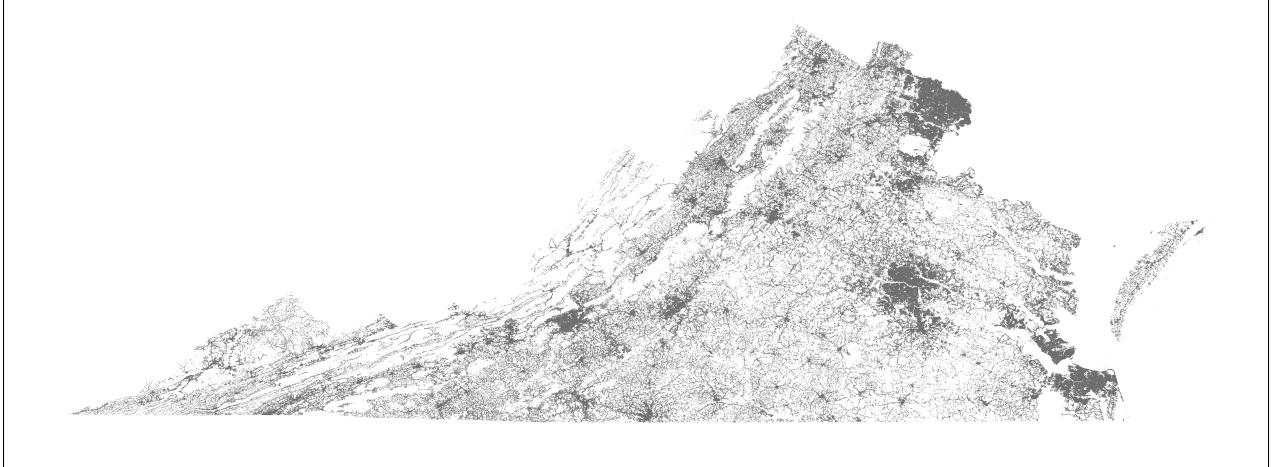


Figure 12: The illustration shows the locations for the synthetic population of Virginia used in this paper. Each location's centroid (longitude and latitude) is shown as a point.

For each person $p \in \mathcal{P}$, the **location assignment model** assigns a location $\ell = \ell(a_i)$ to each of their activities a_i . We denote the sequence of locations visited by p as $\lambda_p = (\lambda_i)_p$. The location assignment model uses American Community Survey (ACS) commute flow data [63] to assign the target county c for Work activities, and a particular location randomly within c work weights assigned to each location in c . School activity locations are assigned based on NCES data, with remaining activities anchored near home and work locations.

Finally, the **contact network model** uses the location assignment to derive the bipartite *people location graph* G_{PL} with vertex sets $V_1 = \mathcal{P}$ and $V_2 = \mathcal{L}$ and a labeled edge (p, ℓ) whenever p visits ℓ where the label includes activity type, time for start of visit, and duration of visit. From this, we derive the list of visitors to each location and the *co-location graph* G_{\max} with vertex set \mathcal{P} and edges all $e = (p, p')$ for people p and p' that are simultaneously present at the same location. Merely being present at a location at the same time does not imply a contact, and sub-location contact modeling is applied at each location to determine which of the edges of G_{\max} should be retained to form the *contact network* G which is also referred to as the *person-person contact network* and denoted as by G_{PP} (rather than simply G) to make this explicit. In this work, we use a random graph model referred to as the *Min/Max/alpha model* at each location to obtain G . Let ℓ be a location and let $N = N_\ell$ denote the maximal number of simultaneous visits to ℓ . Define the function $p_\ell: \mathbb{N} \setminus \{0, 1\} \longrightarrow [0, 1]$ by

$$p_\ell(N) = \min\left\{1, \left[\text{Min} + (\text{Max} - \text{Min})(1 - e^{-N/\alpha})\right]/[N - 1]\right\}, \quad (1)$$

where $\text{Min} < \text{Max}$ are non-negative numbers and $\alpha > 0$. Given $p = p_\ell(N)$ one samples from this random graph model in the same manner as for the standard model $G_{n,p}$ by independently at random applying to each edge e the probability p corresponding to the location ℓ where $e \in G_{\max}$

⁴<http://www.openstreetmap.org>

⁵<https://sedac.ciesin.columbia.edu/data/collection/gpw-v4>

originate. Thus the parameters Min and Max bound the degree of each vertex locally at ℓ (in expectation) for each visit; note, however, that the degree of person p in the resulting graph G is the accumulation of degrees across their trajectory to locations visited while executing their activity sequence. Thus the choices of Min, Max and α will induce the degree of each vertex in a bottom-up manner, see [47] for full details. Finally, for the applications and scenarios of this paper, we project from G , the week-long contact network, to $G_{\text{Wednesday}}$, representing the contact network on a “typical day”.

We will use agent-based network models to study the epidemic process in this paper. Agent-based networked models (sometimes just called agent-based models) extend metapopulation models by explicitly capturing the interaction structure of the underlying populations. In this class of models, epidemic dynamics are modeled as a diffusive process on a specific undirected person-person contact network $G(V, E) = G_{PP}(V, E)$ on a population V (see Section A.1 for precise definitions) – the existence of an edge $e = (u, v) \in E$ implies that individuals (also referred to as nodes) $u, v \in V$ come into contact⁶. Let $N(v)$ denote the set of neighbors of v .

Epidemic process over networks. The SIR model on the person-person contact network G is a dynamical process in which each node is in one of three states: S , I or R . Infection can potentially spread from u to v along edge $e = (u, v)$ with a probability of $\beta(e, t)$ at time t after u becomes infected, conditional on node v remaining uninfected until time t —this is a discrete version of the rate of infection for the classical compartmental mass action models discussed earlier. We let $I(t)$ denote the set of nodes that become infected at time t . The (random) subset of edges on which the infections spread is referred to as a *dendrogram*. This dynamical system starts with a configuration in which there are one or more nodes in state I and reaches a fixed point in which all nodes are in states S or R . *In our simulations, the disease models are significantly more complicated than simple SIR processes; this is described in the Appendix in Section G.*

A.2 Interventions and vaccine allocation policies

Interventions are implemented to inhibit disease transmission. Interventions can be thought of as individual behavioral adaptations or policy mandated changes, such as closing certain facilities or reducing their capacity. Of course, policies also lead to further behavioral adaptations. Our agent-based models have a rich set of interventions implemented. The specific ones we use in the study are detailed in Section F. In our simulations, all policy changes and behavioral adaptations except vaccine uptake can be seen as processes that continually and adaptively change the social contact network and disease transmission parameters.

Here we focus on policy concerning vaccine allocation. Given a schedule \mathcal{S} that specifies the amount of each vaccine available at each time, and the characteristics of the vaccines, a prioritization scheme is a policy that assigns at each time period the individuals (nodes) that are to be vaccinated. In other words, a vaccine prioritization scheme can be thought of as a Markov Decision process – at each time step we know the current state of the system and the available vaccines, and we need to decide who gets the vaccine. In this paper, we only consider non-adaptive policies – i.e. policies that do not change how vaccines are allocated at each time step. The amount depends, of course, on the schedule. We focus on four types of policies, each of which partitions the population into priority groups using different characteristics:

⁶Note that though edge e is represented as a tuple (u, v) , it actually denotes the set $\{u, v\}$, as is common in graph theory.

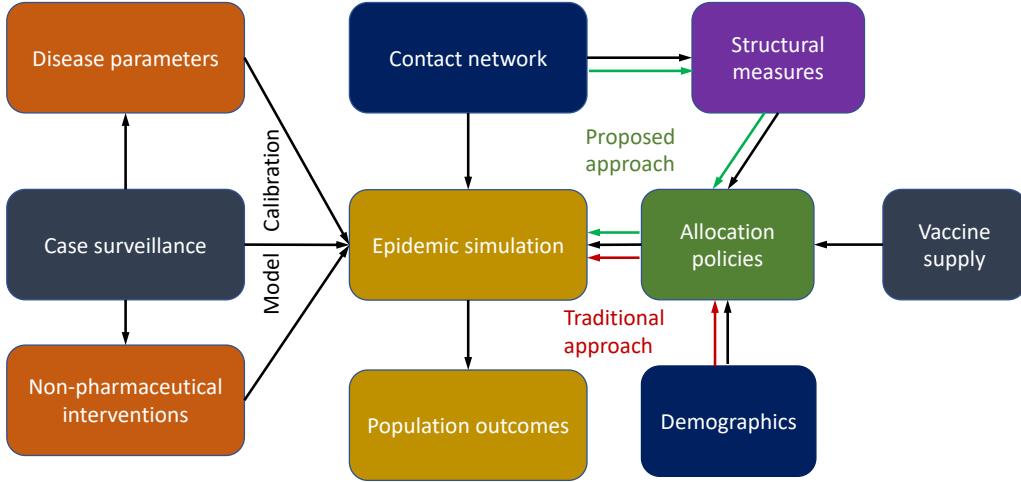


Figure 13: The overall data driven framework used to study various prioritization strategies. The figure shows schematically how real world data is used to drive the modeling and analysis process. This provides a realistic context for the underlying the simulations. For instance, it takes into account the current disease prevalence, vaccine schedule, efficacy, NPIs into account to evaluate the strategies.

1. age group,
2. occupation,
3. number of social contacts (degree), and
4. total duration of social contacts (weighted degree).

Within each type, the policies are distinguished by the fraction allocated to each priority group (further details can be found in Section 2.2).

The criteria for evaluation of policies. A policy's effectiveness will be measured by comparing the total numbers of the following to a baseline case with no vaccines: (i) infections, (ii) hospitalizations, and (iii) deaths.

Note that the simulation output is a random variable and thus we report the empirical expectation and variance of this random variable, where the expectation is taken over all possible initializations of the stochastic process and the probabilistic transmission and intervention process. We follow the discussion in [59]. Formally, at any time t , the state of the system $S_t = (I_t, W_t)$, where I_t and W_t denote the set of infected and vaccinated nodes at time t . The stochastic process is started in some initial condition $S_0 = (I_0, W_0)$. The decision to vaccinate is done over time horizon T and the vaccine schedule is given by \mathcal{S} . Given this for a policy π , its expected utility with respect

to infection (criterion i above) is given by

$$U_{\mathcal{S}}^{\pi} = \mathbb{E} \left[\sum_{t=0}^T |I_t| \mid S_0 \right].$$

An agent-based simulation model is described in Appendix F. The COVID-19 disease model is described in Appendix G.

B Structural Measures for Social Contact Networks

In this section, we define some popular network measures used to characterize real-world networks and use them to study the structure of the contact networks used in this work. These include measures of local connectivity such as degree and clustering coefficient, global connectivity such as diameter, k -core, and graph spectrum. These measures have been used in both structural and dynamical characterization of real-world networks. In particular, these measures are considered important with respect to epidemic simulations, and are frequently reported in the literature [14, 28]. Our objective is two-fold: (i) compare our networks with existing real-world networks, and (ii) use these measures to analyze the efficacy of the different control strategies applied in our work. The latter is discussed in Section E.

The construction of the synthetic contact network of Virginia is described in Appendix A.1 along with definitions of the people-location network G_{PL} and the person-person contact network $G = G_{PP}$. Note that this network was modeled and constructed with epidemics and disease transmission as a target. Generally, what constitutes an interaction (and thus edge) factors through physical proximity, the nature of the interaction, the nature of the dynamics studied (e.g. disease transmission) and other factors such as, for example, air circulation within a building and infection through contaminated inanimate objects. Note first that the person-person contact network G has 7.6×10^6 nodes and 2.0×10^8 edges. Moreover, G has a largest component of (relative) size 0.983.

Diameter. The diameter is the length of the longest shortest path between any two vertices of the network. The diameter of G is $\text{diam}(G) = 12$.

Degree distribution. The degree of a person u in the people-people contact network G is the number of different persons that u has contact with during a day. The average person degree in G is $\bar{d} = 43.5$. In the people-location network G_{PL} , the degree of a location ℓ is the number of distinct visitors to ℓ during a day, and the degree of a person u is the number of distinct locations u visits during a day. Figure 14 shows the degree distributions in G and G_{PL} . The weighted degree of a node u is the total time of contact with its neighbors per day. The weighted degree distribution of the network G is shown in Figure 15a.

Clustering coefficient. This is a measure of the degree to which nodes in a graph tend to locally cluster together. For a node v , its *local clustering coefficient* is the fraction of pairs of its neighbors which have a link between them. It quantifies how close the immediate neighbors are to being a complete graph. If the local clustering coefficient is 1, it means that the node and its neighbors induce a complete graph, and if it is 0, then they induce a star graph with v at the center. The *average clustering coefficient* is the average of all the local clustering coefficients [70]. The average clustering coefficient of G is $\bar{c} = 0.092$; the clustering coefficient distribution is shown in Figure 15b.

k -core. A k -core of a graph G is a maximal connected subgraph of G in which all vertices have degree of at least k . Equivalently, it is one of the connected components of the subgraph of G formed by repeatedly deleting all vertices of degree less than k [71]. The *core number* of a graph is the maximum core it belongs to. The *maximum core* (max. core) of the graph is the set of all nodes with the maximum core number. It is another way of characterizing centrality of nodes. It is considered that for certain epidemiological models, the k -core number is a good predictor of the final outbreak size [50]. Also, the k -core decomposition is in many cases a good predictor of spreading efficiency. The k -core decomposition of G is shown in Figure 15c. The maximum k for which there is a k -core is 57. While there are only 1,700 nodes in the induced subgraph for the 57-core, there are approximately 35,000 nodes (around 0.5% of the total population) in the 45-core, and more than a quarter of the population belongs to the 30-core showing high global connectivity

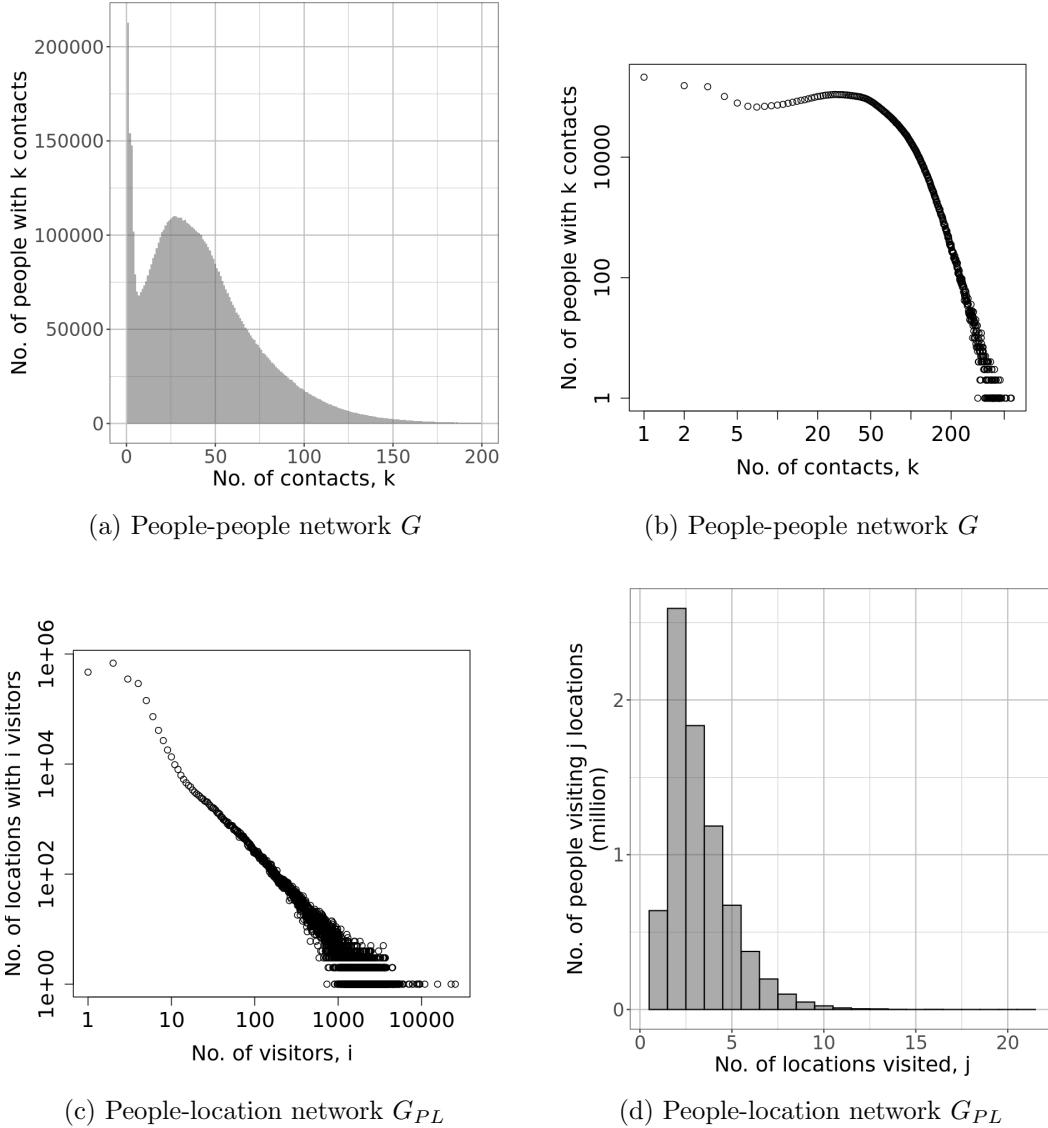


Figure 14: Degree distributions in people-people network G : (a) as a histogram in normal scale, and (b) in log-log scale showing a power-law tail with an exponent of about -7; and degree distribution in people-location network G_{PL} : (c) of locations in log-log scale showing a power-law tail with an exponent of about -1.7, and (d) of people as a histogram in normal scale.

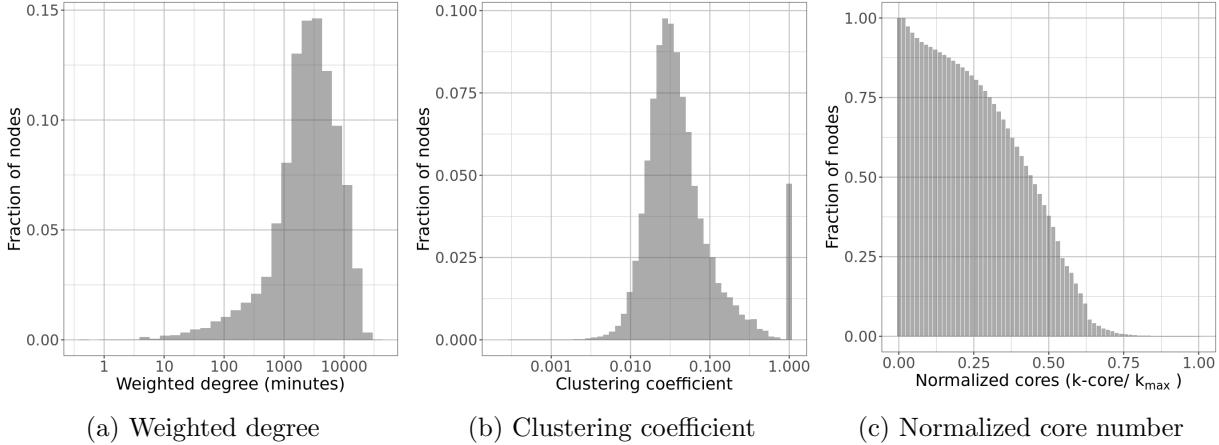


Figure 15: Distributions for connectivity properties of G here including (a) the weighted degree distribution, (b) the clustering coefficient distribution, and (c) the normalized core number distribution for G .

of the people-people network.

Graph spectrum or eigenvalues. The spectrum of a graph is the set of eigenvalues of its adjacency matrix. Also popular is the Laplacian spectrum, which is the set of eigenvalues of the Laplacian matrix of the graph. There are several works that relate spectrum, particularly the first eigenvalue of the adjacency matrix, to disease spread in SEIR-like models [27, 52]. The common result that highlights the impact of the network structure on the dynamics is that epidemics die out “quickly” if $\lambda_1(G) \leq T$, where $\lambda_1(G)$ is the *spectral radius* (or the largest absolute value of an eigenvalue) of graph G , and T is a threshold that depends on the disease model. This relationship has motivated a number of works on epidemic control where the objective is to find an optimal set of nodes (or edges) to remove from the network that leads to maximum reduction in its spectral radius [57, 67, 74].

Activity-based structural analysis. We analyzed the constructed social contact network using the structural measures described above. We considered activity induced sub-networks, where a contact edge is retained only if both individuals corresponding to that edge have the target activity assigned to them. For example, in the case of **School** network, only **School–School** edges are retained. For **School** activity, we observed an average degree of 29.7, which is between the average degrees of 13.5 and 47.3 reported for the school networks of the SocioPatterns collaboration networks [14, 28]. Our average degree for **Work** activity is 16.3. In comparison to the SocioPatterns data, this seems to be on the higher side for a typical office environment (< 7) but comparable to their hospital network (14.0). This is expected, since our **Work** activity includes a wide variety of workspaces that include office environment, factories, restaurants, etc. The max. core for the **School** activity network is 32. In SocioPatterns, the two schools have max. cores of 24 and 47 respectively. For the **Work** network, it is 37, while in SocioPatterns it is 11, 25 and 23 respectively for the three office networks. Again, in our case the max. core is the maximum among all locations.

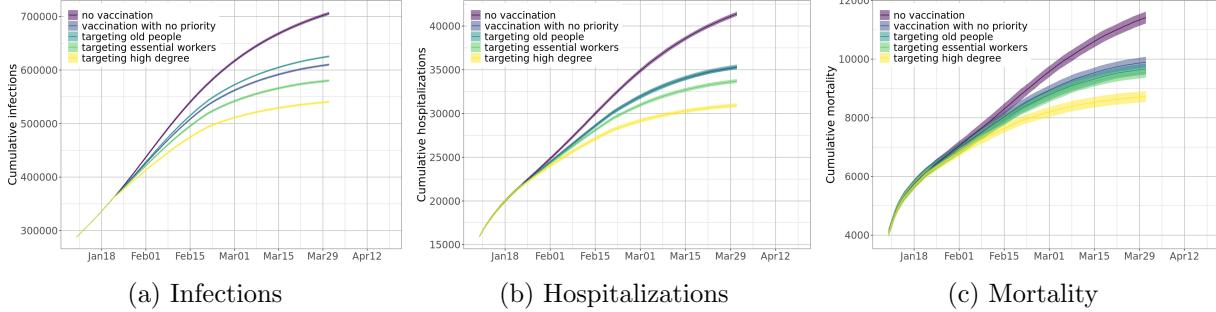


Figure 16: Cumulative counts under different vaccination strategies for (a) total infections, (b) total hospitalizations, and (c) total mortality, assuming current non-pharmaceutical interventions remain at the same level (*as-is*).

C Additional Results and Analysis

We present additional results as discussed in the paper.

C.1 Epidemic curves with various strategies

Figures 16 to 18 show the cumulative numbers of infections, hospitalizations, and mortality in the first three months of 2021 under different vaccination schemes, assuming no relaxation of NPIs (Figure 16), slow relaxation (Figure 17), and fast relaxation (Figure 18). Note that for the age-based, essential worker-based, and degree-based schemes, we only show the moderate level of priority enforcement 80%. All curves show the uncertainty of one standard deviation above and below the mean.

Infections. We find that in all three scenarios of relaxation, while vaccinations can significantly reduce infections, targeting high degree people clearly outperforms all other schemes; and the no-priority scheme performs better than the age-based scheme but not as well as the essential worker targeting scheme.

Hospitalizations. While the no-priority scheme outperforms the age-based scheme on reducing infections, they seem to have similar reductions on hospitalizations. This is because older people have fewer contacts on average, so vaccinating them does not decrease the connectivity of the network as much as the other schemes. But since vaccines can protect the vaccinated against severe illness (we assume $e_D = 50\%$), and the older people have a higher hospitalization ratio, so targeting them is more effective on reducing hospitalizations than on reducing infections.

Mortality. Due to the same protection to the vaccinated against severe illness, targeting old people seems more effective on reducing mortality than on reducing infections. Targeting essential workers is marginally better than the age-based and the no-priority schemes. The degree-based scheme, however, continues to be the most effective on reducing mortality.

We also observe that when NPIs are relaxed, vaccinations become more important. This is discussed in Section 3.3.

C.2 Sensitivity analysis

This section shows some of the variability in network measures for a statistical design over the network construction parameters Min, Max and α . Specifically, we consider the hyperplane given

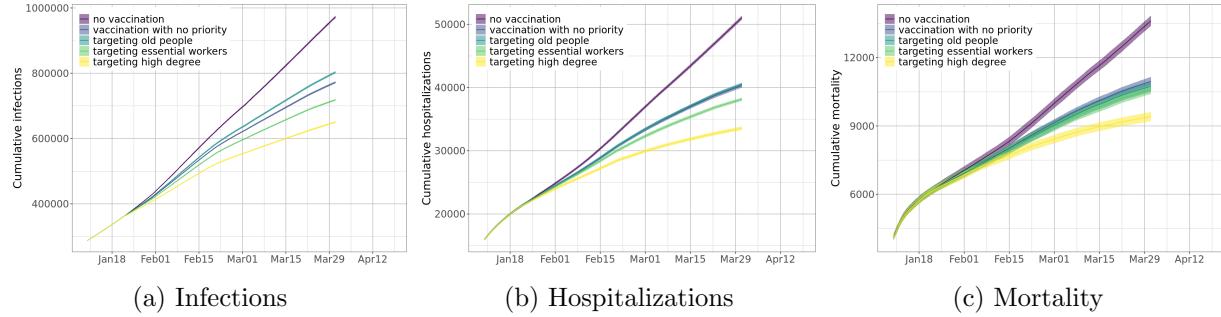


Figure 17: Cumulative counts under different vaccination strategies for (a) total infections, (b) total hospitalizations, and (c) total mortality, assuming slow relaxation of non-pharmaceutical interventions.

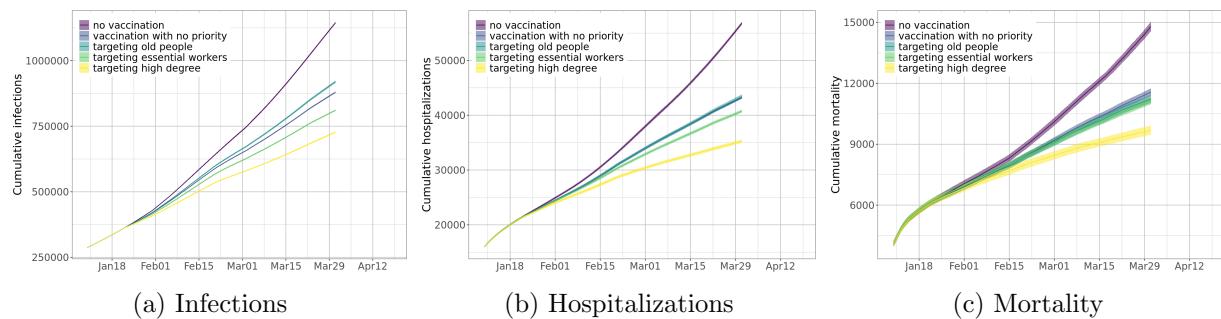


Figure 18: Cumulative counts under different vaccination strategies for (a) total infections, (b) total hospitalizations, and (c) total mortality, assuming fast relaxation of non-pharmaceutical interventions.

by $\text{Min} = 5$, $\alpha = 1000$ and $\text{Max} \in \{35, 40, 45\}$ with $\text{Max} = 40$ representing the base case used in all the simulations of this work. Accordingly, we have three graphs G-35, G-40, and G-45. As can be seen in Table 2, there is a slight increase in properties such as \bar{d} , \bar{T} , and d_{\max} while the diameter drops by 1 for $\text{Max} = 45$. The slightly bigger change occurs in eigenvalues, but we note that the spectral gap (i.e., $\lambda_1 - \lambda_2$) remains nearly constant.

Network	$ V $	$ E $	\bar{d}	\bar{T}	d_{\max}	T_{\max}	\bar{c}	k_{\max}	ρ_{\max}	diam	λ_1, λ_2
G-35	7.6e06	2e+08	41.0	54.4	455	588.5	0.091	51	0.983	12	100.6, 87.8
G-40	7.6e06	2e+08	43.5	59.2	565	1526.9	0.092	57	0.983	12	118.5, 100.6
G-45	7.6e06	2e+08	49.2	64.4	636	655.6	0.096	66	0.983	11	127.1, 110.0

Table 2: Here $|V|$ denotes the number of nodes, $|E|$ the number of edges, \bar{d} the average degree, \bar{T} the average total contact time (unit: hour), d_{\max} the maximal degree, T_{\max} the maximal total contact time, \bar{c} the average clustering coefficient, k_{\max} the maximal core, ρ_{\max} the relative size of the maximal component, diam the network diameter, and λ_1 and λ_2 the two largest eigenvalues.

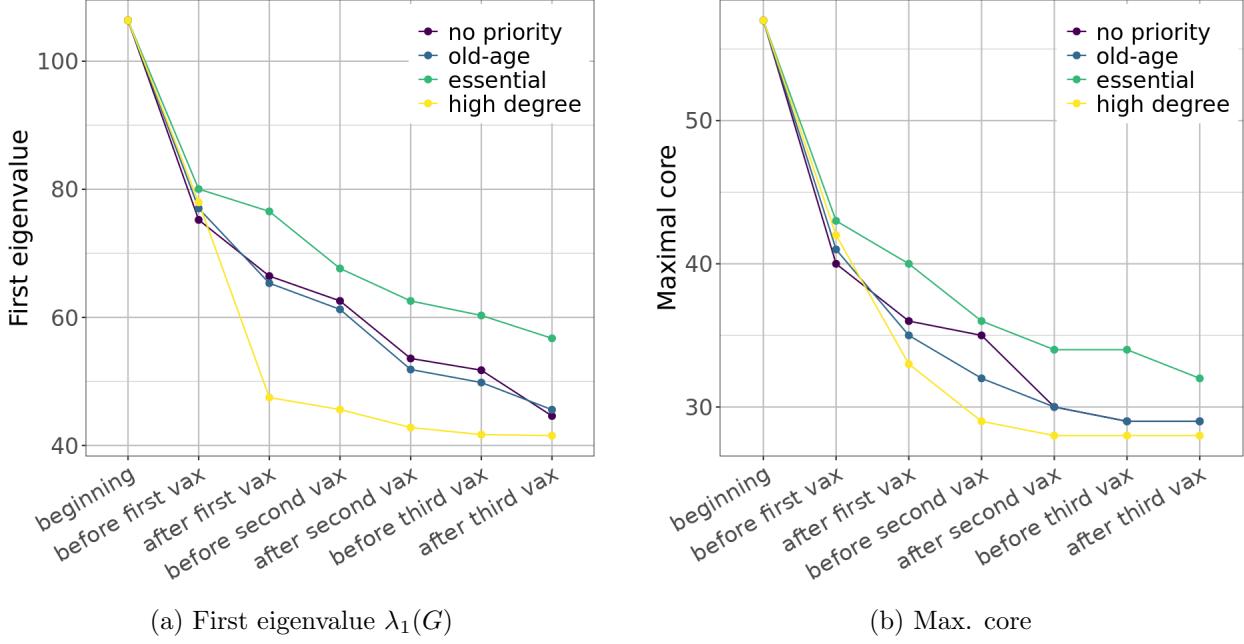


Figure 19: Reduction in the first eigenvalue and max. core of the network at different vaccination phases. As nodes are infected or vaccinated, we remove edges incident on them from the contact network, to form sparser and sparser *snapshot* networks. We consider such snapshots under different prioritization strategies, assuming no relaxation of NPIs and 100% distribution. After the first round of vaccination, the degree-based strategy reduces the first eigenvalue and the max. core the most as compared to other strategies.

D Analysis of the time-varying network

Each time the chosen control strategy is applied to the network, some subset of nodes are removed leading to structural changes in the network. We analyzed this evolution of the network using the various measures introduced in Section B. The network measures for each residual network are listed in Table 3. For ease of comparison, the evolution of the max. core and the spectral radius are plotted in Figure 19. Also, in Figure 20, we have plotted the distribution of core number.

The measures that are generally affected by vaccination are max. degree, avg. degree, max. core and eigenvalues. Diameter and size of giant component remain little changed over time. We note that among all strategies, the most significant change in network properties is observed for the degree-based strategy. The efficacy of targetting old population and targetting essential workers are comparable to random vaccinations with respect to these measures.

Strategy	Snapshot	Nodes	Avg. deg.	Max. deg.	Max. contact hrs.	Avg. clust. coeff.	Max. core (k_{\max})	Giant comp.	Diameter	Top eigenvalues
none	beginning	7605430	43.5	565	3053.7	0.092	57	0.983	12	118.5,100.6
high degree	before first vax	6306549	35.8	400	2161.8	0.086	42	0.984	12	87.8,79.5
essential	before first vax	6305803	35.8	507	2165.2	0.086	43	0.984	13	89.7,80.5
old-age	before first vax	6301976	35.7	512	1664.6	0.086	41	0.983	13	85.1,79.0
no-priority	before first vax	6305637	35.8	504	1948.7	0.086	40	0.983	12	83.2,77.0
high degree	after first vax	5719119	26.6	374	1673.4	0.09	33	0.981	13	48.9,46.2
essential	after first vax	5704209	32.6	492	1999.0	0.082	40	0.984	13	83.6,74.3
old-age	after first vax	5679458	33.1	506	1647.1	0.085	35	0.987	12	72.4,65.0
no-priority	after first vax	5698684	32.8	494	1847.5	0.082	36	0.985	12	73.8,67.5
high degree	before second vax	5605355	25.7	363	1278.2	0.09	29	0.981	13	47.4,44.9
essential	before second vax	5563558	31.3	457	1292.0	0.082	36	0.983	13	71.0,64.7
old-age	before second vax	5500145	31.3	487	1287.9	0.086	32	0.986	13	67.0,61.4
no-priority	before second vax	5534739	31.1	480	1350.5	0.083	35	0.984	13	67.5,62.6
high degree	after second vax	5059583	19.4	179	1259.7	0.097	28	0.977	14	45.3,42.4
essential	after second vax	4961207	28.4	435	1292.0	0.078	34	0.984	12	65.2,58.3
old-age	after second vax	4888575	28.8	380	1287.0	0.086	30	0.989	13	56.7,52.0
no-priority	after second vax	4929293	28.3	363	1347.9	0.079	30	0.986	13	58.2,51.3
high degree	before third vax	5022947	19.1	172	1259.7	0.098	28	0.977	14	43.9,41.3
essential	before third vax	4914302	28.0	423	1232.2	0.078	34	0.983	13	64.0,60.2
old-age	before third vax	4813271	28.1	370	1227.7	0.087	29	0.988	13	54.1,49.8
no-priority	before third vax	4864470	27.7	360	1331.1	0.079	29	0.986	13	56.4,51.6
high degree	after third vax	4520360	18.3	169	1259.7	0.092	28	0.98	14	43.7,41.1
essential	after third vax	4533299	26.3	273	1219.0	0.076	32	0.983	12	61.3,57.3
old-age	after third vax	4212668	25.8	361	1227.7	0.088	29	0.991	13	47.8,45.5
no-priority	after third vax	4256026	25.0	349	1331.1	0.074	29	0.988	13	46.8,44.3

Table 3: Properties of the temporal snapshots of the social contact network before and after each application of vaccination. The “Strategy” column refers to the vaccination prioritization. Each line corresponds to a snapshot network extracted from the contact network by removing edges incident on infected nodes and vaccinated nodes. These snapshot networks are the same ones used in Figure 19.

E Why do social interaction-based heuristics work and how can they be implemented

We briefly discuss the reasons why the degree- and weighted degree-based heuristics work. We also briefly discuss the role of digital devices in estimating degrees of individuals.

Why do degree-based heuristics work. The efficacy of degree-based heuristics has been discussed in several earlier papers, e.g., [3, 10, 22, 51, 73], as discussed in Section H. This phenomenon is also well understood in certain random graph models [10, 51]. Pastor-Satorras et al. [51] derive an immunization threshold in terms of the fraction of high degree nodes that are immunized through an analysis that relies on a degree-based mean field (DBMF) assumption. This is derived more

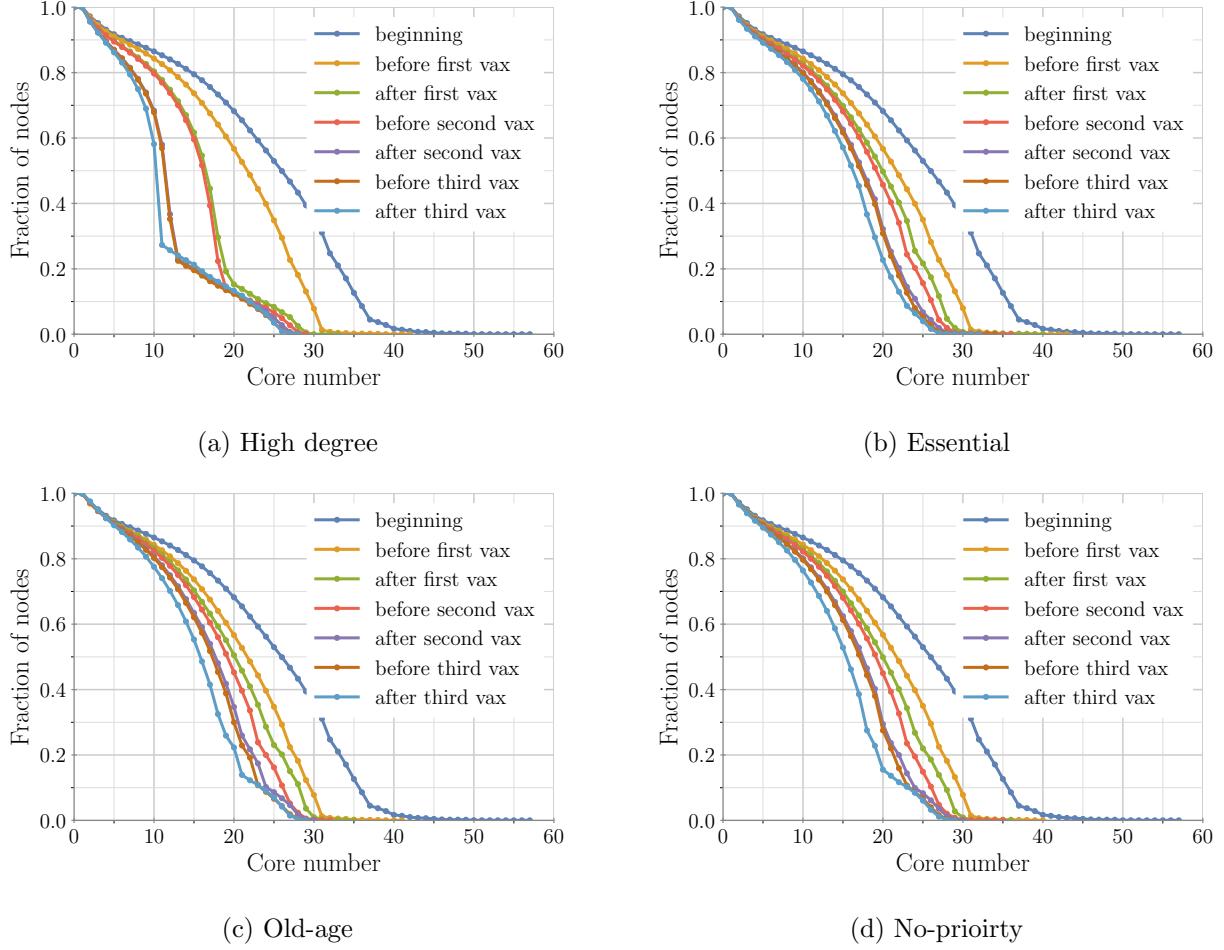


Figure 20: The evolution of k -core distribution in the snapshot networks extracted at various stages of the simulation. Different vaccination prioritization strategies are implemented in the four subfigures. Note that these snapshot networks are the same ones used in Figure 19 and Table 3.

rigorously by Bollobas et al. [10], who show that there is a threshold fraction of high degree nodes whose removal shatters the graph—this can be viewed as an immunization strategy in their random graph model for the case when the transmission probability $p = 1$.

However, it is important to note that the theoretical results are *not directly applicable* in our context for the following reasons: (i) many of the theoretical results show the efficacy of these methods for power law networks – the networks we generate are similar to power law networks but with a very different exponent; and (ii) many of the results are shown when vaccines are applied at the start of the epidemic process, and the results do not say anything of what happens when the vaccine is applied temporally – this is important, because the temporal epidemic process infects individuals already and thus changes the network structure substantially, including the application of the non-pharmaceutical interventions.

As nodes are infected or vaccinated, they are either not susceptible or less susceptible (i.e. lower probability of getting infected). Consider the network structural changes due to such dynamics. For simplicity, we remove all edges incident on nodes who are or have been infected and those who

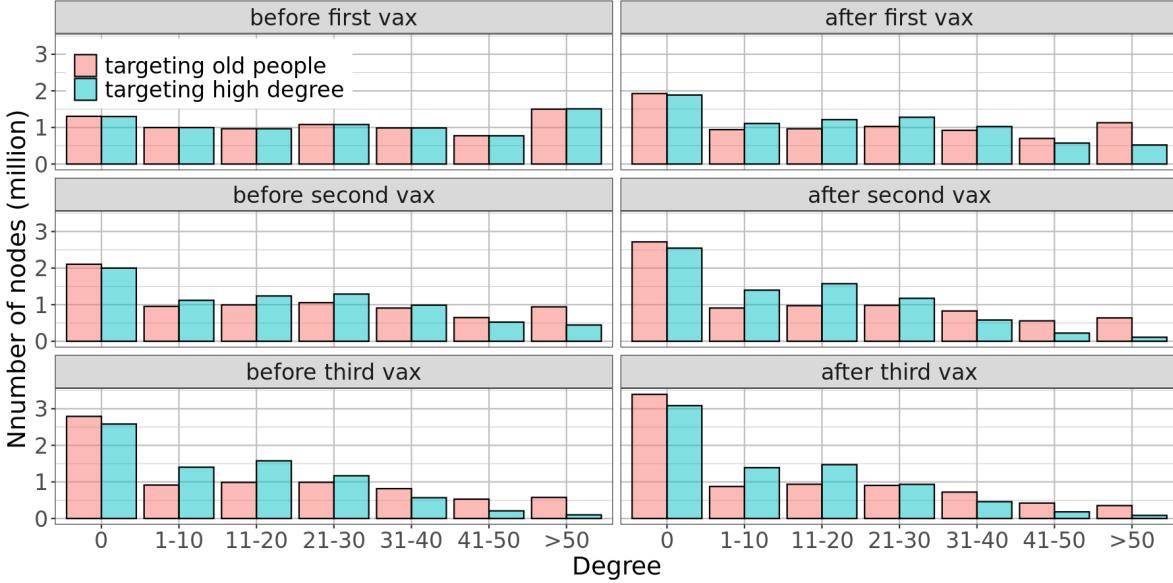


Figure 21: As nodes are infected or vaccinated, we remove edges incident on them from the contact network, to form sparser and sparser *snapshot* networks. We consider such snapshots under age-based and degree-based strategies, assuming no relaxation of NPIs and 100% distribution. After the first round of vaccination, the degree-based strategy reduces degrees more than the age-based strategy: we observe more low degree nodes and fewer high degree nodes in the snapshot networks from the degree-based strategy.

are vaccinated, to form a sparser *snapshot* network at different time point, especially right before and right after each mass vaccination. In Figure 21, we compare the degree distributions of such snapshot networks from age-based and degree-based vaccinations, before and after each vaccination. The networks are extracted from a simulation run under as-is scenario with fast vaccine distribution and 100% prioritization. Before the first vaccination, the snapshot network from both strategies have the same degree distribution. After the vaccination, since more edges are removed with the degree-based strategy, the snapshot network has more low degree nodes and fewer high degree nodes, compared to those from the age-based strategy. Similar is observed between the snapshot network before and after the second vaccination. After that the degree distribution does not change much with vaccination, since the network is already very sparse.

The connection between spectral properties and epidemic thresholds provides another insight into the efficacy of a degree-based strategy. As mentioned in Section B, it has been shown for SIS/SIR models, via different approaches (including different mean field approximations), that if the first eigenvalue (referred to as the *spectral radius*) $\lambda_1(G)$ of the network is below a certain threshold, the epidemic is not very “large” [27, 45, 52]. Figure 19a shows the first eigenvalue $\lambda_1(G)$ of the network at different points in time as interventions are applied. It is very significant to note that $\lambda_1(G)$ shows a substantial drop when the degree intervention is applied. Thus, the intuition behind the efficacy of such methods is clear – high degree nodes protect not only themselves, but also confer a higher level of indirect protection on their neighbors as they interact with many individuals. The fact that real-world social networks have sufficient nodes of high degree ensure that such heuristics are effective. It is important to note, however, that just the presence of high degree nodes themselves does not ensure that degree-based heuristics work.

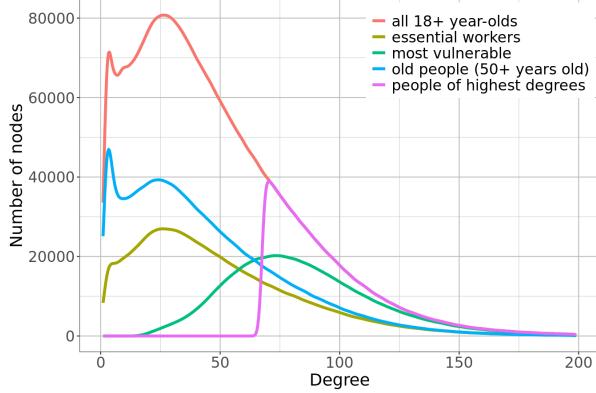


Figure 22: Degree distribution among nodes that are prioritized differently. Essential workers and older people seem to have lower degrees than high-degree people and highly vulnerable people.

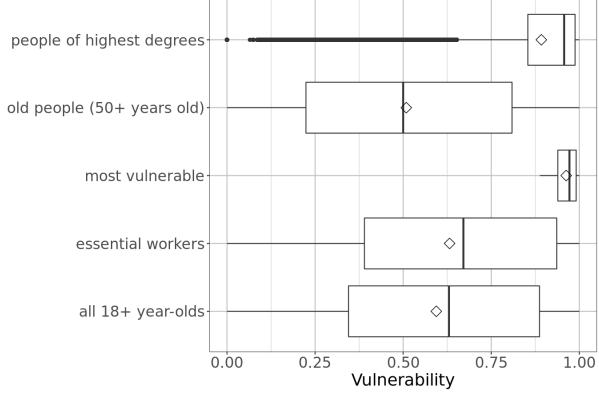


Figure 23: Vulnerability distribution of nodes with different priorities. People of highest degrees have high vulnerability, close to that of the most vulnerable people.

A dynamical property: Vulnerability and its relationship to degree. Given a generalized SEIR system $G(V, E)$ along with a random initial configuration and a node $v \in V$; and an integer $t \geq 1$, we define $t\text{-VUL}(v)$ as the probability that v gets infected *at* time t . Vulnerability of v by time t ($t\text{-TOTVUL}(v)$) is the probability that v gets infected *by* time t . Informally speaking, $t\text{-VUL}(v)$ and $t\text{-TOTVUL}(v)$ estimate the chance of a node getting infected when a random set of nodes are infected to begin with. Vulnerability in essence captures various ways the infection can reach a given node. Degree-based interventions effectively reduce the vulnerability of several nodes at once. Interestingly, nodes of high degree are also highly vulnerable, and this can be empirically observed by noting the plots in Figures 22 and 23. Vulnerability computations are computationally challenging (#P-hard) and our simulations provide good estimates of the same.

How can we estimate degrees. A central question is: can individual node degrees be estimated? We believe that this is possible, and more so as a result of digital devices and apps that have recently been deployed for contact tracing. The current set of apps built for contact tracing keep track of the social interactions of an individual in a privacy-preserving way. These apps can easily be modified to count the number of such interactions. Furthermore, our results show that duplicate counts that might result are okay, as the degrees are likely to be skewed some. These apps can be modified to not only store the keys, but also the encrypted form of the time a person spends with the person that generates such a key.

Furthermore, a person does not need to reveal their degree to anyone but the app can simply provide this information for prioritization. The alerting and scheduling of vaccination appointments can be conducted through the app itself. Additionally, demographic information such as age, or other risk-factor information such as additional comorbidities or race and/or ethnicity, may also be used to further prioritize individuals within their degree-based allocation groups.

How accurate can such measurements be. Let V be partitioned into groups V_1, \dots, V_r , and suppose p_i is the probability that a node in V_i has the app. Think of V_i as an age group. Survey results show that younger individuals have a higher propensity to use digital device as well as download the app. We assume each node $v \in V_i$ decides independently with probability p_i to install the app; if v chooses to install the app, we say that it is “sampled”. Let H denote the

sampled graph, which is induced by the nodes (node induced subgraph) which have the app, and let $V(H)$ denote the set of nodes in H . For a node v , let $N_G(v)$ denote the set of its neighbors in G . Similarly, let $N_H(v) = \{u : u \in N_G(v) \text{ and } u \in V(H)\}$ denote the set of neighbors of v in H . Let $d_G(v) = |N_G(v)|$ denote its degree in G , and let $d_G(v, i)$ denote the number of neighbors it has in group V_i (so $d_G(v) = \sum_i d_G(v, i)$). If v is sampled, let $d_H(v) = |N_H(v)|$ denote its degree in H . Let $X(u) \in \{0, 1\}$ be a random variable which is 1 with probability p_i , if $u \in V_i$, and let $Y(v) = \sum_{u \in N_G(v)} X(u)$. Observe that $\mathbb{E}[d_H(v)] = \mathbb{E}[Y(v)] = \sum_{u \in N_G(v)} \Pr[X(u) = 1] = \sum_i \sum_{u \in N_G(v) \cap V_i} p_i = \sum_i p_i d_G(v, i) = f(v)$. By applying a Chernoff bound to $Y(v)$, we have $\Pr[|Y(v) - f(v)| > \epsilon f(v)] \leq 2e^{-\epsilon^2 f(v)/3} \leq 2/n^2$ if $f(v) \geq \frac{6}{\epsilon^2} \ln n$. Let $V_h = \{v : f(v) \geq \frac{6}{\epsilon^2} \ln n\}$. Therefore, with probability at least $1 - 2/n$, every node $v \in V_h$ has $d_H(v) \in [(1-\epsilon)f(v), (1+\epsilon)f(v)]$. Let $Q \subseteq V$ denote the subset of nodes in the top quartile with respect to $d_G(v)$. Then, with high probability, for every node $v \in V_h \cap Q$, we have $d_H(v) \sim f(v)$. Therefore, if the sets V_h and Q have high overlap, high degree in H corresponds to high degree in G . The expected number of nodes in $V_h \cap Q$ which get sampled is $\sum_i \sum_{v \in V_h \cap Q \cap V_i} p_i$.

Where is the tradeoff? The phased approach is a balance between exposure/infection risk (policy equivalent of high social contact) and disease risk (age-based). Our proposal is a stronger version of the former, but seems to be significantly better across all outcomes. This is one of the counter-intuitive aspects of the proposed approach, since it does not take into account disease risk explicitly, and purely relies on structural measures of the network. While individualized policies aimed at minimizing mortality do exist (e.g., comorbidities) and are part of the phased approach, lack of technology thus far had made it difficult to ‘individualize’ policies targeted on minimizing transmission. Aggregate policies targeting essential workers and other high exposure categories (e.g., school children for influenza) while in a similar vein to our approach, may still have high variance in reducing transmission potential. The advantage of our approach is in leveraging the mechanistic and network based understanding of disease spread, and creating priority categories that cut across age-, risk- and other demographic characteristics. It is possible to construct pathological instances of networks in which vaccinating high degree nodes is not optimal. But as is well known in the field of network science, real world social networks are often characterized by power-law degree distributions and exhibit a hub-spoke structure. In such networks, targeting the hub nodes at individual level, ‘shatters’ the transmission network faster than aggregate policies. Especially such a strategy might be necessary in regions with very limited vaccine supplies and at higher risk of variant induced surges. We believe our work helps advance the case for better integration of social network and digital technologies in swifter public health response.

F The Agent-Based Simulator

The EpiHiper agent-based simulator supports disease models which are comprised of *disease states*, *disease transmissions* (through contacts), and *disease progressions*. The disease models are specified independently of the people and their contact network over which the disease spreads. All individuals have the same infection processes and disease progression dynamics. The infection processes, however, factors in individual attributes such as susceptibility and infectivity which are generally dynamic. All input files to EpiHiper are provided in JSON format with the exception of the contact network which, due to its large size, uses either CSV or binary format.

Disease Transmissions are caused by contacts between a susceptible individual P^s in state X_i and an infectious individual P^i in state X_k . The susceptible individual P^s will transition to an exposed state X_j based on information specified by the transmission configuration $T_{i,j,k}$, the contact $E(P^i, P^s)$, and attributes from the individuals P^s and P^i such as the susceptibility $\sigma(P^s)$ of P^s and the infectivity $\iota(P^i)$ of P^i .

Under the assumption of independence of transitions across contacts for individual P^s with one or more infectious individuals P^i , the *propensity* of the state transition to the exposed state X_j based on the transmission configuration $T_{i,j,k}$ and the single contact $E(P^i, P^s)$ is defined as:

$$\rho(P^s, P^i, T_{i,j,k}) = T \times w_e \times [\sigma(P^s) \cdot \iota(P^i)] \times \omega(T_{i,j,k}), \quad (2)$$

where T is the contact duration, w_e is the edge weight, and $\omega(T_{i,j,k})$ is the transmission rate.

The *propensities* of all state transitions to the exposed state X_j are added, and we use the Gillespie algorithm [30] to determine whether a transition occurs during the simulation interval, and, if it does, which contact to attribute.

Disease Progression covers the health state transitions within an individual P that are independent of other people. For the EpiHiper model, a disease progression diagram describes all the possible health state transitions that take place within a person. The nodes of the diagram are the health states $\mathcal{X} = \{X_i\}$ and directed edges of the form $e = (X_i, X_j)$ with an assigned probability $p_e = \text{prob}(X_i, X_j)$ and a *dwell time distribution* D_e . The sum of all probabilities of transitioning out of a given state X_i must be either 1 or 0. Zero indicates a terminal state.

The **System State** at any point in time is given by the attributes of the individuals (nodes) and contacts (edges) of the network, the simulation time, and the value of user-defined variables.

The values *nodeTrait* and *edgeTrait* are user-defined attributes which may be used to govern interventions (described below). They do not influence the disease transmission or progression directly.

Interventions are external modifications of the state of the simulation where “external” means not governed by the disease model nor the contact network. An intervention comprises of a trigger, a target, and an action ensemble. The action ensemble, which is a set of instructions to apply to the dynamic state of the target (a collection of nodes and/or edges), is only applied when the trigger evaluates to 1 (or `true`). The trigger is a function of the system state and thus may depend on any of the above-mentioned attributes, including the *Person Trait Database*. The operations of an action ensemble are partitioned into the following three categories: (i) those triggered once per intervention (e.g., to update global variables), and which are thus independent of the intervention target, (ii) those that are applied to each element of the intervention target, and (iii) those applied to a sampled subset of the intervention target and optionally to the complement of the sampled subset. Moreover, sampling may be nested, thus permitting chained sub-sampling. Each individual

Object	Value	Access	Description
system	time	r	the current time (mapped from iteration number/tick)
node	id	r	PID of node
node	infectivity	rw	infectivity scaling factor
node	susceptibility	rw	susceptibility scaling factor
node	healthState	rw	health state
node	nodeTrait[traitName]	rw	value of nodeTrait[traitName] of node
edge	sourceID	r	source vertex ID of edge
edge	targetID	r	target vertex ID of edge
edge	sourceActivity	r	activity of source of edge
edge	targetActivity	r	activity of target of edge
edge	active	rw	active flag of edge
edge	weight	rw	weight of edge
edge	edgeTrait[traitName]	rw	value of edgeTrait[traitName] of edge
variable	name	rw	a user-defined numerical variable referenced by name

Table 4: EpiHiper state values of nodes and edges

action of the action ensemble is given a delay specifying the time (or offset) at which the action should execute, which permits flexible, fine-tuned scheduling within the simulation.

Before starting the simulation, the state of all individuals P must be initialized. In EpiHiper, initialization is simply a special case of an intervention where the trigger is omitted (and by convention is **true**).

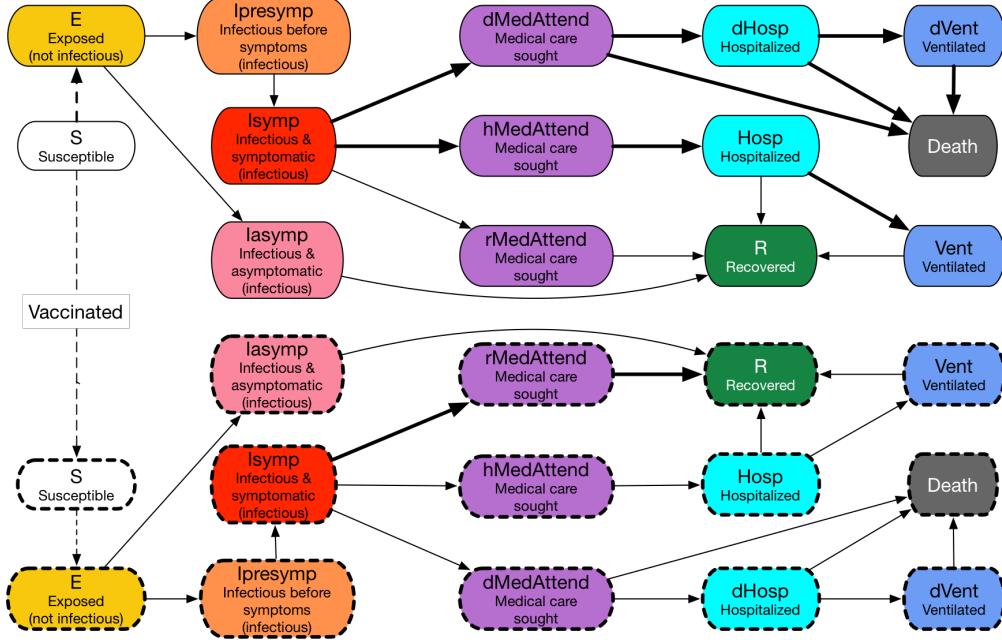


Figure 24: The COVID-19 disease model with both unvaccinated and vaccinated states. This disease progression model is represented as a probabilistic timed transition system (PTTS): the state transitions are probabilistic, and, in many cases, are timed, i.e., transitions after a given time period. An individual starts from the upper S (Susceptible) state. If an individual receives a vaccine the individual enters a new susceptible state represented by a dotted box. The dashed lines represent state transitions triggered by either interactions with infectious individuals or vaccination. The solid lines represent probabilistic timed state transitions. The shapes with a solid border represent states of an unvaccinated individual; those with a dashed border represent states of a vaccinated individual. The thicker lines represent larger probabilities. Therefore, if vaccinated, an individual has a smaller probability of getting infected (protection against infection), and even if infected the individual has smaller probability of being hospitalized or needing ventilation or death (protection against severe illness).

G The Disease Model Parameters

The within-host disease transmission model is shown in Figure 24. **Transmission** may occur when an individual in one of the states *Susceptible* or *RX Failure* comes in contact with one or more individuals in the states *Presymptomatic*, *Symptomatic*, or *Asymptomatic*. The individual transmissions are governed by the parameters in Table 5. **Progression** from one disease state to the next is governed by the parameters in Table 6

State	Attribute	Value
Presymptomatic	transmissibility	0.18
Symptomatic	infectivity	0.8
Asymptomatic	infectivity	1.0
Susceptible	susceptibility	1.0
RX Failure	susceptibility	1.0

Table 5: Disease transmission parameters

Progression	Attribute	Age				
		0-4	5-17	18-49	50-64	65+
Exposed - Asympt	prob			0.35		
Exposed - Asympt	dt-mean			5		
Exposed - Asympt	dt-std dev			1		
Asympt - Recovered	prob			1		
Asympt - Recovered	dt-mean			5		
Asympt - Recovered	dt-std dev			1		
Exposed - Presympt	prob			0.65		
Exposed - Presympt	dt-fixed			1		
Presympt - Sympt	prob			0.65		
Presympt - Sympt	dt-fixed			1		
Symp - Attd	prob	0.9594	0.9894	0.9594	0.912	0.788
Symp - Attd	dt-discrete	1:0.175, 2:0.175, 3:0.1, 4:0.1, 5:0.1, 6:0.1, 7:0.1, 8:0.05, 9:0.05, 10:0.05				
Attd - Recovered	prob			1		
Attd - Recovered	dt-mean			5		
Attd - Recovered	dt-std dev			1		
Symp - Attd(D)	prob	0.0006	0.0006	0.0006	0.003	0.017
Symp - Attd(D)	dt-fixed			2		
Attd(D) - Hosp(D)	prob			0.95		
Attd(D) - Hosp(D)	dt-fixed			2		
Hosp(D) - Vent(D)	prob	0.06	0.06	0.06	0.15	0.225
Hosp(D) - Vent(D)	dt-fixed			2		
Vent(D) - Death	prob			1		
Vent(D) - Death	dt-fixed			4		
Hosp(D) - Death	prob	0.94	0.94	0.94	0.85	0.775
Hosp(D) - Death	dt-fixed			6		
Attd(D) - Death	prob			0.05		
Attd(D) - Death	dt-fixed			8		
Symp - Attd(H)	prob	0.04	0.01	0.04	0.085	0.195
Symp - Attd(H)	dt-fixed			1		
Attd(H) - Hosp	prob			1		
Attd(H) - Hosp	dt-mean	5	5	5	5.3	4.2
Attd(H) - Hosp	dt-std dev	4.6	4.6	4.6	5.2	5.2
Hosp - Recovered	prob,			0.2		
Hosp - Recovered	dt-mean	3.1	3.1	3.1	7.8	6.5
Hosp - Recovered	dt-std dev	3.7	3.7	3.7	6.3	4.9
Hosp - Vent	prob	0.06	0.06	0.06	0.15	0.225
Hosp - Vent	dt-mean			1		
Hosp - Vent	dt-std dev			0.2		
Vent - Recovered	prob			1		
Vent - Recovered	dt-mean	2.1	2.1	2.1	6.8	5.5
Vent - Recovered	dt-std dev	3.7	3.7	3.7	6.3	4.9

Table 6: Disease progression parameters as given by the CDC document [15]. One value per line applies to all age groups. Abbreviations: prob: probability, dt: dwell time, Attd: attended, Hosp: hospitalized, Vent: ventilated, (D): resulting in death, and (H): resulting in hospitalization

H Related Work

There has been a lot of work on analyzing interventions to control epidemics, and this falls into two broad categories. The first involves using a system of coupled differential equations to represent the dynamics, e.g., [42, 43, 55, 69, 72]. Even though closed-form solutions are not available even for simple models, when the system is not very large, it can be solved by brute-force local search methods, e.g., [43]. For some types of models, greedy strategies have been used [68, 72]. The second class of methods is network or agent based, of the form we study here, e.g., [23, 29, 31, 38, 40].

Analyzing interventions to minimize the expected outbreak size (or to optimize other epidemic outcomes) in network models is much harder, and is generally open to problems. Prior work has generally attempted to solve these problems by either simplifying the network (e.g., assuming random graph models), or simplifying the disease model. The simplest setting is that of transmission probability of 1 (modeling a highly contagious disease), with a fixed source. Even this setting is challenging, and work by [24, 32] designs bicriteria approximation algorithms for this problem.

A variation of this setting is when the source is chosen randomly, and, in this case, the problem of minimizing the number of infections corresponds to deleting a subset of nodes such that the sum of squares of the component sizes in the residual network is minimized. A minor modification of the results of [6, 36] gives approximation algorithms for this objective. We note that [58] uses a stochastic optimization approach for minimizing the expected number of infections. While their worst case approximation factor can be quite large, their empirical performance is quite good. The work of [3, 10] on the robustness of networks can be viewed as interventions to reduce the spread of an outbreak.

It is well understood that the network structure has a significant impact on the dynamics of epidemic spread. This has motivated a lot of research on modifying network properties to control epidemic spread. One of the most studied properties is degree, and in many network models, as well as in a broad class of real world networks, it has been found that removing the highest degree nodes (equivalently, vaccinating high degree nodes) turns out to be very effective [3, 10, 19, 22, 51, 73]. Cohen et al [19] show that a simple decentralized strategy of “acquaintance immunization” has the effect of selecting high degree nodes. Another set of properties that has been studied extensively are spectral properties, namely the eigenvalues and eigenvectors associated with the adjacency matrix of the graph and its Laplacian. It has been shown using multiple approaches [27, 45, 52] that epidemic spread exhibits a threshold behavior—if the spectral radius (the largest absolute value of an eigenvalue) is below a certain threshold, the disease dies out. This has motivated a considerable amount of work on reducing the spectral radius to control the outbreak [46, 53, 54, 56, 62]. *In general, the theoretical studies do not apply to temporal vaccine allocation problems – in such cases the network is constantly changing as the epidemic spreads and vaccines are distributed in time.*

In the context of COVID-19, where we have multiple approved vaccine candidates, the role of vaccine efficacy, especially whether it reduces susceptibility to disease or transmission becomes important [37]. A recent study by Bubar et al. [12] identified that under different underlying assumptions, vaccine prioritization policies vary from 20-49 years to adults over 60 years old. They also note that prioritizing seronegative individuals could improve the marginal impact of a given policy. A similar study at a global scale using different supply assumptions was reported in [34]. See [1, 9, 26, 34, 41, 55, 59] for other recent papers on this topic. Multiple studies have also identified the tradeoffs based on the underlying policy objectives [13, 41] using compartmental models. The current allocation policy in the US at the federal level is centered around the framework developed by the National Academies of Sciences, Engineering, and Medicine (NASEM) [48].

Very few papers have studied vaccine allocation problems when there is a vaccine schedule (temporal vaccine allocation). Furthermore, they do not study how robust such methods are against uncertainty in estimating the structural properties; this is a crucial contribution of the present paper. Nevertheless, these results do suggest the potential value of such methods.

Digital apps to estimate network properties. Digital contact tracing apps have recently been deployed in several countries [2, 5, 17, 25, 35] with mixed success. Reasons for the range of outcomes include: (i) low penetration levels, (ii) compliance, and (iii) accuracy of the apps in discovering neighbors accurately. Our allocation method is based on exploiting simple network properties that can be estimated using digital devices. We use two measures here: (i) degree and (ii) weighted degree. Digital contact tracing apps can potentially measure both of these quantities quite accurately.

Boolean Games: Inferring Agents’ Goals Using Taxation Queries

Abhijin Adiga¹, Sarit Kraus², Oleg Maksimov² and S. S. Ravi¹

¹Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, VA, USA

²Department of Computer Science Bar Ilan University, Ramat Gan, Israel

{abhijin, ssravi}@virginia.edu, sarit@cs.biu.ac.il, oleg@maksimov.co.il

Abstract

In Boolean games, each agent controls a set of Boolean variables and has a goal represented by a propositional formula. We study inference problems in Boolean games assuming the presence of a PRINCIPAL who has the ability to control the agents and impose taxation schemes. Previous work used taxation schemes to guide a game towards certain equilibria. We present algorithms that show how taxation schemes can also be used to infer agents’ goals. We present experimental results to demonstrate the efficacy our algorithms. We also consider goal inference when only limited information is available in response to a query.

1 Introduction

Boolean games [Harrenstein *et al.*, 2001] are a class of games where agents’ goals are represented by a propositional logic formula. Each agent i controls a distinct set of Boolean variables Φ_i , and there is a cost associated with each assignment. Its formula or goal γ_i is composed of variables (set Γ_i) that are not necessarily in its control. Each agent’s first priority is to achieve its goal and its second priority is to minimize its total cost. Much of the work on Boolean games is of theoretical nature [Harrenstein *et al.*, 2001; Sauro and Villata, 2013; Wooldridge *et al.*, 2013; Grant *et al.*, 2014; Ågotnes *et al.*, 2013; Bonzon *et al.*, 2007; Levit *et al.*, 2013b]. Boolean games have been used to model some real-world problems such as charging electric vehicles and traffic signalling [Levit *et al.*, 2013b].

Recently, a number of works have emerged where a user either actively queries the system or uses passive observations to infer the functions in multi-agent systems [Kleinberg *et al.*, 2017; Adiga *et al.*, 2018; Narasimhan *et al.*, 2015; He *et al.*, 2016]. Wooldridge *et al.* [2013] introduced the notion of a PRINCIPAL – an external agent– who can influence the agents’ decisions through *taxation schemes* (additional costs for assigning values to variables) to achieve a desirable equilibrium. Under this framework, we study the inference problem where the PRINCIPAL’s objective is to infer the agents’ goals by repeatedly “querying” the system and observing the outcomes.

It is known that pure Nash equilibria (PNE) need not exist for every taxation scheme or there may be multiple PNE [Levit *et al.*, 2013b]. Such scenarios might make it impossible to infer some or all the goals. To overcome this problem, we allow the PRINCIPAL to *inhibit* some agents from achieving their goals. A *taxation query* (or simply a query) specifies (i) a subset of inhibited agents and (ii) an unambiguous taxation scheme (where costs for setting a variable to 0 and 1 are not equal). While this framework seems to provide an unrealistic amount of control to the PRINCIPAL, we show scenarios where without such control inference might not be possible. Our approach is to strategically inhibit some agents and infer the goals of uninhibited agents by simultaneously querying them using taxation schemes. To this end, we construct a graph representing dependencies between agents’ goals and apply vertex coloring to that graph. Then, for each query, the PRINCIPAL observes a PNE. The questions of primary interest here are: Does there exist a set of queries so that the PRINCIPAL can infer all goals? If yes, what is its size?

1.1 Summary of Results

Necessary and sufficient conditions for the existence of an NE for a taxation query. Using an undirected graph (called the goal overlap graph) that captures the overlaps between the sets of variables used in agents’ goals, we establish necessary and sufficient conditions for the existence of an NE for a Boolean game and any {0,1}-taxation query, i.e., a query with only 0 and 1 costs (Section 3).

Evaluating goals of selected agents at an NE. When there is an NE for a {0,1}-taxation query, we show that the value of the goals of selected agents for the zero cost assignment (i.e., the assignment whose cost is zero for a selected agent) can be determined regardless of which NE is reached by the agents (Section 3). We note that the results in this and the previous paragraph hold for any unambiguous taxation query. We use {0,1}-taxation queries for convenience.

Goal inference algorithms. Using the results in the above two paragraphs, we show that taxation queries can be used to infer agents’ goals. Our taxation queries play a role similar to membership queries used to learn Boolean functions (e.g., [Abasi *et al.*, 2014; Angluin and Slonim, 1994]). Further, using a valid node coloring of the goal overlap graph, we show that the goals of many agents can be inferred simultaneously. We point out that the coloring-based scheme can

be significantly better than inferring the goals one at a time. We also obtain more efficient (in terms of the number of taxation queries) goal inference algorithms for two special classes of goal functions, namely threshold functions and symmetric functions (Section 4).

Experimental results. We show experimentally that our coloring-based inference algorithm uses significantly fewer queries compared to inferring goals one at a time. Further, the coloring-based approach also uses significantly less time even for games with 36,000 agents (Section 5).

Inference with limited information. We show that goal inference is possible in a context where the response to a query only provides the values assigned to the variables and not whether agents achieved their goals. In particular, we discuss this result when each goal is a threshold function (Section 6).

For space reasons, most proofs are sketched or omitted; detailed proofs can be found in [Adiga *et al.*, 2020].

1.2 Related Work

Harrenstein *et al.* [2001] introduced Boolean games as a class of two-player games and Bonzon *et al.* [2007] generalized the framework to n players. Structural and computational properties of PNE in Boolean games are well-studied. Bonzon *et al.* [2007] define a “dependency graph” between players to characterize PNE in Boolean games, much like our goal overlap graph. Levit *et al.* [2019; 2013b] study methods for finding a taxation scheme that incentivizes the agents to reach a stable state. Also, Levit *et al.* [2013a] discuss an application of Boolean games to the charging of electric vehicles where some vehicles are not allowed to charge at certain time intervals to avoid overloading. This is similar to our notion of the PRINCIPAL inhibiting agents. Boolean games where players have incomplete information about each other’s goals have also been considered in the literature (e.g., [Clercq *et al.*, 2015; Ågotnes *et al.*, 2013]).

2 Definitions and Preliminaries

Components of a Boolean Game. We follow [Wooldridge *et al.*, 2013] in defining a Boolean game. Let $V = \{1, 2, \dots, n\}$ be a collection of n agents. Let Φ be a finite set of Boolean variables. γ_i is the Boolean function (a propositional formula over the variables in Φ) that represents the goal of agent i ($1 \leq i \leq n$). $\Gamma_i \subseteq \Phi$ denotes the set of variables in γ_i . Φ is assumed to be partitioned into n sets $\Phi_1, \Phi_2, \dots, \Phi_n$, where Φ_i denotes the subset of variables controlled by agent i ($1 \leq i \leq n$). Some of the subsets in the collection may be empty; the corresponding agents do not control any of the variables in Φ . Since the collection of subsets of Φ is pairwise disjoint, each variable in Φ is controlled by exactly one agent. We use \mathbb{B} to denote the set $\{0, 1\}$.

In addition to the set of agents V , it is assumed that there is a special agent (external to the game) called the PRINCIPAL. There is a special Boolean variable, called the **inhibitor variable**, ψ_i for each agent i , $1 \leq i \leq n$. (The inhibitor variables are *not* in Φ .) For any agent i , the PRINCIPAL can set ψ_i to 0 to *inhibit* agent i . The behaviors of inhibited and uninhibited agents are discussed later in this section.

Taxation queries and Nash equilibrium. The PRINCIPAL can influence the goal of any agent i in two ways. Firstly, it can set the inhibitor variable ψ_i . Secondly, as discussed by Wooldridge *et al.* [2013], it can choose a **taxation scheme** τ that specifies the tax values, denoted by $\tau(x = 0)$ and $\tau(x = 1)$, for setting each variable x in Φ to 0 and 1 respectively. While our results hold for any unambiguous taxation scheme (where assignments 0 and 1 have different costs), for convenience, we will use the $\{0, 1\}$ -*taxation scheme* defined by Wooldridge *et al.* [2013]. A taxation scheme τ is a $\{0, 1\}$ -**taxation scheme** if for all $x \in \Phi$ and $b \in \mathbb{B}$, (i) $\tau(x = b) \in \{0, 1\}$ and (ii) $\tau(x = b) = 1 - \tau(x = \neg b)$.

Our algorithms for goal inference rely on **taxation queries**. Formally, a **taxation query** q is a pair (\mathbb{I}, τ) , where \mathbb{I} is the **inhibition vector** that specifies the value of the inhibitor variable ψ_i for each agent i and τ is the taxation scheme. For each taxation query $q = (\mathbb{I}, \tau)$, the response of an agent i is to set each variable in Φ_i to 0 or 1. The **cost** for agent i is the sum of the tax values over all the variables in Φ_i . Each agent i sets the variables in Φ_i in the following manner.

- (a) If agent i is inhibited by the PRINCIPAL, then the agent sets each variable in Φ_i to the value with the lower tax.
- (b) If agent i is *not* inhibited by the PRINCIPAL, then it proceeds as follows.
 - (i) Agent i always prefers assignments to the variables in Φ_i that satisfy the goal γ_i over those which don’t satisfy the goal (given other agents’ assignments).
 - (ii) When there are two or more assignments that satisfy the goal, the agent prefers one with the minimum cost. (If there are two or more minimum cost assignments that satisfy the goal, the agent chooses one of them arbitrarily.)
 - (iii) If none of the assignments to the variables in Φ_i satisfies the goal γ_i , then agent i chooses a (non-satisfying) assignment with the minimum cost.

Definition (Nash Equilibrium). An assignment to the variables in Φ is a pure **Nash Equilibrium** (NE) if no agent i has an incentive to *unilaterally* change the assignment it has chosen for the variables in Φ_i .

Response to a taxation query. For each taxation query q , if the agents reach an NE, the PRINCIPAL receives a response consisting of two items: (i) for each uninhibited agent i , whether i achieved its goal and (ii) the value assigned to each variable $x \in \Phi$. If there are two or more equilibria for a taxation query, the agents may choose any one of them. If there is no NE, the system produces the response “No equilibrium” and does not provide any other information.

We now present examples to illustrate the above concepts.

Example 1. Consider a Boolean game with two agents ($\Phi = \{x, y\}$), and $\Phi_1 = \{x\}$ and $\Phi_2 = \{y\}$. Let $\gamma_1 = x \vee y$ and $\gamma_2 = x \oplus y$, where ‘ \oplus ’ is the exclusive-or operator. Let the PRINCIPAL choose the following $\{0, 1\}$ -taxation scheme τ : $\tau(x = 1) = \tau(y = 1) = 0$. (Thus, $\tau(x = 0) = \tau(y = 0) = 1$.) We consider two inhibition vectors with the taxation scheme τ to show that they may lead to different NE.

- (i) Suppose the PRINCIPAL does not inhibit either of the agents. The assignment $x = y = 0$ is not an NE for this setting since agent 1 has the incentive to change x to 1 to achieve its goal. Likewise, neither of the two assignments

$(x = 0, y = 1)$ and $(x = 1, y = 1)$ is an NE; in the former, agent 1 has an incentive to change x to 1 to reduce its cost while in the latter, agent 2 has an incentive to change y to 0 to achieve its goal. However, the assignment $(x = 1, y = 0)$ is an NE for this setting.

(ii) Suppose the PRINCIPAL inhibits only agent 2. Since agent 2 is inhibited, it sets y to 1 as that choice makes the cost of agent 2 to be zero. Given agent 2's choice for y , agent 1 achieves its goal by setting $x = 1$ since that choice incurs the minimum cost (namely, zero). Thus, the only NE for this situation is $(x = 1, y = 1)$, which is different from the one where neither agent is inhibited. \square

Example 2: The purpose of this example is to show that there are Boolean games which have no NE for some taxation schemes. The three agent game is defined in Table 1. Suppose

Agent	Goal	Control set
1	$x_1 \vee x_3$	$\Phi_1 = \{x_1\}$
2	$x_1 \vee x_2$	$\Phi_2 = \{x_2\}$
3	$x_2 \vee x_3$	$\Phi_3 = \{x_3\}$

Table 1. A game without a NE for some taxation schemes.

the PRINCIPAL chooses the taxation query q which does *not* inhibit any agent and uses the following taxation scheme: for each variable x_i , $\tau(x_i = 0) = 0$ and $\tau(x_i = 1) = 1$, $1 \leq i \leq 3$. The reader can verify that none of the 8 possible choices of values for the three variables is an NE. Thus, there is no NE for this example. \square

Statement of the Goal Inference Problem. The PRINCIPAL knows the number of agents n and for each agent i , the sets Φ_i and Γ_i (i.e., the set of variables used in the goal γ_i). The PRINCIPAL's objective is to infer the goal function γ_i of each agent i . (a) To that end, the PRINCIPAL issues taxation queries to the system a number of times. (b) For each query, the principal receives a response as described earlier. (c) The PRINCIPAL should use as few queries as possible to infer all the goals.

Challenges in developing goal inference algorithms. Algorithms to infer agents' goals must deal with two primary challenges. First, such an algorithm must avoid using queries for which there is no NE. Second, for a chosen taxation query, there may be multiple equilibria and the agents may arbitrarily choose one of them. Thus, the algorithm must be able to extract useful information regardless of which NE is chosen. We now present an example to illustrate this situation.

Example 3. This example shows that a Boolean game may have multiple equilibria. The game has two agents and $\Phi = \{x_1, x_2\}$. Each agent i controls x_i . Their goals are identical: $\gamma_1 = \gamma_2 = x_1 \leftrightarrow x_2$. Suppose both agents are uninhibited. It is easy to see that for any taxation scheme, $x_1 = x_2 = 0$ and $x_1 = x_2 = 1$ are both Nash equilibria. In both cases, each agent's goal is satisfied. Further, if agent i changes the value of x_i , then it will not achieve its goal. For every taxation scheme, if the agents choose $x_1 = x_2 = 0$ as the response, then the PRINCIPAL will never know the response for other assignments. Thus, an inference algorithm must carefully choose inhibition vectors. \square

2.1 Additional Definitions

Definition 1 (Zero cost assignment). *Given a $\{0,1\}$ -taxation scheme τ and agent i , the zero cost assignment with respect to a set $X_i \subseteq \Phi_i$ of variables is obtained by assigning for each $x \in X_i$ the value $b \in \mathbb{B}$ such that $\tau(x = b) = 0$. (Thus, the total cost of this assignment over all the variables in X_i is zero.)*

The following definition of a *goal overlap graph* plays an important role in our inference methods.

Definition 2 (Goal overlap graph). *For any Boolean game, the goal overlap graph $G_O(V, E)$ is an undirected graph constructed as follows. Each node in V represents an agent. For any two distinct agents i and j , if $\Gamma_i \cap \Gamma_j \neq \emptyset$, then E contains the edge $\{i, j\}$.*

We assume that the reader is familiar with graph theoretic concepts such as independent set and (node) coloring (see e.g., [West, 2003]). The smallest integer k for which a graph G is k -colorable is called its **chromatic number** $\chi(G)$. In any valid k -coloring, the set of nodes which are assigned the same color constitute a **color class**. Each color class induces an independent set in G . Thus, a valid k -coloring of G partitions the node set V into k independent sets.

A Preliminary Lemma. A useful property of $\{0,1\}$ -taxation queries for which there is an NE is shown below. (The proof is omitted due to space reasons.)

Lemma 1. (Control of variables using taxation queries) *Suppose there is at least one NE for a $\{0,1\}$ -taxation query. Then, the following statements hold for every agent i in every NE. (1) If agent i is inhibited, then i chooses the zero cost assignment with respect to the set of variables Φ_i . (2) If agent i is uninhibited, then i chooses the zero cost assignment with respect to the set of variables $\Phi_i - \Gamma_i$.*

3 Existence of Equilibria Under Taxation Schemes with Inhibition

In this section, we establish necessary and sufficient conditions for the existence of Nash equilibria under taxation schemes that inhibit certain agents.

A sufficient condition for NE. Our sufficient condition can be stated formally as follows.

Theorem 1. *Consider a Boolean game with goal overlap graph $G_O(V, E)$. Let $W \subseteq V$ be any nonempty independent set of G_O . Let an agent be uninhibited if and only if it belongs to W . For every $\{0,1\}$ -taxation scheme τ under this setting, the following results hold.*

1. **(Existence)** There exists an NE.
2. **(Consistency)** In all the NE, the outcomes are identical; that is, for any agent i and any pair of NE \mathcal{E}_1 and \mathcal{E}_2 , agent i achieves its goal in both \mathcal{E}_1 and \mathcal{E}_2 or does not achieve its goal in both \mathcal{E}_1 and \mathcal{E}_2 .
3. **(Ability to evaluate)** For every agent $i \in W$, it is possible to evaluate whether its goal is satisfied by the zero cost assignment with respect to the set of variables Γ_i .

Proof (sketch). For space reasons, we will only sketch the proof for Part 1 of the theorem. We construct an assignment α of 0/1 values to the variables in Φ and prove that it corresponds to an NE. First, for any inhibited agent i (i.e., $i \in V - W$), we choose the zero cost assignment with respect to Φ_i . Now, consider each uninhibited agent i (i.e., $i \in W$), and let i choose the zero cost assignment with respect to the variables in $\Phi_i - \Gamma_i$. It can be shown that for each agent i , such an assignment sets each variable $x \in \Gamma_i - \Phi_i$ (which are controlled by agents other than i) to the value of zero cost. Now, agent i considers each combination of values to the variables in $\Phi_i \cap \Gamma_i$. If none of these combinations satisfies the goal γ_i , it chooses the zero cost assignment with respect to $\Phi_i \cap \Gamma_i$. Otherwise, it chooses an assignment that satisfies its goal and has the lowest cost among all such assignments. It can be shown that this assignment is an NE. ■

A necessary condition for NE. In proving Part 1 of Theorem 1, we assumed that the uninhibited agents form an independent set in the goal overlap graph and that some agents are inhibited. Here, we point out that these two assumptions are necessary to ensure the existence of an NE.

Proposition 1. *There are Boolean games that satisfy the following properties: (i) If the set of uninhibited agents do not form an independent set in the goal overlap graph, then there is no NE for any taxation scheme. (ii) If no agent is inhibited, then there is no NE for any taxation scheme.*

Proof sketch. We use the same Boolean game to prove both parts of this result. Consider a game with two agents denoted by 1 and 2. Let $\Phi = \{p, q\}$ denote the set of variables. Further, let $\Phi_1 = \{p\}$ and $\Phi_2 = \{q\}$. Let γ_1 be given by $p \leftrightarrow q$ and γ_2 be given by $\neg(p \leftrightarrow q)$. Thus, $\Gamma_1 = \Gamma_2 = \{p, q\}$. Assume that neither of the agents is inhibited. It can be seen that the set of agents $\{1, 2\}$ does not form an independent set. A simple case analysis can be used to argue that regardless of the taxation scheme, there is no NE for this system. The result of Part (ii) also follows since in the above example neither of the agents is inhibited. ■

Proposition 1 points out that there are games in which if the uninhibited agents do not form an independent set in the goal overlap graph, no taxation query has an NE. Therefore, taxation queries in which uninhibited agents do not form an independent set may be wasteful. Moreover, Theorem 1 points out that for any game and taxation scheme, if the uninhibited agents form an independent set, then every taxation query has an NE. Accordingly, our goal inference algorithms choose such taxation queries.

Ability to query many agents simultaneously. Our algorithms reduce the number of taxation queries by querying many agents simultaneously. The following lemma shows how this can be done when the uninhibited agents form an independent set in the goal overlap graph. (The proof is omitted for space reasons.)

Lemma 2. (Simultaneous querying) *Consider a Boolean game with goal overlap graph $G_O(V, E)$. Further, let $W \subseteq V$ be a nonempty independent set in G_O . Let an agent be uninhibited if and only if it belongs to W . There exists a $\{0,1\}$ -taxation scheme τ such that for every agent $i \in W$ and any*

given assignment g_i for the variables in Γ_i , one can decide whether g_i satisfies agent i 's goal γ_i using τ .

4 Goal Inference Algorithms

Suppose \mathcal{A} is an algorithm that learns a Boolean function f of r variables using an oracle in the following manner. For each input α to f , suppose the oracle produces the value $f(\alpha)$. Algorithm \mathcal{A} infers the function f from the values provided by the oracle. We now explain how a Boolean game can simulate an oracle that provides the value $f(\alpha)$ for each assignment α to the variables of f .

Consider the Boolean game where there is only one agent whose goal f is unknown. The variables of f are the variables in the game and the single agent controls all the variables. For this game, any input α to the function f can be converted into a $\{0,1\}$ -taxation query $q = (\mathbb{I}, \tau)$ as follows. The inhibition vector \mathbb{I} of q does not inhibit the agent. If α sets a variable x to the Boolean value b , then the $\{0,1\}$ -taxation scheme τ sets $\tau(x = b) = 0$. This ensures that α is the zero cost assignment under this taxation scheme. Since the uninhibited set consisting of a single agent is trivially an independent set, by Theorem 1, there is at least one NE for the taxation query q . Moreover, by Part (3) of Theorem 1, no matter which NE is reached, we can obtain the value of the goal function f for the zero cost assignment α . Thus, the Boolean game simulates the oracle that provides the value $f(\alpha)$ for the input α . In this manner, the Boolean game in conjunction with Algorithm \mathcal{A} can be used to infer the function f . Thus:

Proposition 2. *If there is an algorithm to learn a Boolean function f using β queries to an oracle, then there is an algorithm that infers the goal function f using β $\{0,1\}$ -taxation queries. ■*

Simultaneous inference of multiple goals. Suppose we have a set of m agents in a Boolean game such that the goal overlap graph for these agents is an independent set. Suppose for each goal function γ_i , we have an algorithm that can learn γ_i using q_i queries to an oracle. As discussed above, each of these queries can be translated into a $\{0,1\}$ -taxation query. Further, using Lemma 2, the $\{0,1\}$ -taxation queries for the m agents can be combined into a single $\{0,1\}$ -taxation query for the Boolean game. Since the agents form an independent set in the goal overlap graph, by Part (3) of Theorem 1, we can obtain the values of all the goal functions from the single taxation query. Thus, if $q^* = \max\{q_1, q_2, \dots, q_m\}$, then the number of $\{0,1\}$ -taxation queries used to infer all the m goal functions is q^* . This is an improvement over the simple method of inferring goals one at a time, for which the number of $\{0,1\}$ -taxation queries used is $\sum_{i=1}^m q_i$; in the worst-case, this sum can be as large as mq^* .

We can extend the above idea of simultaneous inference to an arbitrary Boolean game as follows. Let G_O denote the goal overlap graph for the game. Let χ denote the minimum number of colors needed to obtain a node coloring of G_O . Recall from Section 2 that each of the χ color classes forms an independent set. As discussed above, we can infer the goals of all the agents in a color class simultaneously. If q^* denotes the maximum number of $\{0,1\}$ -taxation queries used to infer the goal of any agent, then the number of $\{0,1\}$ -taxation queries

used by this coloring-based method is at most χq^* . We state this result formally below.

Proposition 3. *Let G_O denote the goal overlap graph for a Boolean game and let χ denote the chromatic number of G_O . Let q^* denotes the maximum number of $\{0,1\}$ -taxation queries used to infer the goal of any agent. Then, the number of $\{0,1\}$ -taxation queries used to infer the goals of all the agents is at most χq^* .* ■

Computing the chromatic number of an undirected graph is an **NP-hard** problem [Garey and Johnson, 1979]. In practice, one can use heuristics to obtain node colorings with a small number of colors. For example, any graph whose maximum node degree is Δ can be efficiently colored using $\Delta + 1$ colors using Brooks' theorem [West, 2003]. Such a method is useful when the maximum node degree is small.

Special classes of Boolean functions. Here, we present inference algorithms when the goals of all the agents are from special classes of Boolean functions. For these classes, the goal inference algorithms use only a small number of $\{0,1\}$ -taxation queries. In particular, we consider two special classes, namely **threshold functions** and **symmetric functions**. We begin with the definitions of these classes of functions (see, e.g., [Kohavi, 1970; Crama and Hammer, 2011]). For any Boolean function with Boolean inputs, the number of 1's in the input is called the **Hamming weight** of that input.

Definition 3. *Let $f(x_1, x_2, \dots, x_r)$ be a Boolean function of r Boolean variables. Function f is **symmetric** if the value of f depends only on the Hamming weight of the input.*

Example 4. Consider the Boolean function $f(x_1, x_2, x_3) = x_1 \oplus x_2 \oplus x_3$, where ' \oplus ' is the exclusive-or operator. This function is symmetric since it has the value 1 iff the Hamming weight of the input is odd. □

Definition 4. *Let $f(x_1, x_2, \dots, x_r)$ be a Boolean function of r Boolean variables. For each integer $k \geq 0$, f is a k -**threshold function** iff the value of f is 1 when the Hamming weight of the input is at least k .*

Thus, any k -threshold function is also a symmetric function. We note that the 0-threshold function is the constant function which has the value 1 for all inputs. If a function f with r inputs is the constant function which has the value 0 for all inputs, we will regard f as the $(r+1)$ -threshold function. For space reasons, we will discuss only the inference algorithm for threshold functions; a similar algorithm can be designed for symmetric functions.

Infering threshold functions. Suppose the goal of each agent i in a Boolean game is the k_i -threshold function for some unknown integer k_i , $0 \leq k_i \leq |\Gamma_i| + 1$. The objective of the inference problem in this case is to determine the value of k_i for each agent a_i , $1 \leq i \leq n$. The algorithm that learns the threshold value of a single agent through queries uses a simple binary search over the range $[0..r+1]$, where r is the number of inputs to the function; thus, the number of queries used is $O(\log r)$. Each query specifies an input with a certain Hamming weight, determined by the binary search procedure. Combining this binary search algorithm with the coloring of G_O and simultaneous inference of thresholds of

Algorithm 1: Inference algorithm for threshold functions.

```

1 Construct the goal overlap graph  $G_O(V, E)$ .
2 Obtain an optimal node coloring for  $G_O$ . Let  $\chi$  denote
   the number of colors used and let  $V_1, V_2, \dots, V_\chi$ 
   denote the color classes themselves.
3 for  $j = 1$  to  $\chi$  do
4   Inhibit every agent in  $V - V_j$ .
5   Using a simultaneous binary search over the agents
      in  $V_j$ , get the threshold of each agent in  $V_j$ .
6 end
7 Output the threshold value of each agent.

```

all the agents in a color class (as discussed above), we obtain the inference procedure which is shown in pseudocode form as Algorithm 1.

Let Γ^* denote a set of maximum cardinality among $\Gamma_1, \Gamma_2, \dots, \Gamma_n$. Thus, the maximum number of taxation queries used for any color class is $O(\log(|\Gamma^*|))$. As a consequence, the number of taxation queries used to infer all the threshold values is $O(\chi \log(|\Gamma^*|))$. We state this result formally below.

Proposition 4. *Suppose the goal of each of the n agents in a Boolean game is a threshold function. Let χ denote the chromatic number of the goal overlap graph and let Γ^* denote a set of maximum cardinality among $\Gamma_1, \Gamma_2, \dots, \Gamma_n$. The threshold values of all the agents can be inferred using $O(\chi \log(|\Gamma^*|))$ taxation queries.* ■

5 Experimental Results

The aim of the experiments is to study two aspects: (a) how the coloring idea reduces the number of queries (compared to inferring the goals one node at a time) and (b) scalability of the approach. Our experimental procedures and results discussed below are for goals which are threshold functions. We use SEQ to denote the (sequential) querying algorithm that infers goals one node at a time and CBQ to denote the coloring-based querying algorithm.

Part (a): Effectiveness of querying using node coloring. We ran an extensive set of experiments creating a large number of networks and agents' goals by varying the number of nodes, the minimum degree, the number of variables in an agent's goal controlled by the agent and by others and the threshold. In each network, nodes represent agents and the network itself represents the goal overlap graph. We ensured this by creating for each edge $\{u, v\}$, a variable $x_{u,v}$ that appears *only* in the goals of u and v . To make the game general, we also created other variables and generated control sets for nodes so that for each node (agent) i , the sets Φ_i and Γ_i have a nonempty intersection. For each agent i , we generated a random threshold value in the range $[1..|\Gamma_i|]$.

Let us first consider the SEQ approach which infers thresholds one node at a time. For each node i , our program finds its threshold by inhibiting all other nodes and constructing $\{0,1\}$ -taxation queries to do binary search over the possible threshold values of i . Thus, we can compute the total number of queries used by SEQ. Next, we generated a Brooks coloring of the graph using the simple greedy algorithm [West,

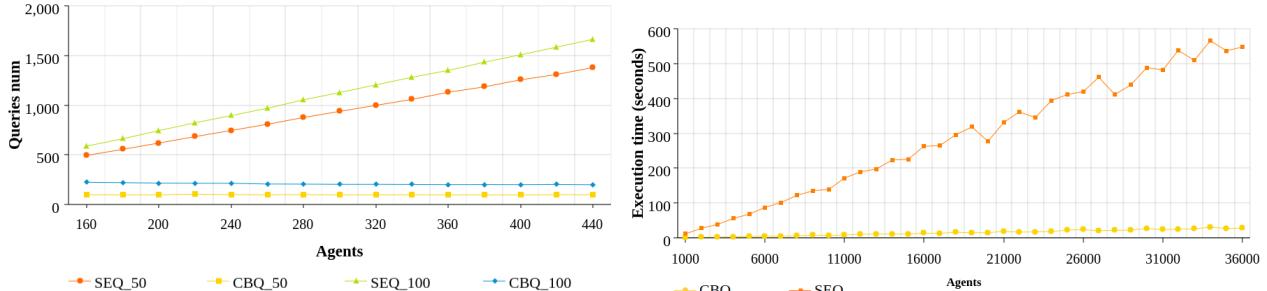


Figure 1. (a) The number of queries as a function of the number of agents for the SEQ algorithm (red and green lines) vs CBQ algorithm (yellow and blue lines) for graphs with minimum node degree of 50 (red and yellow) and minimum degree of 100 (green and blue). (b) Time in seconds as a function of the number of agents for graphs with minimum node degree of 50.

2003]. This method uses $\Delta + 1$ colors, where Δ is the maximum node degree. Now, for each color class, we used queries that simultaneously find the threshold values of all the nodes in the color class. This allows us to compute the total number of queries used by CBQ. For space reasons, we only present a few results from our experiments. In Figure 1(a), we varied the number of agents between 100 and 440 and considered two values for the minimum node degree, namely 50 and 100. We report the number of queries for the two algorithms and the two minimum degree values. It is easy to see from Figure 1(a) that for the sequential algorithm (SEQ), the number of queries increases with the number of agents (red and green lines). However, in the coloring-based algorithm (CBQ), the number of queries depends on the number of color classes, which is closely related to the maximum degree and not to the number of agents (yellow and blue lines). It is important to note that in these experiments, the generation of the Brooks coloring of a graph took on average only 3298 microseconds for graphs with minimum degree of 50 and 5935 microseconds for graphs with minimum degree of 100.

Part (b): Scalability of the coloring-based approach. In Figure 1(b) we demonstrate the scalability of CBQ. We varied the number of nodes between 1,000 and 36,000 and report the computation time for graphs with minimum degree of 50 and under the assumption that there are 20 processors that can be used for running the searches of the same color class, in parallel. It can be seen that the computation time of CBQ increases only at a low rate with the number of agents, while the time for the SEQ algorithm increases at a much higher rate.

6 Inference with Limited Information

Here, we observe that inference is possible for some special classes of goals even when the agents don't indicate whether they achieved their goals. For space reasons, we discuss only an algorithm for inferring thresholds. A similar algorithm can be designed when goals are conjunctions of literals.

Suppose the goal γ_i of each agent i is a threshold function and the threshold value θ_i of agent i is in the range $[1 \dots |\Gamma_i|]$. We also assume that of the $|\Gamma_i|$ variables used in γ_i , at least $q \geq 1$ variables are also in Φ_i ; that is, $q = |\Gamma_i \cap \Phi_i| \geq 1$. For simplicity, we will discuss the algorithm for inferring the threshold θ_i , assuming that all other agents are inhibited.

(The algorithm can be extended to simultaneous inference of goals using the idea of coloring.) The steps are as follows.

1. Issue a $\{0, 1\}$ -taxation query where $\tau(x = 1) = 1$ for all the variables. From the resulting assignment, find the the number ℓ of variables in $\Gamma_i \cap \Phi_i$ which are set to 1. If $\ell \geq 1$, then we conclude that the threshold of agent i must be ℓ (since this is the minimum cost for achieving the goal γ_i) and stop.
2. (Here, the value of ℓ from Step 1 is zero; that is, $q < \theta_i \leq |\Gamma_i|$.) Consider the variables in $A_i = \Gamma_i - \Phi_i$. For each j , $1 \leq j \leq |A_i|$, let the $\{0, 1\}$ -taxation query \mathbb{I}_j be constructed as follows: (i) choose an arbitrary subset A' of j variables from A_i and (ii) for each variable x in A' let $\tau(x = 1) = 0$ and for all other variables, let $\tau(x = 1) = 1$. Using these queries, find the *smallest* j (using binary search) such that for the query \mathbb{I}_j , the response sets *all* the variables in $\Gamma_i \cap \Phi_i$ to 1. Then, it can be seen that the threshold value $\theta_i = q + j$.

When the response doesn't indicate whether agents achieved their goals, one cannot distinguish between the threshold value 0 (i.e., the goal function has the value 1 for all inputs) and threshold value ∞ (i.e., the goal function has the value 0 for all inputs). In both cases, regardless of the taxation query, the uninhibited agent i would set all the variables in Φ_i to the value of minimum cost and no useful conclusion can be drawn from the outcome.

7 Future Work

It will be useful to develop inference algorithms for other goal functions. While we established upper bounds on the number of queries used for inference, it will be of interest to develop appropriate lower bounds. Finally, it is also of interest to identify other classes of goal functions that can be inferred with limited information.

Acknowledgments

We thank the IJCAI 2020 reviewers for their comments. This work was partially supported by NSF Grants ACI-1443054 (DIBBS), IIS-1633028 (BIG DATA), CMMI-1745207 (EAGER), OAC-1916805, IIS-1908530 and by the Ministry of Science & Technology, Israel and the Ministry of Education, Science, Research and Sport of the Slovak Republic.

References

- [Abasi *et al.*, 2014] Hasan Abasi, Nader H. Bshouty, and Hanna Mazzawi. On exact learning monotone DNF from membership queries. *CoRR*, abs/1405.0792:1–16, 2014.
- [Adiga *et al.*, 2018] Abhijin Adiga, Chris J. Kuhlman, Madhav V. Marathe, S. S. Ravi, Daniel J. Rosenkrantz, and Richard E. Stearns. Learning the behavior of a dynamical system via a “20 questions” approach. In *Thirty second AAAI Conference on Artificial Intelligence*, pages 4630–4637, Palo Alto, CA, 2018. AAAI Press.
- [Adiga *et al.*, 2020] Abhijin Adiga, Sarit Kraus, Oleg Maksimov, and S. S. Ravi. Boolean Games: Inferring Agents’s Goals Using Taxation Queries. Tech Report, Biocomplexity Insitute and Initiative, University of Virginia, Charlottesville, VA, USA, 2020.
- [Ågotnes *et al.*, 2013] Thomas Ågotnes, Paul Harrenstein, Wiebe Van Der Hoek, and Michael Wooldridge. Verifiable equilibria in Boolean games. In *Proc. of AAAI*, pages 689–695, 2013.
- [Angluin and Slonim, 1994] Dana Angluin and Donna K. Slonim. Randomly fallible teachers: Learning monotone DNF with an incomplete membership oracle. *Machine Learning*, 14(1):7–26, 1994.
- [Bonzon *et al.*, 2007] Elise Bonzon, Marie-Christine Lagasquie-Schiex, and Jérôme Lang. Dependencies between players in Boolean games. In *European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 743–754, Hidelberg, Germany, 2007. Springer.
- [Clercq *et al.*, 2015] Sophie De Clercq, Steven Schockaert, Ann Nowé, and Martine De Cock. Multilateral negotiation in Boolean games with incomplete information using generalized possibilistic logic. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*, pages 2890–2896, Palo Alto, CA, 2015. AAAI Press.
- [Crama and Hammer, 2011] Yves Crama and Peter L. Hammer. *Boolean Functions: Theory, Algorithms, and Applications*. Cambridge University Press, New York, NY, 2011.
- [Garey and Johnson, 1979] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-completeness*. W. H. Freeman and Co., San Francisco, CA, 1979.
- [Grant *et al.*, 2014] John Grant, Sarit Kraus, Michael Wooldridge, and Inon Zuckerman. Manipulating games by sharing information. *Studia Logica*, 102(2):267–295, 2014.
- [Harrenstein *et al.*, 2001] Paul Harrenstein, Wiebe van der Hoek, John-Jules Meyer, and Cees Witteveen. Boolean games. In *Proceedings of the 8th conference on Theoretical aspects of rationality and knowledge*, pages 287–298, Burlington, MA, 2001. Morgan Kaufmann Publishers Inc.
- [He *et al.*, 2016] Xinran He, Ke Xu, David Kempe, and Yan Liu. Learning influence functions from incomplete observations. In *Advances in Neural Information Processing Systems*, pages 2073–2081, San Diego, CA, 2016. Neural Information Systems Processing Foundation.
- [Kleinberg *et al.*, 2017] Jon Kleinberg, Sendhil Mullainathan, and Johan Ugander. Comparison-based choices. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 127–144, New York, NY, 2017. ACM.
- [Kohavi, 1970] Zvi Kohavi. *Switching and Finite Automata Theory*. McGraw-Hill, New York, NY, 1970.
- [Levit *et al.*, 2013a] Vadim Levit, Tal Grinshpoun, and Amnon Meisels. Boolean games for charging electric vehicles. In *Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 02*, pages 86–93. IEEE Computer Society, 2013.
- [Levit *et al.*, 2013b] Vadim Levit, Tal Grinshpoun, Amnon Meisels, and Ana LC Bazzan. Taxation search in Boolean games. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 183–190, Richland, SC, USA, 2013. International Foundation for Autonomous Agents and Multiagent Systems.
- [Levit *et al.*, 2019] Vadim Levit, Zohar Komarovsky, Tal Grinshpoun, Ana LC Bazzan, and Amnon Meisels. Incentive-based search for equilibria in Boolean games. *Constraints*, 24:288–319, 2019.
- [Narasimhan *et al.*, 2015] Harikrishna Narasimhan, David C. Parkes, and Yoran Singer. Learnability of influence in networks. In *Advances in Neural Information Processing Systems*, pages 3186–3194, San Diego, CA, 2015. Neural Information Systems Processing Foundation.
- [Sauro and Villata, 2013] Luigi Sauro and Serena Villata. Dependency in cooperative Boolean games. *Journal of Logic and Computation*, 23(2):425–444, 2013.
- [West, 2003] Douglas West. *Introduction to Graph Theory*. Prentice-Hall, Inc., Englewood Cliffs, NJ, 2003.
- [Wooldridge *et al.*, 2013] Michael Wooldridge, Ulle Endriss, Sarit Kraus, and Jérôme Lang. Incentive engineering for Boolean games. *Artificial Intelligence*, 195:418–439, 2013.



Networked experiments and modeling for producing collective identity in a group of human subjects using an iterative abduction framework

Vanessa Cedeno-Mieles^{2,4,6} · Zhihao Hu⁵ · Yihui Ren⁷ · Xinwei Deng⁵ · Abhijin Adiga¹ · Christopher Barrett^{1,3} · Noshir Contractor⁹ · Saliya Ekanayake¹² · Joshua M. Epstein¹⁰ · Brian J. Goode² · Gizem Korkmaz¹ · Chris J. Kuhlman¹ · Dustin Machi¹ · Michael W. Macy¹¹ · Madhav V. Marathe^{1,3} · Naren Ramakrishnan^{4,8} · S. S. Ravi¹ · Parang Saraf⁸ · Nathan Self⁸

Received: 6 March 2019 / Revised: 9 December 2019 / Accepted: 16 December 2019 / Published online: 7 January 2020
© Springer-Verlag GmbH Austria, part of Springer Nature 2020

Abstract

Group or collective identity is an individual's cognitive, moral, and emotional connection with a broader community, category, practice, or institution. There are many different contexts in which collective identity operates, and a host of application domains where collective identity is important. Collective identity is studied across myriad academic disciplines. Consequently, there is interest in understanding the collective identity formation process. In laboratory and other settings, collective identity is fostered through priming a group of human subjects. However, there have been no works in developing agent-based models for simulating collective identity formation processes. Our focus is understanding a game that is designed to produce collective identity within a group. To study this process, we build an online game platform; perform and analyze controlled laboratory experiments involving teams; build, exercise, and evaluate network-based agent-based models; and form and evaluate hypotheses about collective identity. We conduct these steps in multiple abductive iterations of experiments and modeling to improve our understanding of collective identity as this looping process unfolds. Our work serves as an exemplar of using abductive looping in the social sciences. Findings on collective identity include the observation that increased team performance in the game, resulting in increased monetary earnings for all players, did not produce a measured increase in collective identity among them.

Keywords Online social experiments · Agent-based models · Abduction · Abductive loop · Collective identity

1 Introduction

1.1 Background and motivation

1.1.1 Collective identity: types, contexts, and applications

Group or collective identity (CI) is an individual's cognitive, moral, and emotional connection with a broader community, category, practice, or institution (Polletta and Jasper 2001).¹ There are several themes of, and implications for, CI, including: (1) an individual's willingness to place the needs

of the group above personal needs [e.g., contributions to Public Goods Games (PGGs) (Charness et al. 2014; Brewer and Gardner 1996)]; (2) a person's susceptibility to positive social influence from group members [e.g., sensitivity to evaluations from a collective group (Charness et al. 2014; Brewer and Gardner 1996)]; (3) one's desire to differentiate from others not in the collective [e.g., allocation between in-group/out-group (Bornstein and Yaniv 1998)]; (4) an individual's willingness to enforce conformity to group norms established by the collective identity (McAuliffe and Dunham 2015; Brewer and Gardner 1996; Kozlowski and

✉ Vanessa Cedeno-Mieles
vcedeno@vt.edu

Extended author information available on the last page of the article

¹ There are other definitions for collective identity (CI). For example, McFarland et al. (2014) state that collective identity means that members become more familiar and equal. Wendt (1994) defines CI as the positive identification with the welfare of another, such that the other is seen as a cognitive extension of the self, rather than independent. See Owens (2006) for a discussion of various definitions of CI.

Ilgen 2006; DeChurch and Mesmer-Magnus 2010); and (5) a person deriving self-esteem from the group (Tajfel 1974; Kozegi 2006). Hence, there are many behavioral and attitudinal manifestations as consequences of CI.

There are many types of, and contexts for, collective identity, including: (1) religious identity (Peek 2005; Benjamin et al. 2016), (2) philosophical identity (Muller 1996; Greene 2000), (3) gender identity (Cameron 1997; Butler 1988), (4) (sports) fan identity (Snow 2001), (5) labor movements (Goldberg 2003), (6) social movements such as African-American civil rights, women's suffrage, gay rights (Taylor and Whittier 1992; Polletta and Jasper 2001; Snow and McAdams 2000), (7) political identities (Plutzer and Zipp 1996; Juergensmeyer 2003), (8) racial and ethnic identities (Tatum 2003; Alexander et al. 2004; Nagel 1996; Eriksen 2010), (9) national and cultural identities (Manchester 1993; Alexander et al. 2004; MacGregor 2018), and (10) ideologies (van Dijk 2000).

CI is a widely studied concept across academic disciplines. Extensive experimental research in social science, political science, psychology, biology, geography, anthropology, religion, criminology, philosophy, and economics shows that CI influences human decision making (Kahn and Ryen 1972; Paris et al. 1972; Goldman et al. 1977; Worchel et al. 1977; Erikson 1980; Oldenquist 1982; Brewer and Gardner 1996; Perdue et al. 1990; Brody 2000; Rousseau and van der Veen 2005; Currarini et al. 2009; van Zomeren et al. 2008; Lustick 2000; Silke 2008; Eriksen 2010; Pilny et al. 2017; Suri and Watts 2011; Shank et al. 2015; Brewer 1991; Benjamin et al. 2016).

There is a host of applications for which CI is important, including team formation, maintenance, and behavior in organizations and communities (Kozlowski and Ilgen 2006; DeChurch and Mesmer-Magnus 2010). The ability to generate identity within (marginalized) groups, e.g., through sacred values, is an important aspect of violent group formation (Silke 2008; Atran et al. 2014a, b). These are compounded by effects of culture and ethnicity (Gilwhite 2001; Atran et al. 2007; Ginges and Atran 2013). International relations are affected by CI among independent states (Wendt 1994). Political leaders of minority or marginalized groups may control identity narratives to persuade their constituents of posturing with governments (Choup 2008). Relatedly, CI is a cohesive force for groups fighting governments to secure rights and indigenous lands (Snow and McAdams 2000; Brody 2000). Religious identity can be a source of stability for immigrants assimilating into a new country (Peek 2005). Language and preservation of culture are intimately tied to collective or group identity (Brody 2000). Ramifications of a lack of identity are studied in Stout et al. (2017).

Individuals may possess several group identities, with different degrees of salience (strength of affinity and association), such that multiple identities may be simultaneously

operative (Snow 2001; Peek 2005; Benjamin et al. 2016). There may be a hierarchy of identities, with different identities coming to the fore in different situations (Stryker 1980). [The ability to use different identities in different situations has been referred to as *freedom* in a philosophical context (Heller 2019)]. Multiple identities may also be negatively correlated, e.g., religious and national identities (Verkuyten and Yildiz 2007). Furthermore, identities and their saliences may be transient over short time scales, and may ebb and flow over longer time scales (Butler 1988; Snow 2001; Vryan et al. 2003; Owens 2006; Benjamin et al. 2016). Consequently, a person's identity may include a combination of dynamically changing, hierarchical collective identities.

Relationships between CI and other phenomena can be intricate. We take collective action (CA), for which there is a massive literature (e.g., Olson 1965; Granovetter 1978; Schelling 2006; Tarrow 2010), as an example. Causal relations between CI and CA are very complicated, with the causal direction between the two changing for different circumstances (Wendt 1994; Polletta and Jasper 2001; Snow 2001; Choup 2008; Fominaya 2010).

1.1.2 Dimensions of collective identity to study

These issues make the study of CI both interesting and challenging. It is the generality of the concept of CI, its application in a wide range of contexts, its many types and its ramifications for humans and their behaviors that have led to myriad CI studies since the term *collective identity* was first coined by Durkheim some 65-plus years ago (Durkheim 1951). *Our focus here is the CI formation process: how CI is formed among a group of people.* We now overview work on the CI formation process to set up what is entailed in our “focus on the CI formation process”.

CI formation is studied in several works (Wendt 1994; Brewer and Gardner 1996; Peek 2005; Choup 2008; Greenhill 2008; Chen and Li 2009; Ackland and O'Neil 2011; Charness et al. 2014; Swanson 2015; Brunson 2017; Pilny et al. 2017). See Related Work, Sect. 3.3. All of these works, except one, are empirical, examining events in the field, under varying circumstances. Surveys, questionnaires, and interviews with human subjects are used to establish, through expert judgment, whether CI has formed within a group.

The work by Charness et al. (2014) also studies CI formation, but is quite different from these other works. They use controlled experiments to produce CI among human subjects through priming using team anagram games, wherein players work cooperatively to form words from a collection of letters that they are given. For example, letters *t, c, a*, and *s* can be used to form words such as *cat* and *cats*. There are many other aspects to their game. Group identity was then

measured after the anagram game using a public goods game (PGG). The greater the PGG contributions of individuals to the team, the greater the collective identity of these individuals. It was found that the priming activity increased PGG contributions. To the best of our knowledge, these are the only controlled experiments that seek to produce CI through priming (in an anagram game) and measure CI quantitatively (through the proxy of PGG contributions). Charness et al. (2014) influence our work herein.

We note in passing that priming tasks are central in social and economic experiments (e.g., Drouvelis et al. 2010; Feher et al. 2016; Smith et al. 2017) and are therefore worthy of study for this reason alone.

In addition to the references above, CI formation is discussed and theorized about in Melucci (1995), Melucci (1989), Snow and McAdams (2000), Snow (2001), Polletta and Jasper (2001), Tarrow (2010), Fominaya (2010). We note that these theoretical works are descriptive and qualitative in nature and are *not* concerned with computational modeling.

1.1.3 Dimensions of our work for collective identity

From the foregoing, we specify what is entailed in “focusing on the CI formation process” for this work. First, there is no agreed-upon method for generating CI. So specifying a priming or CI-producing activity is non-trivial. Our priming game is motivated by the one in Charness et al. (2014), but our game is significantly different, and we describe how, and the technical challenges that exist, in Sect. 1.4. Since we want to study CI formation, we must measure it quantitatively. This, also, is non-trivial and is discussed in Sect. 1.4. Only Charness et al. (2014) measures CI quantitatively in a game; other CI-producing works resort to qualitative methods, which is surely an indication of its difficulty. We want not only to experimentally study CI, but also to model the process. Despite all of the work on CI (described here and in Related Work, Sect. 3), we know of no works that quantitatively model any CI formation process.² In summary, a focus on the CI formation process includes specifying experiments; conducting and analyzing experimental data; quantitatively measuring CI; and modeling both the CI formation and measurement processes. We also use CI as an exemplar for using abductive iterations in the CI formation process. *This work makes contributions in all of these areas.*

² We use the term *model* to mean a representation of equations and algorithms to compute some result. In contrast, in the social and some other sciences, *model* often refers to a qualitative (textual) description of some process that is much more conceptual and not computational. Our models that we present herein are of the first type. We use the term *model* in the former (quantitative) sense, unless otherwise specified.

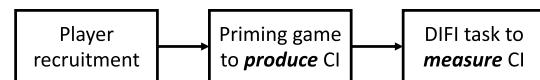


Fig. 1 The main components of the experimental setup: (i) players are recruited through Amazon Mechanical Turk; (ii) players collectively participate as a team in a *priming activity* in the form of a *group anagram game* with the goal of *producing* collective identity among the players; and (iii) players individually state their affinity for (i.e., identity with) the team through a *dynamic identity fusion index* (DIFI) task (Swann et al. 2009) that we take as quantifying or *measuring* the level of CI formed in the anagram game. The modeling effort is to model the second and third steps of this process. Note that because the group anagram game is more complex than the DIFI task, more work is required to produce the experimental platform, to conduct experiments and to analyze the experimental data, and to build models of player actions in this priming activity. Furthermore, it is in the priming activity that game conditions are altered, and it is these conditions that we seek to correlate with CI

After our work is overviewed in the next two subsections, we address, in turn, technical challenges, the novelty of our work, and our contributions, and these address all of the topics just itemized. In the next section, we translate these topics into work tasks.

1.2 Overview of work scope

Our work has three broad elements. First, we develop an online experiment, motivated by the work of Charness et al. (2014), that is designed to produce CI within a group of participants, through priming, and then measure the CI produced. See Fig. 1. Specifically, the priming activity consists of players cooperating in a group anagram game, where participants share letters with their neighbors in order to help all players form more words. Specifically, the main player actions are: (1) requesting letters from neighbors, (2) replying to letter requests of neighbors, and (3) forming words. This activity is intended to *produce* CI among a group of players. This priming activity is accompanied by a dynamic identity fusion index (DIFI) task that measures how much a person associates with the team or group as a result of playing the group anagram game (it is a proxy for CI). Unlike the group anagram game, the DIFI task is done individually, in isolation. This DIFI task is to *measure* the CI formed in the group anagram game.

For the second element of our work, we construct models of the CI priming process (the group anagram game) and of the DIFI task, and compare predictions of player behavior against experiments in the group anagram game. We develop and evaluate three agent-based models for the group anagram game, and a statistical model for the DIFI task. Because the group anagram game is much more intricate than the DIFI task, more effort is devoted to the former activity.

As the third element of our work, we use abduction as our framework for this study where both experimental work and modeling work take place within an abductive

loop framework. We now address abductive iterations, since they tie together experiments and modeling.

1.3 Overview of our experiment and modeling approach: abductive iterations

Abduction is an inference approach that uses data and observations to identify plausible (preferably, best) explanations for phenomena (Pierce 1931; Flach and Kakas 2010). That is, abduction is reasoning from effects to causes (Chamiak and Santos 1992). Effects are often generated by results from (laboratory) experiments or in situ observations of systems. One then constructs hypotheses and identifies or develops theories that explain these observations.

Much of the work on abduction has focused on topics such as producing explanations for different logic settings (e.g., Echenim et al. 2013); determining the computational complexity of abduction problems (e.g., Wei-Kleiner et al. 2014); and generating solutions for special problems or transformations that are useful in obtaining solutions (e.g., Pfandler et al. 2013). Abduction has broad application in robotics, genetics, automated systems, and image understanding (Shanahan 2005; Andrews and Bonner 2011; Vanderhaegen and Caulier 2011; Juba 2016).

However, in contrast to the above notion of abduction, our focus is the specification and implementation of an abductive *looping* process, wherein abduction is executed in successive iterations. Every iteration builds off of all previous ones, so that explanations may evolve from accumulated data from experiments and observations. As a differentiator from previous work, our interests are behaviors and human interactions within networked groups in the social sciences (Contractor 2019). In particular, our exemplar is to understand whether a cooperative game can produce collective identity (CI) within a group.

The abductive loop (AL) process that we employ is described in Sect. 2, but among its components are experiments and modeling, and we make note of works on coupling experiments and modeling here. There have been several controlled experimental studies of comparable size to our experiments (e.g., Kearns et al. 2009; Judd et al. 2010; Kearns et al. 2012). Also, empirically grounded, data-driven modeling of human behavior is done (Mason and Watts 2012; Li et al. 2014; Nguyen et al. 2017; Zhang et al. 2016). We combine these two ideas, in a particular way that is guided by abduction, and perform them iteratively. The proposed abductive analysis is to form hypotheses to evaluate theories as part of the looping process, and develop new insights about CI. Our approach provides an exemplary case of coupling theory development/evaluation with real problems.

1.4 Technical challenges

There are several technical challenges in our work. The first two, and foremost, challenges are generating and measuring CI in a group setting. CI can be a transient phenomena, so it can be generated in a group that is newly formed. A group anagram game has been successfully used to produce CI among people that do not know each other (see Sect. 3.6 of Related Work). This is the only previous work in a group setting where CI was produced over a single game or encounter of relatively short duration and then quantitatively measured. To our knowledge, ours is the first work to attempt CI formation under similar conditions, but in addition, in an online setting. We also use a group anagram game (our game is different from the one referred to above, but is motivated by that work). In our setting, however, a team's or group's members do not have the benefit of observing facial expressions and body language as was done in Charness et al. (2014). By comparison, other works seek to form CI within groups that meet regularly over periods weeks or months [see Sect. 3.3, e.g., Swanson (2015)]. Still other works [see Sect. 3.3, e.g., Peek (2005), Choup (2008)] rely on CI being formed over years, based on cultural and other factors. The point is that there are many works that seek to produce CI in different ways, so studying methods of producing CI is important. A second point is that our requirement to produce CI quickly in an online game, with no use of visual cues or body language, is novel and extremely challenging.

Also, our method of measuring CI quantitatively has never been used in relatively short-duration settings. Specifically, we use the DIFI score (Jiménez et al. 2016) to measure CI formed in the group anagram game. Previously, DIFI score has been used to *pre-select* people from a larger population that have *already formed* CI, often based on cultural or other long-lived attributes that germinate and grow over years, and that was formed *in situ* [see Sect. 3.4, e.g., Swann et al. (2010a)]. In Charness et al. (2014), a public goods game was used to quantify CI among group members. Several other works (e.g., Swanson 2015) use a binary scale of determining the existence of CI (i.e., “no CI formed,” or “CI formed”) based on opinions of experts that, for example, base their subjective determinations on language used by group members. Our challenge here is to determine whether we can use DIFI score to characterize CI; we are using the DIFI method for a different class of CI, for which it has not been used. Quantitative methods of measuring CI, are few, and it has very rarely been done.

From this description thus far, it should be clear that we have unique and challenging setups and goals for producing and measuring CI. There are other challenges.

Our goal is to study quantitatively CI within an abductive loop setting. Looping over abductive analyses is relatively rare (see the robotics work Shanahan (2005) as an

exception), and the use of abduction and abductive iterations in the social sciences is very rare.

Furthermore, our abductive analyses involve both experiments and modeling. Most abductive work focuses solely on using experiments within loops. Here, we make modeling a first-class element of the abductive process. In fact, we know of no work whatsoever in modeling a priming activity designed to produce CI among interacting members of a group. Developing models is all the more challenging because our game consists of multiple player actions (versus binary choice games) that can be repeated over time (versus one-shot games) in any user-defined sequence of repeated actions, where neighboring player actions (or inaction) can influence action choices among players.

Also, there are finite resources and a finite pool of candidate players for conducting experiments, using the constraint that we want participants to play the game only one time. (Amazon Mechanical Turk does not provide a gateway into an infinite pool of players). A challenge is to specify game conditions that produce measurable CI while also providing data over a range of input conditions to enable development of models that are more general than the experimental conditions.

We believe that this work pushes the state of the art in all of these dimensions. Moreover, we attempt to interpret our CI results in a way that focuses on player interactions, and not specifically on our group anagram game. Consequently, while we believe that this work achieves several firsts (see the novelty section immediately below), this work is not the final word on CI, abductive analyses, and conducting and modeling group CI experiments. Rather, we view it as a beginning: a beginning that opens new avenues for producing and quantitatively measuring CI.

1.5 Novelty of our work

The novelty of our work is in the areas of research process (through abductive iterations), experiments, modeling, and social science (through CI). Specific novelties include:

1. Performing the first online group anagram game (for producing CI) and using a dynamic fusion index (to measure CI);
2. Conducting experiments where players can choose actions from a candidate set; these actions can be repeated any number of times over a specified time duration; players interact, cooperate, and can affect others' subsequent actions (versus binary choice, one-shot games);
3. First-of-their-kind experimental results and implications for CI from online experiments;
4. First modeling of this temporal, multi-action, interacting group anagram game;

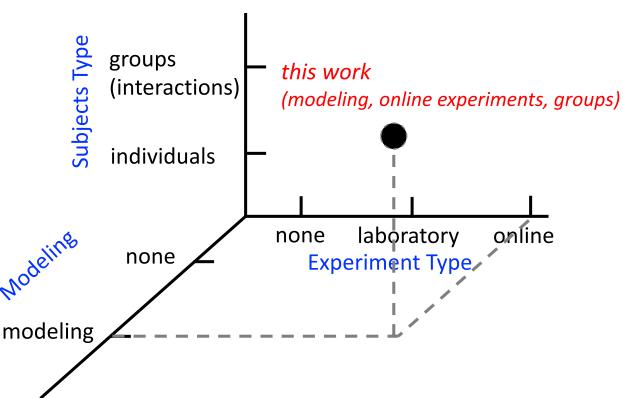


Fig. 2 Conceptual view of three dimensions of this work, illustrating how our group anagram game study (for producing CI) and our approach for measuring the CI produced is situated. Our experiments consist of online web-based human subjects experiments. Our modeling component consists of model and algorithm development, and agent-based modeling. We study groups of interacting individuals. To our knowledge, this combination of study features is unique. These dimensions are used in Table 1 to compare our work with those of others

5. First evaluation of these models by comparing model predictions with experimental results;
6. One of the first uses of abductive looping in a social science context, and the first use in the study CI.
7. One of the first demonstrations of abductive looping to test (social) theories.

Figure 2 shows the unique context of our work along three dimensions of experiments, modeling, and types of experimental subjects. Table 1 makes this more concrete by presenting representative works along various combinations of values along the three dimensions of Fig. 2. It is clear that our work—identified at the bottom right of the table—is unique.

1.6 Contributions

Our major contributions follow.

1. *Insights on the collaborative anagram game (Sect. 4)* We present novel experimental data that illustrate how players interact in group anagram games played through an online game/experimental platform. We focus on experimental data that are useful in modeling. We find that letter requests and letter replies are made throughout the game, rather than solely at the outset. However, if there are few neighbors ($k = 2$) and consequently fewer available letters (3 letters per neighbor), there are fewer letter requests and letter replies near the end of the game. Also, players generally respond relatively quickly to their neighbors' letter requests: replies are typically made within 30 s of the request. In the same

Table 1 A hierarchy of different collective identity (CI) experiments in the literature, which puts the uniqueness of our work on anagram game experiments and modeling into the context of the works of others. For each of in-laboratory and online environments, there are categories of works on individual subjects and groups of subjects. With groups of subjects, we are interested in interactions among these subjects. Then, for each of these four categories, we break those works that are experiments only (i.e., Exp. Only), and those works that combine experiments and modeling (i.e., Exp. & Modeling). The

In-laboratory				Online			
Individual subjects		Group of subjects		Individual subjects		Group of subjects	
Exp. Only	Exp. & Modeling	Exp. Only	Exp. & Modeling	Exp. Only	Exp. & Modeling	Exp. Only	Exp. & Modeling
Brewer and Silver (1978)	Rousseau and van der Veen (2005)	Worchel et al. (1977)	No work	Pilny et al. (2017)	Ackland and O'Neil (2011)	No work	This work

way as letter requests and letter replies, word submissions are made throughout the 5-min game, but the numbers of neighbors and available letters do not affect this type of action.

2. Data-driven networked agent-based models (ABMs) of experiments: design, construction, and evaluation (Sect. 5) We design, construct, and evaluate three data-driven ABMs of the group anagram game experiment. We adapt a conditional random fields (CRF) (Sutton and McCallum 2011) modeling approach with four parameters to flexibly incorporate history effects on agent actions that evolve in time. That is, our models predict time histories of player actions in the group anagram game. These actions are: (1) requesting letters from neighbors, (2) replying to letter requests of neighbors, (3) forming words, and (4) thinking (or idling). We capture these activities through a state transition matrix approach, where, in our most sophisticated model, the action at time $(t + 1)$ is based on the action at time t and on a feature vector that captures an individual's state. Our approach can alleviate the overfitting problem that would arise with, e.g., a static Markov model that would require capturing many more state transitions.

ABM is used as our simulation modeling approach because of its fine granularity and for its generative properties (Epstein 2007). That is, local interactions produce population-level dynamics. We use inductive and deductive inference in three ways, use KL divergence to compare model predictions with experimental data, and compare results across multiple ABMs. For example, our KL-divergence evaluations are broken down by ABM, player action, and number of neighbors in a game. For each combination, we use overall data at the end of the 5-min group anagram game and at 1-min intervals during the game to evaluate temporal effects. All of these are used to demonstrate that the ABMs

last row has one representative work within each category; some categories have no work. Our work (labeled **This Work**) studies online experiments of collective identity with modeling and interaction among subjects. It is the *only* work that combines experiments and modeling of human subjects in a group setting. (These references are not exhaustive; more detail is included in the related work of Sect. 3. However, for the categories labeled with *No Work*, there is no work in the literature, to the best of our knowledge)

successively improve with the process of incorporating more data that enables greater modeling sophistication.

Our three successive ABMs are named M0, M1, and M2. Our work in evaluating the ABMs shows that ABM M1 reduces KL-divergence values by $4\times$ or more, over those for ABM M0, in many cases. (Smaller KL-divergence values are better; they indicate that model predictions are in better agreement with experimental data. A KL-divergence value of zero means that a distribution from a model and from data are interchangeable.) Our work also shows that in many cases, ABM M2 has KL-divergence values that are $4\times$ or more reduced from those of ABM M1. Interestingly, ABM M1 does slightly better than our most sophisticated model (ABM M2) for a small range of parameters that were used in generating M1, but M2 does much better over the entire input parameter space.

3. Specification and demonstration of iterative abductive analysis process (Sects. 2 and 7) We perform experiments (Contribution 1), and modeling and evaluation (Contribution 2), within an iterative abductive process. Using Haig (2005), Timmermans and Tavory (2012) as a starting point, we explicitly incorporate modeling and iterations into the abductive process. The latter necessitates specifying what is to be done in the next iteration. The iterative process is successfully demonstrated through the group anagram experiments, agent-based modeling, and hypothesis generation and testing. *The proposed abductive process can be considered as a general methodology for other social science researches. Although the methodology is general, we provide considerable detail in both the experiments and modeling, and their interactions (e.g., how they complement each other), that we believe will be helpful for other social science studies.* For example, our method of model construction from data (see Contribution 2 above) can be

used to capture other temporal human action sets among interacting agents.

4. Statistical analysis of numbers of samples required for modeling (Sect. 6) We evaluate the quality of our state transition matrices of our ABMs using a root of mean squared errors (RMSE) approach. Specifically, we are interested in how many test samples are required to achieve a specified small error in predicted transition probabilities, which are an integral part of our ABMs, as compared to measured transition probabilities. We use our feature vector from the ABM and break each element down into bins, and add to it the dimension of number of neighbors that a player has in a game. By evaluating all of the successive state transition pairs among the actions, i.e., action $a(t)$ at time t and the next action $a(t+1)$ at time $(t+1)$, within each of the resulting 324 distinct bins of data, we find that the minimum number of observations (samples) for each state transition clearly demarcates small from large RMSE. The data show that small RMSE values result when a state transition has at least 100 observations.

5. New experimental understanding of the formation of collective identity (CI), (Sects. 4 and 7) First, we demonstrate that our group anagram game can prime game players to form CI (Sect. 7.3). Second, we discover three novel insights on the formation of CI by coupling the team anagram game and DIFI score. (a) Players' DIFI scores increase with increasing numbers of neighbors in the anagram game. Since DIFI score is our proxy for CI, this implies that a player's sense of CI increases with increasing numbers of neighbors (Sect. 7.4). (b) The number of interactions increases as number of neighbors of a player increases from two to four. However, the numbers of interactions, relatively speaking, saturates with further increases in degree, up to a degree of eight (Sect. 7.4). (c) Despite this saturation in numbers of interactions, the DIFI score continues to increase with degree, suggesting complicated interactions among game parameters (Sect. 7.4).

Third, what we did not find is interesting (Sect. 7.3). Specifically, the number of words formed does not correlate with players' feelings of CI (as measured by the DIFI score). We conjectured that since players' earnings are directly tied to the number of words that they generate, players would deem this outcome important. That is, the more words formed by the group, the greater their earnings and hence the greater the success achieved by the group. Success breeds cohesion. We therefore hypothesized that the greater the number of words formed, the greater the CI among team members. We found that this is not the case: number of words formed is not significantly correlated with DIFI score. Rather, numbers of interactions correlated more strongly with DIFI score (Sect. 7.3).

Fourth, these experimental observations, and hypotheses that go along with them, are made in the context of the abductive loop (see Sect. 7). To the best of our knowledge, these are the first experimental results of this kind. In this process of conducting two iterations, we demonstrate: constructing hypotheses, testing theories, falsifying hypotheses, finding support for hypotheses in full or partially, and finding support for multiple hypothesis from the same observations that require disambiguation in more abductive iterations. We cast our hypotheses in general terms, using only degrees of players in networks, numbers of interactions in games, and game rewards, i.e., abstracting away our particular game conditions, thus enabling testing of our findings by other researchers, potentially using different games.

1.7 Extensions from the conference paper

This paper was originally published as Ren et al. (2018). Extensions of that work, presented herein, are summarized as follows. (1) Introduction has been expanded to give fuller treatment of background, motivation, and problem context. (2) Related work is expanded with more detail and new topics. (3) Game description has more detail. (4) Experimental data from the game are given with new insights on player behavior. (5) Fuller treatment of the development of each of the three ABMs (M0, M1, and M2). (6) Fuller treatment of comparisons of model predictions with experimental data. (7) Additional model evaluation and data, comparing model predictions to experimental results across games. (8) Enhanced description and results in error analysis, comparing experiments and models.

1.8 Paper organization

Figure 3 shows the technical sections of this paper, and their relationships. An overview of the abductive loop process is presented in Sect. 2, providing a framework for the rest of the paper. Related work is in Sect. 3. The group anagram experiments and results are described in Sect. 4. Models of the experiments are developed in Sect. 5. Model predictions are compared to experimental data. Section 6 contains error analyses of the models. Sections 4 through 6 contain the major technical components of the abductive loop that is overviewed in Sect. 2. These analyses and results enable a more streamlined description of the particular abductive loops executed in this work in studying CI in Sect. 7, so that the abductive process is clear. In Sect. 7, the relevant sections of the experiments and modeling are referenced. This also makes the experiments and modeling more clear. Furthermore, we discuss the generalizable knowledge gained from the CI study, how CI is formed and not formed, and because we state our hypotheses in terms of interactions, how other experiments can be undertaken to confirm or

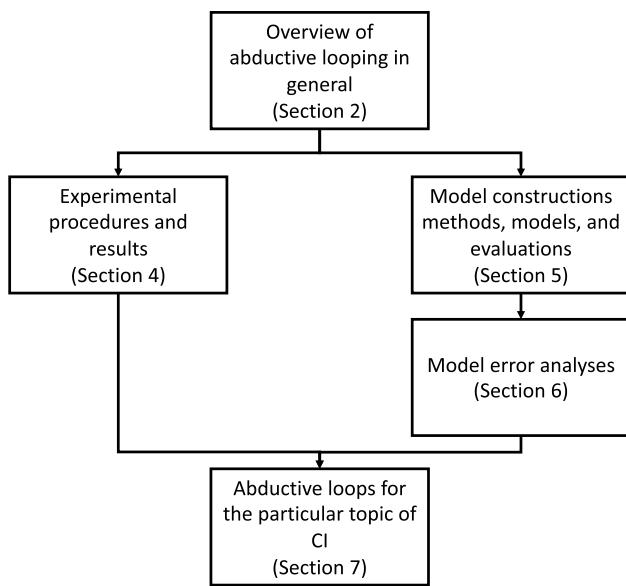


Fig. 3 Relationships among the sections of the paper that describe the technical work. Section 4 presents both the CI formation game (group anagram game) and the CI measurement task (DIFI). Section 5 describes the CI formation models. The model for predicting CI measurement (DIFI score) is given within the abductive iterations in Sect. 7. Section 6 analyzes errors in the anagram model predictions, compared to experimental data. Section 7 takes key points from the experiments and modeling and presents them within the framework of abductive looping. This section also contains additional analyses of experimental data, hypotheses about CI and their evaluations, and selected modeling results. Section 7 is a culmination of all the work in the preceding sections. Sections not appearing in the figure are Related Work (Sect. 3), Limitations of the work (Sect. 8), and Summary (Sect. 9)

contradict our results. Limitations of this work are presented in Sect. 8. Section 9 summarizes. Sections 4 and 5 are substantial in size. Consequently, we provide tables within these sections to organize the work and guide the reader, and we present many of the results in the Ph.D. dissertation of Cedeno (2019).

2 Overview of Abductive Loop

Figure 4 illustrates our iterative abductive process, which includes inductive and deductive steps and hypothesis testing. All work in this paper takes place within this framework. This structure follows that of Haig (2005), Timmermans and Tavory (2012), which are based on Piercian abduction (Pierce 1931), but augments it in key areas. Note that in contrast to confirmatory (deductive) analyses, where theories, hypotheses, and models are developed *first*, and used to predict results of future candidate experiments, one-step abduction first generates data through experiments or observations. (Abduction uses data to drive the scientific

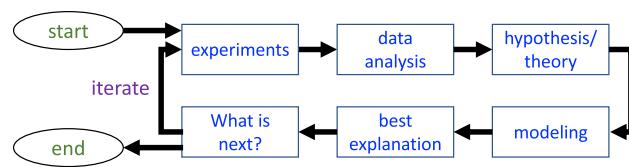


Fig. 4 Steps in our iterative abductive analysis/loop

discovery process.) Then, data analysis consists of searching for *patterns* and generalizing these into *phenomena*, which is an inductive step. These results are used to formulate hypotheses based on theories whose purpose is to explain the data. Hypotheses may exist (e.g., from a previous loop) or may be proposed in this step, and can be removed (e.g., via falsification). Multiple candidate theories may be posed for a given phenomena. Models are developed from the data, with the objective of generating outputs that help evaluate hypotheses and theories, and/or help guide experiments for the next loop. The best explanation, or hypothesis/theory appraisal, is the process of identifying the best explanation for the phenomena (Thagard 1989); this includes hypothesis falsification. Finally, the last step in an iteration is to determine what to do next, in terms of designing new experiments. The iterative process may terminate for any number of reasons; e.g., a best explanation has been found.

This description provides the structure for the rest of the paper. The experimental work of Fig. 4 is described in Sect. 4, after related work. The modeling work in Fig. 4 is presented in Sects. 5 and 6. We provide the experimental and modeling methodologies, data and results in these sections because they are too large to fit within a discussion of results from the abductive iterations. Following these sections, we return to the abductive loop and reference experimental and modeling results as appropriate, to make the looping process and results more streamlined and cogent (and provide additional results).

3 Related work

Related work topics are provided in Table 2, along with each topic's relevance to our work. Each subsection below provides research for one row in the table.

3.1 Overviews of CI

Overviews of CI are provided in Tajfel (1974), Abrams and Hogg (1990), Owens (2006), Vryan et al. (2003), Fiske et al. (2010), Hogg and Abrams (2007), Snow (2001). Peek (2005) provides an interesting view of CI as a combination of social structure (through roles) and processes (via perceptions and interactions) (Melucci 1995).

Table 2 Research literature topics addressed in Sect. 3, related work. Selected sections are presented here, while others appear in “Appendix A”

Section of related work	Name	Relevance
3.1	Overviews of CI	CI is a broad topic. These are surveys of CI for the interested reader
3.2	Individual Anagram Games: Experiments	Individual anagram games are precursors to group anagram games and have been extensively studied for more than 60 years to analyze the effects of goal setting, compensation types, internal-external attributions, and test anxiety. It includes a broad range of disciplines like sociology, economics, management science, and (social) psychology. For our work, anagram games are priming activities
A.1	Individual Anagram Games: Modeling	With all of the experimental work on anagram games, it is surprising that very little work has been done in modeling and simulating these games
A.2	Individual Anagram Games: Experiments and Modeling	Few works combining experiments and modeling of individual anagram games exist (Feather 1969; Feather and Simon 1971a, b)
3.3	Collective Identity-Based Experiments: Formation of CI	Our work is motivated by CI, and in particular the CI formation process. These works study different methods from ours in generating CI
3.4	Collective Identity-Based Experiments: Implications of CI	Along with the Introduction, this section provides works that demonstrate the implications of CI, thus motivating why we study it
3.5	Measurement of CI	Methods used in research to measure (quantify) CI are important.
3.6	Combined Group Anagram and CI Experiments	This section emphasizes that there is only one work on group anagram game. That work motivated our work. However, there are differences between that work and ours
A.3	Modeling of CI	Demonstrates that there are few modeling studies of CI, and no works like ours
A.4	Agent-Based Models of Anagram Games and Formation of CI	This puts our preliminary results into context. The first and only work, to our knowledge, in modeling human group anagram games is our work Ren et al. (2018)
A.5	Studies of Phenomena Related to CI	As described in the Introduction, CI is relevant for and closely related to, many other phenomena like cooperation and collective action. These works provide some background on these works
A.6	Data-Driven: Combining Experiments and Data-Driven Modeling	Demonstrates that combined experimental and modeling studies, as we do here, are used for other phenomena besides CI
3.7	Modeling of Time Sequences of Actions	These are studies that investigate time series models. Our modeling and ABMs are essentially time series models
3.8	Evaluation of Model Predictions	Methods for comparing experimental and model prediction distributions, as we do here, are presented
3.9	Abduction and Abductive Loop	We use abductive iterations as a framework for our experimental and modeling work. We survey other abductive works

3.2 Individual anagram games: experiments

Over 20 experimental works use anagram games—with *individual* players (e.g., Mayzner and Tresselt 1958; Russell and Sarason 1965; Tresselt 1968; Warren and Thomson 1969; Dominowski 1969; Feather 1969; Feather and Simon 1971a, b; Davis and Davis 1972; Sarason 1973b, a; Miller and Ross 1975; Goldman et al. 1977; Gilhooly and Johnson 1978; Stones 1983; Locke and Latham 1990; Latham and

Locke 1991; Vance and Colella 1990; Schweitzer et al. 2004; Cadsby et al. 2007, 2010)). An individual game means no interactions (e.g., sharing letters) between subjects playing a game at the same time.

We review anagram game studies that are purely experimental. In Stones (1983), experiments of anagram games are used to test player’s specification of causality for their performance (e.g., did a player attribute good performance to skill or luck?). It was found that people more likely to be responsible

for their own actions attributed success or failure to their own behavior, versus assigning outcomes to chance. Miller and Ross (1975) analyzed how individuals engage in attributions of causality. Situational factors were studied through anagram games in Davis and Davis (1972).

Effects of goal setting are analyzed with anagram tasks in Locke and Latham (1990), Vance and Colella (1990), Latham and Locke (1991), Schweitzer et al. (2004). In Vance and Colella (1990) players played the anagram game and their assigned goals became increasingly difficult. For example, for each goal trial, subjects were assigned a goal for the number of words they have to form. After each goal trial, subjects recorded their performance (i.e., the number of words formed) as well as their assigned goal for the next trial. Difficulty of assigned goal was increased by two words per trial. Before beginning the next trial, subjects completed a form on which they calculated their GDF (goal discrepancy feedback: performance minus assigned goal) and PDF (performance discrepancy feedback: performance this trial minus performance last trial). Assigned goals were rejected when GDF became sufficiently negative. GDF and PDF differed both in sign and magnitude of effects on acceptance and personal goals, indicating that subjects used these feedback discrepancies differently in the goal evaluation process. Unusually, personal goals and performance remained high even after assigned goals were rejected. In Locke and Latham (1990), Latham and Locke (1991), theories of goal settings are developed. In Schweitzer et al. (2004), it was found that people with unmet goals were more likely to engage in unethical behavior than people attempting to do their best.

Goldman et al. (1977) use the anagram task to examine three factors and their effects on group performance: intergroup competition or cooperation, intragroup competition or cooperation, and task means interdependence. In Russell and Sarason (1965), Sarason (1973b), Sarason (1973a), studies look at anxiety generated from performing a task, where the task is the anagram game. In Cadsby et al. (2007), pay-for-performance and fixed-salary compensation were compared using an anagram task. In Cadsby et al. (2010), an anagram game was employed as the experimental task to evaluate a target-based compensation system, a linear piece-rate system and a tournament-based bonus system. Larger amounts of cheating occurred under target-based compensation. In Mayzner and Tresselt (1958), Dominowski (1969), Warren and Thomson (1969), Gilhooly and Johnson (1978), the effects of letter order and word frequency on anagram game performance are analyzed.

3.3 Collective identity-based experiments: formation of CI

The following references study or theorize on the CI formation process. That is, they study processes by which a group

of individuals that does not possess CI can form CI by, for example, interacting or undergoing a priming task.

In Brewer and Gardner (1996), laboratory experiments of CI with no interactions between subjects are performed using priming. They argue that the personal, relational, and collective levels of self-definition (shift from personal to collective) represent distinct forms of self-representation with different origins, sources of self-worth, and social motivations. They suggest the concept “we” primes social representations of the self that are more inclusive than that of the personal self-concept. In a preliminary investigation of the implications of different levels of the social self-concept, a set of three experiments were conducted to explore the effects of priming various “we” schemas on individual judgments and self-descriptions. In the priming task, participants read a descriptive paragraph with instructions to circle all the pronouns that appeared in the text, as part of a proofreading and word search task. After completing this word search task, participants were escorted to another room and asked to judge, as quickly as possible, whether the statements were similar or dissimilar to their own views by pressing a number key on the keyboard, ranging from 1 (very dissimilar) to 4 (very similar). They found that individuals primed with “we” would entail an expanded sense of self that would lower thresholds for agreement and assimilation.

In Chen and Li (2009), laboratory experiments with no interactions between subjects measure the effects of induced group identity on participant social preferences. They show that participants are more altruistic toward an in-group match. They evaluate different ways of creating group identity in the laboratory, to explore the formation of groups and to investigate the foundation of what group identity is. When participants are matched with an in-group member (as opposed to an out-group member) they show a 47% increase in charity concerns when they have a higher payoff and a 93% decrease in envy when they have a lower payoff. Also, participants are 19% more likely to reward an in-group match for good behavior, but 13% less likely to punish an in-group match for misbehavior. Participants are significantly more likely to choose social-welfare-maximizing actions when matched with an in-group member.

In Pilny et al. (2017), online experiments with no interactions between subjects are performed. To expand upon perspectives on the commons dilemma (e.g., do I contribute to the common resource or do I free ride), Pilny et al. (2017) developed an online experiment grounded on group decision making. They create manipulations based on three modalities of structure: dense versus sparse networks (domination), collective versus individual identity (signification), and social sanction versus non-social sanction (legitimation). The online experiments reveal that modalities of signification positively influence contribution rates on the commons dilemma, when participants were provided

information meant to stimulate a CI. This is analogous to the findings of Charness et al. (2014); see Sect. 3.6. They mention how challenging it is for an online experiment to create CI, because the individual is sitting alone playing the game on a computer. In their experiments they try to stimulate CI by communicating three additional pieces of information regarding collective outcomes: (1) total collective score, rather than just an individual collective score, (2) collective rank compared to previous sessions, and (3) the score of the highest collective score from previous sessions. By including more collective, rather than individual, information, the user may come to behave more in a collective fashion and contribute to the public good.

In Ackland and O’Neil (2011), an online experiment using data collected from the websites of over 160 environmental activist organizations is developed. A model is presented where social movement actors exchange practical and symbolic resources in the guise of website text content and hyperlinks, as part of a process of online CI formation. The hyperlink and online frame networks are compared on three measures of centralization: degree, betweenness and closeness.

Wendt (1994) argues that international cooperation among independent states can be fostered through CI. He describes different mechanisms that may lead to CI, and takes examples from past events or general ideas. For example, he states that trade relations among states can foster CI through the emergence of the feeling of a common fate, but there are no experiments nor historical observations. It focuses more directly on identities and interests as the dependent variable and investigates whether, how, and why identities change.

Greenhill (2008) evaluates self and recognition theories—recognition theory states that an individual or group places recognition of itself by others as a very high-priority goal—to determine whether these two ideas can combine to produce CI. The reasoning is that as the self acknowledges others, and this process is replicated by all participants, a collective identity is formed. However, social identity theory-based experiments do not support this line of reasoning. This work is more akin to a meta-study, summarizing existing results.

Peek (2005) studies CI generation among Muslims in the USA. It is an empirical study of the formation of religious CI among 127 subjects, using focus groups, individual interviews, and participant observations. She presents three consecutive steps to form CI: religion as an ascribed identity; religion as chosen identity; and finally religion as declared identity.

Choup (2008) studies the relationships among specific (poor) constituent groups and governments, and how these groups use their shared (collective) identity to position themselves. She also uses observations (of group meetings)

and interviews of group leaders to produce a model of CI formation and its effect on collective action.

Swanson (2015) uses small groups of music students (sizes of 2–5 students) to study the formation of CI. Again, as with several previous works, surveys, interviews, and observational studies are used to document the CI formation processes as students work together.

Brunsdon (2017) examines South Africa and the fracturing of the nation among different societal groups. Factors contributing to the lack of a national CI (e.g., a lack of trust among sub-groups and misunderstandings) are also discussed. Finally, the article posits that one way to heal these divisions and form a national CI is through religious understanding.

A final work in CI formation is experiments with interactions among subjects performed in Charness et al. (2014). This work, in a general way, motivated our anagram game experiments (although there are many differences between our work and that in Charness et al. (2014)). Consequently, we address this work separately below.

Dismissing for the moment this last reference, it is clear that none of the above works on CI formation are like ours. In contrast, our work uses controlled online laboratory experiments to produce CI through priming groups of subjects using a cooperative anagram game.

3.4 Collective identity-based experiments: implications of CI

The effects of religious (group) identity on individual behavior is studied in Benjamin et al. (2016). Subjects (self-identified as Protestant, Jew, Catholic, or agnostic/atheist) were primed or not primed with respect to religion. Priming consisted of having players unscramble a set of words that form a sentence, and that sentence has religious content. The unprimed subjects unscrambled words to form a sentence with no religious content. The purpose of priming is to make salient the religious identities of players, if they exist. Subjects then played a number of games, including public goods games, risk aversion games, discount rate elicitation games (i.e., delayed gratification games), among others. In a public goods game, players are given some amount of money. They have the option of contributing a portion of their money to the group. The pooled money that is contributed to the group by all members is then typically multiplied by some factor and redistributed to the players. Hence, there may be some incentive to contribute to the group. There are several interesting results. Among them is that religious identity salience (i.e., priming) produced an increase in Protestant subjects’ contributions to Public Goods Games (PGG), while it generated a decrease in Catholic subjects’ contributions.

In related experimental economics work using Indian caste and other nonreligious identities, Eckel and

Grossman (2005), Hoff and Pandey (2006, 2014), Charness et al. (2007), Chen and Chen (2011), Cohn et al. (2014), Chen et al. (2014), Cohn et al. (2015) find that group identity effects on behavior strengthen with the salience of group membership. Chen and Yeh (2014) manipulate the norms (expressed by legal rulings) that subjects are exposed to and study how these norms affect their self-identification.

The following works study the implications of *identity fusion*, where individuals may feel fused with (i.e., strongly connected to) a group (Swann et al. 2009, 2010a, b; Gomez et al. 2011a; Swann et al. 2014; Gomez et al. 2011b). We interpret identity fusion to be synonymous with, or very similar to, CI.

In Swann et al. (2009), the authors use online experiments to test the notion that fusion represents a distinctive form of allegiance to groups. They propose that when people become fused with a group, their personal and social identities become functionally equivalent. To measure identity fusion they used a modified version of a fusion scale developed by Schubert and Otten (2002). They prove that activating either personal or social identities of people who were fused with their group increased the extent to which they were willing to fight or even die for the group. Thus, even when people become deeply aligned with a group, their personal identities remain potent.

In Swann et al. (2010a), using an intergroup version of the trolley problem, the authors explored participants' willingness to sacrifice their lives for their group. Studies showed that nonfused participants expressed reluctance to sacrifice themselves, and identification with the group predicted nothing. To measure identity fusion they used the same scale as in Swann et al. (2009).

In Swann et al. (2010b), they assume that autonomic arousal will increase agency (i.e., the capacity to initiate and control intentional behavior) for fused and nonfused persons. In four experiments, increasing autonomic arousal through physical exercise elevated heart rates among all participants. Fused participants, however, uniquely responded to arousal by translating elevated agency into endorsement of pro-group activity. To measure identity fusion they used the same scale as in Swann et al. (2009).

In Gomez et al. (2011b), online experiments showed that when people are ostracized (i.e., rejected and excluded) by either an out-group or an in-group, they may either withdraw or engage in compensatory activities designed to reaffirm their social identity as a group member. The authors proposed that individual differences in identity fusion (an index of familial orientation toward the group) would moderate the tendency for people to display such compensatory activity. Four experiments showed that irrevocable ostracism increased endorsement of extreme, pro-group actions (fighting and dying for the in-group) among fused persons but not

among nonfused persons. To measure identity fusion they used the same scale as in Swann et al. (2009).

In Gomez et al. (2011a), the authors determine what fusion is and the mediating mechanisms that lead fused individuals to make extraordinary sacrifices for their group. For measure of group identification, they proposed a seven-item verbal scale with greater fidelity than the earlier pictorial measure of identity fusion from Swann et al. (2009).

In Swann et al. (2014), online experiments explored the cognitive and emotional mechanisms that underlie the endorsement of self-sacrifice. Using participants responses to moral dilemmas, they found that only those who were strongly fused with the group preferentially endorsed self-sacrifice. Identity fusion was measured using the seven-item verbal fusion scale from Gomez et al. (2011a).

3.5 Measurement of CI

Researchers measure or declare the existence of CI in different ways. This is in part because there are many definitions for, and types of, CI (see Sect. 1.1).

Many references on CI formation (Wendt 1994; Peek 2005; Choup 2008; Greenhill 2008; Swanson 2015), presented in Sect. 3.3, pronounce that CI has been formed based on expert evaluation of textual comments of participants, survey responses, and interviews. These are subjective approaches for determining the existence of CI. They require an expert to interpret the data, and multiple experts may arrive at different conclusions.

In PGGs (Ledyard 1994), players are given some amount of money. They have the option of contributing a portion of their money to the group. The pooled money that is contributed to the group by all members is then typically multiplied by some factor and redistributed to the players. Hence, there may be some incentive to contribute to the group. Charness et al. (2014), Chen and Li (2009) use PGG contributions as a proxy for CI. In Charness et al. (2014), the percentage of a persons money that they contribute to the group is taken as their identification with the group: those with greater group identity contribute more of their money to the team.

In Swann et al. (2009) a modified version of a fusion scale developed by Schubert and Otten (2002) is proposed. To capture fusion in a manner that emphasized perceived overlap and nothing else, participants choose from five pictures which best represented the way they perceived their relationship with the group. Each figure in the scale shows two circles of different sizes. The small circle represents "the self", the big circle represents "the group". When participants need to choose from the scale, five figures with symmetrical degrees of overlap (0%, 25%, 50%, 75%, and 100%) are presented. For example, the first figure shows the two circles not intercepting, the second figure show a 25% interception and the fifth figure show a 100% interception

with the small circle. To measure identity fusion, the following works use this scale (Swann et al. 2009, 2010a, b; Gomez et al. 2011a).

In Gomez et al. (2011a), a seven-item verbal scale is proposed to obtain greater fidelity in the measurement of identity fusion, compared to the pictorial measure from Swann et al. (2009). The levels in the verbal scale are represented with the following sentences (1) “I am one with my group”, (2) “I feel immersed in my group”, (3) “I have a deep emotional bond with my group”, (4) “My group is me”, (5) “I’ll do for my group more than any of the other group members would do”, (6) “I am strong because of my group”, (7) “I make my group strong”. Swann et al. (2014) use this scale to measure identity fusion.

In Jiménez et al. (2016) the DIFI is introduced to combine the simplicity of the single pictorial item (Swann et al. 2009) with the higher predictive fidelity of the verbal scale (Gomez et al. 2011a). The scales presented in Swann et al. (2009), Gomez et al. (2011a) are not dynamic. In Jiménez et al. (2016) the DIFI is defined as a continuous measure of identity fusion, introducing a dynamic behavior for web-based questionnaires. The DIFI shows a figure formed by two circles of different sizes in the screen of the computer. The small circle represents “the self”, and the big circle represents “the team”. The player can move the small circle by clicking and dragging with the mouse to measure the degree to which the player feels part of the team.

3.6 Combined group anagram and CI experiments

A group anagram game entails cooperation in requesting and receiving letters, with the goal of forming more words with additional letters received from teammates. The only *face-to-face* cooperative team-play of an anagram game is reported in Charness et al. (2014). Their goal, like ours, is to foster CI among teammates. While this motivated our experiment, there are several differences in procedures and context. Major differences include (1) the game setup: we used larger fixed team compositions, while in Charness et al. (2014), the four-person team composition varied in time (by people voting themselves and others onto and off of teams); (2) in Charness et al. (2014), games were played face-to-face among participants in the same room cooperatively manipulating Scrabble-like tiles on a table, while we used remote players interacting in a game through a web application; and (3) in Charness et al. (2014), they measure CI with the proxy of PGG contributions, while we use DIFI score.

3.7 Modeling of time sequences of actions

We review modeling of time sequences because our ABMs are essentially in this class of models.

Many complex action sequences from human behavior are being collected from different environments, like sensors (Guralnik and Haigh 2002; Aipperspach et al. 2006; Bergmann et al. 2014; Tanaka et al. 2018) or computer-based applications (Kinnebrew et al. 2013; Chierichetti et al. 2014; Kurashima et al. 2018). Sequence mining techniques to model and predict human behavior in the real world can be used in different types of applications to improve a person’s life (e.g., mobile health (Kurashima et al. 2018), education patterns (Kinnebrew et al. 2013), smart-home optimization (Guralnik and Haigh 2002; Aipperspach et al. 2006)).

Sequence analysis is an important task to understand human behavior (Abbott 1995). The sequential pattern mining problem was first introduced by Agrawal and Srikant (1995), where the main focus is on the patterns present in the sequential order of different transactions. But the complexity of human behavior with time-varying, interdependent and periodic action sequences (Kurashima et al. 2018) makes accurate analysis and predictions a challenging task.

Kurashima et al. (2018) use activity data from logging applications to model the task of predicting future user actions and their timing through a mixture of Gaussian intensities. The model captures short-term and long-term periodic interdependencies between actions through Hawkes process-based self-excitations (Hawkes 1971). Accurate recommendations could improve a person’s health through the personalization of these applications. In Kinnebrew et al. (2013), a combination of sequence mining techniques uses data from computer-based learning environments to model students learning behavior patterns. Guralnik and Haigh (2002) use sequential pattern learning to model an agent-based system to aid elderly people in living longer in their homes. Aipperspach et al. (2006) use pervasive home sensors, like motion sensors, door close sensors, and floor pressure pads, to model and predict discrete human actions with smoothed n -grams.

We are not modeling specific actions in our work. Rather, we are modeling the sequencing of actions during an anagram game. The above works use primarily data from *in situ* environments, while our data come from human subjects experiments.

3.8 Evaluation of model predictions

Predictive models can have many forms. For example, simple classifier algorithms try to predict discrete class labels. Another technique used in predictive modeling is regression analysis, which tries to predict the mean value of a quantitative response variable. Also, the factor analysis approach, tries to predict the distribution of a set of correlated quantitative variables (i.e., predicts the values of some variables from knowing the values of others). The evaluation of prediction models can be developed using a

variety of different methods and metrics. For classification, the usual measure of error is the fraction of cases mis-classified, called the mis-classification rate or the error rate. For linear regression, the measure of accuracy is R^2 and the measure of error is the sum of squared errors or $1 - R^2$. For the method of factor analysis, when a model predicts a whole distribution, the negative log-likelihood is the usual measure of error, but sometimes a direct measure of the distance between the predicted and the observed distribution is used (Hand et al. 2001).

In this work, we are primarily concerned with using well-known measures to characterize the difference between two statistical distributions. In our work, one distribution is generated from experimental data, and one distribution is generated from predictions of models from Sect. 5. Gibbs and Su (2002) list ten metrics on probability measures: (1) Discrepancy, (2) Hellinger distance, (3) Relative entropy (or Kullback–Leibler divergence), (4) Kolmogorov (or Uniform) metric, (5) Lévy metric, (6) Prokhorov metric, (7) Separation distance, (8) Total variation distance, (9) Wasserstein (or Kantorovich) metric, and (10) χ^2 distance.

It is clear that there are many measures for comparing two probability distributions, and different ones are used in different settings. For our needs, we have chosen to use KL divergence (also called relative entropy). The KL divergence was introduced by Solomon Kullback and Richard Leibler in 1951 as the directed divergence between two distributions (Kullback and Leibler 1951).

The most important measure in information theory is called entropy and measures the uncertainty associated with a random variable. The entropy of a random variable X denoted $H(X)$ is a lower bound on the average length of the shortest description of the random variable (Cover and Thomas 1991). The concept of information entropy was introduced by Shannon (1948). The Shannon entropy, defined in Shannon (1948), measures how close a random variable is to being uniformly distributed. Shannon entropy estimates the average minimum number of bits needed to encode a string of symbols based on an alphabet size, and the frequency of the symbols is calculated using the following formula $H(X) = - \sum_{x \in X} P(x) \log P(x)$. The KL divergence measures the discrepancy between two probability distributions, and from which Shannon entropy can be constructed. For discrete probability distributions P and Q defined on the same probability space, the KL divergence between P and Q is defined to be

$$D_{KL}(P||Q) = - \sum_{x \in X} P(x) \log \left(\frac{Q(x)}{P(x)} \right).$$

In the simple case, a KL divergence of 0 indicates that the two distributions in question are identical. The KL



Fig. 5 Steps for the overall online game include: recruitment of players from Amazon Mechanical Turk (AMT), directions for the use of the platform, DIFI1 score procedure, anagram game, and DIFI2 score procedure. This figure directly maps onto the experimental components in Fig. 1

divergence is not symmetric. The evaluations of our models are described in Sect. 6.

3.9 Abduction and abductive loop

Works on constructive procedures for implementing abductive analyses include Haig (2005), Timmermans and Tavory (2012). We extend those works for abductive looping by making modeling a first-class process, and by adding the task of determining what to do in the next iteration. In addition to the applications cited in the Introduction, abduction was used to understand emergency room personnel's efforts to save injured people in terms of "social viability" (Timmermans 1999). Perhaps the work closest to ours is Singla and Mooney (2011) in that they develop models and make predictions based on data. However, their data are either artificially generated or address isolated individuals, and they use abduction rather than abductive iterations. Several additional works are provided in Sect. 1.3.

4 Experiments

In Sect. 4.1, we provide a description of the experiment and overview the web application (web app) software system for running games. An experiment consists of an anagram game and two executions of the dynamic identity fusion index (DIFI) procedure. We present analyses of the experimental data that illustrate how players interact in the anagram games in Sect. B.

4.1 Experiment description

The elements of an experiment, as specified in Fig. 5, are:

1. Players are recruited from Amazon Mechanical Turk (AMT), to play our anagram game.
2. Players receive directions on how to use the platform, including a description of the game and how to play it, and information about remuneration at the end of the game.
3. Players play the Dynamic Identity Fusion Index (DIFI) 1, DIFI1, procedure individually.

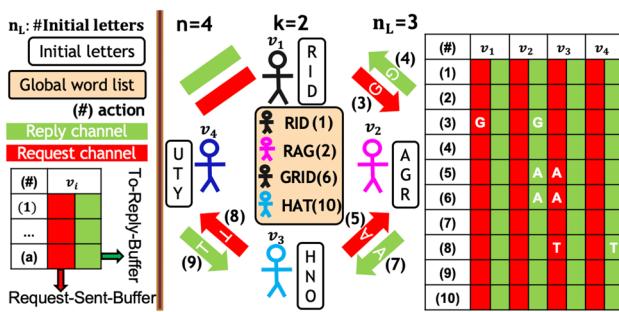


Fig. 6 Group anagram game configuration with a $k = 2$ regular graph on $n = 4$ players (v_1, v_2, v_3, v_4) with number of initial letters $n_L = 3$ assigned to each player, as shown in the boxes next to the players. Requests for letters and replies are sent across the channel links (red to request letters, green to reply with letter). Request-Sent-Buffer keeps track of player v_i 's letter requests. To-Reply-Buffer contains letter requests from other players to v_i . Numbers (#) denote actions by players during a game, and a table illustrating the time sequencing of the actions appears at right. Table 3 shows a detailed description of these actions (color figure online)

4. Players play the anagram game in a cooperative group setting.
5. Players play the DIFI2 procedure individually.

The terms DIFI1 and DIFI2 are used to indicate the first and second uses of the DIFI procedure (Fig. 5). The two DIFI procedures are the same.

4.1.1 Group anagram game description

The group anagram game is a word construction game, where n players cooperate in sharing letters to form and submit words of length ≥ 3 letters. Communication channels between pairs of agents mean that they can request and share letters with each other. An edge between nodes (players) v_i and v_j means that v_i and v_j can share letters with each other; v_i and v_j are neighbors. We use random regular graphs of degree k on the n players so that everyone has the same number of neighbors. A random k -regular graph on n players is an undirected graph such that each player has k neighbors assigned uniformly at random. Over all abductive loops, experiments are run in groups with nominal values of $10 \leq n \leq 20$ and with regular degrees $2 \leq k \leq 8$. We describe our motivation for using random regular graphs in Sect. 4.1.2.

An example game configuration and system states are provided in Fig. 6. The game configuration can be represented as a graph $G(V, E)$ where V is the set of nodes that represent players and E is the set of edges that are communication channels between pairs of nodes. Red channels are for letter request and green channels are for letter replies. The number of players is $n = 4$ with players v_1, v_2, v_3 , and v_4 , the degree of each player is $k = 2$, and the number of initial

Table 3 Action table detailing the sequences of actions by all players during the group anagram game example from Fig. 6. The first column defines the number of the sequence of actions during the game, appearing in Fig. 6. For this example, the duration of the game is 10 actions. The second column shows the player initiating the action. The third column shows the name of the action. The fourth column provides a description of the action

(#)	Player	Action	Description
(1)	v_1	Form word	v_1 forms word "RID"
(2)	v_2	Form word	v_2 forms word "RAG"
(3)	v_1	Request letter	v_1 requests v_2 for letter "G"
(4)	v_2	Reply letter	v_2 replies v_1 with letter "G"
(5)	v_3	Request letter	v_3 requests v_2 for letter "A"
(6)	v_1	Form word	v_1 forms word "GRID"
(7)	v_2	Reply letter	v_2 replies v_3 with letter "A"
(8)	v_3	Request letter	v_3 requests v_4 for letter "T"
(9)	v_4	Reply letter	v_4 replies v_3 with letter "T"
(10)	v_3	Form word	v_3 forms word "HAT"

letters per player is $n_L = 3$. The players have the following initial letters: $L_{v_1}^{\text{init}} = \{\text{RID}\}$, $L_{v_2}^{\text{init}} = \{\text{AGR}\}$, $L_{v_3}^{\text{init}} = \{\text{HNO}\}$, and $L_{v_4}^{\text{init}} = \{\text{UTY}\}$. Key (#) shows the sequence of actions by all the players during a game. In Fig. 6, the sequence of actions is detailed in Table 3, which narrates the actions. The To-Reply-Buffer and the Request-Sent-Buffer of Fig. 6 are buffers, per player, that contain outstanding requests-to-be-fulfilled and requests of letters, respectively. For example, in step (5) of Table 3, v_2 has a request from v_3 for the letter A. Therefore, v_3 has an entry A in its Request-Sent-Buffer and v_2 has an entry A in its To-Reply-Buffer. If/when v_2 fulfills that request [in the example this happens in step (7)], v_3 's "received letters" will contain an A, A will be removed from v_3 's Request-Sent-Buffer, and v_2 's To-Reply-Buffer will become empty.

Team members earn money by forming as many words as possible. Players are told that the total team earnings e_t are split evenly; each player receives e_t/n , so that it is in their interests to assist their neighbors. Players must form words with at least three letters. A single letter can be used any number of times in a word, e.g., a player can form the word TOT if she has a T and an O among her current letters (own letters and those received from neighbors) because the T can be used twice. Moreover, players do not lose letters that they use. Hence, a player has infinite multiplicity of each letter they possess so that letters can be reused any number of times. This means that a player only has to request a letter (and receive it) one time. Therefore if a player forms TOT, she still possesses T and O with which to form more words. A player can only share their initial letters with her neighbors; letters received from neighbors cannot be shared with others. These rules were designed to foster word

Table 4 Description of anagram game configurations played with players recruited from Amazon Mechanical Turk (AMT). There were 47 games with 289 players, of which 34 games and 224 players were used in analysis and modeling. Others were scoping experiments

Degree, k	No. Players, n	No. Games
2	10	18
2	20	10
3	15	1
4	15	9
5	15	2
6	15	3
8	15	4

construction, to increase earnings potential, and to foster team cohesion.

A total of 105 players participated in 47 games. The anagram game is played for 5 min; Table 4 shows all the game configurations played.

We provide an overview of the web application (app) game platform that we built. The web app software platform consists of the oTree infrastructure (Chen et al. 2016) for recruiting players from Amazon Mechanical Turk (AMT) and interactions during the game; Django Channels for player interactivity; and JavaScript and HTML for generating the screens for a consent form, instructions, information, a survey, and game interactions. Experiments and data analyses are part of the abductive loop of Sects. 2 and 7 and Fig. 4. This game platform was constructed as part of our work.

A screen shot of one player's screen at one point in time is shown in Fig. 7. Each player is given $n_L = 3$ letters that she can use to form words and that she can share with others. She has an infinite supply of letters so that sharing letters does not inhibit her own use of letters. A player can also request letters from her neighbors and if the neighbors provide those letters, then she can use those letters in words, but she cannot pass on the received letters.

Initially, a player sees her n_L own letters and those of all of her neighbors, but has access only to her own letters. Over the 5-min anagram game duration, players can form words, request letters from their neighbors and reply to requests.

4.1.2 Choice of random regular networks for experiments

There may be many possible network configurations to explore in some games. These include Erdos–Renyi random graphs, small-world networks, and scale-free networks. We select random regular networks for the following reasons.

First, since all players in a game have the same degree, all players have the same number of neighbors and the same

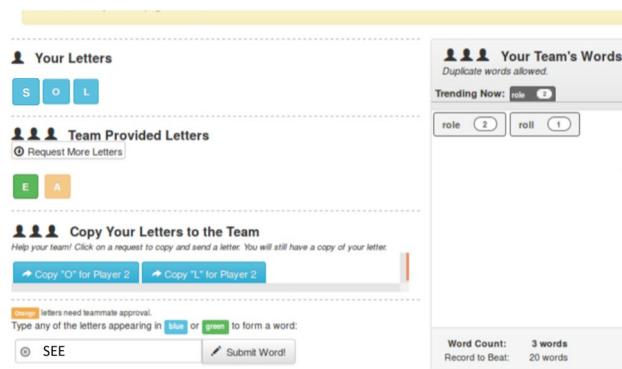


Fig. 7 The anagram game screen of the web app for one player. This player has own letters “S,” “O,” and “L” and has requested an “E” and “A” from neighbors. The “E” is green, so this player’s request has been fulfilled and so “E” can be used any number of times in forming words. But the request for “A” is still outstanding so cannot be used in words. Below these letters, it shows that player 2 has requested “O” and “L” from this player; this player has to reply to these requests, if this player so chooses. Below that is a box where the player types and submits new words, like “SEE” (color figure online)

maximum number of interactions. No player can be viewed as special.

Second, our experiments use between five and twenty players per game, consistent with other studies of this kind. With these numbers of players, one cannot construct Erdos–Renyi, scale-free, and exponential decay graphs with the requisite degree distributions because there is an insufficient number of nodes. For example, to generate a scale-free or power law degree distribution, one needs about 5000 to 10,000 nodes in a graph. (One can demonstrate this by generating a scale-free network on 1000 nodes in NetworkX, and you will see that the degree distribution is quite “choppy”). Hence, instances of these latter classes of graphs cannot be generated with roughly five to twenty nodes. Critically, though, we can model these classes of network. That is, we can specify, say, a scale-free network on $n = 5000$ players and model each player based on their degree in the graph, as described below.

Third, our analyses and modeling in Sect. 5 are based on representations of individual agents. We account for the local structure of a node (agent) in a game by considering its closed neighborhood (i.e., its degree, plus one), which consists of itself and its immediate neighbors. Consequently, our models account for each player, and each player’s closed neighborhood. This is, in particular, the Model M2 of Sect. 5.6. In this way, we can “connect” agents and their neighborhoods to form larger networks for modeling than the sizes of graphs we were able to test.

Fourth, by controlling degree in each experiment, we can conduct experiments at $k = 2, 4, 6$, and 8, and build models that interpolate for intermediate values of k (see Sect. 5). This enables a systematic approach to model building for

individual agent behavior. Note that this systematic approach includes the ability to model systems where numbers of neighbors is *heterogeneous* across agents, as in Erdos–Renyi, scale-free, and other network classes, precisely because we generate individual agent models. For example, with our models, we can simulate a single game where players have arbitrarily assigned node degrees, and the model assigned to each player is based on that node's degree.

Fifth, eight neighbors is used as the upper end of the degree range. With each player sharing three letters, a player with degree eight has access to 27 letters (including their own three letters), and hence a high probability of having access to almost all letters of the alphabet. In our models, as a first approximation, if a player has degree greater than eight, we use the $k = 8$ model for that agent.

Sixth, using a uniform degree for all players in a game enables us to generate multiple (replicate) sets of data across multiple individuals within one game. The importance of this replication in degree values is demonstrated in Sect. 6. In that section, we conduct rigorous analyses to identify the number of experimental observations required to drive down errors between model predictions and experimental data. An important result is that had we experimented with different network structures and hence a greater number of values for k (node degree), then the errors in our models would have been greater because of the “dispersion” (i.e., wider range of degree k) in experimental conditions that would have resulted. For example, if we had created one network with node degrees of 2, 3, 5, and 7 (among others), then we would have generated data over more k values, which would have dispersed the data over more k , and led to fewer data points at each k . This would result in larger errors in model predictions.

Seventh and finally, there is a real-time, pragmatic issue. We recruit game players through AMT. But not all players who promise to show up for the game do so at the appointed time and date. By using even regular degree values, we can produce valid random regular graphs as long as $n > k$. With graph structures such as scale-free (were it even possible), if players do not show up, one is faced with the ambiguous decision about which nodes in the communication graph to delete.

For all of these reasons, we chose to use random regular graphs in our study: it enables us to study behavior based on player degree and to construct and evaluate models. Additional graph structures (i.e., heterogeneous degree graphs) may be studied as part of future work.

4.1.3 DIFI description

The DIFI procedure precedes and follows the anagram game. Each player executes individually the DIFI procedure (Swann et al. 2010b), to measure the degree to which

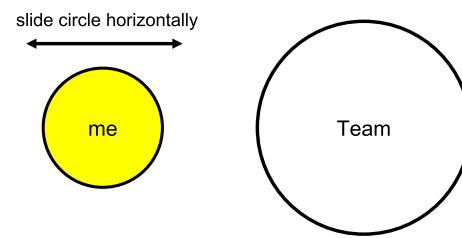


Fig. 8 DIFI game where player v_i moves the smaller circle, representing the player, either over (partially), or away from, the bigger circle that represents the team. The team circle is stationary. The distance δ between centroids of circles is measured. The distance is such that $\delta = 0$ corresponds to the small and large circles just touching; $\delta < 0$ means that the two circles are disjoint; and $\delta > 0$ means the two circles overlap. The distance δ is transformed into a DIFI value. The range in DIFI value is: $-100 \leq \delta \leq 125$. The DIFI score is a proxy for CI. This is an individual player game, played in isolation; results are never shared among players, so there is no concern over reprisals by other players, to foster honest actions

a player feels part of a team (i.e., associates their identity with that of a team). Each player does this individually by moving a circle in a browser, relative to a fixed team circle. The DIFI score is in the range $[-100, 125]$, with a score < 0 representing no overlap of circles, and therefore indicating no CI; $= 0$ representing the circles just touching; and > 0 indicating overlap of the two circles and hence formation of some level of CI. See Fig. 8. There are screens in the web app that also step each player through the steps in the DIFI game/procedure.

4.2 Experimental data

Experimental data are provided in “Appendix B”.

5 Agent-based models (ABMs) of the group anagram game and modeling results

We present three progressively more sophisticated ABMs of the anagram game that are used in the abductive loop analyses to follow in Sect. 7. All models were developed as part of the abductive loop process, but are presented here to emphasize their construction and evaluation, and to obviate the need for a large digression for the models in the description of the AL process in Sect. 7. Each model represents the behavior of one player or agent. The models are data-driven, and hence *inductive inference* is used with data in three ways: to inform model structure, model parameters, and to compute parameter values. Some of the figures appear in “Appendix C”.

In all models, we represent the set V of players and the set E of their communication channels (edges) as an undirected graph $G(V, E)$. The game is modeled as a discrete-time

Table 5 Actions of players in the model. The set A of actions is $A = \{a_1, a_2, a_3, a_4\}$

Item	Variable	Name	Description
1	a_1	Idling	Thinking
2	a_2	Reply	Replying to a neighbor with a requested letter
3	a_3	Request	Requesting a letter from a neighbor
4	a_4	Words	Forming and submitting a word

stochastic process, where at each time step, a player performs one of the actions from the action set A , consisting of: (1) a_1 : idling (i.e., thinking); (2) a_2 : replying to a neighbor with a requested letter, (3) a_3 : requesting a letter from a neighbor, and (4) a_4 : forming and submitting a word. Table 5 shows the actions.

5.1 Discrete-time stochastic process

In all ABMs, actions are taken at integer numbers of seconds; that is, simulations of interacting agents take place as time advances in discrete 1-second increments from 0 to 300. This time increment is based on the experimental data where no player takes two or more actions in one second.

We chose ABMs for their generative properties, fine granularity, and ability to model temporal effects. These enable us to more readily quantify “what if” scenarios (counterfactuals) as part of parametric studies and sensitivity analyses. Also, ABM maps well onto the actual experiments: players have connections in a network arrangement and they interact through their edges, taking actions at discrete times as in Fig. 6.

The choice of discrete time or discrete event simulation arises. If we selected discrete event simulations, then we would also have to predict the time at which the next action for a player takes place (at some Δt into the future). However, with discrete time, we know we are always predicting for the next time unit (here, one second). We also used a multinomial logistic regression model; other approaches could have been employed.

5.2 KL-divergence

To measure the performance of our models, we use *Kullback–Leibler divergence* between our model prediction on x and the experimental observation of x , *throughout this manuscript*. That is, we are comparing distributions of data: distributions of experimental data against distributions of model predictions. Most relevant for our work is Boltzmann’s (Bach 1990) concept of generalized entropy, where the entropy of a physical system is a measure of disorder related to it. Kullback and Leibler (1951) derived

an information measure, now referred to as the KL divergence, the negative of Boltzmann’s entropy. The motivation for Kullback and Leibler’s work was to provide a rigorous definition of information. The Kullback–Leibler distance can be conceptualized as a directed distance between two models, say a and b (Kullback 1959). This is a measure of discrepancy. It is not a simple distance because the measure from a to b is not the same as the measure from b to a . It is a directed, or oriented, distance. The KL divergence $D_{KL}(a, b)$ is always positive, except when the two distributions a and b are identical (i.e., $D_{KL}(a, b) = 0$ if and only if $a(x) = b(x)$ everywhere). Entropy is zero if there is unit probability at a single point. If the distribution is widely dispersed over a large number of individually small probabilities, then the entropy is high (e.g., $D_{KL} > 1$).

5.3 Overview of the three agent-based models

ABM M0 is a baseline model, where each player makes a probabilistic transition from action $a_i \in A$ to action $a_j \in A$. The transition matrix is time invariant and is the same for all players. Data from the experiments is used to infer the model parameters using a ring topology (degree of each node is 2) of player connectivity within an anagram game. Model M1 is similar to M0 but with the crucial difference that the transition matrix is time variant. Model M2 is similar to M1 but now instead of a ring topology, we used other topologies and infer model parameters (degree from 2 to 8). Models M0, M1 and M2 predict the actions of A for a player but are generic in that letter request a_3 , letter reply a_2 , and submit word a_4 are not associated with particular letters. For example, if the player action is a_4 , then the model assumes that the player can form a word. Table 6 shows a description of the three progressively sophisticated models.

Models M0, M1, and M2 are presented in Sects. 5.4, 5.5, and 5.6, respectively. In each of these subsections, model development and results are provided.

Throughout, we use k to denote the number of neighbors (degree) of an agent $v \in V$. Also, we evaluate five variables and their distributions, across all players in a set of games, in comparing models and experiments: $x = (x_1, x_2, x_3, x_4, x_5)$, where x_1 is the number of letter replies received ($RplR$); x_2 is the number of replies sent ($RplS$); x_3 is the number of letter requests received ($RqsR$); x_4 is the number of requests sent ($RqsS$); and x_5 is the number of words formed ($Wrds$). Table 7 summarizes these variables.

In the results sections for each model, simulations are performed using ABMs that implement each of the described models. These simulations produce, for each player, time histories of the actions in Table 7. One hundred simulations are run and results are averaged across these simulations, i.e., are averaged across all players in each simulation. These data are post-processed to generate distributions of

Table 6 Progressively sophisticated models of the group anagram game are developed in this work. Models were constructed in order M0, M1, and M2. The incremental improvements in models are given in columns two and three, in terms of transition probabilities and degrees of players in the games

Model	Transition probabilities	Degree k
M0	Fixed	2
M1	Temporal	2
M2	Temporal	2, 4, 6, 8

Table 7 Variables that are measured in experiments for each player, and predicted with models for each agent, where vector $x = (x_1, x_2, x_3, x_4, x_5)$. All $x_i, 1 \leq i \leq 5$, are time dependent

Item	Variable	Name	Description
1	x_1	RplR	Number of replies received
2	x_2	RplS	Number of replies sent
3	x_3	RqsR	Number of requests received
4	x_4	RqsS	Number of requests sent
5	x_5	Wrds	Number of words formed

the variables in Table 7. These distributions from ABM predictions are compared against corresponding distributions generated from experiments.

Note that fixing $n = 10$ in all simulations does not introduce errors because the distributions that we use are density distributions, not counts. Thus, the number of players is normalized out of all comparisons of distributions of experimental data and model predictions.

Table 8 shows the structure of comparisons of results for each of the models M0, M1, and M2. First, comparisons are made between distributions of experimental results and model predictions, for each x_i of Table 7, at the end of a game (i.e., over all 5 min of an anagram game). Then, these data are broken down into 1-min intervals to assess temporally the distributions of data and predictions. Next, we compute KL-divergence values that provide a scalar representing how well the model predictions of the distributions of x_i compare with those of the experimental data. From Sect. 5.2, $D_{KL} = 0$ means the model distribution agrees very well with the corresponding experimental distribution. As D_{KL} increases from zero, model predictions worsen. Table 8 denotes that these comparisons are performed over all 5 min of the anagram game (number 3), corresponding to the end of the group anagram game, and for each 1-min interval over the game (number 4) of Table 8. Finally, we compare these sets of computed D_{KL} across all x_i of Table 7. The reason for the temporal breakdown is to

Table 8 Summary of the model comparison plots for each of the models M0, M1, and M2. For each model, we collect the data into the five groups shown. See the text for details and justification. The

fifth column indicates the time period, in minutes, over which experimental data are compared to model predictions. These plots facilitate comparisons of final outcomes and temporal performance of models

No.	Method	Plot	Variables, player actions	Time
1	Comparisons of distributions at end of game	(a)	x_1	0–5
		(b)	x_2	0–5
		(c)	x_3	0–5
		(d)	x_4	0–5
		(e)	x_5	0–5
2	Temporal comparisons of distributions	(a)	x_1, x_2, x_3, x_4, x_5	0–1
		(b)	x_1, x_2, x_3, x_4, x_5	1–2
		(c)	x_1, x_2, x_3, x_4, x_5	2–3
		(d)	x_1, x_2, x_3, x_4, x_5	3–4
		(e)	x_1, x_2, x_3, x_4, x_5	4–5
3	Comparisons of KL-divergence distributions at end of game		x_1, x_2, x_3, x_4, x_5	0–5
4	Temporal comparisons of KL-divergence distributions	(a)	x_1, x_2, x_3, x_4, x_5	0–1
		(b)	x_1, x_2, x_3, x_4, x_5	1–2
		(c)	x_1, x_2, x_3, x_4, x_5	2–3
		(d)	x_1, x_2, x_3, x_4, x_5	3–4
		(e)	x_1, x_2, x_3, x_4, x_5	4–5
5	Comparisons of KL-divergence distributions combining all variables		x_1, x_2, x_3, x_4, x_5	0–5

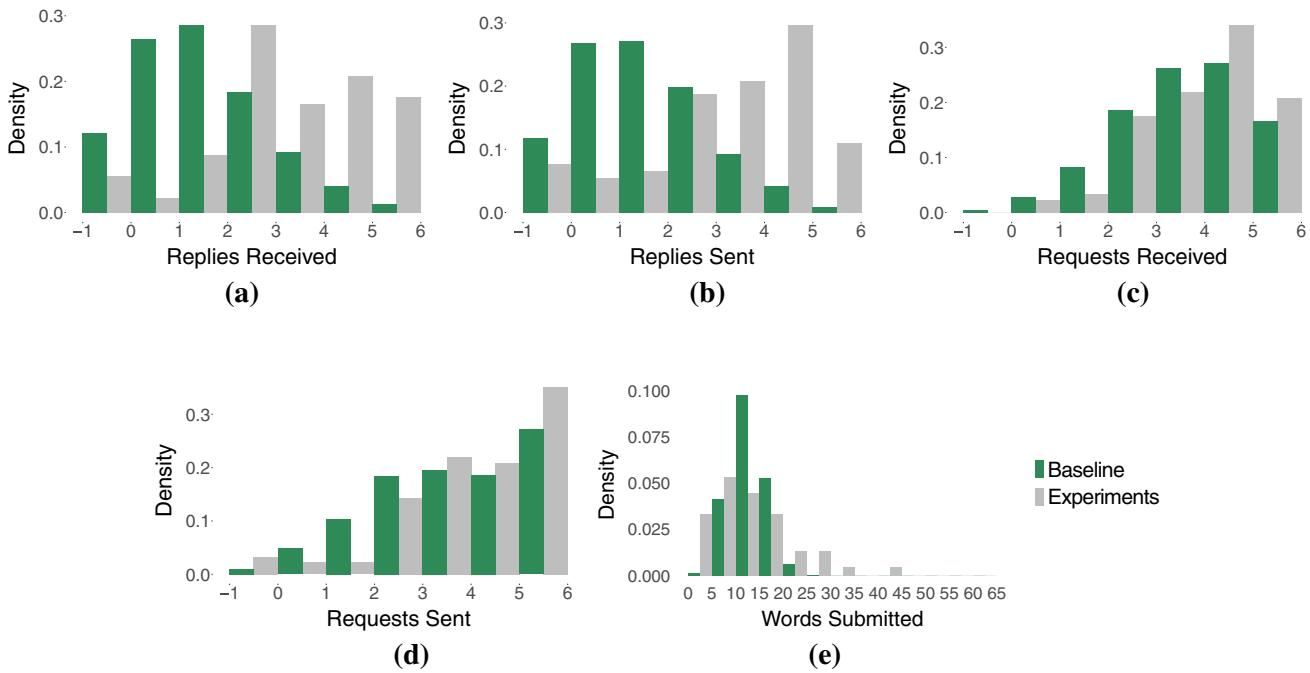


Fig. 9 ABM baseline model M0 predictions of the $k = 2$ experiments (in green) and experimental data (in gray), over the entire 5-min group anagram game. The probability density function is show for **a** distribution of replies received, **b** distribution of replies sent, **c** distribution of requests received, **d** distribution of requests sent, and **e** distribution of words formed, each at the end of the 5-min anagram game (gray bars are experimental data) for all $k = 2$ experiments. The Baseline model M0 predictions are from 100 simulations of a $n = 10$ player game. It is clear from visual inspection that model M0 predictions are in better agreement with the experiment data for the requests received and requests sent variables. We make this comparison more precise using KL divergence below in Fig. 10 (color figure online)

examine model predictions over time. Temporal comparisons are hidden in numbers 1, 3 and 5 of Table 8, which examine aggregated data.

5.4 Baseline agent-based model M0

5.4.1 ABM M0 development

The goal is to accurately quantify the transition probability from one action $a(t) = a_i$ at time t to the next action $a(t+1) = a_j$ for each agent $v \in V, i, j \in [1..4]$ and $a(t) \in A$. For clarity, we use i and j to represent the actions a_i and a_j . Agent v executes a stochastic process driven by transition probability matrix $\Pi = (\pi_{ij})_{m \times m}$, where $m = |A|$ (here, $m = 4$) and

$$\pi_{ij} = Pr(a(t+1) = j | a(t) = i) \quad \text{with} \quad \sum_{j=1}^m \pi_{ij} = 1. \quad (1)$$

The transition matrix Π is formed from the data by using successive pairs of actions of players in experiments so that the 16 values of π_{ij} in Eq. (1) are *constant*, i.e., time invariant. The matrix in Eq. (2) shows the transition probabilities for Model M0 (the baseline model) generated from experiment data with $n = 10, k = 2$. For example, given that the action of a player v_i at time t is a_2 (replying to a letter

request), the probability that v_i 's next action, at time $(t + 1)$, is a_1 (thinking) is 0.84.

$$\Pi = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ a_1 & 0.93 & 0.01 & 0.02 & 0.04 \\ a_2 & 0.84 & 0.16 & 0 & 0 \\ a_3 & 0.98 & 0.01 & 0.01 & 0 \\ a_4 & 0.93 & 0.01 & 0 & 0.06 \end{bmatrix} \quad (2)$$

5.4.2 ABM M0 (baseline) results

We address all of the results in Table 8 for model M0.

Comparisons of distributions between model and experiments for individual variables at the end of the anagram game Figure 9 shows the ABM M0 predictions of the $k = 2$ experiments. Figure 9a shows the distribution of replies received, Fig. 9b shows the distribution of replies sent, Fig. 9c shows the distribution of requests received, Fig. 9d shows the distribution of requests sent, and Fig. 9e shows the distribution of words formed, each at the end of the 5-min anagram game (gray bars) for all $k = 2$ experiments, compared to Baseline M0 predictions (green) for 100 simulations of an $n = 10$ player game. It is clear from visual inspection that model M0 predictions



Fig. 10 KL-divergence values for the Baseline Model M0 across the five parameters of x : lower values are better. M0 does a better job predicting the number of Requests Received and Requests Sent. Analyses are based on the data of Fig. 9, over 5 min, at the end of a game

are in better agreement with the experimental data for the requests received and requests sent variables. We make this comparison more precise using KL divergence in Fig. 10.

Temporal comparisons of distributions between model and experiments for individual player actions Appendix B.1.5 in Cedeno (2019) shows the figures resulting from the temporal analysis by minute of distributions between Model M0 and experiments for $k = 2$. Each plot contains data over a time window for each variable of x from Table 7. Often, but not always, the largest discrepancies between the model predictions and experiments occur in the first minute of the game.

Comparisons of KL-divergence values between model and experiments for individual variables at the end of the anagram game. Figure 10 shows the KL-divergence values for the baseline M0 across the five parameters of x : lower values are better. M0 does a better job predicting the number of requests received and requests sent at the end of a game. These data span the entire 5-min game. That is, the request-related operations are better predicted than reply operations.

Temporal comparisons of KL-divergence values between model and experiments for individual player actions. Figure 26 shows the temporal KL-divergence values for the baseline M0 across the five parameters of x , at 1-min intervals: lower values are better. Each figure contains data over a time window: Fig. 26a shows the 0–1 min, Fig. 26b shows the 1–2 min, Fig. 26c shows the 2–3 min, Fig. 26d shows the 3–4 min, and Fig. 26e shows the 4–5 min results of the 5-min anagram game. These plots show that request-related predictions are better than reply-related predictions for the first 3 min, but are worse for the last 2 min, based on KL-divergence. Reply-related predictions are better in the second half of the 5-min anagram games, but Fig. 23 shows that in experiments, there are fewer replies in the second half of the games.

Comparisons of KL-divergence values between model and experiments for combining all variables. Figure 3.15 in

Cedeno (2019) shows the distribution of KL-divergence values for comparing distributions of model output with corresponding distributions of experimental data for the anagram game. The model is the ($n = 10$, $k = 2$) baseline. The data sets used in the comparison are ($n = 10$, $k = 2$). There are 30 values in the distribution, with five values for variables x_i over the 5-min game, at the end of the game; and 25 values for the five variables of x over five intervals of 1 min duration. It shows that for model M0, some KL-divergence values are high (e.g., > 0.5), indicating poor agreement between model predictions and the experiment data. As we see in Figs. 10 and 26, M0 does not do a good job predicting the number of replies received, replies sent, and words formed.

5.5 Agent-based model M1

Model M1 is similar to M0 but with the important enhancement that the transition matrix Π is time variant.

5.5.1 ABM M1 development

To make Π [and its components π_{ij} in Eq. (1)] dynamic in time and account for history effects, four variables are introduced in Eq. (3): number $z_L(t)$ of letters that v has available to use (i.e., in hand) at t ; number $z_W(t)$ of valid words that v has formed; size $z_B(t)$ of the buffer of letter requests that v has yet to reply to; and number $z_C(t)$ of consecutive time increments that v has taken the same action. See Table 9. Thus, letting $z = (1, z_L, z_W, z_B, z_C)_{g \times 1}$, we can model π_{ij} as a function of these covariates, among other variables.

We use a multinomial logistic regression to model π_{ij} —the probability of a player taking action a_j at time $t + 1$, given that the player took action a_i at time t —as

$$\pi_{ij} = \frac{\exp(z' \beta_j^{(i)})}{1 + \sum_{l \neq i} \exp(z' \beta_l^{(i)}),} \quad (3)$$

where $\beta_j^{(i)} = (\beta_{j1}^{(i)}, \dots, \beta_{j,g}^{(i)})'$, for $j \neq i$, and $\beta_i^{(i)} = \mathbf{0}$, prime indicates vector transpose, and $\beta_{j,h}^{(i)}$ are the elements of $\beta_j^{(i)}$, with $1 \leq h \leq g$ being the index of the element of the z vector. For a given i , the parameter set can be expressed as

$$\mathbf{B}^{(i)} = \begin{pmatrix} \beta_{11}^{(i)} & \beta_{12}^{(i)} & \dots & \beta_{1,g}^{(i)} \\ \beta_{21}^{(i)} & \beta_{22}^{(i)} & \dots & \beta_{2,g}^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{41}^{(i)} & \beta_{42}^{(i)} & \dots & \beta_{4,g}^{(i)} \end{pmatrix}, \quad (4)$$

with matrix entries $\beta_{j,h}^{(i)}$. Parameters in Eq. (4) are inferred from the $k = 2$ experimental data using the framework of maximum likelihood estimation for the multinomial distribution.

Table 9 The feature vector $z = (z_L(t), z_W(t), z_B(t), z_C(t))$ used in the models M1 and M2. These capture history effects in determining the next action of a player

Variable	Name	Description
z_B	Size of reply buffer	Number of current letter requests to which this player may reply. Captures the notion that the more letter requests that have not been replied to, the more likely v is to reply
z_L	Number of letter in hand	Number of unique letters in hand to form words. Captures the idea that the more letters v has in hand, the more likely the agent is to form words
z_W	Number of words formed	Number of words formed. Captures the notion that the more words that have been formed, the larger the vocabulary of the player
z_C	Number of consecutive actions	Number of consecutive time steps at which player takes the same action. Captures the notion that the more time v is idle (thinking), the more likely v will take some other action at the next timestep

5.5.2 Inductive inference

We address the three dimensions of inference stated above: (1) model structure; (2) model parameters; and (3) parameter values. First, the model structure is informed by the $k = 2$ data, by design, as described above. Second, the parameters identified in the feature vector z are described and justified in Table 9. In fact, we claim that identifying this feature vector has elements of art. Third, parameters in Eq. (4) are inferred from the $k = 2$ experimental data using the framework of maximum likelihood estimation for the multinomial distribution.

The reason to emphasize inductive inference is because this is an integral part of the abductive looping process, and of abduction itself: the data drive the model and theory development and hypothesis identification, and not the other way around.

5.5.3 ABM M1 results

Results for Model M1 are provided according to Table 8 as was done for Model M0. In many cases, we compare KL-divergence values for M0 and M1 to show improvements in performance. These results, like those for model M0, are compared against the $k = 2$ data in Table 4.

Comparisons of distributions between models and experiments for individual variables at the end of the anagram game Figure 11 shows M0 and M1 model predictions and experimental data distributions for all variables in Table 7. These data are over all 5 min of the anagram game for all $k = 2$ experiments. Model predictions are averages over 100 simulations with $n = 10$ players. Figure 11a shows the distributions of replies received, Fig. 11b shows the distributions of replies sent, Fig. 11c shows the distributions of requests received, Fig. 11d shows the distributions of requests sent, and Fig. 11e shows the distributions of words formed. It is clear from visual inspection that model M1 predictions are in better agreement with the experiment data than are M0 predictions. We make this comparison more precise using KL divergence in Fig. 12.

Temporal comparisons of distributions between models and experiments for individual variables Appendix B.1.6 in Cedeno (2019) shows the figures resulting from the temporal analysis by minute of distributions between Models M0, M1 and Experiments for $k = 2$. Each plot contains data over a 1-min time window for each variable of x from Table 7. It is clear from visual inspection that model M1 predictions are in better agreement with the experiment data than are M0 predictions.

Comparisons of KL-divergence values between models for individual variables at the end of the anagram game Figure 12 shows KL-divergence values for comparing distributions of model outputs with corresponding distributions of experimental data for the anagram game. The models are (baseline) M0 and M1 for the ($n = 10$, $k = 2$) experiments. The comparisons are at the end of the game, i.e., at $t = 5$ min, over the entire game. For each experiment/model combination, the variables (and hence distributions) compared are: number of replies received, number of replies sent, number of requests received, number of requests sent, and number of words formed. Lower values are better. This figure shows that M1 generates predictions much closer to the experimental data than does M0. For example, M1 significantly reduces the reply-related and words formed KL-divergence values (weaknesses of model M0 as shown in Fig. 10).

Temporal comparisons of KL-divergence values between models for individual player actions Figure 27 shows the temporal KL-divergence values for the baseline M0 and M1 across the five parameters of x : lower values are better. Each plot contains data over a time window: Fig. 27a for 0–1 min, Fig. 27b for 1–2 min, Fig. 27c for 2–3 min, Fig. 27d for 3–4 min, and Fig. 27e for 4–5 min time intervals of the 5-min anagram game.

The plots demonstrate that KL-divergence values for the model M1 predictions are closer to the experimentally-determined data distributions than are those from model M0. While Model M0 has good predictions for the minute 3 and minute 5 (with the exception of the words formed), Model M1 has better predictions for the minute 3 and minute 5 for

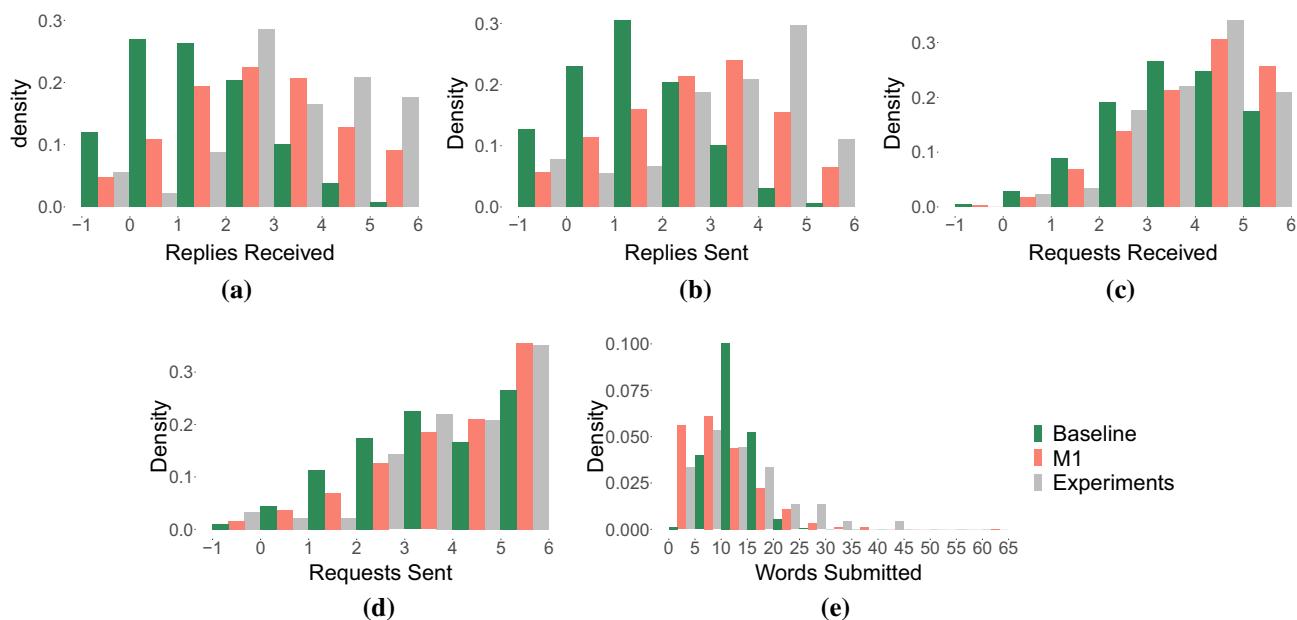


Fig. 11 Baseline Model M0 and Model M1 predictions of the $k = 2$ experiments, along with the experimental data. The probability density function is show for **a** distribution of replies received, **b** distribution of replies sent, **c** distribution of requests received, **d** distribution of requests sent, and **e** distribution of words formed, each at the end of the 5-min anagram game (gray bars) for all $k = 2$ experiments, compared to M1 predictions (red) for 100 simulations of an $n = 10$ player game. M1 predictions (red) for 100 simulations of an $n = 10$ player game. The Baseline Model M0 is shown in green for comparison. It is clear from visual inspection that model M1 predictions are in better agreement with the experiment data than are M0 predictions. We make this comparison more precise using KL divergence in Fig. 12 (color figure online)



Fig. 12 KL-divergence values for the Baseline Model M0 and Model M1 across the five parameters of x : lower values are better. The modeling conditions are those of the experiments with $k = 2$. This figure shows that M1 greatly improves a weakness of model M0 in poorly representing RplR (number of replies received), Rpls (number of replies sent), and Wrds (number of words formed)

all five x variables of Table 7. These data are significant because they evaluate the quality of the models to predict behavior temporally. That is, just because a model can produce predictions at the end of some scenario, this does not mean that it can capture the trajectory (or time evolution) of phenomena. With these types of plots, we demonstrate that our models do capture temporal behavior.

Comparisons of KL-divergence distributions between models and experiments for combining all variables Figure 3.19 in Cedeno (2019) shows the distribution of KL

divergence for comparing distributions of model output with corresponding distributions of experimental data for the anagram game. The models are ($n = 10, k = 2$) M0 and M1. The data sets used in comparison are experiments: ($n = 10, k = 2$). There are 30 values in the distribution, with five values for each variable x at the end of the game, and 25 values for the five variables x over five intervals of 1-min increment. It shows that for Model M1, the great majority of KL-divergence values are less than 0.2, while they can be much greater for Model M0.

Summary of M0 and M1 model comparisons. Clearly, ABM M1 is in better agreement with the experimental data compared to the baseline model. From KL-divergence values in Fig. 12, it is clear that the predictions of M1 represent the experimental data better than those of the baseline model.

In addition, we use M1 to make predictions for anagram games with $k > 2$, resulting in more interactions. Counterintuitively, as shown in Fig. 13, the number of replies does not change as k increases. These results call for more experiments at larger k . Note that we exercise M1 learned from experiments with $k = 2$. The results in Fig. 13 indicate that M1 predicts no changes in the number of letter replies received as k increases, which seems counter intuitive. One would expect more letter requests and replies with increasing numbers of neighbors. These types of data lead us to construct model M2.

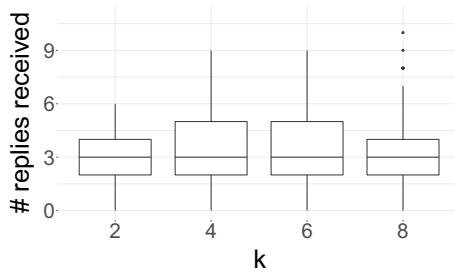


Fig. 13 M1 model distributions predicted for the number of replies received at the end of game ($n = 10, 100$ simulations), for different regular degrees k of the game network G . This partially motivated our development of ABM M2, since model M1 predictions do not vary significantly with the number of a player's neighbors

We remark that we also fitted M1 using experimental data with $k = 4$ (call this Model M1b), and consequently made predictions for the case of $k = 2$. We compared the distributions of x between prediction and experimental results using KL-divergence, and determined values in the range 0.11 to 0.46, indicating good predictions. Note that Model M1b is interpolating when it predicts $k = 2$ experimental data, while Model M1 is extrapolating to predict $k = 4$ experimental data.

5.6 Agent-based model M2

5.6.1 ABM M2 development

Model M1 was developed with data where all game players have the same degree $k = 2$. To generalize M1 to incorporate various k , we conducted additional experiments with $2 < k \leq 8$ as a part of the second AL (Sect. 7.4). We resume from the description of Model M1 in Sect. 5.5 and Eqs. (3) and (4).

Recall that for Model M1, $\beta_{jh}^{(i)}$ in $\mathbf{B}^{(i)}$ denotes parameters that are used to compute the transition probability π_{ij} based on a player taking action a_i at time t and action a_j at $(t+1)$. Since there are $m = 4$ possible player actions in action set A , $1 \leq i, j \leq m$. As in Sect. 5.5, h is the index of an element of the z vector with $1 \leq h \leq g$. The $\beta_{jh}^{(i)}$ are determined from analyses of the transitions in the experimental data.

Here, we build a hierarchical model to incorporate the effect of agent degree k . For different values of k , the parameter coefficients in $\mathbf{B}^{(i)}$ of Eq. (4), used in Eq. (3), are now a function of k , denoted as $\mathbf{B}^{(i)}(k)$. We use an orthogonal polynomial basis to construct a continuous and smoothing function for $\beta_{jh}^{(i)}(k)$ for any given i, j, h , as

$$\beta_{jh}^{(i)}(k) = \alpha_{j,h}^{(i,0)} + \alpha_{j,h}^{(i,1)}\xi_l(k) + \alpha_{j,h}^{(i,2)}\xi_q(k), \quad (5)$$

where ξ_l and ξ_q are the linear and quadratic functions of the orthogonal basis in terms of k . We have

$$\mathbf{B}^{(i)}(k) = \mathbf{C}_0^{(i)} + \mathbf{C}_1^{(i)}\xi_l(k) + \mathbf{C}_2^{(i)}\xi_q(k), \quad (6)$$

where

$$\mathbf{C}_r^{(i)} = \begin{pmatrix} \alpha_{11}^{(i,r)} & \alpha_{12}^{(i,r)} & \dots & \alpha_{1,g}^{(i,r)} \\ \alpha_{21}^{(i,r)} & \alpha_{22}^{(i,r)} & \dots & \alpha_{2,g}^{(i,r)} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{41}^{(i,r)} & \alpha_{42}^{(i,r)} & \dots & \alpha_{4,g}^{(i,r)} \end{pmatrix}, \quad r = 0, 1, 2, \quad (7)$$

correspond to the constant, linear, and quadratic coefficient matrices in Eq. (6), with $\alpha_{ih}^{(i,r)} = 0$ for any r and h . Here, $\alpha_{jh}^{(i,r)}$ in Eqs. (5) and (7) are the elements of $\mathbf{C}_r^{(i)}$ determined from the successive pairs of actions a_i at time t and a_j at time $t+1$, with current values of z .

Equation (4) is a special case of Eq. (6) when $\mathbf{C}_1^{(i)} = \mathbf{0}$ and $\mathbf{C}_2^{(i)} = \mathbf{0}$, i.e., when the coefficient matrix $\mathbf{B}^{(i)}$ is not a function of k .

5.6.2 Inductive inference

We address the three dimensions of inference, as for M1: (1) model structure; (2) model parameters; and (3) parameter values. In this case, the model structure we employ to capture the effect of k was identified a priori. However, if the model structure was found lacking, we would have tried another approach. The model parameters given in Eqs. (6) and (7) were also anticipated owing to the development of ABM M1. Hence, these first two steps were not solely driven by the data. To estimate the parameters sets $\mathbf{C}_0^{(i)}, \mathbf{C}_1^{(i)}, \mathbf{C}_2^{(i)}$, we use maximum likelihood estimation across the experimental observations for $k = 2, 4, 6$, and 8. For a given i and k , denote the corresponding observational data as $\mathcal{D}_k^{(i)}$. Then we conduct parameter estimation by

$$\hat{\mathbf{C}}_0^{(i)}, \hat{\mathbf{C}}_1^{(i)}, \hat{\mathbf{C}}_2^{(i)} = \arg \max \sum_{k=d_{min}}^{d_{max}} \log L(\mathbf{C}_0^{(i)}, \mathbf{C}_1^{(i)}, \mathbf{C}_2^{(i)} | \mathcal{D}_k^{(i)}),$$

where $L(\mathbf{C}_0^{(i)}, \mathbf{C}_1^{(i)}, \mathbf{C}_2^{(i)} | \mathcal{D}_k^{(i)})$ is the likelihood function with respect to the data $\mathcal{D}_k^{(i)}$ collected under the setting of k neighbors in the experiments of Sect. 4.

5.6.3 ABM model M2 results

Results for model M2 are provided according to Table 8. Results are often compared to those for model M1.

Comparisons of distributions between models and experiments for individual variables at the end of the anagram game Figure 14 shows data distributions at the end of the 5-min anagram game (gray bars) for all $k = 2$ experiments, compared to M2 predictions of distributions (blue) for 100 simulations of an $n = 10$ player game. These results are over all 5 min of the group anagram game. Figure 3.22 in Cedeno

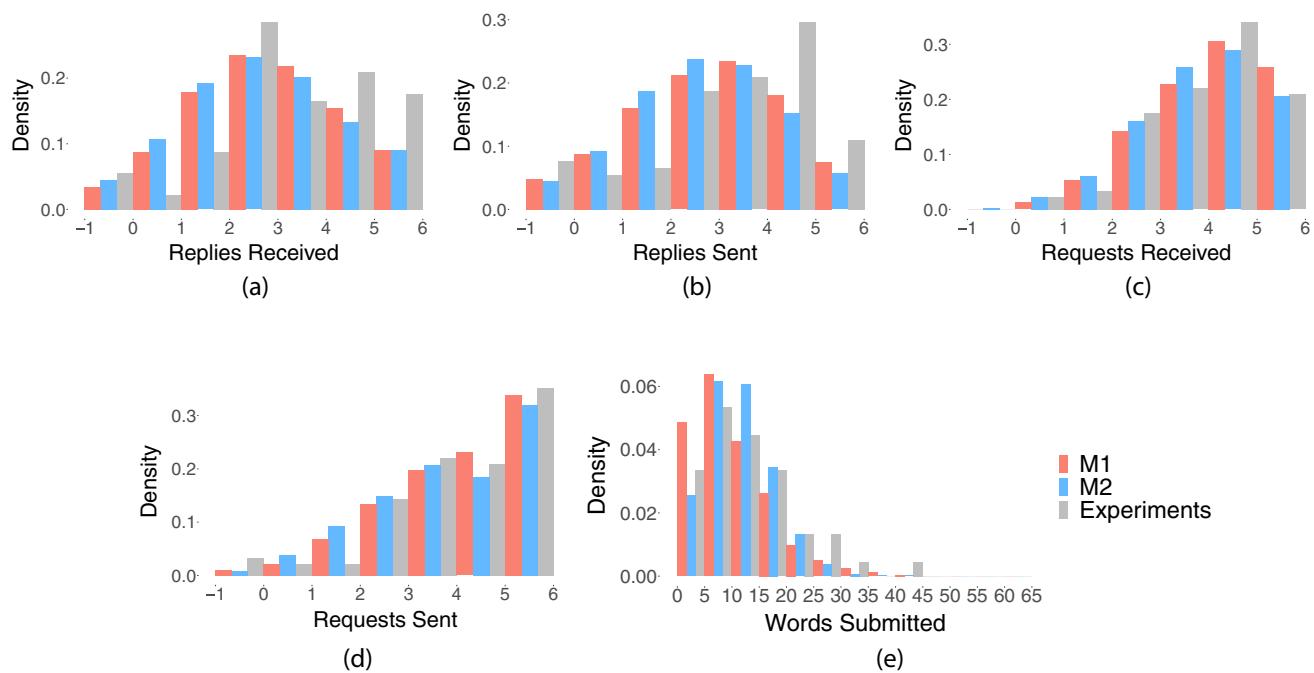


Fig. 14 Model M1 and Model M2 predictions of the $k = 2$ experiments, and experimental data, over all 5 min of the group anagram games. The probability density distributions are shown for **a** distribution of replies received, **b** distribution of replies sent, **c** distribution of requests received, **d** distribution of requests sent, and **e** distribution of words formed, each at the end of the 5-min anagram game (gray bars are experimental data) for all $k = 2$ experiments, compared to M2 predictions (blue) for 100 simulations of an $n = 10$ player game. The Model M1 predictions are shown in red for comparison. It is clear from visual inspection that models M1 and M2 generate similar predictions, in agreement with the experiment data, as M1 is learned solely from $k = 2$ experimental data. We make this comparison more precise using KL divergence in Fig. 15 (color figure online)

(2019) shows data distributions at the end of the 5-min anagram game (gray bars) for all $k = 4$. In Appendix B.1.7, in Cedeno (2019), Figure B.15 shows data distributions at the end of the 5-min anagram game (gray bars) for all $k = 6$. Figure B.16 Model M1 is shown in red for comparison. In all of these figures, Figure (a) shows the distributions of replies received, Figure (b) shows the distributions of replies sent, Figure (c) shows the distributions of requests received, Figure (d) shows the distributions of requests sent, and Figure (e) shows the distributions of Words Formed. M2 gives much better performance, as expected, as it explicitly accounts for agent degree. As expected, M1 and M2 perform equally well for $k = 2$ as M1 is learned from $k = 2$ experimental data.

Temporal comparisons of distributions between models and experiments for individual variables Appendix B.1.8 in Cedeno (2019) shows the temporal analysis by minute of distributions for models M1 and M2 and experiments for $k = 2$. Each plot contains data over a time window of 1 min. For $k = 2$ experiments, Figure B.17 shows temporal analysis for the number of Replies Received at the end of each minute. Figure B.18 shows temporal analysis for the number of Replies Sent at the end of each minute. Figure B.19 shows temporal analysis for the number of Requests Received at the

end of each minute. Figure B.20 shows temporal analysis for the number of Requests Sent at the end of each minute. Figure B.21 shows temporal analysis for the number of Words Formed at the end of each minute.

Collections of plots for each of $k = 4, 6$, and 8 are analogously provided in Appendix B.1.8 in Cedeno (2019). As expected, M1 and M2 perform equally well for $k = 2$, as M1 is learned from $k = 2$ experimental data. For $k > 2$, M2 performs better. We make this comparison more precise using KL divergence below.

Comparisons of KL-divergence distributions between models and experiments for individual variables at the end of the anagram game Figures 15 and 16 in this section, and Figures B.37 and B.38 in Appendix B.1.9 in Cedeno (2019) show KL-divergence values for comparing distributions of model outputs with corresponding distributions of experimental data, for the group anagram game. The figures are for, respectively, $k = 2, k = 4, k = 6$, and $k = 8$ experiments. The models are M1 (red) and M2 (blue). These four figures show clear and interesting behavior. Model M1 agrees better with experiments than does Model M2 for $k = 2$, since Model M1 was specifically developed with $k = 2$ data. However, for larger k ($4 \leq k \leq 8$), Model M2 does better than M1. This is because Model M2 was developed using data

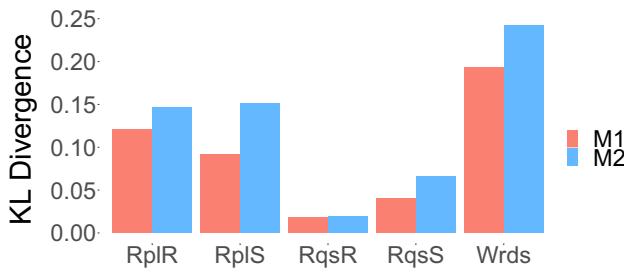


Fig. 15 The plot shows on the x -axis KL-divergence values for the M1 and M2 model predictions at the end of the 5 min anagram game. Here we compare $k = 2$ data for M1 and M2 model predictions with the experiments across the five parameters of x : lower values are better. This figure shows that model M1 and M2 generate similar predictions to the experimental data, with Model M1 slightly better (color figure online)

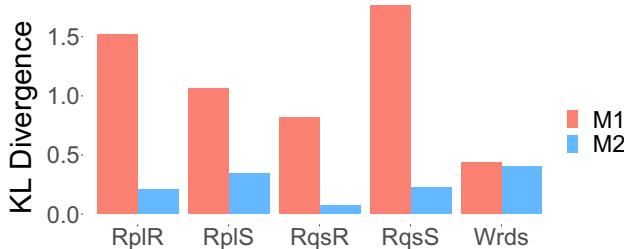


Fig. 16 The plot shows on the x -axis KL-divergence values for the M1 and M2 model predictions at the end of the 5 min anagram game. Here we compare M1 and M2 model predictions of the $k = 4$ data with the experiments across the five parameters of x : lower values are better. This figure shows that M2 gives much better performance than M1 in predicting the time to generate an action for an agent. M2 gives better performance, as expected, as it explicitly accounts for agent degree (color figure online)

across all of these k values. Hence, to obtain a wider range in input space for simulations, our Model M2 does slightly worse for a particular k ($k = 2$).

Temporal comparisons of KL-divergence distributions between models and experiments for individual player actions This section shows the temporal KL-divergence values for the model M1 and M2 predictions across the five parameters of x . Lower values are better. Figure 3.25 in Cedeno (2019) shows $k = 2$ experiments, and Figure 3.26 in Cedeno (2019) shows $k = 4$ experiments. In Appendix B.1.10 in Cedeno (2019) Figure B.39 shows $k = 6$ experiments, and Figure B.40 shows $k = 8$ experiments. Each plot contains data over a 1-min time window, as in previous analyses. It is clear from visual inspection that model M2 predictions are in better agreement with the experiment data than are M1 predictions for $k > 2$. As noted above, however, Model M1 does slightly better for $k = 2$. That is, the comparisons between models M1 and M2, for temporal variations in 1-min time intervals over the 5-min group anagram

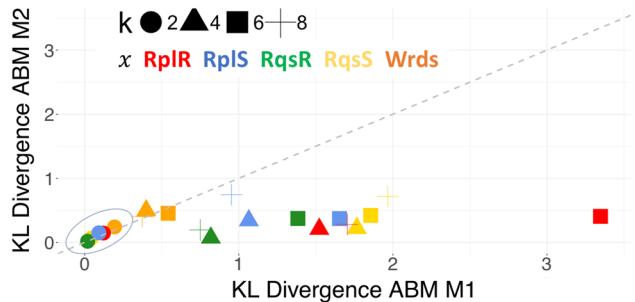


Fig. 17 A scatter plot of KL divergence for M1 (x-axis) and M2 (y-axis) for four k values and five x variables. For $k > 2$, M2 performs better than M1, as M2 incorporates experimental data with $2 \leq k \leq 8$. Interestingly, M1 and M2 perform equally well (highlighted) for $k = 2$ as M1 is learned from $k = 2$ experimental data (M1 is slightly better). These are data over the total 300 s anagram game (color figure online)

game, are similar to those comparisons when combining all data into one analysis over the entire 5-min game.

Summary of M1 and M2 model comparisons In addition to Figure 3.27 in Cedeno (2019) discussed immediately above, Fig. 17 compares Model M1 and Model M2 for each of the five actions in x , accumulated through the 5-min group anagram game. Model M2 does not perform quite as well as Model M1 for the $k = 2$ data, but does better than M1 for $k = 4, 6$, and 8. Thus we sacrifice some quality for $k = 2$ and get in return capabilities over a range of k . Hence, Model M2 is of greater value, since it covers a broader range of inputs for simulations.

In Appendix B.1.11 in Cedeno (2019), Figure B.41 shows the boxplots grouped by type of $k = 2, 4, 6, 8$, where each box contains five values of KLD corresponding to the five x variables at the end of each minute. The plot show that our models show highest median values on the first 2 min of the game.

6 Model evaluation

This section contains evaluations of Model M2 from Sect. 5. Our goal is to understand the conditions for which our estimated model transition probabilities π_{ij} are sufficiently accurate.

To evaluate the goodness of fitting for the proposed hierarchical model, we compare the estimated (model) transition probability matrix $\hat{\Pi} = (\hat{\pi}_{ij})$ with the empirical (data) transition probability matrix $\tilde{\Pi} = (\tilde{\pi}_{ij})$ under different settings of covariates (the z vector of Table 9). Here, the empirical transition probability matrix $\tilde{\Pi}$ is obtained under the settings by grouping the value of each covariate into three levels, as described in Table 10, to obtain comparable numbers of samples across bins.

Table 10 Three bins and ranges of values for the z variables from Sect. 5.5. These bins are created for each of the four values of $k \in \{2, 4, 6, 8\}$

Level	Buffer (z_B)	Letter (z_L)	Word (z_W)	Consec. (z_C)
1	0	0–3	0–1	0–3
2	1	4–6	2–8	4–11
3	≥ 2	≥ 7	≥ 9	≥ 12

For each setting, there is a level combination of the four covariates. We compute a counting matrix $\mathcal{N} = (n_{ij})$, where n_{ij} is the number of data instances observed for the transition from action i to next action j across all players in group anagram games. (Here, actions i and j refer to actions $a_i, a_j \in A$ in Table 5.) We then calculate the empirical probability $\hat{\pi}_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}$. There are 324 settings in total from the grouping of variables in Table 10 (three settings of each of four variables, and four k values), and 279 of them have valid empirical transition probability matrices. For the estimated transition probability matrix $\hat{\Pi} = (\hat{\pi}_{ij})$, the value of $\hat{\pi}_{ij}$ is estimated by the proposed model under each setting of covariates, where the averaged value at each level of the covariate is used in the estimated model.

The squared Root of Mean Squared Errors (RMSE) is used to quantify the difference between $\hat{\Pi} = (\hat{\pi}_{ij})$ and $\tilde{\Pi} = (\tilde{\pi}_{ij})$. RMSE is calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{4|\mathcal{I}|} \sum_{i \in \mathcal{I}} \sum_{j=1}^4 (\hat{\pi}_{ij} - \tilde{\pi}_{ij})^2} \quad (8)$$

where $\mathcal{I} = \{i : \min_j n_{ij} > 0\}$ is the index set of the rows where the empirical probability can be obtained.

Figure 18 shows the scatter plot between the RMSE and n_{\min} for the 279 settings for which there are sufficient data, where the plot is in log10-log10 scale. From the figure, the proposed method generally provides an accurate estimation of probability transition matrix in most of settings. Clearly, the value of RMSE decreases as the Min.Count n_{\min} increases. When $n_{\min} \geq 100$, the value of RMSE is smaller than 0.069, showing a very good model fitting. When n_{\min} is small, the RMSE is relatively high. One explanation is that the empirical probabilities cannot be calculated accurately when n_{\min} is small.

7 Abductive loop analyses and results

7.1 Overview

In this section, we present the results of iterative abductive analyses, described in Fig. 4. First, we “unroll” the abductive

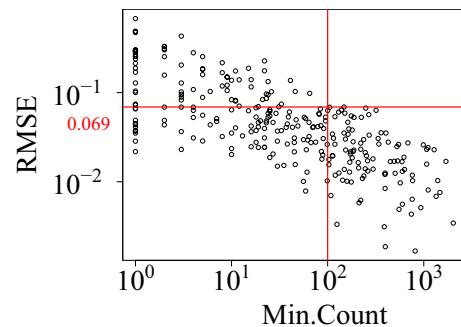


Fig. 18 Scatter plot of RMSE against Min.Count in different settings of covariates in Table 10. See Eq. (8) for RMSE and text for Min. Count. One hundred observations in a category drives RMSE down to 0.069

loop to illustrate several iterations of abduction and different analysis paths that can be taken, depending on results generated up to that point; see Sect. 7.2. Then, we present two ALs in the next two subsections. At the end of loop-2, we describe how ABM M2 can be used in further loops. We then summarize the findings of the abductive iterations, and discuss their generality, and candidate research questions for future work. We note that the experiments (Sect. 4) and modeling (Sect. 5) are major components of the abductive looping process, and were separated out to make this section more streamlined.

7.2 Abductive iterations with hypotheses

Figure 19 provides a tree structure representation of several *candidate* abductive loops. Specifically, an iteration or loop of the abductive looping process, shown in Fig. 4, is represented by a node in the graph of Fig. 19. Each loop (Fig. 4) specifies and evaluates at least one hypothesis. Figure 19 emphasizes a hypothesis H_{ij} at each node. These hypotheses are stated in Table 11. The last step in a loop is to determine what is to be done next, and the options are represented by the edges out of a node, extending toward its child nodes in Fig. 19 (a node may have any number of children). Hence, the tree graph in Fig. 19 is motivated by the usefulness of representing several abductive loops and their dependencies in a compact fashion, as opposed to writing down Fig. 4 for each loop.

The hypotheses in Fig. 19 and Table 11 are candidates because they depend on the data generated as successive abductive iterations are completed (Haig 2005; Timmermans and Tavory 2012). Hence, it may be modified with iterations. The tree here is not unique. Different analysts may compose different hypotheses and different trees, and a tree will generally need to be modified as analyses unfold. Nonetheless, it is a useful exercise to construct such trees as part of reasoning about a problem.

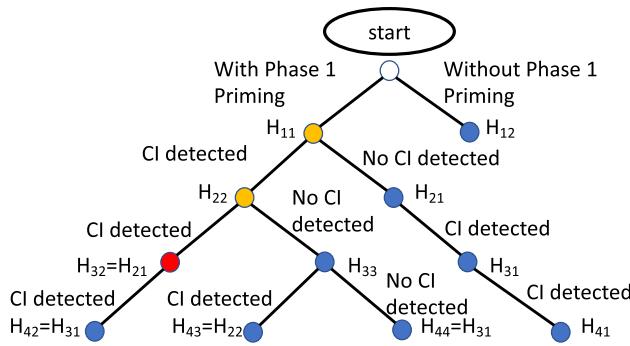


Fig. 19 An abductive tree representing candidate abductive loops with dependencies. Each node in the figure is one loop (see Fig. 4). Since each loop has at least one hypothesis, we label the nodes here with hypotheses that are provided in Table 11. Edges to child nodes are labeled with outcomes from hypothesis evaluations within a loop, and indicate, based on this evaluation, which abductive loop to perform next. A node can have any number of outgoing edges to child nodes. The orange colored nodes correspond to abductive iterations presented herein. The red node is a candidate next loop. This tree is not unique; different analysts can devise different trees, and they can be modified as analyses proceed. The purpose of this construction is to provide a succinct representation of multiple candidate abductive loops, and their dependencies (color figure online)

We now overview the two abductive iterations detailed in subsequent sections. The root node of the tree in Fig. 19 is the starting point. We conduct experiments with the Phase 1 (group anagram priming game), so we take the left branch (edge) from the root node, labeled “With Phase 1 Priming”. We perform an abductive loop, where we form and evaluate hypothesis H_{11} . This loop is detailed in Sect. 7.3. Since it is clear that CI was formed, we follow the “CI detected” path out of H_{11} to arrive at the node hypothesis H_{22} . This AL-2

Table 11 Candidate hypotheses to evaluate in abductive iterations. Not all of the hypotheses are evaluated herein. The goal of these hypotheses, coupled with Fig. 19, is to illustrate that there are many

Hypothesis Number	Description
H_{11}	In the team-based anagram game, the collective identity formed is driven more by the number of words a player forms than the number interactions of a player (requests and replies)
H_{12}	Playing the group anagram game will produce greater individual DIFI scores than not playing the group anagram game
$H_{21} = H_{32}$	As the number and quality of letters assigned to a person decreases (i.e., as the letters assigned to a player occur less frequently in common words), collective identity of the player will increase
$H_{22} = H_{43}$	(a) As the number of neighbors of a player increases in the anagram game, the number of interactions of a player increases, but only up through degree $k = 4$. For further increases in k , there will be no increase in numbers of interactions (b) The trend in numbers of interactions with k in hypothesis (a) will be reflected in the CI produced in group anagram games: CI will increase for $2 \leq k \leq 4$, but will saturate thereafter because the numbers of interactions saturate
$H_{31} = H_{42} = H_{44}$	Playing the game with players face to face will produce greater individual DIFI scores (by enabling the players to communicate and pick up on visual and verbal cues)
H_{33}	Lesser payouts in the group anagram game means that players do not have enough incentive to engage their neighbors
H_{41}	Having the group anagram game score of another team displayed during the game will increase CI because it will create a stronger in-group/out-group paradigm

is described in Sect. 7.4. Since we do obtain a CI signal from these experiments, we follow edge “CI detected” to hypothesis H_{32} . Details are provided below, and we note that modeling results guide decisions about what experiments to perform in the next abductive iteration, illustrating the value of modeling. This is one reason why our abductive loops promote modeling to a central role. We also note that a hypothesis can appear at multiple nodes within the abductive tree, e.g., the group anagram game conditions may be altered after particular loops. Finally, a node need not have two children; e.g., fewer or more children are possible.

7.3 Abductive Loop 1 (AL-1)

We describe the steps of the abductive loop in Fig. 4, in turn.

Experiments A set of 18 experiments with a total of 87 players was completed where $k = 2$. See Sect. 4.

Data Analysis For this loop, data analysis and modeling are intertwined and so both are described under the *models* step below. It is critical to note that the data analyses below came *before* the specification of hypotheses, because a critical element of abduction is that patterns in the data drive the hypotheses—not the other way around.

Hypothesis/Theory Hypothesis H_{11} : *In the team-based anagram game, the CI formed is driven more by the number of words a player forms than the number interactions of a player (requests and replies).* Social Exchange Theory (Homans 1961) focuses on the individual and suggests that the number of words resonates more in forming CI because they are directly related to reward in the game. Theory of Social Interactions (Becker 1974) indicates that interactions are important for forming an interdependent

possible hypotheses that can be formulated, and it is up to analysts to decide which ones to pursue. An analyst will be guided by the results of completed iterative abductive analyses

organization. Reciprocity Theory suggests that v_i will respond to v_j 's requests because v_i wants v_j to respond to hers, so that interactions are important.

Models There are two types of models constructed. One type is the models of the group anagram game: Baseline Model M0 and M1 from Sect. 5. The other is a regression model to predict DIFI2 score as a function of outputs from the group anagram games (e.g., number of requests sent n_{RqsS} , number of replies received n_{RplR}).

The group anagram game models M0 and M1 of Sects. 5.4 and 5.5 were constructed from the time histories of actions of players for experiments with $k = 2$. The results relevant to this abductive iteration are provided in Figs. 11 and 12. Model M1 is much better at capturing the dynamics in the experiments than is Baseline Model M0.

From data of the actions $a_i \in A$ from the anagram games, and the measured DIFI2 scores after the group anagram games, a linear regression was performed to correlate DIFI2 score with the numbers of actions of each kind for game players. The DIFI2 score is given as

$$\widehat{\text{DIFI2}} = c_1 + c_{RplR} n_{RplR} + c_{RplS} n_{RplS} + c_{RqsR} n_{RqsR} + c_{RqsS} n_{RqsS} + c_{Wrds} n_{Wrds} \quad (9)$$

where Table 12 provides the equation coefficients and the definitions of variables.

Table 13 provides the regression results that identify the player actions that correlate with DIFI score. At the 0.05 level, replies received, replies sent, and requests sent are all significant.

Best Explanation Results of a linear regression in Table 13 indicate that hypothesis H1 is falsified because the number Wrds, i.e., the number of words formed, is not significant, while numbers of RplR, RplS and RqsS (i.e., interactions) are significant. Thus, Social Exchange Theory can be eliminated as a theory of CI formation in this experiment. It is somewhat surprising that Wrds is not significant because it is the variable that is most closely associated with the reward (earnings). A conjecture was made that the greater the monetary reward given to players, the greater their affinity would be for the team; these data do not support this conjecture. In the social sciences, and in many domains, eliminating candidate theories is a valuable result (that is, an analysis does not always have to identify the best theory). Thus, at this point, the best explanation is Reciprocity Theory and Theory of Social Interactions because the analyses results in Table 13 show that interactions correlate most strongly with DIFI score.

A key result indicated by this first iteration is that the group anagram game can produce CI (as measured by the proxy DIFI score).

Table 12 Constants in the regression of Eq. (9) to predict DIFI2 score of a player from the player actions in the team anagram game

Coefficient	Value
Intercept c_1	102.7
c_{RplR} on number of replies received n_{RplR}	14.95
c_{RplS} on number of replies sent n_{RplS}	-12.99
c_{RqsR} on number of requests received n_{RqsR}	6.406
c_{RqsS} on number of requests sent n_{RqsS}	-16.43
c_{Wrds} on number of words formed n_{Wrds}	-0.2134

What is Next? Figure 13 indicates that Model M1 predicts player behavior that is invariant with respect to the degrees k of players [and hence the number of letters that neighbors possess] (plots of other variables of x are similar). We want to determine whether there is an effect of k , and hence the next experiments are specified to study increasing k (i.e., $k > 2$). Thus, the ABM M1 (driven by the data) is guiding what to do next. While Social Exchange Theory was eliminated in this loop, Reciprocity Theory and Theory of Social Interactions are carried forward into the next loop(s), where they may be supported or refuted.

7.4 Abductive Loop 2 (AL-2)

We execute the steps of the abductive loop in Fig. 4, as described next.

Experiments A set of 16 experiments with a total of 137 players was completed where $k = 4, 6$, and 8 in turn. See Sect. 4.

Data Analysis We continued the same types of analyses described in AL-1, but with the added dimension of k . Figure 20 shows the frequency distributions for replies received, for the four values of k . Note the large change in distributions in going from $k = 2$ to $k = 4$, but relatively minor changes for further increases in k . This indicates two regimes of behavior: (1) $2 \leq k \leq 4$, and (2) $k > 4$. In the first regime, numbers of interactions increases with k , and in the second regime, the numbers of interactions does not appreciably increase with further increases in k . Additional data analyses are presented in step *best explanation* below.

Hypothesis/Theory Two hypotheses are formed based from the preceding data. Hypothesis H2 ($= H_{22}$): (a) As the number of neighbors of a player increases in the anagram game, the number of interactions of a player increases, but only up through $k = 4$. For further increases in k , there will be no increase in numbers of interactions. (b) The trend in numbers of interactions with k in hypothesis (a) will be reflected in the CI produced in group anagram games: CI will increase for $2 \leq k \leq 4$, but will saturate thereafter because the numbers of interactions saturate.

Table 13 Results of linear regression of variables in x (see Table 7) against dependent variable DIFI2 score, indicating that interactions are more significant than number of words formed in producing CI

Var.	Interc.	RplR	RplS	RqsR	RqsS	Wrds
est.	103.	15.0	-13.0	6.41	-16.4	-0.213
p val.	0.001	0.019	0.011	0.332	0.011	0.735

The values in bold indicate that the number of replies received (RplR), number of replies sent (RplS) and number of requests sent (RqsS) are significant

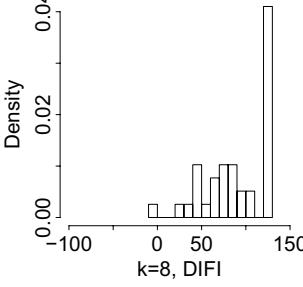
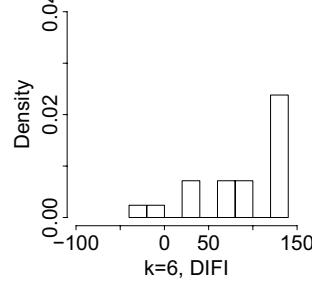
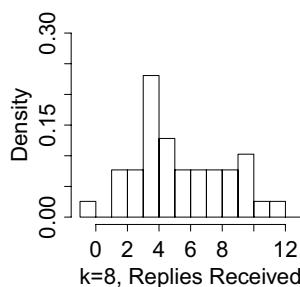
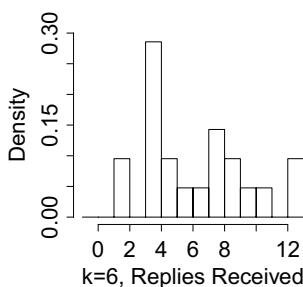
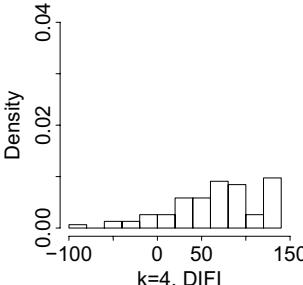
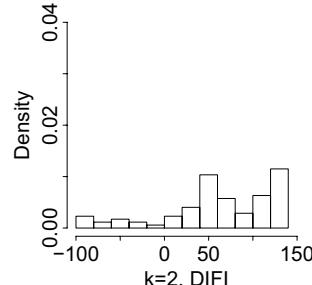
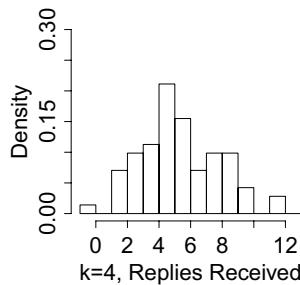
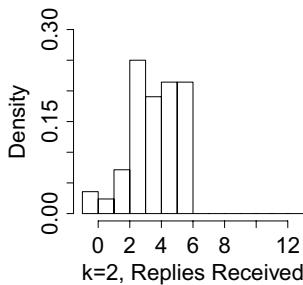


Fig. 20 Statistical analysis correlation results of the anagram game parameters. The probability density of replies received changes markedly from $k = 2$ to $k = 4$, but relatively little for further increasing k up to eight

Model Model M2 of Sect. 5.6 was constructed from the time sequences of actions of players, from the combined data from *both* iterations. Model results relevant to this iteration are provided in Figs. 14 through 17. ABM M2 captures trends in degree k more effectively than ABM M1, for all parameters of x .

Best Explanation Figures 20 and 21 provide results that address the hypotheses.

For $H_{22}(a)$, we return to Fig. 20 and the observations under *data analysis*. We note a saturation in the distributions of replies received (distributions for other actions are similar). As the number of neighbors of a player increases from two to four, the numbers of interactions increases; but for $k > 4$, the number of interactions does not change appreciably with further increases in k . This can be explained by Utility Theory and by Cognitive Load Theory. Utility theory argues that 15 letters is enough for a player to form words (three letters from each of four neighbors and three own letters of a player), so a player derives no marginal utility

Fig. 21 Statistical analysis correlation results of the anagram game parameters and DIFI2 score. The probability density of DIFI score moves to larger DIFI score with increasing k , between two and eight

from more neighbors and more letters. Cognitive load theory states that a player cannot reason about forming words with more than 15 letters, so no attempt is made to acquire more letters. Hence, hypothesis $H_{22}(a)$ is not falsified. It is possible in future work to conduct more experiments, possibly by modifying the priming procedures, to disambiguate these two explanations.

We note the agreement between our hypothesis and the results of the experiments. This is a direct result of the abductive approach: the data guide the hypotheses. We stated in $H_{22}(a)$ that the transition between the two regimes occurred at $k = 4$ because the data in Fig. 20 indicated this. It is the goal of hypothesis and theories to *explain* the behavior.

For hypothesis $H_{22}(b)$, we use Fig. 21, showing how the probability density of DIFI score moves to higher scores as k increases from two through eight. Thus, Figs. 20 and 21 together support $H_{22}(b)$, but only for $k \leq 4$. The explanation in the hypothesis—that greater DIFI scores is caused by greater numbers of interactions—is not supported by the

combined view of Figs. 20 and 21 for $k > 4$. The latter figure shows an ever increasing probability density of greater DIFI score with increasing k , but the former figure shows an essential saturation of numbers of actions with increasing k beyond $k = 4$. Hence, the numbers of interactions may be contributing to increasing DIFI scores for $2 \leq k \leq 4$, but appears not to be not the reason for increasing DIFI scores for $k > 4$. Consequently, the Theory of Social Interactions can explain the results for $2 \leq k \leq 4$, but this theory is falsified for $k > 4$.

What is Next? At this point we halt the iterative abduction process for this paper. In a next iteration, we could try to isolate the effects of number of interactions versus the number of neighbors in different experiments on DIFI score, to more fully explain the results in AL-2. Several other directions are possible, guided by the hypotheses in Table 11: evaluating the quality of letter assignments, varying the number of letters per neighbor, playing face-to-face games, adding competition for the team. We could also perform a deductive (confirmatory) analysis by making specific quantitative predictions for experiments using ABM M2 as part of AL-2, and running corresponding experiments in AL-3.

7.5 Summary of experimental contributions to the understanding of CI and possible extensions

Abductive loops one and two are presented in Sects. 7.3 and 7.4. Here, we summarize our findings and identify how our findings might be tested or extended by presenting a series of research questions for others to consider. First, we consider the generality of our specified hypotheses.

Hypotheses H_{11} and H_{22} are stated in Table 11. Since the number of words that players form is directly proportional to the monetary reward of players in the game, H_{11} may be restated in more general terms as *Hypothesis H_{11}^* : The collective identity formed is driven more by player reward in a game than the number interactions of a player with its neighbors*. Note that we have not removed the notion of interactions—which we effect by inducing a network on the players, but many games use networks to control interactions, so this is not a problem in our view. The hypotheses of H_{22} can be restated in more general terms as *Hypothesis H_{22}^* : (a) As the number k of neighbors of a player increases, the number of interactions of a player increases, but only up to a (saturation) point denoted by a critical value of degree k^* . For further increases in $k > k^*$, there will be no increase in numbers of interactions. (b) The trend in numbers of interactions with k in hypothesis (a) will be reflected in the CI produced: CI will increase up to the saturation point, but will remain relatively constant thereafter because the numbers of interactions saturate*.

There is no mention of the group anagram game or our DIFI score task in these restated hypotheses. Specifically, our hypotheses and findings need not be specified in terms of the particulars of our game, such as particular letters requested of neighbors or specific words formed. Rather, our hypotheses and findings may be stated in terms of game rewards (i.e., earnings), numbers of connections (i.e., network degree) of players, and numbers of player interactions. These are basic features. One reason to state the hypotheses in terms of the games, as we did in Sects. 7.3 and 7.4, is because it makes clearer how the analyses tie in with the hypotheses. We return to the implications of these restatements after summarizing our findings.

In Sect. 7.3, we find that CI is *not* produced owing to the parameter most closely aligned with the payout (i.e., earnings) to the group: the number of words formed. Since a player's monetary benefit is most closely and directly tied to the number of words formed, one might conjecture that this variable would most closely correlate with the DIFI score (i.e., our measure of CI). That is, we speculated that success in the group anagram game, in terms of increases in monetary reward, would result in increases in players' affinities for the team. Our analyses indicate that this is not the case. Hence, financial rewards did not translate into greater CI. We found instead that numbers of interactions are more significantly correlated with greater DIFI scores, our proxy for CI, for the $k = 2$ data.

In Sect. 7.4, where we study the effects of $2 \leq k \leq 8$, we find that the numbers of interactions increase most rapidly as number of neighbors increases from two to four. As player degree increases from four to six, and then to eight, the numbers of interactions remains essentially constant. However, the DIFI score increases as player degree increases from two to eight. Thus, DIFI score increases with increasing numbers of interactions, as described in the abductive loops above, but increases further as numbers of interactions remain relatively constant, but numbers of neighbors increases.

As stated earlier in this subsection, our hypotheses and their evaluations can be stated in general terms. Consequently, it should be possible to test our findings using other group interaction approaches. These could be online or offline settings, computer-based or not computer-based. Player (participant) interactions can be of various forms: verbal communication, written communication, transactions or partial transactions in the form of actions, or selection (i.e., choosing a subset of neighbors from a player's complete set of neighbors). We use a network setting; this is not required, although it does provide the ability to control who might interact with whom. Also, we use DIFI score as a proxy for CI, based on the literature, but other means of inferring the existence of CI are also viable. These decisions are of course the prerogative of the researcher.

Nonetheless, we provide some specific questions that others may consider in their work. This list is not exhaustive, but illustrates how our findings may be tested and/or extended by others:

1. Do other graph structural properties, such as clustering coefficient, influence CI formation?
2. Does the proxy or measurement of CI increase with the number of neighbors (interactors) that a player has?
3. Does the proxy or measurement of CI increase with the number of interactions that a player has?
4. Is a one-shot, binary decision game as effective as a multi-action (or multi-decision) game where actions may be repeated over time, in producing CI?
5. Can the CI priming process (for us, the group anagram game) be more effective by introducing competition with another team? (There are many ways to do this, including using a fictional team, using data from a previous experiments where other teams did well, or poorly.)

As we stated in Sect. 1.4 when describing our technical challenges, we believe that our work is not the final word on CI, but rather is closer to a starting point, providing detailed methods and results for others to scrutinize and extend.

7.6 Abductive loops: role of analyst and bigger picture

Two ALs have been demonstrated. Many additional loops are possible, as illustrated in Fig. 19, which depicts several hypotheses, including the two addressed above (in orange). These additional loops would require more experiments. Figure 19 and Table 11 make clear the important role of an analyst in this process, as she guides the direction of the looping. So, while a plan such as that in Fig. 19 may be useful, the actual tree structure will evolve with analyst decisions as the looping progresses and as data are generated, because hypotheses are based on newly-generated data in abduction.

8 Limitations and additional work

More experiments, particularly at greater k would be useful. Also, we would like to alter the number of letters and to control the “quality” of letters that are assigned to players (e.g., e is a more desirable letter than q) in additional experiments, so that we can run experiments that will more stringently test the models. We would like to study more network structures (i.e., connectivity among players in a game), such as a clique structure. Beyond collecting more data, we could test more conditions, and also specify and evaluate more hypotheses about CI. We attempted to correlate player behavior with

survey information in the online experimental platform. For example, we tried to correlate DIFI score with player age, gender, nationality, ethnic group, and education level. We did not get a strong signal in any of these correlation studies. This would be a huge step forward if such correlations exist and can be found because it would relate macro-player features with player behavior. With respect to modeling, we can improve the models for the player actions (e.g., the process of forming words); this work is in progress. We can improve the modeling in translating results in the group anagram game to the DIFI scores, to better understand the connection between priming and CI formation.

9 Summary

We formalize an abductive loop, implement it computationally, and exercise it in an experimental setting (the group anagram game) designed to induce CI, as operationalized by Swann’s DIFI score. However, our abductive looping process is not tied to CI. As part of the abductive iterations, we provide novel experimental insights into CI and build and evaluate three ABMs. This work establishes the potential of iterative abductive looping for the (computational) social sciences.

Acknowledgements We thank the anonymous reviewers for their helpful feedback. We thank our colleagues at NSSAC and computer system administrators: Dominik Borkowski, Jason Decker, Miles Gentry, Jeremy Johnson, William Marmagas, Douglas McMaster, and Kevin Shinpaugh. This work has been partially supported by NSF CRISP 2.0 (Grant No. 1832587), DARPA Cooperative Agreement D17AC00003 (NGS2), DTRA CNIMS (Contract HDTRA1-11-D-0016-0001), DTRA Comprehensive National Incident Management System Contract HDTRA1-17-D-0023, NIH 1R01GM112938-01. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

A Supplemental related work

Related work topics that augment those in Sect. 3 are provided here. See Table 2 for a listing of all related work topics.

A.1 Individual anagram games: modeling

In Tresselt (1968), problem solving and verbal cues are analyzed with an anagram game. Tresselt (1968) modifies the H. Kendler and S. Kendler (1962) mediational model of problem-solving behavior (introducing word length and letter position), to understand anagram problem solving. This is a theoretical model of individual anagram games.

A.2 Individual anagram games: experiments and modeling

These works combine experiments and modeling. In Feather (1969), it was found that subjects who were initially confident of passing an anagram game test tended to attribute success to ability and failure to bad luck. However, subjects who were initially not confident tended to attribute success to good luck and failure to lack of ability. Results are discussed in terms of Heiderian theory and a valence-difficulty model. In Feather and Simon (1971a, b), two individuals played anagram games simultaneously but independently to test whether a person attributed her success (if she performed better) to skill versus good fortune, and failure to inferior skill or bad luck. Attributions were found to be dependent on expectations of players. Results are discussed in terms of models involving Heider's principle of balance and his analysis of the causes of action, in terms of positivity biases in social perception, and as indicating effects of the social context of performance upon attribution and valence.

A.3 Modeling of CI

Lustick (2000), Rousseau and van der Veen (2005) use ABMs to study identity diffusion. An agent adopts (changes) her type of identity to that of a neighbor with a stronger (higher valued) type of identity. Hence, these are contagion processes and are implemented much like voter models (de Oliveira 1992; Pereira and Moreira 2005). Other works modeling collective identity (van Zomeren et al. 2008; Chen and Li 2009; Benjamin et al. 2016; Ackland and O'Neil 2011) are presented in Sect. 3.3.

A.4 Agent-based models of anagram games and formation of CI

The Charness et al. (2014) work in Sect. 3.6 has no modeling for the group anagram game. This motivated the *online* experiments and ABMs in Ren et al. (2018). This article is an expansion of Ren et al. (2018). In this work, we model the priming process of producing CI, which is the group anagram game. There are no ABMs (or models of any kind) of group anagram games, to our knowledge, other than ours.

A.5 Studies of phenomena related to CI

Many phenomena, such as in-group and out-group effects are related to CI. In Brewer and Silver (1978), Perdue et al. (1990), laboratory experiments with no interactions between subjects are performed. In Brewer and Silver (1978), it was found that bias in favor of the in-group on a reward allocation task was unaffected by the arbitrariness of classification into groups. An effort was made to assure

that subjects in the arbitrary condition would not perceive the out-group as dissimilar. They found that similarity-dissimilarity of the out-group did not affect allocation bias as long as the in-group was perceived as similar to the subject. Subjects were divided clearly into groups labeled "dark" and "light". Subjects then were asked to indicate their ratings first of "the other members of my group" and then of "the members of the other group" on a series of six-point bipolar scales (friendly-unfriendly; trustworthy-untrustworthy; cooperative-competitive; intelligent-stupid; weak-strong; generous-stingy; likeable-unlikeable). In Perdue et al. (1990), classical conditioning in-group and out-group descriptors (e.g., "us" and "them") are used to establish evaluative responses to novel, unfamiliar targets. Nonsense syllables unobtrusively paired with in-group designating pronouns (e.g., "we") were rated as more pleasant than syllables paired with out-group designators (e.g., "they").

Paris et al. (1972) study how the anticipated interaction between groups determines the representations that groups have of each other. When students are categorized into groups, discrimination occurs such that the in-group is more favorably represented than the out-group before interaction takes place and also when no interaction is anticipated. Such discrimination is stronger when competitive interaction is anticipated in an important situation. In this condition, intergroup differences are also more easily projected on physical traits. Categorization is shown to be not only an independent variable but also a dependent variable in intergroup relations.

In Kahn and Ryen (1972), Own Group Bias (OGB) was measured by differences in pre and postgame scores on the evaluative scales of the Semantic Differential (SD).

In Shank et al. (2015), an experiment on Amazon Mechanical Turk was used to develop an agent-based simulation to understand how people's motivations and behaviors within public goods dilemmas interact with the properties of the dilemma to lead to collective outcomes. They predict how the public good's benefit and size, combined with controlling individual versus group properties, produce different levels of cooperation in public goods dilemmas.

In Sethi and Somanathan (2006), a simple model of collective action is presented as a framework for empirical research into the issue of when collective action in the commons will be successful.

In van Zomeren et al. (2008), an integrative social identity model of collective action (SIMCA) is developed that incorporates three socio-psychological perspectives on collective action. Instructions for coders were to answer different questions like "Does the measure of identification (used in this study) refer to a disadvantaged group or a social movement?", "Is this group incidentally disadvantaged or structurally disadvantaged?". Coders also rated the

extent to which collective disadvantage was structural on a 5-point Likert-type scale ranging from 1 (not at all) to 5 (very much).

In Salganik and Watts (2009), new insights into the role of individual behavior on collective outcomes are obtained using a multiple-worlds experimental design in a web-based experiment in which 2930 participants listened to, rated, and download 48 songs by up-and-coming bands.

In Suri and Watts (2011), laboratory experiments with interactions between subjects are performed. Web-based experiments are conducted where 24 individuals played a local public goods game arranged on one of five network topologies that varied between disconnected cliques and a random regular graphs. It was found that although players did generally behave like conditional cooperators, they were as likely to decrease their contributions in response to low contributing neighbors as they were to increase their contributions in response to high contributing neighbors. They also found that positive effects of cooperation were contagious only to direct neighbors in the network.

In Capraro (2013), online experiments using Amazon Mechanical Turk were used to develop a predictive model of human cooperation able to organize a number of different experimental findings that are not explained by the standard model.

In Rousseau and van der Veen (2005), an agent-based computer simulation of identity change explores how changes in the attributes of the individual and/or elements of the environment influence the dependent variable: the degree of shared identity in a population.

There is a host of other studies that investigate phenomena such as cooperation and a person's affinity for a group that are closely related to CI. In Worchel et al. (1977), Charness et al. (2007) laboratory experiments with interactions between subjects are performed. They study concepts such as group attraction and salience, respectively, which are related to CI. In Worchel et al. (1977), study groups worked cooperatively on two tasks and results were interpreted as showing that both previous interaction and success of combined effort are important variables in determining when intergroup cooperation will increase intergroup attraction. In Charness et al. (2007), groups perform two stage games as priming tasks, the Battle of the Sexes and Prisoner's Dilemma. Results show that the salience of the group affects behavior of members, as well as the behavior of people in another group, and that participants anticipate these effects.

A.6 Data-driven: combining experiments and data-driven modeling

This section reports on works that combine experiments with data-driven modeling. These works cover explore-exploit networked experiments with limited modeling (Mason and

Watts 2012); individual models of single-choice (i.e., one-shot) evacuation decisions (Nguyen et al. 2017); ABM of emotion and information contagions spreading on a network and comparisons with a single event (Li et al. 2014); and ABM of solar panel adoption and comparisons with data in San Diego county (Zhang et al. 2016). See Zhang and Vorobeychik (2019) for a review of innovation diffusion models. None of these works use ABMs to model networked experiments where individuals take a series of actions (that may be repeated) over time, to study CI, as we do.

In Luhmann and Rajaram (2015), small-scale laboratory experiments and an ABM were used to analyze the dynamics of collaborative inhibition. In Gates et al. (2017), the model in Luhmann and Rajaram (2015) was tested against human data collected in a large-scale experiment to find that participants demonstrate non-monotonicities not evident in the predictions. These unexpected results motivate more recent work in elucidating the algorithms underlying collaborative memory. In Paxton et al. (2018), using real-time online social experiments data, a statistical model is used to study interpersonal coordination in a “minimally interactive context” to explore how people become coupled in their perceptual and memory systems while performing a task together.

In contrast to the above works, where *controlled experiments* are used to produce data that are then used for modeling, there are many models based on *observational* data. We survey some of these works here.

In Korolov et al. (2016), the possibility of predicting a social protest (planned, or unplanned) based on social media messaging is studied. In Nguyen et al. (2016), to help increase the performance of retweet prediction, a flexible model under the framework of Random Forest classifier captures a number of behavior signals affecting user's retweet decision. In Hu et al. (2014), a semantic model that can naturally represent various academic social networks, especially various complex semantic relationships among social actors, is presented. In Qin et al. (2017), the proposed method integrates topology and content of networks, and introduces a novel adaptive parameter for controlling the contribution of content with respect to the identified mismatch degree between the topological and content information. In Attema et al. (2015), data-driven multi-agent models predict Twitter trends. In van Maanen and van der Vecht (2013), a method that implements, validates, and improves an individual behavior model is proposed. The multi-agent model contains the social network structure, individual behavior parameters, and the scenario that are obtained from empirical data. In Lee and Oh (2013), emergence and propagation of reputations in social networks is modeled with a distributed algorithm. In Chierichetti et al. (2014), using several Twitter data sets, focusing in particular on the tweets sent during the soccer World Cup of 2010, a model of how users switch between producing information or sentiments

Table 14 Summary of the analyses in the Experimental Data Sect. B, and the questions we answer. Section B.1 presents histograms for the timestamps for letter requests. Section B.2 presents histograms for the timestamps for letter replies. Section B.3 presents histograms for the

timestamps of the time duration between replies received and requests sent. Section B.4 presents histograms for the timestamps for words formed

Section	Histograms	Questions for analysis
B.1	Timestamps for letter request	When do players request letters during the game? How does the number of neighbors affect the behavior of a player to request a letter in the game?
B.2	Timestamps for letter reply	When do players reply to letter requests during the game? How does the number of neighbors affect the behavior of a player to reply a letter in the game?
B.3	Timestamps for time duration(reply received–request sent)	How long does it take players to reply to a letter request? How does the number of neighbors affects the time duration between the timestamps of the letter reply action and the letter request action?
B.4	Timestamps for word formed	When do players submit words during the game? How does the number of neighbors and the number of available letters affects the number of words formed by a player?

and sharing others news or sentiments is developed. In Korolov et al. (2015), a theoretical analysis is developed for how social-chatter quantitatively relates to action via a superlinear scaling law.

Other works include using data from geotagged social media messages and data from mobile health applications (Tran and Lee 2016; Kurashima et al. 2018). In Tran and Lee (2016), to understand citizen reactions regarding Ebola, a large-scale data-driven analysis of geotagged social media messages is performed. In Kurashima et al. (2018), data from mobile health applications is used to develop a statistical model, called TIPAS (Time-varying, Interdependent, and Periodic Action Sequences). This approach is based on personalized, multivariate temporal point processes that model time-varying action propensities through a mixture of Gaussian intensities. Their model captures short-term and long-term periodic interdependencies between actions through Hawkes process-based self-excitations.

Clearly, much of the modeling of observational data is motivated by social media.

B Experimental data

This Appendix describes data from the game experiments of Sect. 4. In this section we present an analysis of the experimental data that illustrates how players interact in the anagram games. We focus on experimental data that are useful in modeling. We identify four main actions $a_i \in A, 1 \leq i \leq 4$, in the set A of actions for a player during the game: (1) request letter from neighbor, (2) reply with letter to a request from a neighbor, (3) form and submit valid word, and (4) think (i.e., a no-op).

We define the following variables for the actions in the game:

- When v_i sends a requests for a *letter* to v_j , a **request sent** occurs.
- When v_j receives the *letter* request from v_i , a **request received** occurs.
- When v_j replies with the *letter* requested from v_i , a **reply sent** occurs.
- When v_i receives the *letter* reply from v_j , a **reply received** occurs.
- When v_i uses its own letters to form a word, a **word formed** occurs.

Table 14 shows a summary of the section plots and the questions we answer with the analyses.

B.1 Timestamp for letter request

The number of letters a player can request through a game depends on the number of its neighbors. Each neighbor can share up to three letters (the initial three letters), so if a player has $k = 2$ neighbors, then six letters can be requested throughout the game. If a player has $k = 8$ neighbors, then 24 letters can be requested. We want to analyze the behavior of players with reference to the letter request action and answer the following questions. When do players request letters during the game? How does the number of neighbors affects the behavior of a player to request a letter in the game?

Figure 22 shows a histogram with 10 bins of 30-s each of timestamps for **request sent**, for 47 experiments with $k = 2, 3, 4, 5, 6, 8$. A kernel-density estimation with Gaussian kernels is used to estimate the probability density function. It indicates that more letters are being requested during the first half of the 300-second anagram game. To analyze whether the number of neighbors affects the letter request, Figure B.1 in Appendix B.1.1 from Cedeno (2019) shows

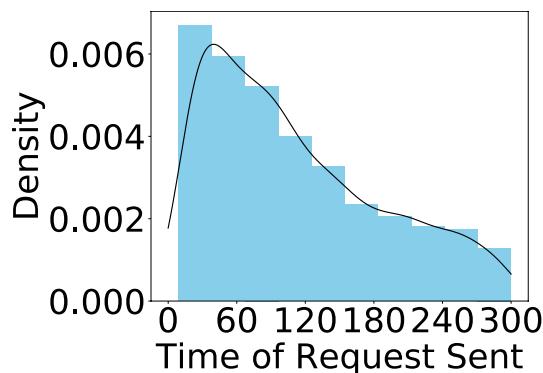


Fig. 22 Probability density distribution for time of request sent over the 300-s anagram game. Each of the bins on the x -axis corresponds to 30-s intervals. It shows experiments with $k = 2, 3, 4, 5, 6, 8$. A kernel-density estimation with Gaussian kernels is used to estimate the probability density function. Letter requests are made throughout the game, rather than solely at the outset

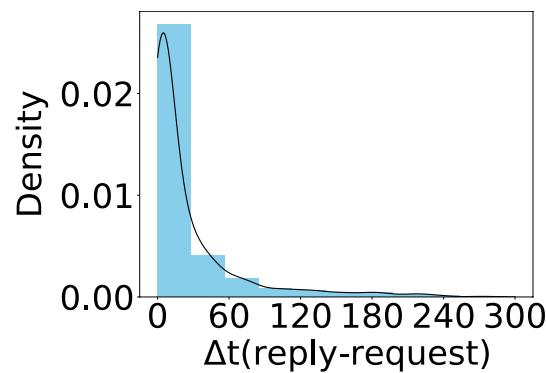


Fig. 24 Probability density distribution for time duration between requesting a letter and replying to the request, over the 300-s anagram game. Each of the bins on the x -axis corresponds to 30-s intervals. It shows experiments with $k = 2, 3, 4, 5, 6, 8$. A kernel-density estimation with Gaussian kernels is used to estimate the probability density function. Players generally respond relatively quickly to their neighbors letter requests, with replies typically made within 30 s of the request

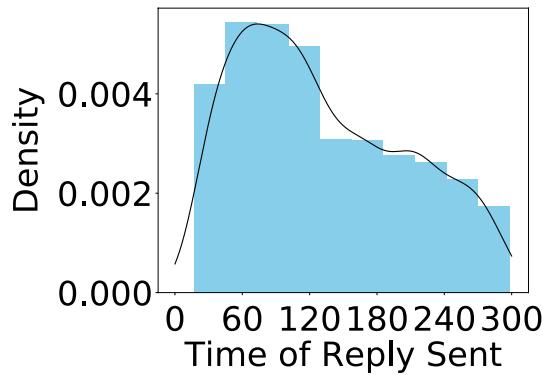


Fig. 23 Probability density distribution for time of reply sent over the 300-s anagram game. Each of the bins on the x -axis corresponds to 30-s intervals. It shows experiments with $k = 2, 3, 4, 5, 6, 8$. A kernel-density estimation with Gaussian kernels is used to estimate the probability density function. Letter replies are made throughout the game, rather than solely at the outset

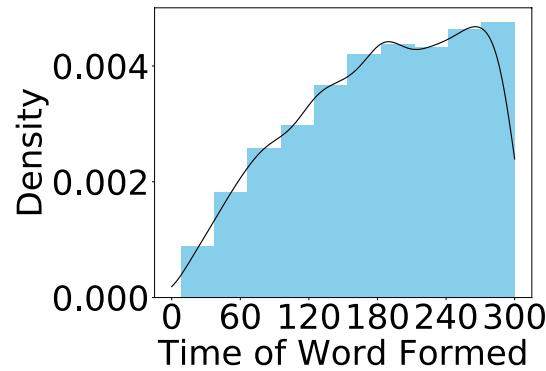


Fig. 25 Probability density distribution for time of forming words over the 300-s anagram game. Each of the bins on the x -axis corresponds to 30-s intervals. It shows experiments with $k = 2, 3, 4, 5, 6, 8$. A kernel-density estimation with Gaussian kernels is used to estimate the probability density function. Word submissions are made throughout the game, and the numbers of neighbors and available letters do not affect this type of action

histograms with 10 bins of 30-s each for **request sent** for experiments with $k = 2, 3, 4, 5, 6, 8$. The same trends exist for each value of k . However, if there are few neighbors ($k = 2$) and consequently fewer available letters (3 letters per neighbor), there are fewer letter requests and letter replies near the end of the game.

B.2 Timestamp for letter reply sent

The number of letters a player can reply with, in response to letter requests, through a game depends on the number of its neighbors. Each neighbor can share up to 3 letters, so if a player has $k = 2$ neighbors, then 6 letters can be replied (when requested) at any time through the game, since the

number of letters assigned initially is three. We want to analyze the behavior of players with reference to the letter reply action and answer the following questions. When do players reply letters during the game? How do the number of neighbors affects the behavior of a player to reply a letter in the game?

Figure 23 shows a histogram with 10 bins of 30 s each, for **reply sent**, for 47 experiments with $k = 2, 3, 4, 5, 6, 8$. A kernel-density estimation with Gaussian kernels is used to estimate the probability density function. It indicates that letter requests are being replied to throughout the game, but more so at the earlier stages of the game. To analyze whether the number of neighbors affects the letter request, Figure

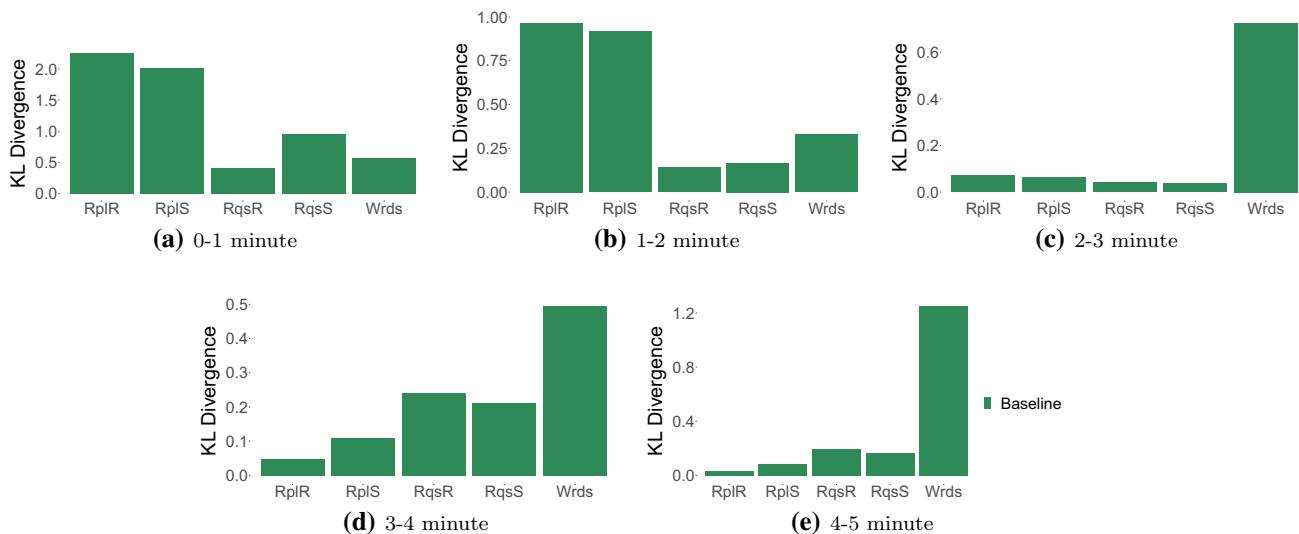


Fig. 26 Within each subfigure we show KL-divergence values for Baseline Model M0 across the five parameters of x at 1-min intervals: lower values are better. Each plot contains data over a time window: **a** 0–1 min, **b** 1–2 min, **c** 2–3 min, **d** 3–4 min and **e** 4–5 min, of the 5-min anagram game. The data are for conditions ($n = 10, k = 2$)

These plots show that request-related predictions are better than reply-related predictions over all time intervals. The reply-related predictions are better in the second half of the 5-min anagram games, but Fig. 23 shows that in experiments, there are fewer replies in the second half of the games

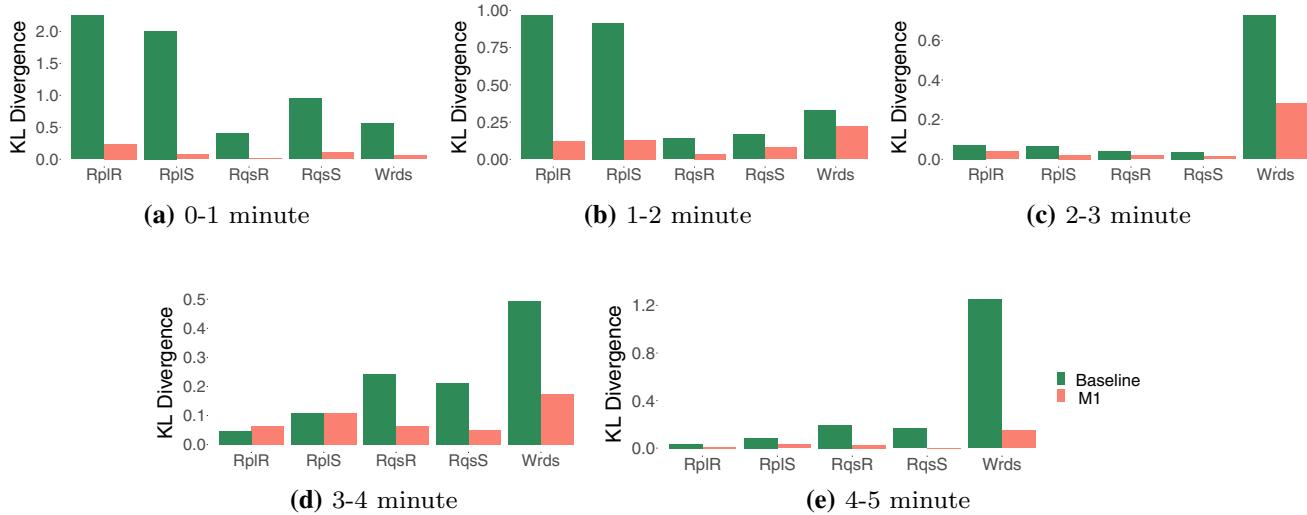


Fig. 27 Within each subfigure we show KL-divergence values for the Baseline Model M0 (in green) and Model M1 (in red) across the five parameters of x : lower values are better. The modeling conditions are experiment with $k = 2$. Each plot contains data over a time window: **a** 0–1 min, **b** 1–2 min, **c** 2–3 min, **d** 3–4 min and **e** 4–5 min, of the

5-min anagram game. While Model M0 has good predictions for minute 3 and minute 5 (with the exception of the words formed), Model M1 has better predictions for minute 3 and minute 5 for all five x variables (color figure online)

B.2 in Appendix B.1.2 in Cedeno (2019) shows histograms with 10 bins of timestamp for **reply sent** for experiments with $k = 2, 3, 4, 5, 6, 8$. Similar trends are obtained when data are broken down by k . We find that letter reply are made throughout the game, rather than solely at the outset.

B.3 Time duration from sending a letter request to receiving the requested letter

When v_i requests a letter of v_j , it has to wait for v_j to respond. Once v_j replies with the letter, then v_i is allowed to use the received letter and form words to contribute to the team. This time duration between request sent and reply received

reveals how long players take to reply to their neighbors' requests. A player only has to request a letter (and receive it) on one occasion to use it as any number of times in forming words. Remember that these rules were designed to foster word construction, to increase earnings potential, and to foster team cohesion. We want to analyze the behavior of players with reference to the time duration between the timestamps of the letter reply action and the letter request action, to answer the following questions. How long does it take for players to reply to a letter request? How does the number of neighbors affect the difference between the timestamps of the letter reply action and the letter request action?

Figure 24 shows a histogram with 10 bins of 30-s each, for the time difference between **reply received** and **request sent**, for 47 experiments with $k = 2, 3, 4, 5, 6, 8$. A kernel-density estimation with Gaussian kernels is used estimate the probability density function. Players generally respond relatively quickly to their neighbors letter requests with replies typically made within 30 s of the request.

To analyze whether this behavior is common while increasing the number of k neighbors in a game, Figure B.3 in Appendix B.1.3 in Cedeno (2019) shows histograms with 10 bins of 30-s each of timestamp change between reply received and request sent for experiments with $k = 2, 3, 4, 5, 6, 8$. The number of neighbors doesn't affect this type of action, players generally respond relatively quickly to their neighbors letter requests with replies typically made within 30 s of the request.

B.4 Timestamp for word formed

At any time during a game, a player can form a word and submit it for validation to our web application. If a player possesses letters to form a valid word, then she forms and submits a word, the application validates it, and the word is added to the game screen. We want to analyze the behavior of players with reference to the action of word formed and answer the following questions. When do players submit words during the game? How does the number of neighbors and the number of available letters affects the number of words formed by a player?

Figure 25 shows a histogram with 10 bins of 30-s each for **word formed**, for 47 experiments with $k = 2, 3, 4, 5, 6, 8$. A kernel-density estimation with Gaussian kernels is used estimate the probability density function. It suggests that words are being formed throughout the game, and even up through the end of the game. This justifies a 5-min anagram game duration. To analyze whether the number of neighbors affects the word formation, Figure B.4 in Appendix B.1.4 in Cedeno (2019) shows histograms with 10 bins of 30-s each for timestamp of **word formed** for experiments with $k = 2, 3, 4, 5, 6, 8$. Word submissions are made throughout the game, and the number of neighbors and available letters does not affect this behavior.

C Modeling supplement

This Appendix contains several figures that support the modeling of Sect. 5. See Figs. 26 and 27.

References

- Abbott A (1995) Sequence analysis: new methods for old ideas. *Ann Rev Sociol* 21:93–113
- Abrams DE, Hogg MA (1990) Social identity theory: constructive and critical advances. Springer, New York
- Ackland R, O’Neil M (2011) Online collective identity: the case of the environmental movement. *Soc Netw* 33:177–190
- Agrawal R, Srikant R (1995) Mining sequential patterns. In: Proceedings of the eleventh international conference on data engineering, IEEE Computer Society, ICDE’95, Washington, pp 3–14
- Aipperspach R, Cohen E, Canny J (2006) Modeling human behavior from simple sensors in the home. In: Proceedings of the IEEE conference on pervasive computing
- Alexander JC, Eyerman R, Giesen B, Smelser NJ, Sztompka P (2004) Cultural trauma and collective identity, 1st edn. University of California Press, Berkeley
- Andrews EAM, Bonner AJ (2011) Explaining genetic knock-out effects using cost-based abduction. In: Proceedings of the 22nd international joint conference on artificial intelligence (IJCAI), pp 1635–1640
- Atran S, Axelrod R, Davis R (2007) Sacred barriers to conflict resolution. *Science* 317:1039–1040
- Atran S, Sheikh H, Gomez A (2014a) Devoted actors sacrifice for close comrades and sacred cause. *Proc Natl Acad Sci* 111(50):17702–17703
- Atran S, Sheikh H, Gomez A (2014b) For cause and comrade: devoted actors and willingness to fight. *Cyclodynamics* 5:41–57
- Attema T, van Maanen PP, Meeuwissen E (2015) Development and evaluation of multi-agent models predicting twitter trends in multiple domains. In: 2015 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pp 1133–1140
- Bach A (1990) Boltzmann’s probability distribution of 1877. *Arch Hist Exact Sci* 41:1–40
- Becker GS (1974) A theory of social interaction. *J Political Econ* 82:1063–1093
- Benjamin DJ, Choi JJ, Fisher G (2016) Religious identity and economic behavior. *Rev Econ Stat* 98(4):617–637
- Bergmann JHM, Langdon PM, Mayagoitia RE, Howard N (2014) Exploring the use of sensors to measure behavioral interactions: An experimental evaluation of using hand trajectories. *PLoS One* 9(2):1–10
- Bornstein G, Yaniv I (1998) Individual and group behavior in the ultimatum game: Are groups more rational players? *Exp Econ* 1:101–108
- Brewer MB (1991) The social self: on being the same and different at the same time. *Personal Soc Psychol Bull* 17:475–482
- Brewer MB, Gardner W (1996) Who is this “we”? Levels of collective identity and self representations. *J Personal Soc Psychol* 71:83–93
- Brewer MB, Silver M (1978) Ingroup bias as a function of task characteristics. *Eur J Soc Psychol* 8(3):393–400
- Brody H (2000) The other side of Eden: hunters, farmers, and the shaping of the world. Farrar, Straus and Giroux, New York

- Brunsdon AR (2017) #MisconstruedidentitiesMustFall collective: identity formation in the current South African context: a practical theological perspective. *HTS Theol Stud* 73:1–7
- Butler J (1988) Performative acts and gender constitution: an essay in phenomenology and feminist theory. *Theatre J* 40(4):519–531
- Cadsby CB, Song F, Tapon F (2007) Sorting and incentive effects of pay for performance: an experimental investigation. *Acad Manag J* 50:387–405
- Cadsby CB, Song F, Tapon F (2010) Are you paying your employees to cheat? An experimental investigation. *BE J Econ Anal Pol* 10:1–30
- Cameron D (1997) Performing gender identity: Young men's talk and the construction of heterosexual masculinity. In: Johnson S, Hanna U (eds) *Language and masculinity*. Blackwell, Oxford, pp 47–64
- Capraro V (2013) A model of human cooperation in social dilemmas. *PLoS One* 8:e72427–1–e72427–6
- Cedeno VI (2019) Pipelines for computational social science experiments and model building. Ph.D. dissertation, Virginia Polytechnic Institute and State University, https://vtworks.lib.vt.edu/bitstream/handle/10919/91445/Cedeno_VI_D_2019.pdf. Accessed 8 Jan 2019
- Chamiak E, Santos E (1992) Dynamic map calculations for abduction. In: Proceedings AAAI conference on artificial intelligence, pp 552–557
- Charness G, Rigotti L, Rustichini A (2007) Individual behavior and group membership. *Am Econ Rev* 97:1340–1352
- Charness G, Cobo-Reyes R, Jimenez N (2014) Identities, selection, and contributions in a public-goods game. *Games Econ Behav* 87:322–338
- Chen DL, Yeh S (2014) The construction of morals. *J Econ Behav Organ* 104:84–105
- Chen R, Chen Y (2011) The potential of social identity for equilibrium selection. *Am Econ Rev* 101(6):2562–89
- Chen Y, Li S (2009) Group identity and social preferences. *Am Econ Rev* 99:431–457
- Chen Y, Li SX, Liu TX, Shih M (2014) Which hat to wear? Impact of natural identities on coordination and cooperation. *Games Econ Behav* 84:58–86
- Chen DL, Schonger M, Wickens C (2016) otree—an open-source platform for laboratory, online and field experiments. *J Behav Exp Financ* 9:88–97
- Chierichetti F, Kleinberg J, Kumar R, Mahdian M, Pandey S (2014) Event detection via communication pattern analysis. In: Proceedings of the 8th international conference on weblogs and social media, ICWSM, pp 51–60
- Choup AM (2008) The formation and manipulation of collective identity: a framework for analysis. *Soc Mov Stud* 7(2):191–207
- Cohn A, Fehr E, Maréchal MA (2014) Business culture and dishonesty in the banking industry. *Nature* 516:86–89
- Cohn A, Marechal MA, Noll T (2015) Bad boys: how criminal identity salience affects rule violation. *Rev Econ Stud* 82(4):1289–1308
- Cover TM, Thomas J (1991) Elements of information theory. Wiley, Hoboken
- Contractor N (2019) How can computational social science motivate the development of theories, data, and methods to advance our understanding of communication and organizational dynamics? In: Foucault Welles B, González-Bailón S (eds) *The Oxford Handbook of Networked Communication*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190460518.013.7>
- Curranini S, Jackson M, Pin P (2009) An economic model of friendship: homophily, minorities and segregation. *Econometrica* 77:1003–1045
- Davis WL, Davis DE (1972) Internal-external control and attribution of responsibility for success and failure. *J Personal* 40:123–136
- de Oliveira M (1992) Isotropic majority-vote model on a square lattice. *J Stat Phys* 66:273–281
- DeChurch LA, Mesmer-Magnus JR (2010) The cognitive underpinnings of effective teamwork: a meta-analysis. *J Appl Psychol* 95(1):32–53
- Dominowski RL (1969) The effect of pronunciation practice on anagram difficulty. *Psychon Sci* 16(2):99–100
- Drouvelis M, Metcalfe R, Powdthavee N (2010) Priming cooperation in social dilemma games. In: IZA Discussion Papers 4963, Institute for the Study of Labor (IZA)
- Durkheim E (1951) Suicide. Free Press, Mumbai
- Echenim M, Peltier N, Tourret S (2013) An approach to abductive reasoning in equational logic. In: Proceedings of the 23rd international joint conference on artificial intelligence (IJCAI), pp 531–537
- Eckel C, Grossman P (2005) Managing diversity by creating team identity. *J Econ Behav Organ* 58:371–392
- Epstein JM (2007) Generative social science: studies in agent-based computational modeling. Princeton University Press, Princeton
- Eriksen TH (2010) Ethnicity and nationalism: anthropological perspectives, 3rd edn. Pluto Press, London
- Erikson EH (1980) Identity and the life cycle. W. W. Norton & Company, New York
- Feather NT (1969) Attribution of responsibility and valence of success and failure in relation to initial confidence and task performance. *J Personal Soc Psychol* 13:129–144
- Feather NT, Simon JG (1971a) Attribution of responsibility and valence of outcome in relation to initial confidence and success and failure of self and other. *J Personal Soc Psychol* 18:173–188
- Feather NT, Simon JG (1971b) Causal attributions for success and failure in relation to expectation of success based upon selective or manipulative control. *J Personal Soc Psychol* 39:527–541
- Feher O, Wonnacott E, Smith K (2016) Structural priming in artificial languages and the regularisation of unpredictable variation. *J Mem Lang* 91:158–180
- Fiske ST, Gilbert DT, Lindzey G (2010) *Handbook of social psychology*, 5th edn. Wiley, Hoboken
- Flach PA, Kakas AC (2010) Abduction and induction: essays on their relation and integration. Springer, New York
- Fominaya CF (2010) Collective identity in social movements: central concepts and debates. *Sociol Compass* 4:393–404
- Gates MA, Suchow JW, Griffiths TL (2017) Empirical tests of large-scale collaborative recall. In: CogSci
- Gibbs AL, Su FE (2002) On choosing and bounding probability metrics. *Int Stat Rev* 70:419–435
- Gilhooly KJ, Johnson CE (1978) Effects of solution word attributes on anagram difficulty: a regression analysis. *Q J Exp Psychol* 30(1):57–70
- Gilwhite FJ (2001) Are ethnic groups biological species to the human brain? Essentialism in our cognition of some social categories. *Curr Anthropol* 42(4):515–553
- Ginges J, Atran S (2013) Sacred values and cultural conflict. In: Advances in culture and psychology, pp 273–305
- Goldberg CA (2003) Haunted by the specter of communism: collective identity and resource mobilization in the demise of the Workers Alliance of America. *Theory Soc* 32(5/6):725–773
- Goldman M, Stockbauer JW, McAuliffe TG (1977) Intergroup and intragroup competition and cooperation. *J Exp Soc Psychol* 13:81–88
- Gomez A, Brooks ML, Buhrmester MD, Vazquez A, Jetten J, Swann WB (2011a) On the nature of identity fusion: Insights into the construct and a new measure. *J Personal Soc Psychol* 100:918–933
- Gomez A, Morales J, Hart S, Vazquez A, Jetten J, Swann WB (2011b) Rejected and excluded forevermore, but even more devoted:

- irrevocable ostracism intensifies loyalty to the group among identity-fused persons. *Personal Soc Psychol Bull* 37:1574–1586
- Granovetter M (1978) Threshold models of collective behavior. *Am J Sociol* 83(6):1420–1443
- Greene B (2000) Letters to a young poet. New World Library, Novato
- Greenhill B (2008) Recognition and collective identity formation in international politics. *Eur J Int Relat* 14(2):343–368
- Guralnik V, Haigh KZ (2002) Learning models of human behaviour with sequential patterns. In: Proceedings of the AAAI-02 workshop “Automation as Caregiver”, aAAI Technical Report WS-02-02, pp 24–30
- Haig BD (2005) An abductive theory of scientific method. *Psychol Methods* 10:371–388
- Hand DJ, Smyth P, Mannila H (2001) Principles of data mining. MIT Press, Cambridge
- Hawkes AG (1971) Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58(1):83–90
- Heller N (2019) The philosopher redefining equality: Elizabeth Anderson thinks we've misunderstood the basis of a free and fair society. *The New Yorker* 07 January. <https://www.newyorker.com/magazine/2019/01/07/the-philosopher-redefining-equality>. Accessed 2 Feb 2019
- Hoff K, Pandey P (2006) Discrimination, social identity, and durable inequalities. *Am Econ Rev* 96(2):206–211
- Hoff K, Pandey P (2014) Making up people—the effect of identity on performance in a modernizing society. *J Dev Econ* 106(C):118–131
- Hogg MA, Abrams D (2007) Intergroup behavior and social identity. In: Hogg MA, Cooper J (eds) *The sage handbook of social psychology*. Sage, New York, pp 335–360
- Homans G (1961) Social behavior: its elementary forms. Harcourt Brace, San Diego
- Hu J, Liu M, Zhang J (2014) A semantic model for academic social network analysis. In: 2014 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM 2014), pp 310–313
- Jiménez J, Gomez A, Buhrmester MD et al (2016) The dynamic identity fusion index: a new continuous measure of identity fusion for web-based questionnaires. *Soc Sci Comput Rev* 34:215–228
- Juba B (2016) Learning abductive reasoning using random examples. In: Proceedings of the thirtieth AAAI conference on artificial intelligence
- Judd S, Kearns M, Vorobeychik Y (2010) Behavioral dynamics and influence in networked coloring and consensus. *PNAS* 107:14978–14982
- Juergensmeyer M (2003) Terror in the mind of god: the global rise of religious violence, 3rd edn. University of California Press, Berkeley
- Kahn A, Ryen AH (1972) Factors influencing the bias towards one's own group. *Int J Group Tens* 2:33–50
- Kearns M, Judd S, Tan J, Wortman J (2009) Behavioral experiments on biased voting in networks. *PNAS* 106:1347–1352
- Kearns M, Judd S, Vorobeychik Y (2012) Behavioral experiments on a network formation game. In: Economics and computation (EC), pp 690–704
- Kandler H, Kendler ST (1962) Vertical and horizontal processes in problem solving. *Psychol Rev* 69:1–16
- Kinnebrew JS, Loretz KM, Biswas G (2013) A contextualized, differential sequence mining method to derive students' learning behavior patterns. *J Educ Data Min* 5(1):190–219
- Korolov R, Lu D, Wang J, Zhou G, Bonial C, Voss C, Kaplan L, Wallace W, Han J, Ji H (2016) On predicting social unrest using social media. In: Proceedings of the 2016 IEEE/ACM international conference on advances in social networks analysis and mining, ASONAM 2016, Institute of Electrical and Electronics Engineers Inc., pp 89–95
- Korolov R, Peabody J, Lavoie A, Das S, Magdon-Ismail M, Wallace W (2015) Actions are louder than words in social media. In: 2015 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pp 292–297
- Kozogi B (2006) Ego utility, overconfidence, and task choice. *J Eur Econ Assoc* 4:673–707
- Kozlowski SW, Ilgen DR (2006) Enhancing the effectiveness of work groups and teams. *Psychol Sci Publ Interest* 7(3):77–124
- Kullback S (1959) Information theory and statistics. Wiley, New York
- Kullback S, Leibler RA (1951) On information and sufficiency. *Ann Math Stat* 22(1):79–86
- Kurashima T, Althoff T, Leskovec J (2018) Modeling interdependent and periodic real-world action sequences. In: World Wide Web Conference WWW, ACM, pp 803–812
- Latham GP, Locke EA (1991) Self-regulation through goal setting. *Organ Behav Hum Decisi Process* 50:212–247
- Ledyard JO (1994) Public goods: a survey of experimental research. Public economics, University Library of Munich, Germany
- Lee JY, Oh JC (2013) A model for recursive propagations of reputations in social networks. In: Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining, ASONAM '13, ACM, New York, pp 666–670
- Li B, Sun D, Lin Z, Ou C (2014) Agent-based simulation research on group emotion evolution of public emergency. In: ASONAM
- Locke EA, Latham GP (1990) A theory of goal setting and task performance. Prentice-Hall, Englewood Cliffs
- Luhmann CC, Rajaram S (2015) Memory transmission in small groups and large networks: an agent-based model. *Psychol Sci* 26(12):1909–1917
- Lustick I (2000) Agent-based modelling of collective identity: testing constructivist theory. *Journal of Artificial Societies and Social Simulation*, vol 3, no. 1. <http://jasss.soc.surrey.ac.uk/JASSS.html>
- MacGregor N (2018) Living with the gods: on beliefs and peoples, 1st edn. Knopf, New York
- Manchester W (1993) A world lit only by fire: the medieval mind and the renaissance: portrait of an age. Little, Brown and Company, Boston
- Mason W, Watts DJ (2012) Collaborative learning in networks. *Proc Natl Acad Sci* 109(3):764–769
- Mayzner MS, Tresselt ME (1958) Anagram solution times: a function of letter order and word frequency. *J Exp Psychol* 56(4):376
- McAuliffe K, Dunham Y (2015) Group bias in cooperative norm enforcement. *Philos Trans R Soc B* 371:20150073-1–20150073-9
- McFarland DA, Moody J, Diehl D, Smith JA, Thomas RJ (2014) Network ecology and adolescent social structure. *Am Sociol Rev* 79:1–34
- Melucci A (1989) Nomads of the present: social movement and identity needs in contemporary society. Temple University Press, Philadelphia
- Melucci A (1995) The process of collective identity. In: Klandermans B, Johnston H (eds) *Social movements and culture*. University of Minnesota Press, Minneapolis, pp 104–130
- Miller DT, Ross M (1975) Self-serving biases in the attribution of causality: fact or fiction? *Psychol Bull* 82:213–225
- Muller W (1996) How then, shall we live? Four simple questions that reveal the beauty and meaning of our lives. Bantam Books, New York
- Nagel J (1996) American Indian ethnic renewal: red power and the resurgence of identity and culture. Oxford University Press, New York
- Nguyen DA, Tan S, Ramanathan R, Yan X (2016) Analyzing information sharing strategies of users in online social networks. In: 2016 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pp 247–254

- Nguyen C, Schlesinger KJ, Carlson JM (2017) Data-driven models for individual and group decision making. In: ASONAM, pp 852–859
- Oldenquist A (1982) Loyalties. *J Philos* 79:173–193
- Olson M (1965) The logic of collective action: public goods and the theory of groups. Harvard Univ. Press, Cambridge
- Owens TJ (2006) Self and identity. In: Delamater J (ed) *Handbook of social psychology*. Springer, New York, pp 205–232
- Paris WD, Budapest GC, Konstan H, Bristol CG, Oregon KL, Sterling AO (1972) An experimental investigation into the formation of intergroup representations. *Eur J Soc Psychol* 2(2):202–204
- Paxton A, Morgan TJH, Suchow JW, Griffiths T (2018) Interpersonal coordination of perception and memory in real-time online social experiments. In: Proceedings of the 40th annual meeting of the cognitive science society, Austin
- Peek L (2005) Becoming Muslim: the development of a religious identity. *Sociol Relig* 66(3):215–242
- Perdue C, Dovidio JF, Gurtman MB, Tyler RB (1990) Us and them: social categorization and the process of intergroup bias. *J Personal Soc Psychol* 59:475–486
- Pereira L, Moreira F (2005) Majority-vote model on random graphs. *Phys Rev E Stat Nonlinear Soft Matter Phys* 71:016123
- Pfandler A, Rümmele S, Szeider S (2013) Backdoors to abduction. In: Proceedings of the 23rd international joint conference on artificial intelligence (IJCAI), pp 1046–1052
- Pierce CS (1931) Elements of logic. In: Hartshorn C et al (eds) Collected papers of Charles Sanders Pierce. Harvard University Press, Harvard
- Pilny A, Poole MS, Reichelmann A, Klein B (2017) A structuralist group decision-making perspective on the commons dilemma: results from an online public goods game. *J Appl Commun Res* 45(4):413–428
- Plutzer E, Zipp J (1996) Identity politics, partisanship, and voting for women candidates. *Publ Opin Q* 60(1):30–57. <https://doi.org/10.1086/297738>
- Polletta F, Jasper JM (2001) Collective identity and social movements. *Ann Rev Sociol* 27:283–305
- Qin M, Jin D, He D, Gabrys B, Musial K (2017) Adaptive community detection incorporating topology and content in social networks. In: Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017, ASONAM '17, ACM, New York, pp 675–682
- Ren Y, Cedeno-Mieles V, Hu Z, Deng X, Adiga A, Barrett C, Ekanayake S, Goode BJ, Korkmaz G, Kuhlman CJ, Machi D, Marathe MV, Ramakrishnan N, Ravi SS, Saraf P, Self N (2018) Generative modeling of human behavior and social interactions using abductive analysis. In: 2018 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pp 413–420
- Rousseau D, van der Veen AM (2005) The emergence of a shared identity: an agent-based computer simulation of idea diffusion. *J Confl Resolut* 49(5):686–712
- Russell DG, Sarason IG (1965) Test anxiety, sex, and experimental conditions in relation to anagram solution. *J Personal Soc Psychol* 1:493–496
- Salganik MJ, Watts DJ (2009) Web-based experiments for the study of collective social dynamics in cultural markets. *topiCS* 1(3):439–468
- Sarason IG (1973a) Test anxiety and cognitive modeling. *J Personal Soc Psychol* 28:58–61
- Sarason IG (1973b) Test anxiety and social influence. *J Personal* 41:261–271
- Schelling TC (2006) *Micromotives and macrobehavio*, revised edn. W. W. Norton & Company, New York
- Schubert TW, Otten S (2002) Overlap of self, ingroup, and out-group: pictorial measures of self-categorization. *Self Identity* 1(4):353–376
- Schweitzer M, Ordonez L, Dumaz B (2004) Goal-setting as a motivator of unethical behavior. *Acad Manag J* 47:422–433
- Sethi R, Somanathan E (2006) A simple model of collective action. *Econ Dev Cultural Change* 54(3):725–747
- Shanahan M (2005) Perception as abduction: turning sensor data into meaningful representation. *Cognit Sci* 29:103–134
- Shank DB, Kashima Y, Saber S, Gale T, Kirley M (2015) Dilemma of dilemmas: how collective and individual perspectives can clarify the size dilemma in voluntary linear public goods dilemmas. *PLoS One* 10:1–19
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423
- Silke A (2008) Holy warriors: exploring the psychological processes of jihadi radicalization. *Eur J Criminol* 5(1):99–123
- Singla P, Mooney RJ (2011) Abductive markov logic for plan recognition. In: Proceedings of the twenty-fifth AAAI conference on artificial intelligence
- Smith K, Feher O, Culbertson J (2017) The influence of word-order harmony on structural priming in artificial languages. In: Proceedings of the 39th annual meeting of the cognitive science society, CogSci 2017, London, 16–29 July 2017
- Snow D (2001) Collective identity and expressive forms. In: Smelser NJ, Baltes PB (eds) International encyclopedia of the social and behavioral sciences. Elsevier, Amsterdam, pp 2212–2219
- Snow D, McAdams D (2000) Identity work processes in the context of social movements: clarifying the identity/movement nexus. In: Stryker S, Owens T, White RW (eds) *Self, identity and social movements*. University of Minneapolis Press, Minneapolis, pp 2212–2219
- Stones CR (1983) Self-determination and attribution of responsibility: another look. *Psychol Rep* 53:391–394
- Stout CT, Kretschmer K, Ruppanner L (2017) Gender linked fate, race/ethnicity, and the marriage gap in American politics. *Polit Res Q* 70(3):509–522
- Stryker S (1980) Symbolic interactionism: a social structural version. Benjamin/Cummings, San Francisco
- Suri S, Watts DJ (2011) Cooperation and contagion in web-based, networked public goods experiments. *PLoS One* 6:1–18
- Sutton C, McCallum A (2011) An introduction to conditional random fields. *Found Trends Mach Learn* 4(4):267–373
- Swann WB, Gomez A, Seyle DC, Morales JF, Huici C (2009) Identity fusion: the interplay of personal and social identities in extreme group behavior. *J Personal Soc Psychol* 96:955–1011
- Swann WB, Gomez A, Dovidio JF, Hart S, Jetten J (2010a) Dying and killing for one's group: Identity fusion moderates responses to intergroup versions of the trolley problem. *Psychol Sci* 21:1176–1183
- Swann WB, Gomez A, Huici C, Morales JF (2010b) Identity fusion and self-sacrifice: Arousal as a catalyst of pro-group fighting, dying, and helping behavior. *J Personal Soc Psychol* 99:824–841
- Swann WB, Gomez A, Buhrmester MD, Lopez-Rodriguez L, Jimenez J, Vazquez A (2014) Contemplating the ultimate sacrifice: identity fusion channels pro-group affect, cognition, and moral decision making. *J Personal Soc Psychol* 106:713–727
- Swanson MS (2015) Composing democracy: Collective identity formation in small group composition. Ph.D. thesis, University of Washington, Washington
- Tajfel H (1974) Social identity and intergroup behavior. *Soc Sci Inf* 13:65–93
- Tanaka Y, Iwata T, Kurashima T, Toda H, Ueda N (2018) Estimating latent people flow without tracking individuals. In: Proceedings of the twenty-seventh international joint conference on artificial

- intelligence, IJCAI-18, international joint conferences on artificial intelligence organization, pp 3556–3563
- Tarrow S (2010) Mentalities, political cultures, and collective action frames: constructing meanings through action. *Sociol Compass* 4:174–202
- Tatum BD (2003) Why are all the black kids sitting together in the cafeteria? And other conversations about race, 2nd edn. Basic Books, New York
- Taylor V, Whittier NE (1992) Collective identity in social movement communities: lesbian feminist mobilization. In: Morris AD, Mueller CM (eds) *Frontiers of social movement theory*. Yale University Press, New Haven, pp 104–130
- Thagard P (1989) Explanatory coherence. *Behav Brain Sci* 12:435–502
- Timmermans S (1999) Social death as a self-fulfilling prophecy: David Sudnow's ‘passing on’ revisited. *Sociol Q* 39:453–472
- Timmermans S, Tavory I (2012) Theory construction in qualitative research: from grounded theory to abductive analysis. *Sociol Theory* 30:167–186
- Tran T, Lee K (2016) Understanding citizen reactions and ebola-related information propagation on social media. In: Proceedings of the 2016 IEEE/ACM international conference on advances in social networks analysis and mining, ASONAM '16, IEEE Press, Piscataway, pp 106–111
- Tresselt ME (1968) Reexamination of anagram problem solving. *Trans N Y Acad Sci* 30:1112–1119
- van Dijk TA (2000) Ideology: a multidisciplinary approach. Sage, Thousand Oaks
- van Maanen PP, van der Vecht B (2013) An agent-based approach to modeling online social influence. In: Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining, ASONAM '13, ACM, New York, pp 600–607
- van Zomeren M, Postmes T, Spears R (2008) Toward an integrative social identity model of collective action: a quantitative research synthesis of three socio-psychological perspectives. *Psychol Bull* 134(4):504
- Vance RJ, Colella A (1990) Effects of two types of feedback on goal acceptance. *J Appl Psychol* 75:68–77
- Vanderhaegen F, Caulier P (2011) A multi-viewpoint system to support abductive reasoning. *Inf Sci* 181:5349–5363
- Verkuyten M, Yildiz AA (2007) National (dis)identification and ethnic and religious identity: a study among Turkish-Dutch Muslims. *Personal Soc Psychol Bull* 33(10):1448–1462
- Vryan KD, Adler EA, Adler P (2003) Identity. In: Reynolds LT, Herman-Kinney NJ (eds) *Handbook of symbolic interactionism*. Altamira Press, Lanham, pp 367–390
- Warren MW, Thomson WJ (1969) Anagram solution as a function of transition probabilities and solution word frequency. *Psychon Sci* 17(6):333–334
- Wei-Kleiner F, Dragisic Z, Lambrix P (2014) Abduction framework for repairing incomplete ontologies: complexity results and algorithms (extended version). In: Proceedings of the 28th AAAI conference on artificial intelligence
- Wendt A (1994) Collective identity formation and the international state. *Am Polit Sci Rev* 88(2):384–396
- Worchel S, Andreoli VA, Folger R (1977) Intergroup cooperation and intergroup attraction: the effect of previous interaction and outcome of combined effort. *J Exp Soc Psychol* 13:131–140
- Zhang H, Vorobeychik Y, Letchford J, Lakkaraju K (2016) Data-driven agent-based modeling, with application to rooftop solar adoption. *Auton Agents Multi-Agent Syst* 30:1023–1049
- Zhang H, Vorobeychik Y (2019) Empirically grounded agent-based models of innovation diffusion: a critical review. *Artif Intell Rev* 52:707–741

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Vanessa Cedeno-Mieles^{2,4,6} · **Zhihao Hu**⁵ · **Yihui Ren**⁷ · **Xinwei Deng**⁵ · **Abhijin Adiga**¹ · **Christopher Barrett**^{1,3} ·
Noshir Contractor⁹ · **Saliya Ekanayake**¹² · **Joshua M. Epstein**¹⁰ · **Brian J. Goode**² · **Gizem Korkmaz**¹ ·
Chris J. Kuhlman¹ · **Dustin Machi**¹ · **Michael W. Macy**¹¹ · **Madhav V. Marathe**^{1,3} · **Naren Ramakrishnan**^{4,8} · **S. S. Ravi**¹ ·
Parang Saraf⁸ · **Nathan Self**⁸

Zhihao Hu
huzhihao@vt.edu

Yihui Ren
yihui.ren@gmail.com

Xinwei Deng
xdeng@vt.edu

Abhijin Adiga
abhijin@virginia.edu

Christopher Barrett
cbarrett@virginia.edu

Noshir Contractor
ncontractor@gmail.com

Saliya Ekanayake
esaliya@lbl.gov

Joshua M. Epstein
je65@nyu.edu

Brian J. Goode
bgoode@vt.edu

Gizem Korkmaz
gkorkmaz@virginia.edu

Chris J. Kuhlman
cjk8gx@virginia.edu

Dustin Machi
dmachi@virginia.edu

Michael W. Macy
mwm14@cornell.edu

Madhav V. Marathe
mmarathe@virginia.edu

Naren Ramakrishnan
naren@cs.vt.edu

S. S. Ravi
ssravi0@gmail.com

Parang Saraf
parang@cs.vt.edu

Nathan Self
nwself@vt.edu

- ¹ Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, VA, USA
- ² Biocomplexity Institute, Virginia Tech, Blacksburg, VA, USA
- ³ Department of Computer Science, University of Virginia, Charlottesville, VA, USA
- ⁴ Department of Computer Science, Virginia Tech, Blacksburg, VA, USA
- ⁵ Department of Statistics, Virginia Tech, Blacksburg, VA, USA
- ⁶ Escuela Superior Politécnica del Litoral, ESPOL, Guayaquil, Ecuador
- ⁷ Computational Science Initiative, Brookhaven National Laboratory, Upton, NY, USA
- ⁸ Discovery Analytics Center, Virginia Tech, Blacksburg, VA, USA
- ⁹ Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, USA
- ¹⁰ Department of Epidemiology, New York University, New York, NY, USA
- ¹¹ Department of Sociology, Cornell University, Ithaca, NY, USA
- ¹² Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Bounds and Complexity Results for Learning Coalition-Based Interaction Functions in Networked Social Systems

Abhijin Adiga¹ Chris J. Kuhlman¹ Madhav V. Marathe^{1,2} S. S. Ravi^{1,3}
Daniel J. Rosenkrantz^{1,3} Richard E. Stearns^{1,3} Anil Vullikanti^{1,2}

¹Biocomplexity Institute and Initiative, University Virginia, Charlottesville, VA 22904

²Computer Science Dept., University Virginia, Charlottesville, VA 22904

³Computer Science Dept., University at Albany – SUNY, Albany, NY 12222

{abhijin, cjk8gx, marathe}@virginia.edu, ssravi0@gmail.com, drosenkrantz@gmail.com,
thestearns2@gmail.com, vskumar@virginia.edu

Abstract

Using a discrete dynamical system model for a networked social system, we consider the problem of learning a class of local interaction functions in such networks. Our focus is on learning local functions which are based on pairwise disjoint coalitions formed from the neighborhood of each node. Our work considers both active query and PAC learning models. We establish bounds on the number of queries needed to learn the local functions under both models. We also establish a complexity result regarding efficient consistent learners for such functions. Our experimental results on synthetic and real social networks demonstrate how the number of queries depends on the structure of the underlying network and number of coalitions.

1 Introduction

Motivation. Learning the nature of interactions in networked physical and social systems is a challenging problem (see e.g., (Laubenbacher and Stigler 2004; Romero, Meeder, and Kleinberg 2011; González-Bailón et al. 2011)). We use a graphical dynamical systems model, called a synchronous dynamical system (SyDS) (see e.g., (Barrett et al. 2006)) to represent these networked systems. Such a system consists of an undirected graph $G(V, E)$, where the nodes represent entities (agents) and the edges represent pairwise interactions. (Formal definitions are provided in Section 2.) Each node v has a time varying state value (assumed to be Boolean) and a local function f_v which determines the next state of the node using the current states of v and its neighbors. The SyDS model assumes that nodes compute and update their state values synchronously. The graph and the local functions determine the dynamics of the system.

The problem of understanding the nature of interactions in a networked system can be formulated as that of inferring the local functions in a SyDS model of the system. We consider inference through interactions with the system where a user may specify each query in the form of a configuration (i.e., the current state values of nodes) and the system provides the successor configuration, i.e., states of the nodes at the next

time instant (Adiga et al. 2018; He et al. 2016). We also consider inference under the Probably Approximately Correct (PAC) learning framework where configuration–successor pairs are independently drawn from an unknown distribution (see similar work in (Narasimhan, Parkes, and Singer 2015; He et al. 2016)). Under both models, we assume that the network is known.

The great majority of prior work focuses on each agent’s individual behavior, where an agent treats each neighbor as an autonomous influencer. However, in several situations, an agent is influenced by groups formed by its neighbors. Here, our focus is on learning a form of interaction based on pairwise disjoint **coalitions** formed by the neighbors of an agent. The motivation for this model comes from the work reported in (Ugander et al. 2012; Laubenbacher and Stigler 2004; Colón-Reyes et al. 2006). The model studied in (Ugander et al. 2012) uses a social network and considers the connected components formed by the one-hop neighbors of a node v . Since the connected components are node disjoint, so are the coalitions. The experimental evidence presented in (Ugander et al. 2012) shows that coalitions are indeed operative in social networked systems. In particular, the results in this reference point out that people consider coalitions of their neighbors in deciding whether to join Facebook. Thus, our model of non-overlapping coalitions has direct relevance to social systems. The model studied in (Laubenbacher and Stigler 2004; Colón-Reyes et al. 2006) considers polynomial interaction functions, where each polynomial is a sum of monomials (products of variables where the degree of each variable is at most 1). The monomials can be thought of as coalitions. For Boolean functions, sums of monomials correspond to monotone functions in disjunctive normal form (DNF); such functions are in the sum of products form where no variable appears negated. Each product term in a DNF represents a coalition. Since coalitions considered in models of social systems generally do not overlap (Branzei, Dimitrov, and Tijs 2005; Ugander et al. 2012), we have the additional requirement that the coalitions must partition the set of inputs. We call such functions **partitioned monotone DNF** (PM-DNF) functions. As an example, the Boolean function of five variables defined by $f(x_1, x_2, x_3, x_4, x_5) = x_1 x_5 + x_2 x_3 + x_4$ consists of

three pairwise disjoint coalitions. The interpretation is that the function takes on the value 1 iff least one of the coalitions is **unanimous**, i.e., all the variables in that coalition have the value 1. Thus, in a social system with PM-DNF functions, a node changes to 1 at time $\tau + 1$ iff there is at least one unanimous coalition among its inputs at time τ .

Our work considers the problem of learning PM-DNF functions under the active query model of (Adiga et al. 2018) and the PAC learning model (Valiant 1984). Two extreme cases of the PM-DNF model are well studied: (i) the Boolean OR function where every coalition has exactly one neighbor (which corresponds to the simple contagion model of (Granovetter 1978)) and (ii) Boolean AND function where all the neighbors form a single coalition (which corresponds to a particular type of complex contagion model of (Centola and Macy 2007)). Our model is a generalization of these two extreme cases. Also, to the best of our knowledge, this is the first work that addresses learning functions that depend on groups of neighbors, rather than individual neighbors. Such a dynamical system can also be viewed as a model for diffusion on hypergraphs (Zhu et al. 2018).

Summary of results.

1. Bounds under the active query model. We present an algorithm that can learn any PM-DNF function with q inputs using $O(q \log q)$ membership queries¹ under the adaptive mode (where a query may depend on the answers to the previous queries). We also show that in the worst-case, $\Omega(q \log q)$ queries are required under the adaptive mode to infer such a function. In addition, we show that $O(\chi \Delta \log \Delta)$ adaptive queries are sufficient to infer all the local functions of a SyDS where χ and Δ represent the number of colors needed to color G^2 (the square graph² of G) and the maximum node degree of G respectively.
2. PAC model upper bound. For any fixed values of the parameters ϵ and δ , we show that for learning the PM-DNF functions at all the nodes of a SyDS, an upper bound on the sample complexity is $O((2m + n) \log(\Delta + 1))$, where $m = |E|$ and Δ is the maximum node degree.
3. Complexity of efficient PAC learning. We show that the class of PM-DNF functions with two or more product terms is not efficiently PAC learnable unless **NP = RP**. (The corresponding problem for one product term is efficiently solvable since a PM-DNF function with one product term is just the AND function.)
4. An algorithm for learning under the PAC model. To cope with the intractability result mentioned in Item 3 above, we present an integer linear programming (ILP)-based algorithm for determining whether there is a PM-DNF function that is consistent with all the given examples (definition in Section 4). This algorithm can be used to construct a PAC learning algorithm for PM-DNF functions in practice.
5. Experimental results. We present experimental results for generating query sets under the adaptive model for both synthetic and real social networks. The number of queries re-

¹A membership query specifies an input to a Boolean function and the response is the value of the function.

²The square $G^2(V, E^2)$ of a graph $G(V, E)$ has the edge $\{u, v\}$ whenever there is a path of length ≤ 2 between u and v in G .

quired depends on the structure of the graph and number of blocks (i.e., coalitions). For example, in the case of scale-free networks, the number of queries required is much less than the theoretical upper bound established in this paper. Under the PAC model, we analyze a single local function with regard to sample distribution, size of the input and number of blocks. Interestingly, the ILP-based algorithm exhibits better performance when the number of blocks is large.

For space reasons, proofs for many of the results are omitted; they appear in (Adiga et al. 2019a).

Related work. Many researchers have addressed the problem of learning components of physical and social systems (see e.g., (Adiga et al. 2018; He et al. 2016; Laubenbacher and Stigler 2004; Romero, Meeder, and Kleinberg 2011; González-Bailón et al. 2011)). As mentioned earlier, the coalition-based interaction model was motivated by the work in (Ugander et al. 2012; Colón-Reyes et al. 2006). The problem of learning Boolean DNF functions has received attention in the learning theory literature under membership query and PAC learning models. For example, bounds on the number of membership queries for learning monotone DNFs are proven in (Abasi, Bshouty, and Mazzawi 2014). In their work, the product terms may not partition the set of variables. The problem of learning discrete distributions over $\{0, 1\}^n$ is considered in (Kearns et al. 1994); some of their results use circuits that compute monotone DNF (but not PM-DNF) functions. Other learning problems for DNF functions have been considered in several papers (e.g., (Angluin and Slonim 1994; Liśkiewicz, Lutter, and Reischuk 2017; Servedio 2004)). The topic of active learning has also been explored in the context of sensor networks (e.g., (Castro and Nowak 2007)).

To our knowledge, the problem of learning PM-DNF functions for networked systems has not been addressed in the literature. In particular, the adaptive query techniques presented in (Adiga et al. 2018) for learning symmetric functions (and threshold functions which are a subclass of symmetric functions) cannot be applied to PM-DNF functions since the latter is not a subclass of the former. For example, the PM-DNF function $f(x_1, x_2, x_3, x_4) = x_1 x_2 + x_3 x_4$ is not a symmetric function since $f(1, 0, 1, 0) = 0 \neq f(1, 1, 0, 0)$; hence, it is also not a threshold function. Learning threshold functions under the PAC model is considered in (Adiga et al. 2019b); they present a complexity result for efficient consistent learners for threshold functions similar to our result for PM-DNF functions. However, our complexity result is not implied by the one in (Adiga et al. 2019b).

2 Definitions and Problem Formulations

Model for networked social systems. Following (Barrett et al. 2006), we use a formalism called a **synchronous dynamical system** (SyDS), to model a networked social system. Let \mathbb{B} denote the Boolean domain $\{0, 1\}$. A SyDS \mathcal{S} over \mathbb{B} is a pair $\mathcal{S} = (G, \mathcal{F})$, where (i) $G(V, E)$, an undirected graph with $n = |V|$ nodes, represents the underlying graph of the SyDS, and (ii) $\mathcal{F} = \{f_1, f_2, \dots, f_n\}$ is a collection of functions, with f_i denoting the **local function** at node v_i , $1 \leq i \leq n$. At any time, each node of G has a state value from \mathbb{B} . The inputs to function f_i are the states

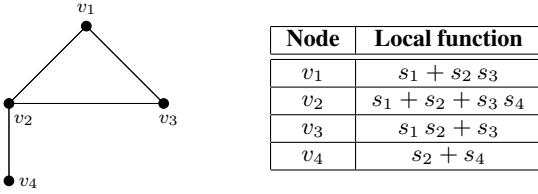


Figure 1: An example of a PM-DNF-SyDS. The local functions (which are all PM-DNF functions) are shown in the table on the right. Variable s_i represents the state of node v_i , $1 \leq i \leq 4$.

of the nodes in the **closed neighborhood** of v_i (i.e., node v_i and the neighbors of v_i in G). For each input, the output of function f_i gives the next state of v_i . In a SyDS, all nodes compute and update their next state *synchronously* (i.e., in parallel). At any time τ , if $s_i^\tau \in \mathbb{B}$ is the state of node v_i ($1 \leq i \leq n$), the **configuration** C of the SyDS is the n -vector $(s_1^\tau, s_2^\tau, \dots, s_n^\tau)$. The system evolves in discrete time steps by repeated application of \mathcal{F} . If C and C' denote two successive configurations of a SyDS, then C' is the **successor** of C .

Partitioned monotone DNF functions. In this paper, each local function f_i is based on coalitions formed by the closed neighborhood of node v_i , $1 \leq i \leq n$. Our focus is on one class of such Boolean functions, called **partitioned monotone DNF** (PM-DNF) functions.

Definition 1. A Boolean function f is a PM-DNF iff it has a disjunctive normal form (DNF) (i.e., sum of products) representation satisfying the following two properties: (i) all the variables appear unnegated in f (i.e., f is monotone) and (ii) the collection of product terms (also referred to as **blocks** or **coalitions**) partitions the set of inputs to f ; i.e., each input appears in *exactly one* block.

Example 1. Suppose we have five Boolean variables, denoted by x_1, x_2, x_3, x_4 and x_5 . One example of a PM-DNF function is $f_1(x_1, x_2, x_3, x_4, x_5) = x_1 x_3 x_5 + x_2 x_4$, which has two product terms (coalitions). Note that the OR function $f_3(x_1, x_2, x_3, x_4, x_5) = x_1 + x_2 + x_3 + x_4 + x_5$ (five coalitions) and the AND function $f_4(x_1, x_2, x_3, x_4, x_5) = x_1 x_2 x_3 x_4 x_5$ (one coalition) are PM-DNF functions. On the other hand, $f_5(x_1, x_2, x_3, x_4, x_5) = x_1 x_2 + x_4 x_5$ is not a PM-DNF function since x_3 doesn't appear in any of the product terms. Likewise, $f_6(x_1, x_2, x_3, x_4, x_5) = x_1 x_2 + x_3 \bar{x}_4 + x_5$ is not a PM-DNF function since x_4 is negated.

For simplicity, we use the abbreviation PM-DNF-SyDS to denote a SyDS in which every local function is a PM-DNF function. We now present an example of such a SyDS.

Example 2. The graph of a PM-DNF-SyDS is shown in Figure 1. Suppose the initial configuration is $(1, 0, 0, 0)$; that is, node v_1 is in state 1 and nodes v_2, v_3 and v_4 are in state 0. It can be seen that the system goes through the following sequence of configurations during the next two time steps: $(1, 0, 0, 0) \rightarrow (1, 1, 0, 0) \rightarrow (1, 1, 1, 1)$. From the configuration $(1, 1, 1, 1)$, no further state changes occur. Such a configuration is a **fixed point** for this system.

Active query model. This query model for SyDSs was proposed in (Adiga et al. 2018). Under this model, each query, which we call a **successor query**, specifies a configuration C ; the response to the query is the configuration C' , the successor of C . One can think of C as specifying an input to each local function and the response C' as specifying the value of each local function for the input specified by C .

For expository purposes, we consider learning each local function separately. Thus, to learn an unknown PM-DNF function f , a query specifies an assignment α of values to the inputs of f ; the response to the query is the Boolean value $f(\alpha)$. In the learning theory literature, such queries are called **membership queries** (see e.g., (Angluin and Slonim 1994)). Since our goal is to use as few membership queries as possible, we will use the **adaptive** query mode considered in (Adiga et al. 2018). In this mode, membership queries are generated one at a time; a query may depend on the responses for previous queries. We also consider learning PM-DNF functions under the PAC model; we refer the reader to (Antony and Biggs 1992; Kearns and Vazirani 1994) for the relevant definitions.

Positive and negative examples. For an unknown PM-DNF f , each example η given to a PAC learner is a pair (α, β) , where α is an assignment of $\{0,1\}$ values to the inputs of f and $\beta \in \{0, 1\}$ is the value $f(\alpha)$ of the function. These are *positive* examples. We need not consider negative examples here since a negative example of the form (α, β) , that is, “ β is not the output of f for input α ”, is equivalent to the positive example $(\alpha, \bar{\beta})$.

The concept class of PM-DNF functions is **PAC learnable** by a learner L using the hypothesis space H if for any target concept c , values ϵ and δ such that $0 < \epsilon, \delta < 1/2$, and distribution \mathcal{D} over the instance space, L outputs with a probability of at least $1 - \delta$, a hypothesis $h \in H$ such that $\text{error}_{\mathcal{D}}(h) \leq \epsilon$. The **sample complexity** of a learner, denoted by $\mathcal{M}(\epsilon, \delta)$, is the number of examples needed by the learner to output an appropriate hypothesis h . We will use the following well-known upper bound (Haussler 1988) on $\mathcal{M}(\epsilon, \delta)$ based on the size of the hypothesis space H :

$$\mathcal{M}(\epsilon, \delta) \leq \frac{1}{\epsilon} (\log |H| + \log(1/\delta)). \quad (1)$$

3 Bounds Under the Active Query Model

Lower bound. We establish the lower bound by pointing out that any algorithm that uses membership queries under the adaptive mode can be viewed as a decision tree, like the one used to establish a lower bound on comparison-based sorting algorithms (Cormen et al. 2009). A proof of the following theorem appears in (Adiga et al. 2019a).

Theorem 1. Every algorithm that uses membership queries under the adaptive mode to learn a PM-DNF function with q inputs must use $\Omega(q \log(q))$ queries in the worst-case.

Upper bound: A query generation algorithm to learn a PM-DNF function. We now discuss our algorithm for generating membership queries under the adaptive mode to learn an unknown PM-DNF function f with q inputs, denoted by x_1, x_2, \dots, x_q . For each block of f , the variable

with the smallest index will be referred to as the **key variable** for that block. For example, if one of the blocks is $x_2 x_7 x_9$, then the key variable for that block is x_2 . For a set of blocks, the **key block** for that set is the block with the largest key variable. For a set of blocks, we define the **superkey variable** for that set of blocks to be the key variable of the key block in the set of blocks.

The algorithm consists of a loop that identifies the blocks of f , one block at a time. For each iteration of this loop, we refer to the already discovered blocks as the **known blocks**, and the remaining blocks as the **unknown blocks**. We refer to the variables in the known blocks as **allocated variables**, and the variables in the unknown blocks as **unallocated variables**. The unallocated variables are sorted by their index, lowest index first. The list of unallocated variables can be considered to be divided into two parts; a left part consisting of **primary unallocated variables**, and a right part (possibly empty) consisting of **secondary unallocated variables**. As the algorithm proceeds, the primary unallocated variables are unallocated variables that are potentially the key variable of some unknown block, whereas secondary unallocated variables are unallocated variables that the responses to previously issued queries have shown are not the key variable of any block.

Initially all the blocks are unknown, and all the variables are primary unallocated variables. At each iteration of the loop, the algorithm finds the key block among the currently unknown blocks. The algorithm does this by first finding the superkey variable for the set of unknown blocks, thereby identifying the key variable of the key unknown block. The current iteration of the loop then proceeds by finding all the remaining variables in the key unknown block, one variable at a time. Once all the variables of the key unknown block have been found, that block is now known, and the status of its variables changes to **allocated**. After changing the status of this block, if all the variables are allocated (i.e., belong to known blocks), then all the blocks of f have been identified, so the algorithm is finished. Otherwise, the algorithm reiterates the loop, to discover the new key unknown block.

Each iteration of the loop consists of two major substeps; the details of these substeps are provided below. Recall that each membership query specifies an assignment α of $\{0,1\}$ values to the variables, and the response to the query is the value $f(\alpha)$. The algorithm uses two types of membership queries: **superkey queries** and **block queries**. Substep 1 of each loop iteration uses superkey queries to find the superkey variable of the unknown blocks. Once this superkey variable is identified, Substep 2 uses block queries to find the other members of the key block.

Substep 1: Finding the superkey variable of the unknown blocks. The algorithm uses a binary search over the primary unallocated variables, using superkey queries to guide the search. The binary search maintains a list L of **candidate variables**, each of which is a primary unallocated variable, and might potentially be the superkey. List L initially consists of all the primary unallocated variables, since the superkey is one of these variables.

If list L of candidate variables contains only one variable, say variable x_k , then the binary search is over, and variable

x_k is the superkey. Otherwise, the binary search to find the superkey proceeds as follows. Let x_j be the $\lceil |L|/2 \rceil^{\text{th}}$ variable on list L . Let α^j be the assignment to the q variables where a given variable x_i is 1 iff x_i is unallocated and $i > j$. The algorithm issues α^j as a query, which we refer to as a *superkey query*.

Suppose $f(\alpha^j) = 0$. Then the unallocated variables to the right of x_j do not contain a complete block, so none of these variables can be the key of any unknown block. So, the status of each primary unallocated variable x_i such that $i > j$ is changed to **secondary**. Also, each candidate variable x_i on list L such that $i > j$ is deleted from L .

Suppose $f(\alpha^j) = 1$. Then the unallocated variables to the right of x_j contain a complete block, so the superkey is a candidate variable x_i such that $i > j$. So, each candidate variable x_i on list L such that $i \leq j$ is deleted from L .

In this manner, the search for the superkey variable is recursively continued on the left or right half of list L , depending on the value of $f(\alpha^j)$, until L contains just one variable. Note that each query reduces the size of L by a factor of 2. (More precisely, if L and L' denote respectively the list before and after the list shortening, then $|L'| = \lceil |L|/2 \rceil$). Thus, the number of queries used to find the superkey variable is at most $\lceil \log(q) \rceil$.

Substep 2: Finding the other variables in the key block. Substep 2 uses a loop that searches for the other variables in the key block, one variable at a time. At the beginning of each iteration of this loop, a nonempty set of key block members (including the superkey) have already been found. We refer to this set of variables as **identified key block members**. The iteration begins by issuing a *block query* α wherein a given variable is 1 iff it is an identified key block member. If $f(\alpha) = 1$, then the identified key block members form the complete key block. The key block is now known, so the status of its members is changed to allocated, and Substep 2 is complete.

If $f(\alpha) = 0$, then the key block contains at least one additional member, and a binary search is used to find the additional member with the lowest index. The binary search maintains a list L of **candidate variables**, each of which is a secondary unallocated variable, and which might potentially be the next member of the key block. List L initially consists of all the secondary unallocated variables to the right of the last member added to the key block, since the next member of the key block is one of these variables. If list L of candidate variables contains only one variable, say variable x_j , then the binary search is over, and variable x_j is the next member of the key block. Variable x_j is now the newest identified key block member, and another iteration of the main loop for Substep 2 begins.

Otherwise, if list L of candidate variables contains more than one variable, the binary search to find the next member of the key block proceeds as follows. Let x_j be the $\lceil |L|/2 \rceil^{\text{th}}$ variable on list L . Let α^j be the assignment to the q variables where a given variable x_i is 1 iff either x_i is an identified key block member or x_i is unallocated and $i > j$. The algorithm issues block query α^j .

Suppose $f(\alpha^j) = 0$. Then the key block has a variable x_i such that $i \leq j$ and x_i is on list L . Thus, each candidate

variable x_i on list L such that $i > j$ is deleted from L .

Suppose $f(\alpha^j) = 1$. Then the next member of the key block is a variable x_i such that $i > j$ and x_i is on list L . Thus, each candidate variable x_i on list L such that $i \leq j$ is deleted from L .

In this manner, the search for the next member of the key block is recursively continued on the left or right half of list L , depending on the value of $f(\alpha_j)$, until L contains just one variable. Since each query reduces the size of L by a factor of 2, the number of queries used to find the next member of the key block is at most $\lceil \log(q) \rceil$.

Overall, the algorithm uses at most $1 + \lceil \log(q) \rceil$ queries per variable. Thus, an upper bound on the number of queries is $q(1 + \lceil \log(q) \rceil) = O(q \log(q))$. Thus, we have:

Theorem 2. *A PM-DNF function f with q inputs can be learned using at most $q(1 + \lceil \log(q) \rceil) = O(q \log(q))$ adaptive membership queries.* ■

Inferring all local functions. The following theorem provides an upper bound on the number of queries needed to learn all local functions of a PM-DNF-SyDS. A proof of the theorem appears in (Adiga et al. 2019a).

Theorem 3. *For a PM-DNF-SyDS with underlying graph G , $O(\chi(G^2) \Delta \log(\Delta))$ successor queries are sufficient to infer all the local functions. Here, Δ is the maximum node degree in G and $\chi(G^2)$ is the minimum number of colors needed for a valid node coloring of G^2 .*

4 Results Under the PAC Learning Model

4.1 Upper bound on the number of queries

We begin with an upper bound on the sample complexity to learn a PM-DNF function under the PAC model. A proof of the following result appears in (Adiga et al. 2019a).

Proposition 1. *Let $\epsilon, \delta > 0$ be fixed. The asymptotic sample complexity $\mathcal{M}(\epsilon, \delta)$ for PAC learning all the PM-DNF local functions for a given graph $G(V, E)$ is $\mathcal{M}(\epsilon, \delta) = O((2m+n)\log(\Delta+1))$, where $m = |E|$ and Δ is the maximum degree of G .*

4.2 A complexity result for efficient PAC learning

We will show that a class of PM-DNF functions is not efficiently PAC learnable unless $\mathbf{NP} = \mathbf{RP}$. To do this, we need to introduce the notion of consistency of a PM-DNF function with respect to a set of examples.

Consistent hypothesis. Given a set \mathcal{E} of examples, we say that a hypothesis (i.e., a PM-DNF function) f is **consistent** with respect to \mathcal{E} if for each example $(\alpha, \beta) \in \mathcal{E}$, $f(\alpha) = \beta$. As is well known in the learning theory literature (see e.g., (Kearns and Vazirani 1994)), algorithms for obtaining consistent hypotheses are useful in constructing PAC learning algorithms.

We now present our complexity result for the class of PM-DNF functions with two or more product terms. (As stated in Section 1, the case of a PM-DNF function with one product term is trivial.) To prove the result, we use the following problem which is known to be **NP**-complete (Garey and Johnson 1979).

Hypergraph 2-Colorability (H2C): Given a set $U = \{u_1, u_2, \dots, u_q\}$ and a collection $Y = \{Y_1, Y_2, \dots, Y_k\}$ of subsets of U (i.e., the hyperedges) with $|Y_j| \geq 2$, $1 \leq j \leq k$, can the elements in U be colored with two colors so that no hyperedge in Y is monochromatic (i.e., each subset in Y contains at least one element of each color)?

Theorem 4. *If $\mathbf{NP} \neq \mathbf{RP}$, the class of PM-DNF functions with two or more product terms is not efficiently PAC learnable.*

Proof (idea). We use a reduction from H2C to show that if there is an efficient PAC learning algorithm for PM-DNF functions with two or more blocks, then there is an **RP**-time algorithm for H2C, contradicting the assumption that $\mathbf{NP} \neq \mathbf{RP}$. Details appear in (Adiga et al. 2019a). ■

We note that Theorem 4 holds for the case of *proper* learning where the hypothesis class and the concept class are the same, namely the class of PM-DNF functions. Whether the result can be extended to the representation-independent setting (see, e.g., (Warmuth 1989)) is left for future work.

4.3 An ILP-based PAC learning algorithm

As is well known, if a hypothesis h (which in this case is a PM-DNF function) that is consistent with all the given examples can be constructed, then the number of examples used to learn h is within a constant factor of the minimum sample complexity needed to learn the hypothesis class (Blumer et al. 1989). Therefore, we focus on developing an algorithm for a consistent learner. We consider the following problem which we call **Consistent Learning of Partitioned Monotone DNF** functions (CL-PMDNF).

Given: A set \mathcal{E} of examples for an unknown PM-DNF function f with q inputs given by $X = \{x_1, x_2, \dots, x_q\}$; \mathcal{E} is partitioned into \mathcal{E}_0 and \mathcal{E}_1 , where \mathcal{E}_0 (\mathcal{E}_1) is the set of examples in which the function value is 0 (1); integer $k \leq q$.

Requirement: Determine whether the variable set X can be partitioned into exactly k blocks, with each block forming a product term of the function, so that the resulting function is consistent with \mathcal{E} . If so, find one such partition.

The above formulation assumes that we know the number of blocks. This can be done without loss of generality since we can try the values $1, 2, \dots, q$ for the number of blocks.

Let B_1, B_2, \dots, B_k denote the blocks (product terms) of the unknown PM-DNF function. To develop our ILP formulation for CL-PMDNF, let z_{ij} be an indicator variable which has the value 1 if variable x_i is in Block B_j and 0 otherwise, $1 \leq i \leq q$ and $1 \leq j \leq k$. We now explain the constraints in our ILP.

The following two sets of constraints enforce the following requirements: (i) each variable appears in exactly one block and (ii) each block is nonempty (since we must have exactly k blocks).

$$\sum_{j=1}^k z_{ij} = 1, 1 \leq i \leq q; \quad \sum_{i=1}^q z_{ij} \geq 1, 1 \leq j \leq k.$$

Consider any example $\eta_p = (\alpha_p, 0)$ in \mathcal{E}_0 . Let $S_p \subseteq X$ be the set of variables that have the value 0 in the input assignment α_p . Since the value of the function is 0, each block

must have at least one of the variables from S_p . This gives rise to the following set of constraints:

$$\sum_{x_i \in S_p} z_{ij} \geq 1, \quad 1 \leq j \leq k.$$

Consider any example $\eta_r = (\alpha_r, 1)$ of \mathcal{E}_1 . Let $S_r \subseteq X$ be the set of variables that have the value 0 in the input assignment α_r . Since the value of the function is 1, there is at least one block which does *not* have any of the variables from S_r . To capture this constraint, we introduce k auxiliary $\{0,1\}$ variables, denoted by $b_{r,1}, b_{r,2}, \dots, b_{r,k}$, and the following constraints. (Note that each example in \mathcal{E}_1 gives rise to a distinct set of auxiliary variables.)

$$b_{r,j} \geq z_{ij}, \quad \forall i \in S_r; \quad \sum_{j=1}^k b_{r,j} \leq k - 1.$$

It can be verified that the last two sets of constraints together imply that there is a block B_j that does not contain any of the variables in S_r .

Thus, the CL-PMDNF problem is represented by the set of constraints given above along with the following constraints on the variables: (i) $z_{ij} \in \{0,1\}$, for $1 \leq i \leq q$, $1 \leq j \leq k$ and (ii) $b_{r,j} \in \{0,1\}$ for each example η_r in \mathcal{E}_1 and $1 \leq j \leq k$. There is a PM-DNF function with k blocks that is consistent with \mathcal{E} iff there is a feasible solution to the above set of constraints.

A PAC learning algorithm from the ILP formulation. Our PAC learning algorithm for PM-DNF functions constructs the ILP from the given set \mathcal{E} of examples and each possible value of k (the number of blocks) and outputs one such function when there is a feasible solution to the ILP.

5 Experimental Results

In this section, we evaluate the algorithms developed in the previous sections and compare the results to the derived bounds. In the case of active query, key aspects that we address are how network and local function structures affect the number of queries. We consider both the size (number of nodes as well as edge density) and structure (regular, scale-free, etc.) of the network. For local functions, we consider different numbers of blocks. For PAC learning, our focus is on the difference between the structures of the inferred partition and the true partition with respect to the number of examples sampled, example distribution, and the number of blocks. We used both synthetic and mined networks from the web for the experiments. As shown in Table 1, five mined networks and three classes of synthetic networks were used. There are five graph instances for each of the random regular (RR) and Barabási-Albert (BA) synthetic (scale-free) networks for a specified edge density.

5.1 Active query

We applied the greedy coloring algorithm in (Kosowski and Manuszewski 2004) to obtain a coloring of the square of the network. The number of colors $C(G^2)$ is shown in Table 1. In each experiment, a PM-DNF function was chosen randomly for each node. Let parameter b denote the maximum

Network	Properties			
	n	d_{ave}	Δ	$C(G^2)$
CitHep	34401	24.46	846	847
CoAstro	17903	22	504	505
Jazz	198	27.69	100	109
NRV	769	11.84	20	35
WikiVote	7115	28.32	1065	1082
Star	1001	1.998	1000	1001
RR ^a	1000	10,50,100	10,50,100	34, 367, 994
BA ^a	1000	10, 50, 100	100,261,374	111,379,780

^a 5 replicates and Δ , d_{ave} (in case of BA) and $C(G^2)$ values are approximate.

Table 1: Table of networks used in our experiments and their properties. Parameters n , d_{ave} and Δ are the number of nodes, average degree and maximum degree respectively. $C(G^2)$ is the number of colors used to color the square graph G^2 by the greedy coloring scheme.

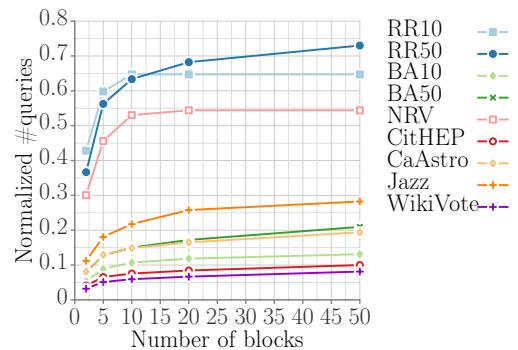


Figure 2: Performance of the active query algorithm on synthetic and mined networks. The Y-axis shows the ratio of the number of queries used to the upper bound given by Theorem 3. The number of queries is averaged over results from 100 repetitions of the experiments. The standard deviation is less than 0.01.

number of blocks possible. The local function was generated using the following iterative process. Blocks were indexed $\{1, 2, \dots, b\}$. In each iteration, the block index was cyclically incremented. We chose a node uniformly at random without replacement and assigned the block index corresponding to that iteration. For example, suppose $q = 5$ and $b = 3$. Then, there are 5 iterations and the block ID assignment happens in the following order: [1, 2, 3, 1, 2]. In the 4th iteration, for example, there are 2 inputs without block ID (since we are sampling without replacement). One of them is chosen randomly and assigned block ID 1. Given this, the algorithm was evaluated using five different values of b (namely 2, 5, 10, 20, and 50) for each network. Each experiment was repeated 100 times.

Effect of network structure. In Figure 2, we note that for a majority of the networks, the number of queries (#queries) required is $< 30\%$ of the upper bound. This is mainly due to the skewed degree distribution of most networks except for random regular networks. Consider the maximum degree of a node in each color class. For scale-free networks, most color classes have a small maximum degree; therefore,

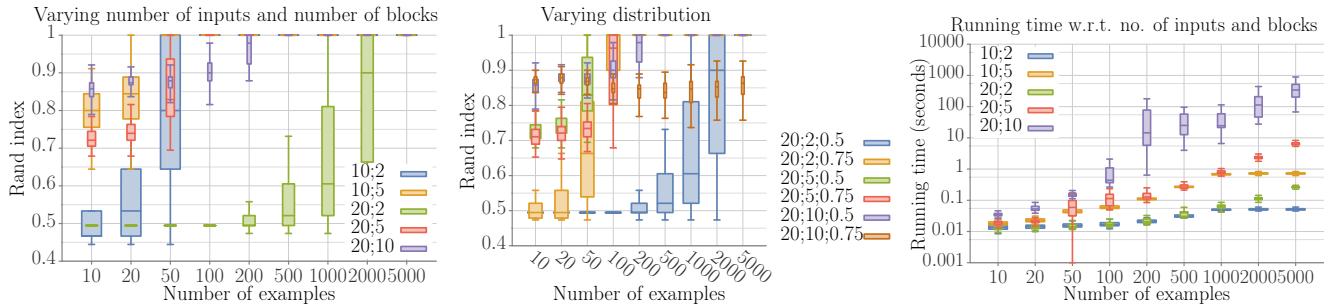


Figure 3: Performance of the ILP algorithm. Each curve in the first and third figures corresponds to $(q; k)$: the number of inputs and true size of the block. In both cases, the results are shown for uniform distribution with $p = 0.5$. The Rand index is averaged over 100 repetitions. In the second figure, each curve corresponds to a distinct (q, k, p) . In all cases, the block size parameter for the ILP was set to k .

few queries are needed to determine the local functions of vertices in these classes. Interestingly, there is no clear correlation between edge density and #queries. For example, the NRV (New River Valley Friendship) network is much smaller in size, average degree and maximum degree when compared to CoAstro (Coauthorship Astrophysics) network. Yet, when compared with the upper bound, the #queries required is much higher for the NRV network.

Effect of the number of blocks b . As the number of blocks increases, the #queries required also increases. This is because most queries are required to discover the beginning of a block. Also, when $b > \Delta$ (the maximum degree), we observe a plateau due to saturation, as all the partitions contain only blocks of size one. For RR10, RR50 and NRV, there is a sharp increase in #queries between $b = 2$ and 10 as in every color class, $b \log_2 \Delta$ additional queries are required. However, for other networks, because of the skewed degree distribution which leads to many nodes with small degrees, saturation occurs at much lower values of b .

5.2 PAC Model

In the PAC model we restricted our attention to a single local function. The objective is to evaluate the ILP-based algorithm with regard to sample distribution, number of inputs and true block size. The true PM-DNF and the inferred PM-DNF were compared using *Rand index* (Rand 1971). Rand index for two partitions X and Y of a set is $\frac{a+b}{a+b+c+d}$ where a , b , c , and d are respectively the number of pairs of elements (x, y) from the set such that x and y are in (i) the same subset in X and in Y , (ii) different subsets in X and in Y , (iii) same subset in X but different subsets in Y , and (iv) different subsets in X but same subset in Y . The examples were sampled from a uniform distribution over configurations. Each element is set to state 1 independently with the same probability p . The values of p considered were 0.25, 0.5 and 0.75. Also, we considered input sizes $q = 10$ and 20 and block sizes $k = 2, 5$ and 10. Each experiment was repeated 10 times. We assumed that the inference algorithm has knowledge of the number of blocks in the true partition. Given the number of inputs and number of blocks, we used a similar method as in the active query case to construct the PM-DNF function.

The results in Figure 3 show a rapid increase in the quality of inference with relatively small increments in the number of queries for the case of uniform distribution (Figures 3(a) and (b)). As expected, the greater the number of inputs, the greater is the number of queries required. Interestingly, unlike the active query case, fewer samples are required to infer the local function as the number of blocks increases. This can be explained as follows. Under the uniform distribution, the probability that all elements of a block of size ℓ are 1 is p^ℓ ; this gives a greater chance of a block being discovered. This is also the reason why as p increases, the chance of discovering a block is higher. However, when p is too high (as in $p = 0.75$), there is a higher chance that a block is hidden in a bigger set of nodes in every example, thus leading to a lower Rand index. Also, we note that as the number of examples is increased, there is an increase in the variance of the Rand index before it is 1 for all repetitions. This variance is higher when the number of blocks is much less than the number of inputs. The running time (Figure 3(c)) steadily increases with the number of examples and blocks.

6 Future Work

It is of interest to extend our results to other types of coalition-based functions; for example, we may require that for the function to have the value 1, at least $k \geq 2$ coalitions must be unanimous. Our complexity result for learning PM-DNF functions is for the case of proper learning where the hypothesis class and the concept class are the same. It is of interest to consider the learning problem under the representation-independent setting. Developing other learning algorithms that can scale to large networks is another direction for future work.

Acknowledgments. We thank the reviewers for carefully reading the manuscript and providing very helpful comments. This work has been partially supported by DTRA CNIMS (Contract HDTRA1-11-D-0016-0001), NSF Grant IIS-1908530, NSF Grant OAC-1916805, NSF CRISP 2.0 Grant 1832587, NSF DIBBS Grant ACI-1443054, NSF BIG DATA Grant IIS-1633028, and NSF EAGER Grant CMMI-1745207. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

References

- Abasi, H.; Bshouty, N. H.; and Mazzawi, H. 2014. On exact learning monotone DNF from membership queries. *CoRR* abs/1405.0792:1–16.
- Adiga, A.; Kuhlman, C. J.; Marathe, M. V.; Ravi, S. S.; Rosenkrantz, D. J.; and Stearns, R. E. 2018. Learning the behavior of a dynamical system via a “20 questions” approach. In *Thirty second AAAI Conference on Artificial Intelligence*, 4630–4637.
- Adiga, A.; Kuhlman, C. J.; Marathe, M.; Ravi, S. S.; Rosenkrantz, D. J.; Stearns, R. E.; and Vullikanti, A. 2019a. Learning coalition-based interactions in networked social systems. Technical report, Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, VA.
- Adiga, A.; Kuhlman, C. J.; Marathe, M.; Ravi, S. S.; and Vullikanti, A. 2019b. PAC learnability of node functions in networked dynamical systems. In *Proc. ICML 2019*, 82–91.
- Angluin, D., and Slonim, D. K. 1994. Randomly fallible teachers: Learning monotone DNF with an incomplete membership oracle. *Machine Learning* 14(1):7–26.
- Antony, M., and Biggs, N. 1992. *Computational Learning Theory*. Cambridge, UK: Cambridge University Press.
- Barrett, C. L.; Hunt, H. B.; Marathe, M. V.; Ravi, S.; Rosenkrantz, D. J.; and Stearns, R. E. 2006. Complexity of reachability problems for finite discrete dynamical systems. *Journal of Computer and System Sciences* 72(8):1317–1345.
- Blumer, A.; Ehrenfeucht, A.; Haussler, D.; and Warmuth, M. K. 1989. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM (JACM)* 36(4):929–965.
- Branzei, R.; Dimitrov, D.; and Tijs, S. 2005. *Models in cooperative game theory*. Springer.
- Castro, R., and Nowak, R. 2007. Active learning and sampling. In *Proc. Foundations and Applications of Sensor Management*, 177–200.
- Centola, D., and Macy, M. 2007. Complex contagions and the weakness of long ties. *American Journal of Sociology* 113(3):702–734.
- Colón-Reyes, O.; Jarrah, A. S.; Laubenbacher, R. C.; and Sturmfels, B. 2006. Monomial dynamical systems over finite fields. *Complex Systems* 16(4).
- Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; and Stein, C. 2009. *Introduction to Algorithms*. Cambridge, MA: MIT Press and McGraw-Hill, Second edition.
- Garey, M. R., and Johnson, D. S. 1979. *Computers and Intractability: A Guide to the Theory of NP-completeness*. San Francisco: W. H. Freeman & Co.
- González-Bailón, S.; Borge-Holthoefer, J.; Rivero, A.; and Moreno, Y. 2011. The dynamics of protest recruitment through an online network. *Scientific Reports* 1:7 pages.
- Granovetter, M. 1978. Threshold models of collective behavior. *American Journal of Sociology* 1420–1443.
- Haussler, D. 1988. Quantifying inductive bias: AI learning algorithms and valiant’s learning framework. *Artificial intelligence* 36(2):177–221.
- He, X.; Xu, K.; Kempe, D.; and Liu, Y. 2016. Learning influence functions from incomplete observations. In *Advances in Neural Information Processing Systems*, 2073–2081.
- Kearns, M. J., and Vazirani, V. V. 1994. *An Introduction to Computational Learning Theory*. Cambridge, MA: MIT Press.
- Kearns, M.; Mansour, Y.; Ron, D.; Rubinfeld, R.; Schapire, R.; and Sellie, L. 1994. On the learnability of discrete distributions. In *Proc. ACM STOC*, 273–282.
- Kosowski, A., and Manuszewski, K. 2004. Classical coloring of graphs. *Contemporary Mathematics* 352:1–20.
- Laubenbacher, R., and Stigler, B. 2004. A computational algebra approach to the reverse engineering of gene regulatory networks. *J. Theoretical Biology* 229:523–537.
- Liśkiewicz, M.; Lutter, M.; and Reischuk, R. 2017. Proper learning of k-term DNF formulas from satisfying assignments. *Electronic Colloquium on Computational Complexity (ECCC)* 24:114.
- Narasimhan, H.; Parkes, D. C.; and Singer, Y. 2015. Learnability of influence in networks. In *Advances in Neural Information Processing Systems*, 3186–3194.
- Rand, W. M. 1971. Objective criteria for the evaluation of clustering methods. *J. American Statistical Association* 66(336):846–850.
- Romero, D.; Meeder, B.; and Kleinberg, J. 2011. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web*, 695–704. ACM.
- Servedio, R. A. 2004. On learning monotone DNF under product distributions. *Inf. Comput.* 193(1):57–74.
- Ugander, J.; Backstrom, L.; Marlow, C.; and Kleinberg, J. 2012. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences* 109(16):5962–5966.
- Valiant, L. G. 1984. A theory of the learnable. *Communications of the ACM* 18(11):1134–1142.
- Warmuth, M. K. 1989. Towards representation independence in PAC learning. In *Proc. International Workshop on Analogical and Inductive Inference (AII’89)*, 78–103.
- Zhu, J.; Zhu, J.; Ghosh, S.; Wu, W.; and Yuan, J. 2018. Social influence maximization in hypergraph in social networks. *IEEE Transactions on Network Science and Engineering*.

VALIDATING AGENT-BASED MODELS OF LARGE NETWORKED SYSTEMS

Abhijin Adiga Chris Barrett Stephen Eubank Chris J. Kuhlman
Madhav V. Marathe Henning Mortveit S. S. Ravi
Daniel J. Rosenkrantz Richard E. Stearns Samarth Swarup Anil Vullikanti

Biocomplexity Institute and Initiative
University of Virginia
Charlottesville, VA 22904, USA

ABSTRACT

The paper describes a systematic approach for validating real-world biological, information, social and technical (BIST) networks. BIST systems are usually represented using agent-based models and computer simulations are used to study their dynamical (state-space) properties. Here, we use a formal representation called a graph dynamical system (GDS). We present two types of results. First we describe two real-world validation studies spanning a variety of BIST networks. Various types of validation are considered and unique challenges presented by each domain are discussed. Each system is represented using the GDS formalism. This illustrates the power of the formalism and enables a unified approach for validation. We complement the case studies by presenting new theoretical results on validating BIST systems represented as GDSs. These theoretical results delineate computationally intractable and efficiently solvable versions of validation problems.

1 INTRODUCTION

Validation and verification (V&V) are long sought after goals of every useful model (Forrester and Senge 1980; Robinson 1997; Kleijnen 1995; National Research Council 2008; Oreskes 2000; Oberkampf and Trucano 2002; Yilmaz 2006; Bharathy and Silverman 2013; Carley 1996). These two concepts can be easily defined. Model verification involves the assessment of the fact that the model implemented is actually what one intended to implement. Model validation is the assessment that the model, up to some measure of comparison, mimics the system it was intended to capture. It should be noted that a model is typically validated to some degree of fidelity or intended use (aka purpose).

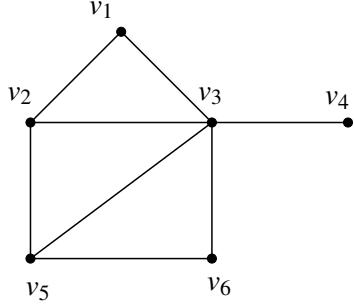
BIST networks and agent-based models. In this paper, we focus on V&V issues as they pertain to networked biological, social, information and technical networks, also referred to as **BIST networks**. These systems lack inherent symmetries and are highly heterogeneous. Individual agents (represented as nodes in the network) are often selfish or adversarial and their behavior co-evolves in response to their perception of system dynamics. As a result, traditional methods rooted in physics and based on predictive validity are often not applicable. For one, in large simulations involving socio-technical networks, data matching exercises are usually postdictions of historical information such as matching epidemiological model output to an infection time series of a flu season. Although this is useful, it can also be misleading and is often inadequate. The configuration space (which captures the causal structure for an evolving system) is important for understanding the system being represented as well as the modeled representation of that system. However, any measured real world data is incapable of capturing this structural range—only those modes that took place in the real world appear in the measured data. Furthermore, this space is extremely large and is not enumerable in practice. Thus, the process of postdiction (retrospective validation) alone is inadequate. Additionally, while postdicting, the available information about the context is often insufficient to properly specify the initial and structural conditions that would allow the model to predict. As a result, high dimensional models are often fitted to relatively sparse data. In this sense, the occurrence of a fit can

be misleading because the inverse problem (i.e., the model matching the real data) does not have a unique solution. Predictive validity is also useful but again of limited value. Predictions based on past behavior that do not account for adaptive behavioral changes of individual agents often do not do well. Moreover, prediction alone does not provide insights into the underlying causal processes. For e.g., a simple time series model might provide accurate predictions of influenza dynamics but might have very little to say why a given season might have fewer cases than past few seasons. *In both cases, explanatory power is really at issue.* In decision-making, a causal basis for the choice of the best option is more relevant than any particular kind of detailed prediction of state. But, simply matching the model output with data collected in the field is of limited value, especially because such data is necessarily sparse, noisy, incomplete and not aligned in time.

Here we focus on V&V of agent-based models—such models have become popular for representing complex real-world BIST networks. Informally, such models are comprised of a collection of agents that represent the underlying BIST system at a certain level of granularity. Agents interact with other agents and modify their behavior as a result of the interactions. Usually the interaction is constrained by an underlying network that captures the agent neighborhood. Agent-based models are very expressive but come with two costs. First, computing all dynamical outcomes (the state-space) as specified by the agent-based models is usually expensive and as such one resorts to simulations. Second, the richness of agents and interaction structure is a strength but also leads to a more critical questioning and interrogation of such models as they pertain to believability and understandability. As a result, validation of agent-based models for BIST networks is even more important, and not surprisingly a subject of several papers (e.g., (Yilmaz 2006; Bharathy and Silverman 2013; Carley 1996)) in the last decade.

External and internal validation. Following (Carley 1996), we will use the term *real data* to mean information (including nominal and procedural information) gathered about the real system that is being modeled via experimental, field, survey or archival analysis. In this paper, we will focus on two distinct types of validation—internal and external. See (Yilmaz 2006; Bharathy and Silverman 2013; Carley 1996) for in depth discussions on various forms of validation. *External validation* aims to compare model output data with real life, in-situ and in-vivo measurements where the state-space data produced by the model are *matched* with measured data. Two key forms are: retrospective validity (matching with already measured data as used in machine learning for example) and predictive validity (matching with the outcomes that occur in the future). Furthermore, Carley (1996) classifies various levels of external validity, including: *face, parameter, process, pattern, point and distributional* validity. External validity connects model outcomes with observations pertaining to the real world problem that is being modeled. *Internal (aka structural) validation* aims to ensure that the model has been put together *correctly*. For agent-based models, this implies that interaction patterns (or networks), individual processes or rules for agents and model parameters are correct and adequate; furthermore, one ensures that the model is consistent with the specific and prevalent physical and social theories. In this sense, verification can be thought of as internal validation.

A mathematical framework. Here we use graph dynamical systems (GDSs) as a mathematical abstraction for agent-based models of BIST networks. Graph dynamical systems provide a powerful abstraction for a large class of BIST networks—these include transportation systems (Barrett et al. 2001), spread of epidemics (Eubank et al. 2004), systems biology (Shmulevich and Kauffman 2004), and immune systems (Alam et al. 2015). Such systems consist of a large number of interacting entities/agents, and the complex global dynamics of the system are the result of local interactions of each agent with its neighbors in an interaction structure, represented as a network. In this paper, we study a special form of GDSs, namely synchronous GDSs, where all agents compute and update their states in parallel. Such GDSs have been successfully used in a number of applications (Barrett et al. 2006). GDSs are *universal*, in that they can represent all other computing models, such as cellular automata (Wolfram 1987), Boolean networks (Ribeiro et al. 2007), neural networks, Hopfield networks, and graph automata; see (Barrett et al. 2007) for details. We will present the formal definitions associated with GDSs in Section 2.



Initial Config.:	(0, 0, 1, 0, 0, 0)
Config. at time 1:	(1, 0, 0, 0, 1, 1)
Config. at time 2:	(1, 1, 1, 0, 1, 1)

Note: The local function at each of the nodes v_1 , v_5 and v_6 is the OR function. The local functions at v_2 , v_3 and v_4 are 2-threshold functions whose value is 1 iff at least two of the inputs are 1. Each configuration has the form $(s_1^t, s_2^t, s_3^t, s_4^t, s_5^t, s_6^t)$, where s_i^t is the state of node v_i at time t , $1 \leq i \leq 6$. The configuration at time 2 is a fixed point.

Figure 1: An example of a GDS.

2 Graph Dynamical Systems: Formal Definitions

We now define a formal model that enables us to develop rigorous formulations of issues related to validation of BIST systems. We follow (Barrett et al. 2006) in presenting these definitions. Let \mathbb{B} denote the Boolean domain $\{0,1\}$. A **synchronous Graph Dynamical System** (GDS) \mathbb{S} over \mathbb{B} is specified as a pair $\mathbb{S} = (G, \mathbb{F})$, where (a) $G(V, E)$, an undirected graph with $n = |V|$, represents the underlying graph of the GDS, with node set V and edge set E , and (b) $\mathbb{F} = \{f_1, f_2, \dots, f_n\}$ is a collection of functions in the system, with f_i denoting the **local function** associated with node v_i , $1 \leq i \leq n$.

Each node (or agent) of G has a state value from \mathbb{B} . Each function f_i specifies the local interaction between node v_i and its neighbors in G . The inputs to function f_i are the state of v_i and those of the neighbors of v_i in G ; function f_i maps each combination of inputs to a value in \mathbb{B} . This value becomes the next state of node v_i .

At any time t , the **configuration** \mathbb{C} of a GDS is the n -vector $(s_1^t, s_2^t, \dots, s_n^t)$, where $s_i^t \in \mathbb{B}$ is the state of node v_i at time t ($1 \leq i \leq n$). In a GDS, all nodes compute and update their next state *synchronously*.

Example: Consider the graph of a GDS shown in Figure 1. Suppose the local transition functions at each of the nodes v_1 , v_5 and v_6 is the OR function. The function at each of the nodes v_2 , v_3 and v_4 is the 2-threshold function whose value is 1 iff at least two of the inputs are 1. Assume that initially, v_3 is in state 1 and all other nodes are in state 0. During the first time step, the states of nodes v_1 , v_5 and v_6 change to 1 since each f_i ($i \in \{1, 5, 6\}$) is the OR function and each of these nodes has a neighbor (namely, v_3) in state 1. Also, the state of v_3 changes to 0 since it requires at least two inputs with value 1. The states of v_2 and v_4 do not change; they continue to be 0. During time step 2, v_2 and v_3 change to 1 but v_4 remains at 0. Once the system reaches the configuration $\mathbb{C} = (1, 1, 1, 0, 1, 1)$ at time step 2, it remains in that configuration forever; that is, \mathbb{C} is a **fixed point** for this system. \square

Additional dynamical systems terminology and notation. If a given GDS can transition in one step from a configuration \mathbb{C}' to a configuration \mathbb{C} , then \mathbb{C} is a **successor** of \mathbb{C}' and \mathbb{C}' is a **predecessor** of \mathbb{C} . Since our local functions are deterministic, each configuration has a unique successor; however, a configuration may have zero or more predecessors. As mentioned above, a **fixed point** is a configuration \mathbb{C} for which the successor is \mathbb{C} itself. A configuration with no predecessors is called a **Garden of Eden** (GE) configuration.

The **phase space** $\mathbb{P}_{\mathbb{S}}$ of a GDS \mathbb{S} is a directed graph defined as follows. There is a node in $\mathbb{P}_{\mathbb{S}}$ for each configuration of \mathbb{S} . There is a directed edge from a node representing configuration \mathbb{C} to that representing configuration \mathbb{C}' if there is a one step transition of \mathbb{S} from \mathbb{C} to \mathbb{C}' . For a GDS with n nodes and a Boolean node state set (here, $\{0, 1\}$), the number of nodes in the phase space is 2^n ; thus, the size of phase space

is *exponential* in the size of a GDS. Each node in the phase space has an outdegree of 1 (since our GDS model is deterministic). Also, in the phase space, each fixed point of a GDS is a self-loop and each GE configuration is a node of indegree zero. We use the symbol **0** (**1**) to denote a configuration in which every node is in state 0 (state 1).

Some terminology regarding Boolean functions. Throughout this paper, we will consider several classes of Boolean functions. We now define these classes. Given an integer k , a Boolean function f is a **k -threshold** function iff f is true when at least k of its inputs are 1. A **symmetric** Boolean function (Crama and Hammer 2011) is one whose value does not depend on the order in which the input bits are specified; that is, the function value depends only on how many of its inputs are 1. Thus, any k -threshold function is symmetric. A symmetric function with q inputs can be specified using a table with $q + 1$ rows, where row i of the table specifies the value of the function when the number of 1-valued inputs is equal to i , $0 \leq i \leq q$. A Boolean function f is **r -symmetric** (Barrett et al. 2007) if the inputs to f can be partitioned into at most r classes such that the value of f depends only on how many of the inputs in each of the r classes are 1. It can be seen that an r -symmetric function with $q \geq r$ inputs can be represented by a table with $O(q^r)$ entries. Note that any Boolean function with d inputs is d -symmetric. We will be mainly concerned with r -symmetric functions where r is a *fixed* integer. We say that a *GDS* is r -symmetric if each of its local functions is r' -symmetric for some $r' \leq r$.

We will explain in Section 5.1 how the GDS model can be used to formalize some research issues regarding validation of BIST networks.

3 SUMMARY OF CONTRIBUTIONS & RELATED WORK

Summary of contributions. In this paper, we describe an approach taken by our group to address issues as they pertain to validation of agent-based models of BIST networks. We present computational studies as well as theoretical results.

(1) *Example case studies.* In Section 4, two case studies spanning a range of BIST networks are presented. Each study illustrates complementary types of validations as necessitated by the application. All of them use the GDS framework to represent the underlying agent-based models.

(2) *Theoretical results.* We complement the discussion with new theoretical results in Section 5. Here, validation problems are formulated as model checking problems over GDSs and rigorous computational bounds are established. Each result compares properties of two BIST networks, represented as GDSs. One system can be considered as *ground truth* and the other as an *inferred* system built from data, and the goal is to evaluate the quality of the inferred system. An important practical consequence of these problem formulations and results is that if the similarity conditions *do not* hold between the two GDSs, then the inferred model does not fully capture the ground truth. Due to space considerations, only deterministic systems are considered for establishing rigorous results. These results are relevant for the following reasons. First, the results are applicable to real systems. The networked experiments of Centola (2010), Centola (2011), for example, seek to infer thresholds for humans in a health care setting. There are rigorous GDS representations for both the threshold model of Granovetter (1978) and the results of these aforementioned experiments that can be compared within our theoretical framework. Second, the results help us to delineate validation questions that are computationally intractable from those that can be solved efficiently. Finally, some of the results identify validation questions that can be answered in practice using open-source software SAT solvers (see, e.g., Gomes et al. (2008)).

Related work. Validation of complex systems has been a subject of extensive research in the modeling and simulation community; see (Balci and Nance 1985; Forrester 1971; Forrester and Senge 1980; National Research Council 1998; National Research Council 2008; Oberkampf and Trucano 2002; Robinson 1997; Sterman 2006). Philosophical discussions on these topics can be found in (Oreskes 2000; Carnap 1936; Popper 1963). The debate on validation can be summarized eloquently by a quote due to G. E. P. Box:

“Essentially, all models are wrong, but some are useful.” (Box and Draper 1987). As discussed above and in the recent reports of National Academies (National Research Council 1998; National Research Council 2008), validation is only meaningful when one states the purpose for which the model is used. Validation of agent-based models of BIST systems has also been actively studied. This includes our theoretical as well as practical work (Barrett et al. 2001; Eubank et al. 2004; Barrett et al. 2006; Barrett et al. 2007) as well as work done by others (Hahn 2013; Hahn 2017; Yilmaz 2006; Klügl 2008; Bruch and Atwell 2015; Bharathy and Silverman 2013; Bianchi et al. 2007; Lux and Zwinkels 2018; Macal 2016; Carley 1996). Additional references can be found in the above papers.

4 CASE STUDIES FOR MODEL VALIDATION

In this section, we describe two case studies to illustrate approaches for validating models of BIST networks. The first study develops models to capture human behaviors in a networked game. The second study models commodity flows among produce markets. The meanings of interactions vary from purposeful two-way communication, to crop flows, to unaided (i.e., human-oblivious) transmission, to passive observations of others’ actions. In all cases, the goal is to develop validated models to quantitatively predict system behavior for a wider range of conditions than were studied in experiments or obtained through observations.

4.1 Case Study 1: Validation of a Word Construction Model

An online game platform was designed and constructed for playing a word construction game. (Word construction games are often called *anagram* games. The board game Scrabble is one such example.) In our particular version of the game, (remote) players are recruited through Amazon Mechanical Turk (AMT) and play the game through their web browsers. Our game is motivated by the work of Charness et al. (2014). Each player $v_i \in V$ ($n = |V|$) is situated in a network $G(V, E)$ of interacting players, where V is the set of players and E is the set of edges or communication channels $\{v_i, v_j\} \in E$ between players $v_i, v_j \in V$. Each player is assigned an initial set of alphabetic letters (called her *own letters*). Players work cooperatively to form words from their own letters and from those letters that they request, and then receive, from their distance-1 neighbors. In this way, players assist each other in forming more words, as a team, than individuals could produce on their own. A word corpus determines the validity of each word formed and submitted by players. The goal of the game is for the team to form as many words as possible over a 5-minute time window.

Observed data gathered from the games performed on this platform are the following actions and the time of each action by each player: (i) submit word; (ii) send letter request; (iii) receive letter request; (iv) send letter reply; and (v) receive letter reply. From these data, we built a set of *action type-time sequence* (ATTS) models that predict the time-ordered sequence of pairs (action type, time). Here, action types are three of the five actions identified immediately above: submit word, send letter request, and send letter reply. The ATTS model is the local function f_i for each $v_i \in V$ (Section 2), and uses multinomial logistic regression that predicts each of the three action types, in time, for a player; see (Ren et al. 2018).

To evaluate our models, we use Kullback-Leibler (KL) divergence in the following way. We use initial conditions for each experiment of a particular class (e.g., the class of experiments where each player has degree $d = 2$ in the network, i.e., each player has two neighbors). We run the (agent-based) ATTS model, for each node, in time to simulate each game in the fashion shown in Figure 1. Thereafter, for each player, we sum up the counts of each of the five action types across all players in the computational games of a particular class. We form frequency distributions from these data—one for each action type—and compare these predicted distributions (which we convert to density distributions) against corresponding distributions from experimental data, using KL-divergence.

Figure 2 summarizes these ideas and some results for two models, M0 and M1, where M0 is based on a simple state transition matrix determined directly from the experimental data and M1 is based on a multinomial logistic regression model (Ren et al. 2018). The first two plots show results from experiments and from model predictions for $d = 2$ experiments, for the distributions of numbers of (letter) replies received and numbers of words formed. One can see by inspection that model M1 provides predictions (in distribution) that are in closer agreement with the data than are those from baseline model M0. This observation is made more formal with the KL-divergence values in Figure 2c. The lesser values of KL divergence for model M1 indeed confirm that this model is better than model M0, for not only replies received and form words, but also for the other three action types.

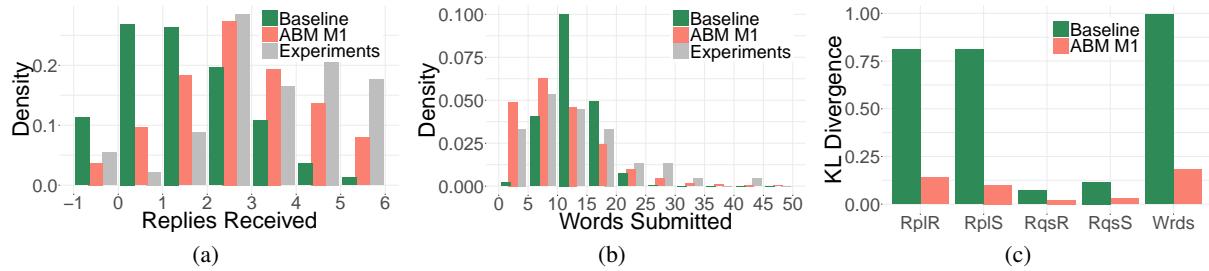


Figure 2: (a) Distributions of numbers of replies received by players from three sources: experimental data (gray), baseline model M0 (green), and regression model M1 (red). (b) Distributions of numbers of words formed by players from three sources: experimental data (gray), baseline model M0 (green), and regression model M1 (red). In these first two plots, the data come from summing the counts of the number of players that take the given number of actions of the specified type over the game duration. (c) KL-divergence values that are computed from the distributions in the first two plots and other similar plots for the remaining action types. The along the x-axis, are respectively, replies received, replies sent, requests received, requests sent, and words formed. These results come from Ren et al. (2018).

These results provide a combination of data and structural validation. Data validation is achieved through the favorable comparisons between model predictions and experimental results. At the same time, structural validation is achieved because our hierarchical logistic regression model (local function) generates the predictions. Note that structural validation does not preclude the possibility of other satisfactory formulations for local functions. Finally, these comparisons are self-consistency checks in that all data are used to build the models, and the predictions in Figure 2 are also over all data. Ten-fold cross validation is work in progress.

4.2 Case Study 2: Validation of a Model for the Spread of an Invasive Species

While trade and transport of goods is widely accepted as a primary pathway for the dispersal of invasive species that affect agricultural crops, these human mediated pathways are not well understood. Few, and often inaccurate, incidence reports and lack of knowledge of trade flows are some of the major hurdles in understanding their role. We modeled realistic spatio-temporal networks of seasonal agro-products between major markets (Venkatramanan et al. 2017) from diverse, multi-type, and noisy datasets. The methodology was applied to develop a spatio-temporal domestic tomato trade network in Nepal and investigate its role in the spread of the South American tomato leafminer or *Tuta absoluta*, a devastating pest of the tomato crop (Campos et al. 2017). Through dynamical analysis of the networks and a novel rank-based Bayesian inference approach, we showed that tomato trade has facilitated the rapid spread of the pest in the region.

We modeled the flow of agricultural produce among markets based on the following premise: major wholesale markets serve as key locations facilitating agricultural commodity flow, and the total outflow from a market depends on the amount of produce in its surrounding regions, and the total inflow is a function of the consumption linked to the market. These assumptions are driven by studies of the pest dynamics in other countries and the fact that tomato is a commercial crop in Nepal. The commodity flow is modeled as a temporal network with markets as nodes and directed weighted edges representing volumes of host crop being traded between the end points. The flows are estimated using a doubly constrained gravity model (Kaluza et al. 2010). The flow F_{ij} from location i to j is given by $F_{ij} = a_i b_j O_i I_j f(d_{ij})$ where, O_i is the total outflow of the commodity from i , I_j is the total inflow to j , d_{ij} is the time duration required to travel from i to j , $f(\cdot)$ is the *distance deterrence function*, and coefficients a_i and b_j are computed through an iterative process to ensure flow balance.

For **structural validation**, we used yearly data from the largest wholesale market of Nepal. In Figures 3a–3c, we compare this data with the network flows. Given a set of network parameters (β, κ, γ) , we obtained the inflow from a particular district to Kathmandu as follows: We combined the weights of all edges of the corresponding network with destination node “Kathmandu” and source nodes belonging to that district. As seen in Figure 3a, for γ values between 0.5 and 1, the flows from the networks are comparable to the Kalimati data except for two districts: Dhading (the top contributor) and Sarlahi (third highest). Upon further investigation we find that Dhading, which is a major producer west of Kathmandu, serves the Mid Hills and Terai regions of the Central Development Region in the flow networks (Figure 3b). While the gravity model predicts that these flows will be directly delivered to these regions, in reality, it is possible that Dhading’s produce is routed through the Kalimati market as there are several traders from Dhading registered in the Kalimati market. A similar argument holds for Sarlahi (Figure 3c).

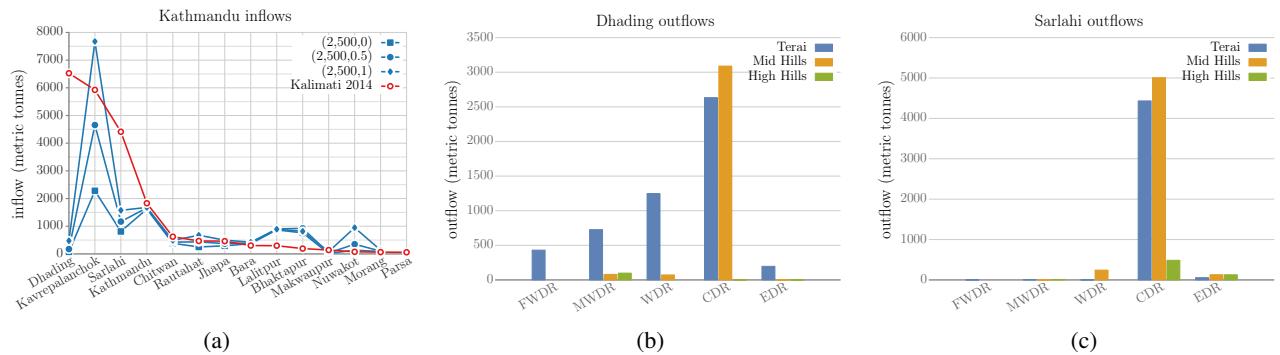


Figure 3: Flow validation (from (Venkatramanan et al. 2017)).

5 FORMALIZING VALIDATION USING THE GDS FRAMEWORK

In this section, we define some similarity relationships between a pair of GDSs to formalize some questions regarding validation. We then present complexity and algorithmic results for testing these relationships.

5.1 Examples of Equivalence and Similarity Relations Between a Pair of GDSs

As mentioned earlier, we will use equivalence and similarity relationships between pairs of GDSs to formulate validation questions for models of BIST networks. Our definitions rely on the following assumptions:

- (i) The underlying graphs of the two GDSs have the same set of nodes. (The set of edges for the two GDSs

may be different.) (ii) The nodes of each GDS are labeled using integers from 1 to n so that the corresponding nodes in the two GDSs have the same number. (iii) The domain of state values for both the GDSs is $\{0, 1\}$. (iv) Each configuration in the phase space of one GDS corresponds to the same configuration in the phase space of the other GDS. Given two GDSs \mathbb{S}_1 and \mathbb{S}_2 satisfying the above assumptions, it is possible to formally represent some validation questions for the corresponding multi-agent systems using suitable similarity relationships between the two GDSs, expressed as logical **predicates**. We now present two examples of such predicates. (Several other examples are presented in (Adiga et al. 2019).) In Section 5.2, we discuss how such predicates are useful in studying validation issues for BIST networks.

- (a) Let the predicate $PSE(\mathbb{S}_1, \mathbb{S}_2)$ be true iff the phase spaces of \mathbb{S}_1 and \mathbb{S}_2 are *identical* (i.e., for every configuration C , the successor of C is the same configuration in both \mathbb{S}_1 and \mathbb{S}_2). We will refer to predicate PSE as the **phase space equivalence** relationship.
- (b) Let the predicate $CS-FP(\mathbb{S}_1, \mathbb{S}_2)$ be true iff the two GDSs have the *same number of fixed points*. A similar predicate $CS-GE(\mathbb{S}_1, \mathbb{S}_2)$ can be formulated for GE configurations. We will refer to $CS-FP$ and $CS-GE$ respectively as **count similarity** (CS) relationship with respect to fixed point and GE configurations.

5.2 Using Similarity Predicates for Validation

The literature on system verification (see, e.g., (Alur et al. 2015)), provides many examples where one wants to compare a model (e.g., a finite state machine) obtained from the specification of a system with one obtained from an implementation. The purpose of the comparison is to verify whether the implemented model correctly represents the specification model. This comparison is carried out by defining various forms of relationships between two models and determining whether those relationships hold. In this paper, we use a similar approach for validation. We use GDSs as models of both the ground truth and the implementation. As mentioned in Section 3, several studies reported in the literature (e.g., Centola (2010), Centola (2011)) use formalisms similar to GDSs. The relationships identified in Section 5.1 provide some ways by which the constructed model can be compared with the model obtained from ground truth. The most stringent of these relationships is phase space equivalence; when this relationship holds, the two systems become indistinguishable with respect to the desired behavior. Some of the relationships are likely to be easier to test in practice, especially due to the availability of tools such as public domain SAT solvers. When a relationship *does not* hold, it stands to reason that the inferred model does not fully capture the properties of the model derived from ground truth. When this happens, one can try to infer a better model using additional data and then proceed to validate the new model. In this manner, the similarity relationships identified in this paper serve as useful ways of carrying out validation.

5.3 Results for Phase Space Equivalence

Here and in the subsequent subsections, we consider the problems of determining the truth value of the predicates defined in Section 5.1. See (Adiga et al. 2019) for detailed proofs.

Recall that the predicate $PSE(\mathbb{S}_1, \mathbb{S}_2)$ is true iff the phase spaces of \mathbb{S}_1 and \mathbb{S}_2 are identical. We start with a general result on the complexity of the problem.

Theorem 1 Given two GDSs \mathbb{S}_1 and \mathbb{S}_2 whose local functions are specified as Boolean expressions, the problem of finding the truth value of the predicate $PSE(\mathbb{S}_1, \mathbb{S}_2)$ is **NP-hard**. This result holds even if the underlying graphs of the two GDSs are identical.

Proof (sketch): We use a reduction from the Boolean Satisfiability problem (SAT) which is known to be **NP-hard** (Garey and Johnson 1979). Let an instance I of SAT be specified using p variables $X =$

$\{x_1, x_2, \dots, x_p\}$ and m clauses $Y = \{Y_1, Y_2, \dots, Y_m\}$. We construct two GDSs \mathbb{S}_1 and \mathbb{S}_2 as follows. For both the GDSs, the underlying graph $G(V, E)$ is a star graph on $p + 1$ nodes, with $V = \{v_0, v_1, \dots, v_p\}$ and $E = \{\{v_0, v_i\} : 1 \leq i \leq p\}$. (Thus, the underlying graphs of the two GDSs are identical.) Let s_i^1 and f_i^1 denote respectively the state value of node v_i and the local function at node v_i , $0 \leq i \leq p$, for GDS \mathbb{S}_1 . Likewise, let s_i^2 and f_i^2 denote respectively the state value of node v_i and the local function at node v_i , $0 \leq i \leq p$, for GDS \mathbb{S}_2 . For $1 \leq i \leq p$ and $\ell = 1, 2$, the local function f_i^ℓ at node v_i is given by $f_i^\ell(s_0^\ell, s_i^\ell) = 0$. For \mathbb{S}_1 , the local function $f_0^1(s_0^1, s_1^1, \dots, s_p^1) = 0$ and for \mathbb{S}_2 , the local function $f_0^2(s_0^2, s_1^2, \dots, s_p^2) = \bigwedge_{j=1}^m Y_j$, with each variable x_i in Y_j replaced by the state variable s_i^2 , $0 \leq i \leq p$ and $1 \leq j \leq m$. It is easy to see that this construction can be done in polynomial time. It can be verified that the predicate $\text{PSE}(\mathbb{S}_1, \mathbb{S}_2)$ is true iff the instance I is *not* satisfiable. ■

We now mention a result that identifies some special cases for which the $\text{PSE}(\mathbb{S}_1, \mathbb{S}_2)$ predicate can be evaluated in polynomial time. A proof of the result appears in (Adiga et al. 2019).

Theorem 2 The predicate $\text{PSE}(\mathbb{S}_1, \mathbb{S}_2)$ can be evaluated in polynomial time for the following special cases: (i) the local functions for both the GDSs are r -symmetric for a fixed r and (ii) each local function in both the GDSs is linear (i.e., it is from the set {XOR, XNOR}).

When the degree of each node is bounded by a constant d , we note that every local function is d -symmetric. Thus, we obtain the following corollary from Theorem 2.

Corollary 3 The predicate $\text{PSE}(\mathbb{S}_1, \mathbb{S}_2)$ can be evaluated in polynomial time if each of the underlying graphs has a degree of at most d , for a fixed d . ■

5.4 Results for Count Similarity

Here, we consider the count similarity relationships introduced earlier. Recall that the predicate $\text{CS-FP}(\mathbb{S}_1, \mathbb{S}_2)$ is true iff the two GDSs \mathbb{S}_1 and \mathbb{S}_2 have the same number of fixed points. In proving the next theorem, we will use the following observation.

Observation 1 Any GDS where each local function is NOR does not have any fixed point. ■

Theorem 4 Let \mathbb{S}_1 and \mathbb{S}_2 be two given GDSs. The problem of evaluating the predicate $\text{CS-FP}(\mathbb{S}_1, \mathbb{S}_2)$ is **NP-hard**.

Proof: We will prove the result using a reduction from the *fixed point existence* (FPE) problem: does a given GDS \mathbb{S} have a fixed point? This problem was shown to be **NP**-complete in (Barrett et al. 2001). Let an instance of FPE be given by the GDS \mathbb{S}_1 . Let \mathbb{S}_2 be the following GDS: the underlying graph of \mathbb{S}_2 is the same as that of \mathbb{S}_1 and the local function at each node of \mathbb{S}_2 is the NOR function. By Observation 1, \mathbb{S}_2 does not have any fixed point. Therefore, the predicate $\text{CS-FP}(\mathbb{S}_1, \mathbb{S}_2)$ is false iff \mathbb{S}_1 has a fixed point. Thus, the problem of evaluating $\text{CS-FP}(\mathbb{S}_1, \mathbb{S}_2)$ is **NP-hard**. ■

We now present a result to show that the count similarity predicate can be evaluated efficiently for special cases.

Proposition 1 Let \mathbb{S}_1 and \mathbb{S}_2 be two GDSs such that their underlying graphs are treewidth bounded¹ and the local functions are r -symmetric for some fixed r . Then the predicate $\text{CS-FP}(\mathbb{S}_1, \mathbb{S}_2)$ can be evaluated in polynomial time.

Proof: It is shown in (Rosenkrantz et al. 2015) that if a GDS \mathbb{S} has bounded treewidth and each local function of \mathbb{S} is r -symmetric for a fixed r , the problem of counting the number of fixed points of \mathbb{S} can be

¹For the definition of treewidth, we refer the reader to (Bodlaender 1993).

solved in polynomial time. Using that algorithm, one can compute the number of fixed points of \mathbb{S}_1 and \mathbb{S}_2 and thus evaluate the predicate CS-FP($\mathbb{S}_1, \mathbb{S}_2$) in polynomial time. ■

6 SUMMARY, CONCLUSIONS AND FUTURE WORK

We observed that many V&V questions for MASs can be expressed using certain predicates on the GDSs corresponding to the MASs. We showed that many of these problems are computationally intractable in general. However, we observed that restricted versions of the problems can be solved efficiently. Also, some of the problems can be solved in practice using commercial or public domain SAT solvers. When the phase spaces of the GDSs modeling the given MASs are identical, the two MASs have identical global behavior. The other types of relationships defined in Section 5.1 provide useful ways of identifying certain *differences* between the MASs under consideration. One can identify and study many other relationship predicates. Further, it is also of interest to study such questions for stochastic GDSs (Barrett et al. 2011).

ACKNOWLEDGMENTS. This work is supported by DTRA CNIMS (Contract HDTRA1-11-D-0016-0001), DTRA V&V Grant Award No. HDTRA1-11-1-0016 (entitled “Rigorous Approaches for Validation and Verification of Networked Systems”), NSF DIBBS Grant ACI-1443054, NSF EAGER Grant CMMI-1745207, NSF BIG DATA Grant IIS-1633028 and USAID Cooperative Agreement No. AID-OAA-L-15-00001 (entitled “Feed the Future Innovation Lab for Integrated Pest Management”). We thank Dr. Paul Tandy (DTRA) for his encouragement and support of this work. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

REFERENCES

- Adiga, A., C. Barrett, S. E. Eubank, C. J. Kuhlman, M. V. Marathe, H. Mortveit, S. S. Ravi, D. J. Rosenkrantz, R. E. Stearns, S. Swarup, and A. Vullikanti. 2019. “Validating Agent-Based Models of Large Networked Systems”. Technical report, University of Virginia, Charlottesville, VA, USA.
- Alam, M., X. Deng, C. Philipson, J. Bassaganya-Riera, K. Bisset, A. Carbo, S. Eubank, R. Hontecillas, S. Hoops, Y. Mei, V. Abedi, and M. Marathe. 2015, 09. “Sensitivity Analysis of an ENteric Immunity SIMulator (ENISI)-Based Model of Immune Responses to Helicobacter pylori Infection”. *PLoS ONE* 10(9):e0136139.
- Alur, R., T. A. Henzinger, and M. Y. Vardi. 2015. “Survey: Theory in Practice for Design and Verification”. *SIGLOG News* 2(1):99–104.
- Balci, O., and R. E. Nance. 1985. “Formulated Problem Verification as an Explicit Requirement of Model Credibility”. *Simulation* 45(2):76–86.
- Barrett, C., H. B. Hunt III, M. V. Marathe, S. Ravi, D. J. Rosenkrantz, R. E. Stearns, and M. Thakur. 2007. “Predecessor Existence Problems for Finite Discrete Dynamical Systems”. *Theoretical Computer Science* 386(1):3–37.
- Barrett, C. L., R. J. Beckman et al. 2001. “TRANSIMS: Transportation Analysis Simulation System”. Technical Report LA-UR-00-1725, Los Alamos National Laboratory Unclassified Report.
- Barrett, C. L., K. Bisset, J. Chen, S. Eubank, B. Lewis, V. A. Kumar, M. V. Marathe, and H. S. Mortveit. 2007. “Effect of Public Policies and Individual Behavior on the Co-evolution of Social Networks and Infectious Disease Dynamics”. In *In Proceedings of the DIMACS/DyDan Workshop on Computational Methods for Dynamic Interaction Networks*. State College, PA: Citeseer.
- Barrett, C. L., H. B. Hunt III, M. V. Marathe, S. S. Ravi, D. J. Rosenkrantz, and R. E. Stearns. 2006. “Complexity of Reachability Problems for Finite Discrete Dynamical Systems”. *Journal of Computer and System Sciences* 72(8):1317–1345.
- Barrett, C. L., H. B. Hunt III, M. V. Marathe, S. S. Ravi, D. J. Rosenkrantz, and R. E. Stearns. 2011. “Modeling and Analyzing Social Network Dynamics Using Stochastic Discrete Graphical Dynamical Systems”. *Theoretical Computer Science* 412(30):3932–3946.
- Barrett, C. L., H. B. Hunt III, M. V. Marathe, S. S. Ravi, D. J. Rosenkrantz, R. E. Stearns, and P. Tolic. 2001. “Gardens of Eden and Fixed Points in Sequential Dynamical Systems”. In *Proc. Discrete Mathematics and Theoretical Computer Science Conference*, 95–110. Paris, France: HAL Archives.
- Bharathy, G. K., and B. Silverman. 2013. “Holistically Evaluating Agent-Based Social Systems Models: A Case Study”. *Simulation* 89(1):102–135.
- Bianchi, C., P. Cirillo, M. Gallegati, and P. A. Vagliasindi. 2007. “Validating and Calibrating Agent-Based Models: A Case Study”. *Computational Economics* 30(3):245–264.

- Bodlaender, H. L. 1993. "A Tourist Guide Through Treewidth". *Acta Cybernetica* 11(1-2):1–22.
- Box, G. E., and N. R. Draper. 1987. *Empirical Model-Building and Response Surfaces*. New York, NY: Wiley.
- Bruch, E., and J. Atwell. 2015. "Agent-Based Models in Empirical Social Research". *Sociological Methods and Research* 44(2):186–221.
- Campos, M. R., A. Biondi, A. Adiga, R. N. Guedes, and N. Desneux. 2017. "From the Western Palaearctic Region to Beyond: Tuta Absoluta 10 Years After Invading Europe". *Journal of Pest Science* 90(3):787–796.
- Carley, K. M. 1996. "Validating Computational Models". Technical report, Carnegie Mellon University, Pittsburgh, PA.
- Carnap, R. 1936. "Testability and Meaning". *Philosophy of Science* 3(4):419–471.
- Centola, D. 2010. "The Spread of Behavior in an Online Social Network Experiment". *Science* 329:1194–1197.
- Centola, D. 2011. "An Experimental Study of Homophily in the Adoption of Health Behavior". *Science* 334:1269–1272.
- Charness, G., R. Cobo-Reyes, and N. Jimenez. 2014. "Identities, Selection, and Contributions in a Public-Goods Game". *Games and Economic Behavior* 87:322–338.
- Crama, Y., and P. Hammer. 2011. *Boolean Functions: Theory, Algorithms, and Applications*. New York, NY: Cambridge University Press.
- Eubank, S., H. Guclu, V. S. Anil Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang. 2004. "Modelling Disease Outbreaks in Realistic Urban Social Networks". *Nature* 429:180–184.
- Forrester, J. W. 1971. "Counterintuitive Behavior of Social Systems". *Technology Review* 73:53–68.
- Forrester, J. W., and P. Senge. 1980. "Tests for Building Confidence in System Dynamics Models". In *System Dynamics*, edited by A. A. Legasto Jr., J. W. Forrester, and J. M. Lyneis, Volume 14 of *Studies in the Management Sciences*, 209–228. Amsterdam, The Netherlands: North-Holland.
- Garey, M. R., and D. S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-completeness*. San Francisco: W. H. Freeman & Co.
- Gomes, C. P., H. Kautz, A. Sabharwal, and B. Selman. 2008. "Satisfiability Solvers". In *Handbook of Knowledge Representation*, edited by F. van Harmelen, V. Lifschitz, and B. Porter, Chapter 2, 89–134. The Netherlands: Elsevier.
- Granovetter, M. 1978. "Threshold Models of Collective Behavior". *American Journal of Sociology* 83(6):1420–1443.
- Hahn, H. A. 2013. "The Conundrum of Verification and Validation of Social Science-Based Models". *Procedia Computer Science* 16:878–887.
- Hahn, H. A. 2017. "The Conundrum of Verification and Validation of Social Science-Based Models (Redux)". Technical report, Los Alamos National Laboratory, Los Alamos, NM.
- Kaluza, P., A. Kölzsch, M. T. Gastner, and B. Blasius. 2010. "The Complex Network of Global Cargo Ship Movements". *Journal of the Royal Society Interface* 7(48):1093–1103.
- Kleijnen, J. 1995. "Verification and Validation of Simulation Models". *European Journal of Operational Research* 82:145–162.
- Klügl, F. 2008. "A Validation Methodology for Agent-Based Simulations". In *Proceedings of the 2008 ACM Symposium on Applied Computing*, SAC '08, 39–43. New York, NY: Association for Computing Machinery, Inc.
- Lux, T., and R. C. J. Zwinkels. 2018. "Empirical Validation of Agent-Based Models". In *Handbook of Computational Economics*, Vol. 4, edited by C. Hommes and B. LeBaron, 437–488. The Netherlands: Social Science Research Network.
- Macal, C. M. 2016. "Everything You Need to Know About Agent-Based Modelling and Simulation". *Journal of Simulation* 10(2):144–156.
- National Research Council 1998. *Modeling Human and Organizational Behavior: Application to Military Simulations*. Washington, DC: National Academy Press.
- National Research Council 2008. *Behavioral Modeling and Simulation: From Individuals to Societies*. Washington, DC: National Academy Press.
- Oberkampf, W. L., and T. G. Trucano. 2002. "Verification and Validation in Computational Fluid Dynamics". Technical Report SAND2002 - 0529, Sandia National Laboratories, Albuquerque, NM.
- Oreskes, N. 2000. "Why Believe a Computer? Models, Measures, and Meaning in the Natural World". In *The Earth Around Us: Maintaining a Livable Planet*, edited by J. S. Schneiderman, 70–82. San Francisco, CA: W. H. Freeman and Co.
- Popper, K. 1963. *Conjectures and Refutations: The Growth of Scientific Knowledge*. London, UK: Routledge.
- Ren, Y., V. Cedeno-Mieles, Z. Hu, X. Deng, A. Adiga, C. Barrett, S. Ekanayake, B. J. Goode, G. Korkmaz, C. J. Kuhlman, D. Machi, M. V. Marathe, N. Ramakrishnan, S. S. Ravi, P. Saraf, and N. Self. 2018. "Generative Modeling of Human Behavior and Social Interactions Using Abductive Analysis". In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 413–420. New York, NY: Association for Computing Machinery, Inc.
- Ribeiro, A. S., S. A. Kauffman, J. Lloyd-Price, B. Samuelsson, and J. E. S. Socolar. 2007. "Mutual Information in Random Boolean Models of Regulatory Networks". In *arXiv:0707.3642v2*, 27 pages. ArXiv Server.
- Robinson, S. 1997. "Simulation Model Verification and Validation: Increasing the User's Confidence". In *Proc. 1997 Winter Simulation Conference*, 53–59. Columbus, OH: Institute for Operations Research and the Management Sciences.
- Rosenkrantz, D. J., M. V. Marathe, H. B. Hunt III, S. S. Ravi, and R. E. Stearns. 2015. "Analysis Problems for Graphical Dynamical Systems: A Unified Approach Through Graph Predicates". In *Proceedings of the 2015 International Conference on*

- Autonomous Agents and Multiagent Systems, AAMAS, 1501–1509.* Liverpool, UK: International Federation for Autonomous Agents and Multi-Agent Systems.
- Shmulevich, I., and S. A. Kauffman. 2004. “Activities and Sensitivities in Boolean Network Models”. *Physical Review Letters* 93(4):048701:1–4.
- Sterman, J. 2006, March. “Learning From Evidence in the Complex World”. *American Journal of Public Health* 96(3):505–514.
- Venkatraman, S., S. Wu, B. Shi, A. Marathe, M. Marathe, S. Eubank, L. P. Sah, A. Giri, L. A. Colavito, K. Nitin et al. 2017. “Towards Robust Models of Food Flows and Their Role in Invasive Species Spread”. In *2017 IEEE International Conference on Big Data*, 435–444. Piscataway, NJ: Institute for Electrical and Electronics Engineers.
- Wolfram, S. 1987. *Theory and Applications of Cellular Automata*. Singapore: World Scientific.
- Yilmaz, L. 2006. “Validation and Verification of Social Processes Within Agent-Based Computational Organization Models”. *Computational and Mathematical Organizational Theory* 12(4):283–312.

AUTHOR BIOGRAPHIES

Abhijin Adiga is a Research Assistant Professor in the Network Systems Science and Advanced Computing (NSSAC) Division of the Biocomplexity Institute and Initiative (BII) at the University of Virginia (UVA). His research interests include network science, modeling, combinatorial optimization, dynamical systems, machine learning and analysis of algorithms.

Email: abhijin@virginia.edu.

Chris Barrett is the Executive Director of the BII and a Professor of Computer Science at the UVA. His research interests include modeling, simulation, dynamical systems, social science and network science. Email: ChrisBarrett@virginia.edu.

Stephen Eubank is the Deputy Director of NSSAC Division, BII and a Professor of Public Health sciences at UVA. His research interests include modeling, simulation, dynamical systems and network science. Email: eubank@virginia.edu.

Chris J. Kuhlman is a Research Associate Professor in the NSSAC Division, BII at UVA. His research interests include modeling, simulation, software engineering, dynamical systems, computational social science and machine learning.

Email: cjk8gx@virginia.edu.

Madhav Marathe is the Director of NSSAC Division, BII and a Professor of Computer Science at UVA. His research interests include modeling, simulation, machine learning, AI, data science, dynamical systems, network science, computational social science, and analysis of algorithms. Email: marathe@virginia.edu.

Henning Mortveit is an Associate Professor in the NSSAC Division, BII, and an Associate Professor of Engineering Systems & Environment at UVA. His research interests include rigorous methods to analyze and understand massively-interacting systems, modeling, simulation and software engineering. Email: Henning.Mortveit@virginia.edu.

S. S. Ravi is a Research Professor in the NSSAC Division, BII at UVA and a Distinguished Teaching Professor Emeritus of Computer Science at the University at Albany – State University of New York. His research interests include analysis of algorithms, data mining, dynamical systems, network science and fault-tolerant computing. Email: ssravi0@gmail.com.

Daniel J. Rosenkrantz is a Distinguished Institute Professor in the NSSAC Division, BII at UVA and a Professor Emeritus of Computer Science at the University at Albany – State University of New York. His research interests include analysis of algorithms, theory of computation, dynamical systems, machine learning, and fault-tolerant computing. Email: drosenkrantz@gmail.com.

Richard E. Stearns is a Distinguished Institute Professor in the NSSAC Division, BII at UVA and a Distinguished Professor Emeritus of Computer Science at the University at Albany – State University of New York. His research interests include analysis of algorithms, theory of computation, dynamical systems and game theory. Email: thestearns2@gmail.com.

Samarth Swarup is a Research Associate Professor in the NSSAC Division, BII at UVA. His research interests include multiagent systems, agent-based modeling, machine learning, AI and simulation analytics. Email: swarup@virginia.edu.

Anil Vullikanti is a Professor in the NSSAC Division, BII and a Professor of Computer Science at UVA. His research interests include analysis of algorithms, combinatorial optimization, machine learning, dynamical systems, network science and computational social science. Email: vsakumar@virginia.edu.

ON THE MODELING AND AGENT-BASED SIMULATION OF A COOPERATIVE GROUP ANAGRAM GAME

Zhihao Hu, Xinwei Deng

Brian J. Goode, Naren Ramakrishnan

Parang Saraf, Nathan Self

Virginia Tech

Blacksburg, VA 24061, USA

Yihui Ren

Brookhaven National Laboratory

Upton, NY 11973, USA

Abhijin Adiga, Gizem Korkmaz

Chris J. Kuhlman, Dustin Machi

Madhav V. Marathe, S. S. Ravi

University of Virginia

Charlottesville, VA 24061, USA

Vanessa Cedeno-Mieles

Virginia Tech

Escuela Superior Politécnica del Litoral, ESPOL

Guayaquil, ECUADOR

Saliya Ekanayake

Lawrence Berkeley National Laboratory

Berkeley, CA 94720, USA

ABSTRACT

Anagram games (i.e., word construction games in which players use letters to form words) have been researched for some 60 years. Games with individual players are the subject of over 20 published investigations. Moreover, there are many popular commercial anagram games such as Scrabble. Recently, cooperative team play of anagram games has been studied experimentally. With all of the experimental work and the popularity of such games, it is somewhat surprising that very little modeling of anagram games has been done to predict player behavior/actions in them. We devise a cooperative group anagram game and develop an agent-based modeling and simulation framework to capture player interactions of sharing letters and forming words. Our primary goals are to understand, quantitatively predict, and explain individual and aggregate group behavior, through simulations, to inform the design of a group anagram game experimental platform.

1 INTRODUCTION

1.1 Background and Motivation

Anagram games, or word construction games, consist of players forming words from a provided group of letters. Research on anagram games—*individual* anagram games—has a long history that dates back at least to 1958, and encompasses more than 20 works that study a variety of issues. Moreover, there are many popular anagram games that are typically played by competing *individuals*, such as Scrabble, Bananagram, and Upwords. Recently, Charness et al. (2014) introduced a *group* anagram game (GrAG), where players cooperate to form words. See Section 2 for details.

Considering the substantial use of anagram games, it is surprising that almost no work has been done in *modeling* and *simulating* these games. In particular, we are interested in modeling GrAGs, notably player interactions and inter-dependence, and the implications of these interactions. There are several general reasons to prefer computational modeling over (laboratory) experiments, e.g., the ability of a validated

model to perform computational experiments much faster and at lesser cost. Beyond general motivations, there are several reasons that are expressly related to anagram games, including: (i) modeling GrAGs can be a precursor to modeling other phenomena such as team unity (Charness et al. 2014); and (ii) GrAGs have much in common with other situations in which individuals may share resources in order to mutually benefit (e.g., how to cope during crises, such as hurricanes and forest fires, along with others who remain behind (Yang et al. 2019)). Finally, *a primary motivation for our modeling and simulation work is to predict and understand individual and group performance (e.g., aggregate temporal changes in numbers of words formed, letters requested, and letter replies) in this game in order to provide insights for designing a GrAG software platform for conducting GrAG experiments.*

1.2 Our Group Anagram Game (GrAG)

An overview of the GrAG is given here; details are provided in Section 3. The GrAG is a game played among several players that work cooperatively to form words. They share letters with their immediate neighbors (players are arranged in a network) who use them to form more words than they could form using only their own allotment of letters. Figure 1 shows an illustrative conceptual view of the pair-wise interactions among four players over three time steps. The stated goal of the game is for the team to form as many words as possible. This is because the team’s earnings in the game are directly proportional to the number of words that the team forms in total. All players split the earnings evenly, regardless of their performance in the game (e.g., regardless of how many words a particular player forms). This is done to motivate the players to cooperate. Hence, forming the greatest number of words is equivalent to players trying to maximize their earnings.

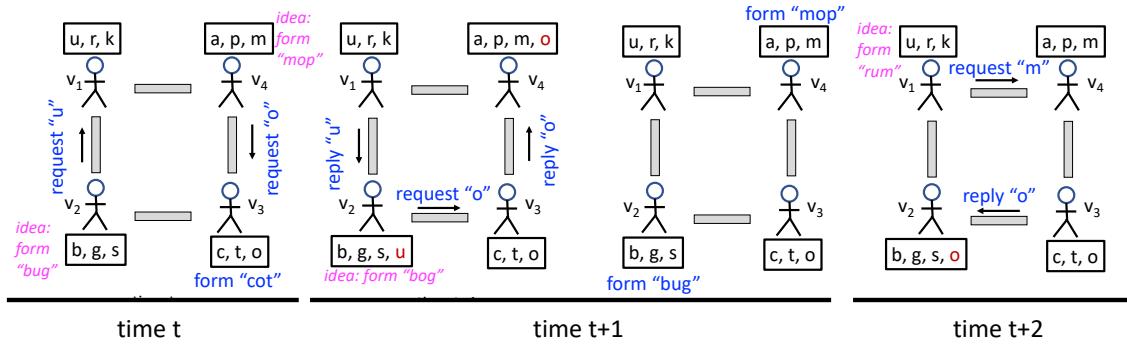


Figure 1: Illustration of the group anagram game (GrAG) setup, and interactions among the four networked players over three time steps. Gray thick lines represent communication channels over which players can request letters from their distance-1 neighbors, and receive them if/when they are sent. Each player v_i has a box containing a sequence L_i^{th} of letters in-hand (i.e., letters that they can use) to form words. Owned letters (i.e., those initially assigned to a player) are in black and received letters from neighbors are in brown. For example, at time $(t + 1)$, v_2 has $L_2^{th} = (b, g, s, u)$; u is a received letter. Player actions (in blue) are form words, request letters from neighbors, reply to neighbor letter requests, and think (think not shown; it is a no-op). Player motivations are in magenta, e.g., at time t , v_2 gets the idea to form *bug*, and so requests u . A letter received from another player vanishes once it is used (see Section 3 for details).

We emphasize that our work is not modeling cognitive processes within agents that determine agent actions. Rather, we model differences among players’ actions by changing their probabilities of taking actions. As will be explained later, greater probabilities result in more actions by players.

1.3 Novelty of Our Work

The novelty of our work is in modeling our GrAG. There is essentially no work in modeling (and thereby predicting) individual behavior in individual anagram games, and there is only one work (Ren et al. (2018)) on predicting player actions in a GrAG. Our GrAG specification differs in an important way from Ren et al. (2018) because in our setup, a player must request a letter for each use, which leads to far more interactions among players. Also, our modeling techniques and the quantities that we predict are also markedly different from this previous work; in our models, a player's actions are based on particular words being formed and the particular letters a player needs (from her neighbors) to form these words (see Section 2). We also propose and evaluate performance measures for game players in our simulations, which has not been done for a GrAG. Even though our model is simplified in some respects (in the absence of data to support more sophisticated models), it gives rise to interesting individual and group dynamics.

1.4 Our Contributions

1. Model of a group anagram game (GrAG). A model for the GrAG has been developed. The model is presented as a set of algorithms in Section 4, useful for understanding the model and for presenting the agent-based simulation (ABS) formulation. The model has different types of parameters that can be tuned to study different behaviors. These are: (i) word corpus for forming valid words; (ii) the number of initial letters assigned to agents (players) and the particular letter assignments; (iii) the network of players (the number of players and the communication channels); and (iv) propensities for players to take particular actions (e.g., form word, request a letter, reply to a letter request).

2. Simulations and results for various parameters. We construct software modules of the game models and exercise them in an high performance computing (HPC) agent-based modeling and simulation (ABMS) framework to understand game player performance (Section 5). We evaluate parameters such as network density and number of letters assigned to each player. Several interesting phenomena arise. For example, *reducing* the number of communication channels for a player, from ten to two, *increases* the number of words generated by the team, for many simulation conditions, because players do not get overloaded with requests that they cannot respond to (in a timely manner).

3. Player performance evaluation. We devise and compute simple performance metrics for agents in the game (Section 4). We study factors that lead to changes in player performance, including activity level of players and numbers of letters assigned to players (Section 5).

4. Use of modeling and simulation results. Modeling results are used in two ways in Section 6. First, the computational results provide insights into parameters and parameter values that help guide specification of game conditions for an online game platform. Second, the computational findings enable hypotheses to be formed that can then be tested (with the aforementioned game platform).

Paper organization. Related work is in Section 2. A detailed description of the experimental setup appears in Section 3. The models of the GrAG and player performance are provided in Section 4. Simulation results of modeling the game are in Section 5. Uses of the modeling results are described in Section 6. A summary with limitations of our study concludes the work.

2 RELATED WORK

Individual anagram games. Research on *individual*-based anagram games has a long history, with over 20 research studies. Mayzner and Tresselt (1958) conducted experiments where players had to unscramble letters to form a unique word. Their goal was to evaluate the degree of scramble of letters, quantified by an edit distance, and its effect on time to form a word. Individual anagram games, where a player forms words with provided letters within a specified time, have been used to study performance anxiety (Russell and Sarason 1965), goal achievement and attributing success or failure (Feather and Simon 1971), and whether people prefer pay that is or is not tied to performance (Cadsby et al. 2007).

Group anagram games. Face-to-face *group* anagram games were performed recently in Charness et al. (2014). That work is purely experimental, where players sit at a table and cooperatively form words with a fixed collection of letters. The GrAG is used as a priming activity to foster a sense of unity within the group. The one modeling work on GrAGs is Ren et al. (2018). They develop a time-sequence model that predicts the *types* of actions a player forms as the game progresses. Their predictions are of the type “player v_i selects action type ‘form word’ at time t .” The action is assumed to take place, without regard for the letters that v_i or v_i ’s neighbors possess. For example, if a player has letters g , b , and x , and the predicted action is form word, then their model will still form an unspecified word from these letters. Our model accounts for letters that players and their neighbors possess and only executes actions if it is possible to do so. Moreover, our GrAG has a key difference. In the game of Ren et al. (2018), neighboring letters only have to be requested once, because each possessed letter is assumed to be in infinite supply. That is, once a player v_i receives the requested letter a , for example, v_i can use a in any number of words; the letter is never exhausted. In our game, however, each use of a neighbor’s received letter is consumed so that letters must be requested and received for each use. This leads to many more player interactions.

3 GROUP ANAGRAM GAME (GrAG) DESCRIPTION

Game configuration. A number n of players and a fixed graph $G(V, E)$ on these players is specified, where $v_i \in V$, $i \in \{1, 2, \dots, n\}$, $n = |V|$. The edge set E represents a set of undirected channels such that edge $e = \{v_i, v_j\}$ means that players (nodes) v_i and v_j can communicate. Each player is initially assigned n_ℓ alphabetic letters, either at random or deliberately, and although each player can have a different number of letters, for exposition, it is assumed that all players receive the same number of (initial) letters. See Figure 1.

Player actions in game. As time marches forward, every player can take the actions identified in Table 1 any number of times and in any order within the game duration t_g . The set A of actions is $A = \{a_1, a_2, a_3, a_4\}$, and includes requesting letters, replying to letter requests, and forming words.

Table 1: Actions $a_i \in A$ of players $v_i, v_j \in V$ in the GrAG. Note that each of actions a_2 *request letter* and a_3 *reply with letter* involves two effects on players. A letter request is *sent* by one agent and *received* by another. A letter reply is *sent* by one agent and *received* by another. These ideas of *send* and *receive* are prominent in the models and simulations, as are the actions.

Item	Variable	Action Name	Description
1	a_1	form word	v_i forms and submits a word.
2	a_2	request letter	v_i requests a letter ℓ from a neighbor v_j .
3	a_3	reply with letter	v_j replies to v_i ’s letter request, with the requested letter ℓ .
4	a_4	thinking	v_i is thinking.

Player use of alphabetic letters. Players use letters in different ways, depending on whether a letter is initially assigned, or has been received from a neighbor. A player has an infinite supply of their initially-assigned letters. This way, a player can share any of its n_ℓ letters with her distance-1 neighbors any number of times, to increase cooperation. At the same time, a player can use their own letters redundantly in any number of words. However, each letter that v_i receives from a distance-1 neighbor cannot be shared with other neighbors of v_i . Also, when a received letter is used to form a word, then this letter is consumed and hence can no longer be used. If v_i wishes to use that letter again, it must request and receive it again.

For example, if v_i has letters $L_i^{ih} = (e, s, t, m)$, and e and s are owned letters, then v_i can form the word *see* because e and s can be used any number of times in a word. After v_i submits *see*, it still has letters in-hand $L_i^{ih} = (e, s, t, m)$. However, if e is a received letter, then v_i cannot form *see*; v_i would need two e ’s from the same or different neighbors. Suppose v_i has $L_i^{ih} = (d, c, o, g, g, j)$ and that owned letters are (d, c, o) . Then, when v_i forms *cog*, L_i^{ih} is updated to (d, c, o, g, j) .

Forming words. A corpus of words C^W is provided for the game. A submitted word is considered valid if it is an element of the corpus. Otherwise it is invalid. A valid word may only be submitted one time by a player. The set of words formed up to time t by v_i is denoted W_i . W_i is the words formed over the entire game if no time is specified. However, multiple players can form the same word.

Illustrative cooperative actions in game play. A series of three time steps in a game is provided in Figure 1, from time t through time $(t+2)$. Time $(t+1)$ is separated into two visuals to make the dynamics clearer. There are four players, v_1 through v_4 , arranged in a “circle” configuration where each node v_i has degree $d_i = 2$. Next to each player is a box containing letters. These represent the letters that a player has in-hand (i.e., in their possession) that can be used to form words. The letters that a player is assigned initially are the initial letter assignments, shown in the players’ boxes of letters in black. For example, if we assume that $n_\ell = 3$ and that the letters in-hand at time t are the initial letters (also called owned letters), then v_1 has initial letters (u, r, k) . Each player in the game knows her distance-1 neighboring players and her neighbors’ letters so that they can be requested.

Figure 1 illustrates several actions in a game. At time t , v_2 thinks to form the word *bug* (magenta text) and therefore requests letter u from v_1 . Player v_3 forms the word *cot* from its three owned letters, and hence maintains all three letters. At time $(t+1)$, v_1 responds to v_2 with letter u . Received letters are shown in brown to distinguish them from owned letters. Also at $(t+1)$, v_2 forms and submits word *bug* (that is why u was requested). (Blue text is an action.) v_2 loses u after forming *bug*. The game ends when $t = t_g$.

4 MODELS and ABM ALGORITHMS

4.1 Group Anagram Game

Figure 2 provides an abstract view, in the form of a graph $G(V, E)$, of the same game configuration as in Figure 1. Here, the four players $v_i \in V$, $i \in \{1, 2, 3, 4\}$ are represented as graph vertices and the communication channels are represented by the edge set E , where $e = \{v_i, v_j\} \in E$ for $v_i, v_j \in V$. (Note that for models, we use the terms player, agent, node, and vertex interchangeably.)

Each agent $v_i \in V$ in the ABM is assigned a set of the data structures in Figure 2. See the caption for B_i^1 and B_i^2 . Each entry in a buffer is a data structure itself. For example, B_i^2 contains letter requests, made by v_i . Fields in a request include: the letter ℓ requested, the particular word w that v_i seeks to form with the letters, the neighbor v_k to whom the request is sent, the requestor v_i , a unique universal identifier, and the time t of the request. Fields of letter requests in buffer B_i^1 , to which v_i may reply (i.e., the reply buffer), include the requestor v_j , the letter requested ℓ , the universal identifier in the request that instigated this reply, the time t of the request, and the agent to which the request is made (which is v_i , used for verification). Another data structure (not shown) contains all letter requests made for a particular word w because a word may require multiple letters from neighbors. Let L_i^{ih} be the sequence of letters currently possessed by v_i in a GrAG. Let L_i' be the combined set of all letters initially assigned to all distance-1 (immediate) neighbors of v_i .

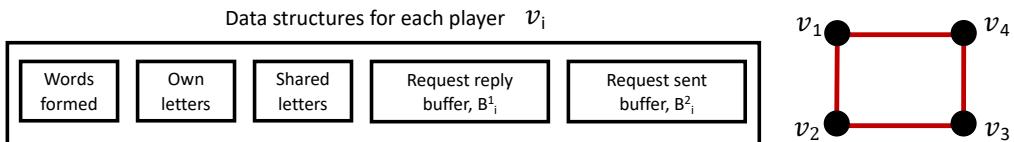


Figure 2: Abstract view of the group anagram game (GrAG) (cf. Figure 1) to support modeling. Each node (player) $v_i \in V$ has the structures depicted to support the player actions in Table 1. Buffer B_i^1 contains letter requests from neighboring players to which v_i may respond. Buffer B_i^2 contains letter requests from v_i to its neighbors.

For each agent in each simulation, $p_{act} = (p_{fw}, p_{sreq}, p_{srep}, p_{think})$ is specified. This is a 4-tuple of fixed probability values, whose sum must be 1.0. These vector elements correspond, respectively, to the probability of an agent v_i taking actions a_1 through a_4 of Table 1 at each time step. At each time, each v_i uniformly samples a probability p^* from [0,1]. An agent deterministically chooses the action $a \in A$ based on p_{act} using p^* , and executes a .

Two algorithms are given in this section for modeling the GrAG. They are presented in a style for understanding both the agent model and the simulation process. Some details are omitted for clarity. For example, the software implementation uses a distributed HPC framework, but specifying aspects of the distributed computations provides no insights into the agent model nor the simulation process.

The controlling algorithm (not shown) loops over individual and independent executions of the game dynamics (each execution is called a *run*), such that a simulation is comprised of a user-specified number of runs. For example, multiple runs may be used to obtain variability for stochastic simulations where each run has the same initial conditions I . For each run, the simulator loops over discrete times, and for each time, the simulator iterates over all agents. For each agent, the VERTEX FUNCTION algorithm (see Algorithm 1) is invoked.

Algorithm 1 is the entrypoint into computations for individual agents $v_i \in V$. An agent first receives all messages from its neighbors that were sent in the previous time step, and buffers B_i^1 and B_i^2 are updated accordingly. Then, the action a that agent v_i takes at t is determined. The algorithm VERTEX ACTION (Algorithm 2) is invoked with v_i and action a . Algorithm VERTEX ACTION executes the action a . All actions are detailed in this algorithm, and may result in updates of buffers, forming a word, forming letter requests and replies, and sending a letter request or reply.

Recall that letters that players have available to form words are used differently, depending on whether the letter is initially assigned to a player or is a letter of a distance-1 (immediate) neighbor. A letter that is initially assigned to v_i can be used any number of times, including multiple times within one word. So a player can form *tot* with initial letters $L_i^{init} = (b, t, o)$, but if v_i 's initial letters are $L_i^{init} = (b, g, o)$, then v_i must request and receive two t 's from neighbors.

Algorithm 1: Steps of the Algorithm VERTEX FUNCTION.

- 1 **Input:** Time t . Agent v_i . Neighbors $n[i]$ of v_i in $G(V, E)$. Letters L'_i of neighbors of v_i . Letters that v_i has in-hand L_i^{ih} . Probabilities of actions in A , $p_{act} = (p_{fw}, p_{sreq}, p_{srep}, p_{think})$. Word corpus C^W . Buffer of v_i 's outstanding letter requests made B_i^2 . Buffer of letter requests to v_i , B_i^1 .
 - 2 **Output:** The next action $a \in A$ and the updated values (state) of all inputs.
 - 3 **Steps:**
 - A. Receive all letter requests from neighbors $n[i]$ of v_i , sent to v_i at the previous time ($t - 1$), and put in buffer B_i^1 . These are requests that v_i may reply to.
 - B. Receive all letter replies from v_i 's neighbors that are in response to v_i 's letter requests, sent to v_i at the previous time ($t - 1$), and put in L_i^{ih} ; mark this letter request in B_i^2 as fulfilled. If received letters enable a visited word to now be formed, then form the word and submit.
 - C. Sample from a uniform distribution [0,1], value p^* .
 - D. From the vector $p_{act} = (p_{fw}, p_{sreq}, p_{srep}, p_{think})$ of probabilities for actions a_1 through a_4 , respectively, of Table 1, and p^* , determine the action $a \in A$ that v_i will take at this time t .
 - E. Call the action routine for vertex v_i (Algorithm 2, VERTEX ACTION).
 - F. Write updated state (variable values) to file for v_i at time t .
-

Algorithm 2: Steps of the Algorithm VERTEX ACTION for vertex v_i .

- 1 **Input:** Time t . Agent v_i . Action $a \in A$. Neighbors $n[i]$ of v_i in $G(V, E)$. Letters L'_i of neighbors of v_i . Letters that v_i has in-hand L_i^{ih} . Probabilities of actions in A , $p_{act} = (p_{fw}, p_{sreq}, p_{srep}, p_{think})$. Word corpus C^W . Buffer of v_i 's outstanding letter requests made B_i^2 . Buffer of letter requests to v_i , B_i^1 .
 - 2 **Output:** The updated values (state) of all inputs.
 - 3 **Steps:**
 - A. **if** a equals a_1 **do** ## Action a is for v_i to form word.
 - i. Select randomly a word $w \in C^W$, where $w \notin W_i$ (i.e., v_i cannot repeat words), and can be formed with letters in $L_i^{ih} \cup L'_i$ (i.e., the union of letters in-hand and the letters of neighbors). If all letters, including multiplicities, of w are in L_i^{ih} (i.e., are in-hand), then form and submit word. Otherwise, there are letters ℓ that need to be requested. Put these letters ℓ of w (each needed letter instance is an individual request) in B_i^2 (to be requested). If there is no such $w \in C^W$, do nothing.
 - B. **if** a equals a_2 **do** ## Action a is for v_i to send letter request.
 - i. If there is a letter request in B_i^2 that has not been sent, send one request using first-in first-out (FIFO) order. Mark request as sent. Otherwise, do nothing.
 - C. **if** a equals a_3 **do** ## Action a is reply (with letter) to a letter requested by a neighbor.
 - i. If there is a letter request from a neighbor of v_i in B_i^1 that is waiting to be fulfilled, then send a letter reply using FIFO ordering, and mark the request in B_i^1 as fulfilled. Otherwise, do nothing.
 - D. **if** a equals a_4 **do** ## Action a is think.
 - i. Do nothing. The process of v_i thinking just consumes time.
 - E. Return updated state (variable values) for input variables (above) of vertex v_i .
-

4.2 Performance Parameters For Individual Players

Let $L_i = L_i^{ih} \cup L'_i$ be the sequence of all letters available to v_i (its own assigned letters and those of its neighbors). The set $W_i^{tot} \subseteq C^W$ of words that v_i can form are the words $w \in C^W$ such that every letter ℓ of w is in L_i .

The performance of v_i in a GrAG is given by three parameters: (i) $\alpha_{w,i}$ in forming words, (ii) $\alpha_{req,i}$ in requesting letters, and (iii) $\alpha_{rpl,i}$ in replying to letter requests of its neighbors, given by

$$\alpha_{w,i} = \overbrace{\frac{n_{words,i}}{n_{words,i}^{max}}}^{\text{words formed}}, \quad \alpha_{req,i} = \overbrace{\frac{n_{reqSent,i}}{n_{reqSent,i}^{max}}}^{\text{letter requests sent}}, \quad \alpha_{rpl,i} = \overbrace{\frac{n_{repSent,i}}{n_{reqRec,i}}}^{\text{letter replies sent}} \quad (1)$$

where $n_{words,i}$ is the number of words that v_i forms in the GrAG; $n_{words,i}^{max}$ is the maximum number of words that v_i can form, i.e., $n_{words,i}^{max} = |W_i^{tot}|$; $n_{reqSent,i}$ is the number of letter requests that v_i sends to its neighbors in $G(V, E)$ in the GrAG; $n_{reqSent,i}^{max}$ is the maximum number of requests that v_i can send to all of its immediate neighbors in $G(V, E)$ in order to form all words in W_i^{tot} ; $n_{repSent,i}$ is the number of letter replies sent by v_i to its neighbors; and $n_{reqRec,i}$ is the number of letter requests received from v_i 's neighbors.

5 MODELING AND SIMULATION RESULTS

5.1 Considerations for Simulation Parameters

Table 2 provides the simulation variables studied. Here, variables are discussed in relation to practical (realistic) considerations for implementing the game we are modeling. The number n of players on the lesser end ($n = 11$) represents smaller teams, and larger n ($= 1000$) simulates a game that is possible to play, for example, by employing students from a large undergraduate course that can be on the order of many hundreds of students. Networks are used to control player degree (number of neighbors) in order to scale up n . For example, it is not practical for $n = 100$ students to all interact with each other.

The use of circle, clique, and random regular graphs means that, for a specific graph, all nodes (agents, players) have the same number of neighbors. Hence, each node is, in a sense, a different replicate instance, because letter assignments vary among the nodes but the numbers of letters assigned to each node in these simulations is the same.

For the probability vector p_{act} of action probabilities for a player, we use \hat{p} values in Table 2; \hat{p} corresponds to probabilities for actions form word, request letter, and reply to requests. Values range from

0.05 to 0.25, to represent a range from inactive (or more deliberate) agents to relatively active (fast thinking) agents. This is, when $\hat{p} = 0.05$, $p_{think} = 0.85$, so that in expectation, a player is thinking—not acting—for 85% of the game time. The mid-range value of $\hat{p} = 0.15$ means that a player is thinking more than one-half of the time ($p_{think} = 0.55$), and the greatest value for \hat{p} means that an agent is only thinking 25% of the time and acting 75% of the time. We study this range to evaluate different levels of mental activity in lieu of a cognitive model of player behavior.

The numbers n_ℓ of own letters assigned to players is designed to control interactions (along with node degree d). The cases $n_\ell \leq 2$ force players to interact with their neighbors in order to form words when the minimum permissible word length is three letters. A player has access to as many as $(d+1)n_\ell$ unique letters (if there are no duplicate letters), if all players are given n_ℓ letters. Owned letter assignments are done such that each player has n_ℓ unique letters; there are no duplicates.

The word corpus of 1015 words considers only 3-letter words. Admitting larger words is a subject for further work, but is unlikely to effect our results, at least qualitatively, because of the way agents step through C^W to select words to try to form.

Table 2: Summary of parameters and their values used in simulations of GrAGs.

Parameter	Description
Networks $G(V,E)$.	Circle graphs, cliques, and random regular graphs. Numbers of nodes are $n = 11, 100$, and 1000. Uniform degrees d of players are two and ten.
Probability vectors for actions p_{act} .	$p_{act} = (p_{fw}, p_{sreq}, p_{srep}, p_{think})$ is the vector of probabilities corresponding to the actions in Table 1 where the probabilities associated with a_1, a_2, a_3 , and a_4 are, respectively $p_{fw}, p_{sreq}, p_{srep}$, and p_{think} . These probabilities are used at each time t to select the player action at that t . We use $\hat{p} = p_{fw} = p_{sreq} = p_{srep}$, where \hat{p} is in the set $\{0.05, 0.10, 0.15, 0.20, 0.25\}$. Consequently, $p_{think} = 1 - 3\hat{p}$. A vector p_{act} of values is assigned to each agent.
Number n_ℓ of owned letters.	This is the number of owned letters assigned to a player as part of initial conditions. Values used are $n_\ell = 1, 2, 3$, and 4. Owned letters for a player purposely contain no duplicates. A value of n_ℓ is assigned to each player.
Letter assignment process.	The initial letter assignments to players are done uniformly at random.
Shared letters.	The letters that a player can share with her neighbors are the same as the owned letters. There may be duplicate letters between pairs of players, including neighbors of an agent v_i .
Duration of GrAG t_g .	The duration of the group anagram game is fixed at $t_g = 300$ seconds.
Word corpus C^W .	The corpus of 1015 3-letter words is taken from http://www.wordfind.com/3-letter-words/ . That is, only 3-letter words are considered in simulations.
Number of runs n_{runs} .	Each simulation is composed of n_{runs} individual dynamics instances, where each instance starts from time $t = 0$, with initial conditions reset, and then the dynamics of the system are executed for t_g discrete time steps. Here, $n_{runs} = 50$.

5.2 Simulation Results for GrAG

Subsections below provide insights into the effects of input variables (see Table 2) on simulation results. Results include all 50 runs (instances) per simulation, either in the form of averages (i.e., the mean values over all runs at each discrete time) with \pm one standard deviation error bars, or boxplots of all data. Observations of behaviors hold only for the conditions of the computational experiments (but are suggestive of more general trends).

Aggregate effects of number of owned letters. Figure 3 shows aggregate time histories for all agents in an 11-agent game where each agent has two neighbors, i.e., $(n, d) = (11, 2)$. Other simulation parameter values are given in the caption. The numbers of letter requests sent by players are in Figure 3a. The numbers of replies (with the letters) to those requests (Figure 3b) are a little less than the numbers of requests. The numbers of words formed (Figure 3c) are still less in number than requests and replies.

The results indicate that there is not much difference between giving players three or four initial letters. However, the differences for $1 \leq n_\ell \leq 3$ are large. Also, $t_g = 300$ seconds is more than adequate for $n_\ell = 1$ because performance saturates, but for $n_\ell \geq 2$, performance is still changing, although it is approaching saturation for $n_\ell = 2$.

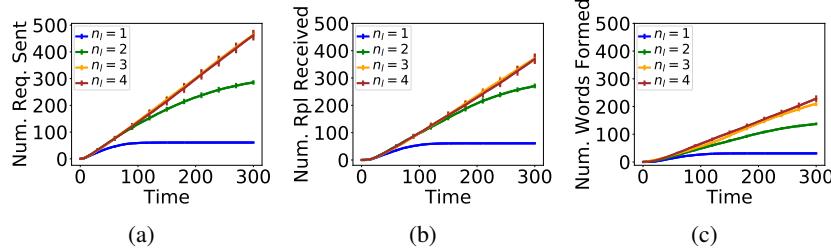


Figure 3: Simulation results for all agents from an $(n, d) = (11, 2)$ graph, for action probabilities $p_{act} = (p_{fw}, p_{sreq}, p_{srep}, p_{think}) = (0.15, 0.15, 0.15, 0.55)$ and number of letters n_ℓ taking values $\{1, 2, 3, 4\}$ in turn (see legend). Standard deviation error bars are shown in each curve, every 30 time units, so as to not overwhelm the curves. Plots (left to right) correspond, respectively, to actions: (a) send (letter) requests, (b) receive (letter) replies, and (c) form words. The ordinate values are the *sum* of all of the respective actions across all 11 agents, so the average number of agent actions is obtained by dividing each ordinate value by 11. Among the results shown are: when a player has only one letter ($n_\ell = 1$), the actions saturate in less than 150 seconds, but for greater numbers of owned letters, actions are taking place at the 5-minute mark; and the differences between results for $n_\ell = 3$ and 4 are small.

Effect of number of neighbors. Figure 4 shows aggregate results for two different values of numbers of neighbors, for experiments with $n = 11$. For all graphs, curves for $d = 2$ are in blue and those for $d = 10$ are in green. For each degree, the solid curves correspond to $n_\ell = 2$, and the dashed curves correspond to $n_\ell = 4$. Where no dashed curve is visible, it is “under” (coincident) with a solid curve.

These results are particularly interesting from the perspective of agents’ “congestion” of communications and performance. For agent v_k that is replying to requests with the desired letters, if v_k has degree $d = 2$, then all of her replies are going to two agents v_1 and v_2 . If v_k has $d = 10$ neighbors, then the replies are being distributed across more neighbors. The main point that the plots convey is that players with large numbers of neighbors can get bombarded with letter requests. That is, when d of a player v_k is sufficiently large, p_{srep} for v_k is sufficiently small, and p_{sreq} of v_k ’s neighbors are sufficiently large, then the letter requests made of v_k “pile up” because v_k cannot reply sufficiently quickly to these requests. The result is that performance, in terms of words formed, can *decrease* as d increases. This is why, in Figure 4, the green curves (for $d = 10$) in Figure 4a for letter requests sent are at or above the blue curves (for $d = 2$), but the green curves fall below the blue curves in Figures 4b and 4c for replies received and words formed, respectively.

Performance of individual (disaggregated) players. Figure 5 disaggregates the end-of-game ($t_g = 300$ seconds) results for $n_\ell = 3$ from Figure 3, and compares player actions with optimal behavior—in terms of numbers of actions. *Optimal behavior* means the best possible performance of players. See Section 4.2 for definitions of performance ratios. Performance in terms of formed word $\alpha_{w,i}$, letter requests *sent* $\alpha_{req,i}$, and letter replies *sent* $\alpha_{rpl,i}$ are in the range $[0, 1]$ in Figures 5a, 5b, and 5c, respectively, per player. There are clear similarities between $\alpha_{w,i}$ and $\alpha_{req,i}$, since letter requests (and replies) will make it possible to form more words. Figure 5d makes it clear that the relatively good performances in words formed and letter requests for Players 0 and 4 are due to the relatively lesser optimal behaviors possible (red and black curves for these players [for requests sent and words formed, respectively] have lesser values). The blue curve, which serves as the denominator for the data in Figure 5c, also explains these data. These data,

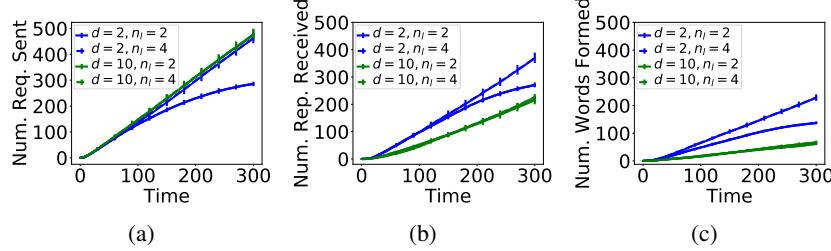


Figure 4: Simulation results aggregated for all agents for graphs with $(n, d) = (11, 2)$ (blue curves) and $(11, 10)$ (green curves), for action probabilities $p_{act} = (p_{fw}, p_{sreq}, p_{srep}, p_{think}) = (0.15, 0.15, 0.15, 0.55)$, and number of letters $n_\ell = 2$ (solid curves) and 4 (dashed curves). Standard deviation error bars are shown in each curve, every 30 time units, so as to not overwhelm the curves. Plots (left to right) correspond, respectively, to actions: (a) send (letter) requests, (b) receive (letter) replies, and (c) form words. The ordinate values are the sum of all of the respective actions across all 11 agents, so the average number of agent actions is obtained by dividing each ordinate value by 11. Taken together, the plots show the interesting result that players that have too many neighbors (so that they have more letters to select from to form words) receive fewer letter replies and form fewer words. See the text for details.

collectively, indicate that initial letter assignments can make a big difference in performance parameters $\alpha_{w,i}$, $\alpha_{req,i}$, and $\alpha_{rpl,i}$.

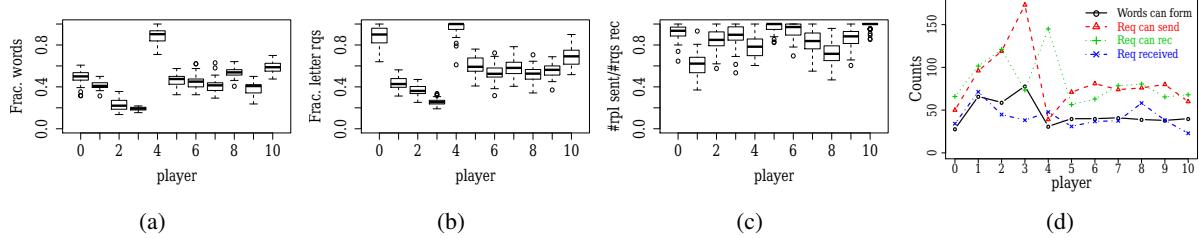


Figure 5: Performance results of individual agents relative to best possible performance for an $(n, d) = (11, 2)$ graph, action probabilities $p_{act} = (p_{fw}, p_{sreq}, p_{srep}, p_{think}) = (0.15, 0.15, 0.15, 0.55)$, and number of letters $n_\ell = 3$ assigned to each player. Plots (left to right) correspond, respectively, to actions: (a) $\alpha_{w,i}$ for words formed, (b) $\alpha_{req,i}$ for letter requests sent, and (c) $\alpha_{rpl,i}$ for replies sent for letter requests. In all cases, boxplots show results for all 50 run instances. (d) Best possible performance in terms of numbers of actions for forming words, requests sent, replies sent (\equiv requests received), and actual requests received. The black and red curves show, for example, that player 4's strong performance is due to the fact that it can form relatively fewer words and request relatively fewer letters than most other players.

Average performance across agents for changes in action probabilities. Figure 6 shows the effects of different player activity levels, implemented by altering \hat{p} in p_{act} ; see Table 2. As \hat{p} increases, the probabilities of each of the actions form word, request letter, and reply to request increase. The figure shows that the probability (i.e., activity) of players has a significant effect on performance, as does the number n_ℓ of letters assigned to players.

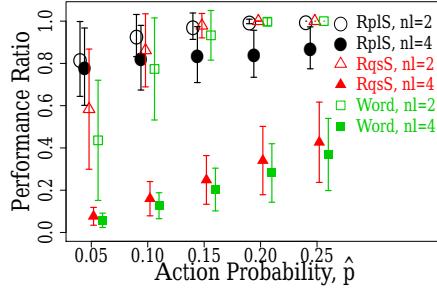


Figure 6: Average performance across all players in $(n, d) = (11, 2)$ experiments for $n_\ell = 2$ and 4, and for all p_{act} vectors with $\hat{p} = 0.05, 0.10, 0.15, 0.20$, and 0.25 (see Table 2). Y-axis values are average values of α_{wi} for form words (“Word” in green), $\alpha_{req,i}$ for letter requests sent (“RqsS” in red), and $\alpha_{rpl,i}$ for replies sent for letter requests (“RplS” in black). Player performance increases as \hat{p} increases and n_ℓ decreases.

6 USES OF MODELING RESULTS

Simulations results of the previous section are used to discuss refinements to game parameters of the envisioned experimental platform and to construct hypotheses to test.

6.1 Game Specifications

Several insights from the simulation results above can be used to specify/refine game conditions. We assume that the probabilities of player actions form word, request letter, and reply to request are in the middle of our investigated range, i.e., $p_{act} = (p_{fw}, p_{sreq}, p_{srep}, p_{think}) = (0.15, 0.15, 0.15, 0.55)$. Figure 3 suggests that a 5-minute game may be adequate for players to form words when $n_\ell \geq 3$ (that is, individual players may form up to 20 or more words). This is not an obvious result. Individual anagram games of a different sort, where a player unscrambles letters to form a *single* unique word, use a duration of four minutes per game (Mayzner and Tresselt 1958). Figure 3, where $d = 2$ and $n_\ell = 3$, suggests that a criterion on d and n_ℓ to produce significant interactions among players is: $(1+d)n_\ell \geq 9$. Results in Figure 4 provide strong evidence for the need to test different network structures. While some may contend that this is an obvious statement, this figure suggests *why* testing different connectivities is important and interesting. Figure 5 indicates that (initial) letter assignments can produce different performance among players. We saw that for Players 0 and 4, performance was aided by the fact that they could form fewer words and could profitably request fewer letters than other players. This is dictated by assignments of a player’s and its neighbors’ letters. (Our ABMS can control the assignment of particular letters to particular players, but this is not reported on here owing to lack of space.) Although not shown, results for different n (= 11, 100, and 1000) players show that numbers of actions scale linearly with n , as expected.

6.2 Illustrative Hypotheses

From the computational results of Section 5, the hypotheses in Table 3 were formulated. These could be tested in a game platform that runs experiments according to the description in Section 3. Conditions not stated in the hypotheses are those represented in the figures of Section 5. For the last one, we did not show results for games with $n = 100$ and 1000 players owing to space limitations. Finally, the hypotheses are quantitative rather than qualitative to provide more stringent tests.

7 CONCLUSION AND LIMITATIONS

We have described a novel GrAG. Novelty and contributions are in Sections 1.3 and 1.4. Future work includes addressing limitations by extending the model to account for how players decide on which actions to take. We need to report on simulations with heterogeneous conditions. Space limitations prevent presentation of these results. Source code is available upon request from Chris Kuhlman at cjk8gx@virginia.edu.

Table 3: Illustrative hypotheses, formulated from the modeling results, to test in an experimental setting.

Topic	Hypothesis
Game duration.	A five-minute GrAG is sufficiently long to produce over 1000 total actions among players in $(n, d) = (11, 2)$ experiments when players have $n_\ell = 3$ letters.
Numbers n_ℓ of assigned letters.	Increasing n_ℓ from 3 to 4 results in no more than a 10% increase in numbers of actions.
Player degree d .	As number of neighbors of a player increases from $d = 2$ to 10, numbers of letter requests increases by 100% but the number of letter replies decreases by 100%.
Numbers of players.	As the number n of players in the game increases from 11 to 100, the numbers of actions will increase disproportionately by 20%.

ACKNOWLEDGMENT

We thank the anonymous reviewers for their helpful feedback. We thank our colleagues at NSSAC and computer system administrators: Dominik Borkowski, Jason Decker, Miles Gentry, Jeremy Johnson, William Marmagas, Douglas McMaster, and Kevin Shinpaugh. This work has been partially supported by NSF CRISP 2.0 (Grant 1832587), DARPA Cooperative Agreement D17AC00003 (NGS2), DTRA CNIMS (Contract HDTRA1-11-D-0016-0001), DTRA Comprehensive National Incident Management System Contract HDTRA1-17-D-0023. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

REFERENCES

- Cadsby, C. B., F. Song, and F. Tapon. 2007. “Sorting and Incentive Effects of Pay for Performance: An Experimental Investigation”. *Academy of Management Journal* 50:387–405.
- Charness, G., R. Cobo-Reyes, and N. Jimenez. 2014. “Identities, Selection, and Contributions in a Public-Goods Game”. *Games and Economic Behavior* 87:322–338.
- Feather, N. T., and J. G. Simon. 1971. “Causal Attributions for Success and Failure in Relation to Expectation of Success Based Upon Selective or Manipulative Control”. *Journal of Personality and Social Psychology* 39:527–541.
- Mayzner, M. S., and M. E. Tresselt. 1958. “Anagram Solution Times: A Function of Letter Order and Word Frequency”. *Journal of Experimental Psychology* 56(4):376.
- Ren, Y., V. Cedeno-Mieles et al. 2018. “Generative Modeling of Human Behavior and Social Interactions Using Abductive Analysis”. In *ASONAM*, 413–420.
- Russell, D. G., and I. G. Sarason. 1965. “Test Anxiety, Sex, and Experimental Conditions in Relation to Anagram Solution”. *Journal of Personality and Social Psychology* 1:493–496.
- Yang, Y., L. Mao, and S. S. Metcalf. 2019. “Diffusion of Hurricane Evacuation Behavior Through a Home-Workplace Social Network: A Spatially Explicit Agent-Based Simulation Model”. *Computers, Environment, and Urban Systems* 74:13–22.

AUTHOR BIOGRAPHIES

ZHIHAO HU, VANESSA CEDENO-MIELES, PARANG SARAF, are graduate students at Virginia Tech. Their email addresses are huzhihao,vcedeno,parang@vt.edu.

XINWEI DENG is a professor in the Department of Statistics at Virginia Tech. His email address is xdeng@vt.edu.

ABHIJIN ADIGA, GIZEM KORKMAZ, CHRIS J. KUHLMAN, DUSTIN MACHI, MADHAV V. MARATHE, S. S. RAVI are faculty in the Biocomplexity Institute & Initiative at University of Virginia. Their email addresses are abhijin,gkorkmaz,cjk8gx,dm8qs,marathe,ssr6nh@virginia.edu.

YIHUI REN is a Research Associate at Brookhaven National Laboratory. His email address is yren@bnl.com.

SALIYA EKANAYAKE is a Postdoctoral Fellow at Lawrence Berkeley National Laboratory. His email address is esaliya@gmail.com.

BRIAN J. GOODE, NAREN RAMAKRISHNAN, NATHAN SELF are, respectively, faculty in the Biocomplexity Institute of Virginia Tech, director of the Virginia Tech’s Discovery Analytics Center (DAC), and faculty in DAC at Virginia Tech. Their email addresses are bjgoode,naren,nwself@vt.edu.

Predicting suitable habitat of an invasive weed *Parthenium hysterophorus* under future climate scenarios in Chitwan Annapurna Landscape, Nepal

MAHARJAN Seerjana^{1,2*}  <https://orcid.org/0000-0002-7907-0145>;  e-mail: seerjana9@hotmail.com

SHRESTHA Bharat Babu¹  <https://orcid.org/0000-0002-9457-2637>; e-mail: shresthabb@gmail.com

JOSHI Mohan Dev²  <https://orcid.org/0000-0001-7680-2585>; e-mail: mojspost@gmail.com

DEVKOTA Anjana¹  <https://orcid.org/0000-0003-1065-3286>; e-mail: devkotaa@gmail.com

MUNIAPPAN Rangaswamy³  <https://orcid.org/0000-0002-2992-2792>; e-mail: rmuni@vt.edu

ADIGA Abhijin⁴  <https://orcid.org/0000-0002-9770-034X>; e-mail: abhijin@virginia.edu

JHA Pramod Kumar^{1*}  <https://orcid.org/0000-0001-7260-9876>;  e-mail: pkjhaprof@gmail.com

* Corresponding author

¹ Central Department of Botany, Tribhuvan University, Kirtipur, Kathmandu, Nepal

² Department of Plant Resources, Thapathali, Kathmandu, Nepal

³ IPM Innovation Lab, Virginia Tech, 526 Prices Fork Road, Blacksburg, VA 24061 USA

⁴ Biocomplexity Institute and Initiative, University of Virginia, USA

Citation: Maharjan S, Shrestha BB, Joshi MD, et al. (2019) Predicting suitable habitat of an invasive weed *Parthenium hysterophorus* under future climate scenarios in Chitwan Annapurna Landscape, Nepal. Journal of Mountain Science 16(10). <https://doi.org/10.1007/s11629-019-5548-y>

© Science Press, Institute of Mountain Hazards and Environment, CAS and Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract: Chitwan-Annapurna Landscape (CHAL) in central Nepal is known for its rich biodiversity and the landscape is expected to provide corridors for species range shift in response to climate change. Environmental assessments have identified biological invasions and other anthropogenic activities as major threats to the biodiversity in the CHAL. One of the rapidly spreading Invasive Alien Plant species (IAPs) in the CHAL is *Parthenium hysterophorus* L., a neotropical invasive weed of global significance. This study aimed to investigate the current and future projected suitable habitat of *P. hysterophorus* in the CHAL using MaxEnt modelling in three “Representative Concentration Pathways” (RCPs 2.6, 4.5 and 8.5) corresponding to different greenhouse gases emission

trajectories for the year 2050 and 2070. A total of 288 species occurrence points, six bioclimatic variables - mean diurnal range, isothermality, annual precipitation, precipitation of driest month, precipitation seasonality, precipitation of driest quarter and two topographic variables (aspect and slope) were selected for MaxEnt modelling. Potential range shift in terms of increase or decline in the suitable habitat areas under the projected scenarios were calculated. Slope and annual precipitation were the most important variables that explained the current distribution of *P. hysterophorus*. Twenty percent of the total area of CHAL was predicted to be suitable habitat for the growth of *P. hysterophorus* in the current climatic condition. Highest gain in the suitable habitat of this noxious weed was found under RCP 4.5 scenario in 2050 and 2070, whereas there will be a loss in the

Received: 23-Apr-2019

Revised: 06-Jul-2019

Accepted: 16-Aug-2019

suitable habitat under RCP 8.5 scenario in 2050 and 2070. Out of four physiographic regions present in CHAL, three regions - Siwalik, Middle Mountain and High Mountain have suitable habitat for *P. hysterophorus* under current climatic condition. The mountainous region is likely to be affected more than the Siwalik region by further spread of *P. hysterophorus* in the future under low (RCP 2.6) to medium (RCP 4.5) emission scenarios. The suitable habitat for this weed is likely to increase in the protected areas of mountain regions (Langtang National Park, Annapurna Conservation Area and Manaslu Conservation Area) in the future. The results have revealed a risk of spreading *P. hysterophorus* from present localities to non-invaded areas in the current and future climatic condition. Such risk needs to be considered by decision makers and resource managers while planning for effective management of this weed to reduce its ecological and economic impacts in the CHAL.

Keywords: Parthenium weed; Ecological Niche Model; MaxEnt; Invasive species; Habitat suitability

Introduction

Biological invasions are considered vital components of global change causing impact on native biodiversity, agriculture, ecosystem processes and ecosystem services (Vitousek et al. 1997; Ricciardi et al. 1998; Pejchar and Mooney 2009; Paini et al. 2016). Climate change and biological invasions have gained much interest because of their emergence as major threats to biodiversity, habitat loss and species extinction (Vitousek et al. 1997; Dukes and Mooney 1999; Thomas et al. 2004). The adverse effect of invasive species is further exacerbated by ongoing climate change (Simberloff 2000). Some elements of global change such as the increased concentration of carbon dioxide and change in land use patterns are likely to have synergistic roles in the spread of invasive species (Dukes and Mooney 1999; IPCC 2014). Rapid growth rate, a typical characteristic of Invasive Alien Plant species (IAPs), supports their quick response to the changing climate and allows invaders to expand into new ranges (Thuiller et al. 2007; Bradley et al. 2010). Increased temperature and CO₂ due to climate change enhance the performance of IAPs more strongly than native plant species (Liu et al. 2016). Altered introduction

and dispersal mechanisms, shifting of the geographical range of invasive species, establishment of new invasive species, altered impact and distribution of existing invasive species, and change in effectiveness of control strategies are the main consequences of climate change for invasive species, and continue to increase the extinction risk of many terrestrial species (Sutherst et al. 2000; Hellmann et al. 2008; IPCC 2014).

Ecological niche models are increasingly being used to develop predictions concerning current and future suitable habitat of species based on the species' current occurrence, which can be useful for many aspects of resource management and conservation planning including invasive species risk assessment (Guisan and Thuiller 2005; Franklin 2009). The possibility of species invasion before their introduction to new areas can also be predicted with the help of modelling procedures (Peterson and Vieglais 2001).

A large number of general and specific software packages and systems of ecological niche modelling techniques are being used to predict suitable habitat of species (Franklin 2009). Among them, maximum entropy (MaxEnt) is a general-purpose machine learning technique that can be applied to presence-background data to generate habitat suitability predictions using environmental variables (Phillips et al. 2006). It is widely used in modelling the potential suitable habitat of invasive species as it has shown higher predictive accuracy using presence only data (Franklin 2009). Absence data for invasive species may not be reliable because the species may be under expansion process and might not have undergone equilibrium state. Such absence data may mislead the modelling procedure if the modelling is being considered to estimate the complete area at risk for invasion (Jimenez-Valverde et al. 2011).

Nepal is ranked as the third most threatened agricultural country in terms of introduction of invasion species to agriculture because of international trade of agricultural products (Paini et al. 2016). *Parthenium hysterophorus* L., a native of tropical America, has become one of the noxious weeds of agriculture in 48 countries and five continents across the world in the last 60 years (Haseler 1976; Shabbir et al. 2019). In Nepal, *P. hysterophorus* is spreading rapidly and has already invaded diverse regions ranging from southern

lowland plain to mountain regions in the northern part of the country up to the elevation of ca. 1900 m asl (Shrestha 2016). It is a major problem in shrubland, grassland, rangelands, agro-ecosystem and residential areas. It alters plant species composition, displace important species and soil nutrient content. It also causes health hazard to human by developing eye inflammation and skin dermatitis. Livestock develop skin lesions and produce bitter milk when *P. hysterophorus* is used as fodder (Shrestha et al. 2015).

Few studies have been conducted on the impact of climate change on the distribution and probable spread of *Parthenium hysterophorus* in broader ecological ranges including Nepal Himalaya in the future (Mainali et al. 2015; Lamsal et al. 2018; Thapa et al. 2018; Shrestha et al. 2018). Chitwan-Annapurna Landscape (hereafter 'CHAL') located in the central Nepal with north-south habitat connectivity is known for its rich biodiversity due to its unique geographic and topographic variation (WWF 2013). The climate change related impacts have been ranked as a very high threat in CHAL with greater droughts and higher rainfall which could become significant threats for spread of diseases and pests in CHAL (MoFSC 2015). The climate change projections showed that most of the lower and mid-hill forests of CHAL are vulnerable to climate change causing extensive conversion into new vegetation types (Thapa et al. 2015). The numbers of IAPs have already been recorded in the landscape and are threatening native biodiversity (WWF 2013). Therefore, the aim of the present study is to identify and develop habitat suitability maps of *P. hysterophorus* under current and projected future climate scenarios in the CHAL. It will also help to identify the areas susceptible to invasion by *P. hysterophorus* in light of changing climate in three Representative Concentration Pathways (RCPs 2.6, 4.5 and 8.5). The result might be helpful to decision makers and resources managers while planning for effective management of this weed to reduce its spread in the CHAL and to minimize its ecological and economic impacts.

1 Method

1.1 Study area

CHAL lies in central Nepal and covers 32,090 km² area including 19 districts. Out of five physiographic zones of Nepal, it represents four physiographic regions – Siwalik (200-1500 m asl), Middle Mountain (1000-2500 m asl), High Mountain (2200-4000 m asl) and High Himalaya (>4000 m asl) (DHM 2017). It has a large elevational gradient from 200 m to >8000 m asl and diverse climatic conditions including tropical, subtropical, temperate, subalpine, and alpine climate. It also comprises three National Parks – Chitwan (also a World Natural Heritage site), Parsa, and Langtang, and two conservation areas – Annapurna and Manaslu (Figure 1B). Due to its unique geographic and topographic variations, it possesses rich biodiversity. Chitwan National Park and Annapurna Conservation Area are iconic protected areas that are globally renowned for their biodiversity. The forested watersheds present in CHAL are important to sustain natural ecological communities, livelihood and economy but the continued fragmentation of subtropical and temperate forests of CHAL can result in decline of species populations and degradation of ecological services (Thapa et al. 2015). The number of naturalized plants is found higher in central Nepal (where the CHAL is located) than in eastern and western parts of Nepal (Bhattarai et al. 2014). Out of 20 IAPs reported in CHAL, *P. hysterophorus* is prioritized as the second and fourth most important invasive weed in the agroecosystem and natural ecosystem respectively by local people in the lowland of CHAL (Shrestha et al. 2019). It is expanding rapidly in urban, peri urban, and now towards the natural habitats including protected areas of CHAL with negative impacts on forage supply, and human and animal health (Shrestha 2012; Shrestha et al. 2015). The increase in temperature due to changing climate leads to encroachment of IAPs, which will ultimately affect the biodiversity of CHAL (WWF 2013).

1.2 Preparation of datasets

1.2.1 Species occurrence data

A total of 318 occurrence points of *P. hysterophorus* in the CHAL were obtained from the previous studies (Shrestha 2014; Shrestha et al. 2016; Siwakoti et al. 2016) and additional 90 occurrence points were collected by one of the

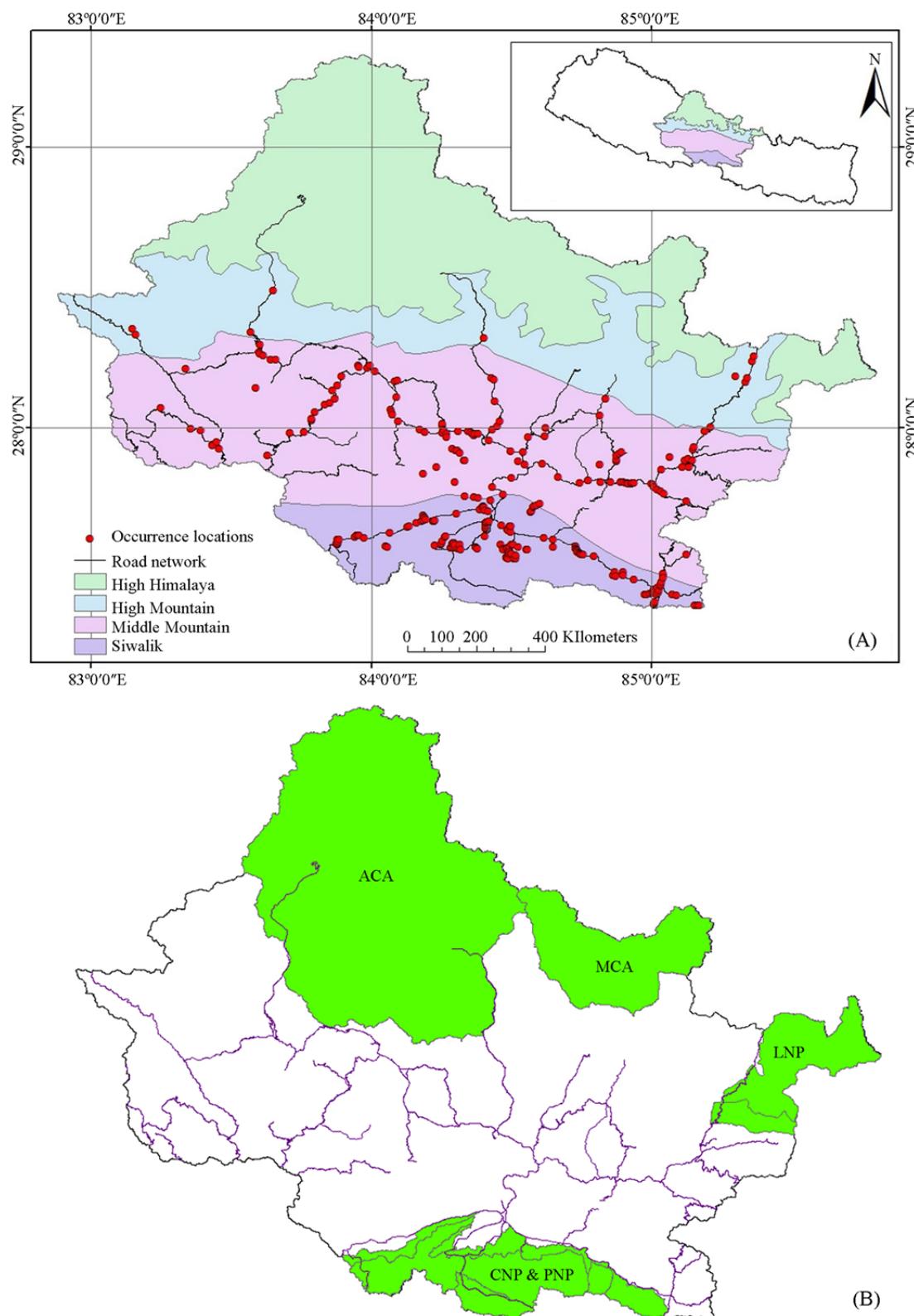


Figure 1 Study area showing (A) Physiographic regions of CHAL with occurrence points (B) Major protected areas in CHAL with buffer zone and their linkage to road network (ACA - Annapurna Conservation Area, MCA - Manaslu Conservation Area, CNP - Chitwan National Park, PNP - Parsa National Park, LNP - Langtang National Park).

(SM) in 2016 - 2018. Occurrence points were mainly collected through surveys along road networks and in residential and urban areas. Therefore, the occurrence points could potentially be biased due to easily accessible or near roadside records, which may influence the model performance by causing spatial autocorrelation (Boria et al. 2014). To remove the spatial autocorrelation and sampling bias, spatial thinning was done in ArcGIS by retaining single occurrence point randomly within $1 \times 1 \text{ km}^2$ grid cells to match the spatial resolution of environmental variables (Elith et al. 2010). Spatial autocorrelation test performed in ArcGIS using Global Moran's I tool expressed the pattern of the occurrence points as the random ($z\text{-score}=1.03$ and $p\text{-value}=0.30$) (Appendix 1). The spatial thinning process resulted in 288 out of 408 collected presence records which were used to predict suitable habitat of *P. hysterophorus* under current and projected future climate (Figure 1A).

1.2.2 Variables selection

Nineteen grid based bioclimatic variables of current climatic conditions (~1960-1990) with spatial resolution of 30 arc sec. (~ 1km^2) (Appendix 2) were downloaded from worldclim datasets (<http://www.worldclim.org/version1>) (Hijmans et al. 2005). Elevation plays an important role in determining the species distribution at topo-scale in a country like Nepal where there is diverse topographic variation (Mackey and Lindenmayer 2001). Hence, Digital Elevation Model (DEM) was obtained from Shuttle Radar Topographic Mission (SRTM). Slope and aspect are important variables that cause variation in topography in determining microclimatic conditions (Bennie et al. 2008). So, these two variables were derived from DEM using ARCGIS. To reduce multicollinearity effect among the bioclimatic variables (Merow et al. 2013), a correlation matrix was built in ArcMap 10.3 using band collection statistics tool within pairs of variables and highly correlated variables having $r^2 \geq 0.8$ value as cut-off threshold were removed (Appendix 3). Hence, six bioclimatic variables - (two temperature variables - Bio2 (Mean Diurnal Range), Bio3 (Isothermality), Bio12 (Annual precipitation); four precipitation variables - Bio14 (Precipitation of Driest Month), Bio15 (Precipitation Seasonality), Bio17 (Precipitation of

Driest Quarter)) and two topographic variables (aspect and slope) were selected for MaxEnt modelling after multicollinearity test (Appendix 3). The main aim of the current study was to examine the effect of bioclimatic and nonchanging variables (altitude, aspect and slope) on the future probable suitable habitat of *P. hysterophorus*. The other anthropogenic variables like road network, landuse, and human population density that might have influences in the current distribution of this weed were excluded in this study due to the unavailability of the projected data for these variables in future.

A global circulation model, Community Climate System Model (CCSM4) adopted by Intergovernmental Panel on Climate Change (IPCC) - AR5, was selected based on an average annual change in means and ranks for Gangetic plain (Lutz et al. 2016) for the projections and to predict the suitable habitat of the species under different projected future climate scenarios. We have chosen three future Greenhouse Gas (GHG) emission scenarios, Representative Concentration Pathways (RCP 2.6, RCP 4.5 and RCP 8.5) for two different time periods (2050 and 2070). RCP 2.6 represents strict mitigating GHG pathway in which there will be a very low increase in radiative forcing and the projected change in global mean temperature is 1°C ; RCP 4.5 represents the intermediate Greenhouse Gas (GHG) emission pathway in which there will be steady increase in radiative forcing and projected change in global mean surface temperature is $1.4\text{-}1.8^\circ\text{C}$; and RCP 8.5 represents the continuous GHG emission pathway in which there will be a highest increase in radiative forcing and projected change in global mean surface temperature is $2^\circ\text{C}\text{-}3.7^\circ\text{C}$ (IPCC 2014). Even though RCP 2.6 is now considered unfeasible (Sanford et al. 2014), it is used in the present study as a control scenario assuming that it will provide information on how more efficient measures against global warming could have influenced to predict the future suitable habitat of *P. hysterophorus*. We downloaded all above data from worldclim datasets.

Raster layers for CHAL were extracted from all the above downloaded present and future scenarios in ArcGIS. Similarly, raster layers for CHAL were prepared for slope and aspect too. All the datasets were converted into ascii raster files with a cell size

of 30 arc sec. ($\sim 1\text{km}^2$) in ArcGIS.

1.3 Modelling

We used an open source software Maximum Entropy (MaxEnt, version 3.4.1) to predict the current and future suitable habitat of *Parthenium hysterophorus*. Logistic output was used to improve model calibration (Phillips and Dudik 2008). The random test percentage was set to 25, which means 75% of the occurrence points were used for training the model and the remaining 25% were used for testing the model. One of the advantages of MaxEnt modelling is that it does not require absence data; instead, it uses background environmental data for the entire study area by setting maximum number of background points to 10,000 (Merow et al. 2013). The number of replicates was set to 15 and the averaged model was built across all the replicates to generate the result. The model was run in subsample replicate type with 5,000 maximum iterations and 10 percentile training presence threshold rule and background predictions was selected to calculate various evaluation statistics. The remaining parameters were run with default settings (regularization multiplier – 1, convergence threshold – 0.00001). The remaining parameters were run with default settings. The resulting maps, showing the predicted probability of species presence for each raster cell of the study area for present and all future climate change scenarios, were converted into binary ‘presence–absence’ maps in ArcGIS through “Reclassify” tool, using 10 percentile training presence threshold value to define suitable and unsuitable areas for species. It was considered as a liberal prediction of species presence and tolerance to the environmental variables by incorporating a larger predicted area and it provides ecologically more significant results for invasive species (Liu et al. 2005; Pearson et al. 2007; Qin et al. 2014).

The suitable maps for current and all future climate scenarios were extracted. The range shift in terms of gain, loss, and stable area for all future climate scenarios compared to current suitable area were extracted from suitable map. Shifts in uppermost elevation limit of *P. hysterophorus* with future climate change scenarios compared to current was obtained by extracting the elevation data from DEM raster for all maps. Similarly, the

change in area in different physiographic regions of CHAL - the Siwalik Hills, the Middle Mountains, the High Mountains and the High Himalayas and major protected areas (Chitwan, Parsa and Langtang National Parks, Annapurna and Manaslu Conservation Area) with respect to current scenarios, were calculated in all future climate scenarios.

1.3.1 Model evaluation

The performance of all models was evaluated through evaluation metrics - Area under Receiver Operating Characteristics (ROC) Curve (AUC), accuracy, sensitivity, specificity, prevalence and True Skill Statistic (TSS) (Swets 1988; Fielding and Bell 1997; Allouche et al. 2006). To calculate all of the above evaluation metrics, MaxEnt output files (sample prediction and background prediction) were processed in ARCGIS to examine error matrix value for all the model outputs (Table 1).

Table 1 Error matrix for the model outputs in different scenarios

$N = a + b + c + d$	Predicted distribution	
	Presence	Absence
	True positive (a)	False positive (b)
Actual distribution	Current	226
	RCP 2.6 2050	231
	RCP 2.6 2070	231
	RCP 4.5 2050	243
	RCP 4.5 2070	239
	RCP 8.5 2050	190
	RCP 8.5 2070	215
Absence	False negative (c)	True negative (d)
	Current	26
	RCP 2.6 2050	21
	RCP 2.6 2070	21
	RCP 4.5 2050	9
	RCP 4.5 2070	13
	RCP 8.5 2050	62
	RCP 8.5 2070	37

Note: N is total number of events to evaluate different performance matrix scores. N, a, b, c and d were used to calculate evaluation statistics.

2 Results

2.1 Models performance and variables' response

The output model performance evaluated by different performance matrix scores is given in

Table 2 Performance of the output model under different evaluation statistics

Evaluation statistics	Current	RCP 2.6		RCP 4.5		RCP 8.5	
		2050	2070	2050	2070	2050	2070
Prevalence (%)	2.46	2.46	2.46	2.46	2.46	2.46	2.46
Accuracy (%)	80.28	79.75	77.36	70.36	71.76	86.31	82.22
Area under ROC curve (AUC)	0.902	0.908	0.901	0.903	0.899	0.899	0.904
True Skill Statistics (TSS)	0.697	0.711	0.687	0.661	0.66	0.62	0.675

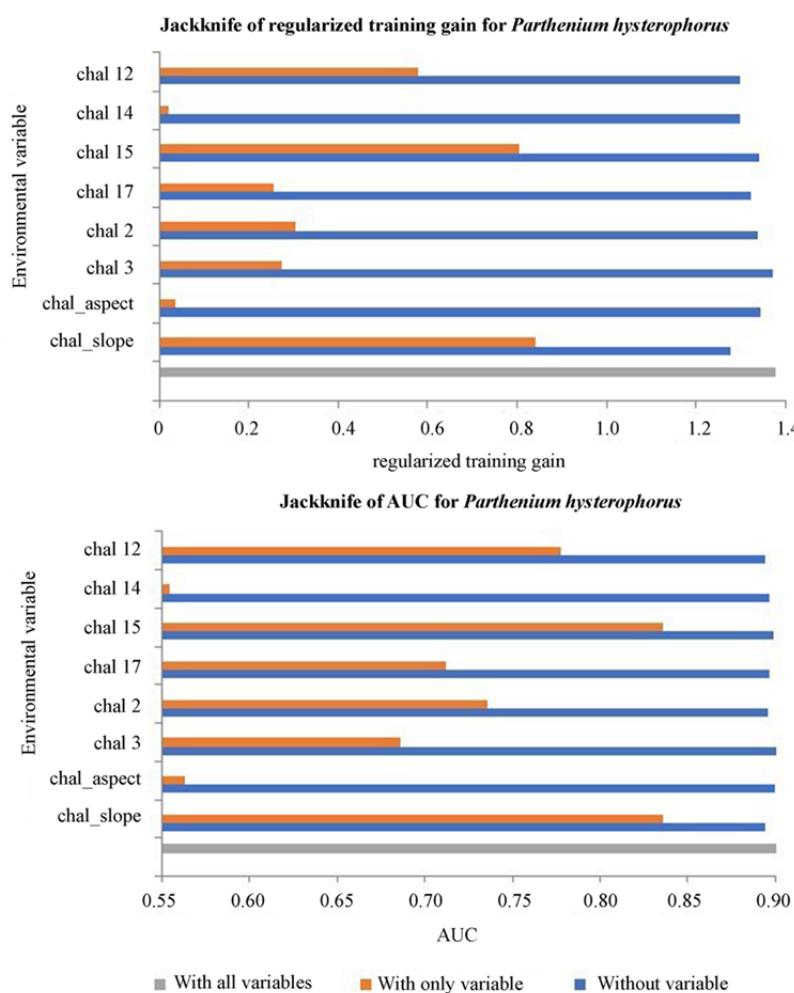
**Figure 2** Jackknife test of variable's contribution in current distribution of *P. hysterophorus*. AUC means Area under Receiver Operating Characteristics (ROC) Curve.

Table 2. AUC and TSS value ranged from 0.89 to 0.9, and 0.63 to 0.69, respectively. Prevalence of all models was 2.6 and accuracy was between 71% to 87%.

Out of eight predictor variables, slope, annual precipitation (Bio 12) and mean diurnal range (Bio2) represented the top three contributing variables, with 43%, 34%, and 8% of the total contribution, respectively, to the MaxEnt model for current predicted suitable habitat (**Table 3**). The results of the Jackknife test of variables'

Table 3 Relative contribution of variables to the Maxent model

Variable	Percent of contribution
Slope	42.8
Bio12 (Annual Precipitation)	33.9
Bio2 (Mean Diurnal Range)	7.7
Bio15 (Precipitation Seasonality)	4.9
Bio14 (Precipitation of Driest Month)	4.1
Bio17 (Precipitation of Driest Quarter)	2.9
Aspect	1.9
Bio3 (Isothermality)	1.8

contribution (Figure 2) also showed that slope had the highest training gain and the highest AUC value compared to other variables when used independently. Similarly, the gain decreased when the ‘slope’ was omitted indicating that the ‘slope’ contained more useful information by itself. The response curve showed that the suitability of *P. hysterophorus* is high in flat areas and the suitability decreased with increase in slope of the area. Similarly, suitability increases with increase in annual precipitation (Bio12) and mean diurnal range (Bio2) (Figure 3). The slope of all occurrence points that were used for model building was below 21.5° , and 90.6% of sample points were lower than 10° (Appendix 4).

2.2 Habitat suitability in current climate and future climate change scenario

Currently, 17 out of 19 districts of CHAL had suitable habitat for *P. hysterophorus*. The habitat suitability map showed 20% of the total area of CHAL is suitable for the growth of *P. hysterophorus* in current climatic conditions (Table 4, Figure 4A). The climatically suitable habitat for the growth of *P. hysterophorus* differs in different GHG emission scenarios. The highest increase (10%) in suitable habitat has been predicted for RCP 4.5 scenario in both 2050 and 2070, but there will be decrease in suitable habitat in RCP 8.5 scenario for both years (Table 4). The effect of future climate on the predicted range of *P. hysterophorus* in CHAL was very noticeable. Area gain increases with increase in radiative forcing from 2.6 to 4.5, whereas area loss is seen in RCP 8.5 in both 2050 and 2070. Area gain is higher in the western region of CHAL than in eastern region for all future climate change scenarios (Figure 4). Some districts of western CHAL like Arghakhanchi, Gulmi, Parbat, Myagdi, Baglung, Syangja might gain suitable habitat for *P. hysterophorus* in RCP 4.5 scenario.

The present occurrence records of *P. hysterophorus* were found within elevational range between 142 and 2032 m asl (Appendix 4) but the current MaxEnt prediction showed the suitable habitat of this weed within an elevational range from 120 to 2490 m. The uppermost suitable elevation limit for *P. hysterophorus* will expand in RCP 2.6 and 4.5 but it will significantly decrease in RCP 8.5. With the changing climate, the weed may

Table 4 Percentage change in suitable area in different climatic scenario

Scenario	Suitable area (km ²)	Area covered in CHAL (%)	Change (%)
current	6374.52	19.88	
RCP 2.6 2050	6444.22	20.1	+0.22
RCP 2.6 2070	7402.80	23.08	+3.21
RCP 4.5 2050	9696.44	30.24	+10.36
RCP 4.5 2070	9299.26	29	+9.12
RCP 8.5 2050	4198.61	13.09	-6.79
RCP 8.5 2070	5564.72	17.35	-2.53

Notes: + gain, - loss; Change: Change in suitable area in CHAL compared to current prediction (%).

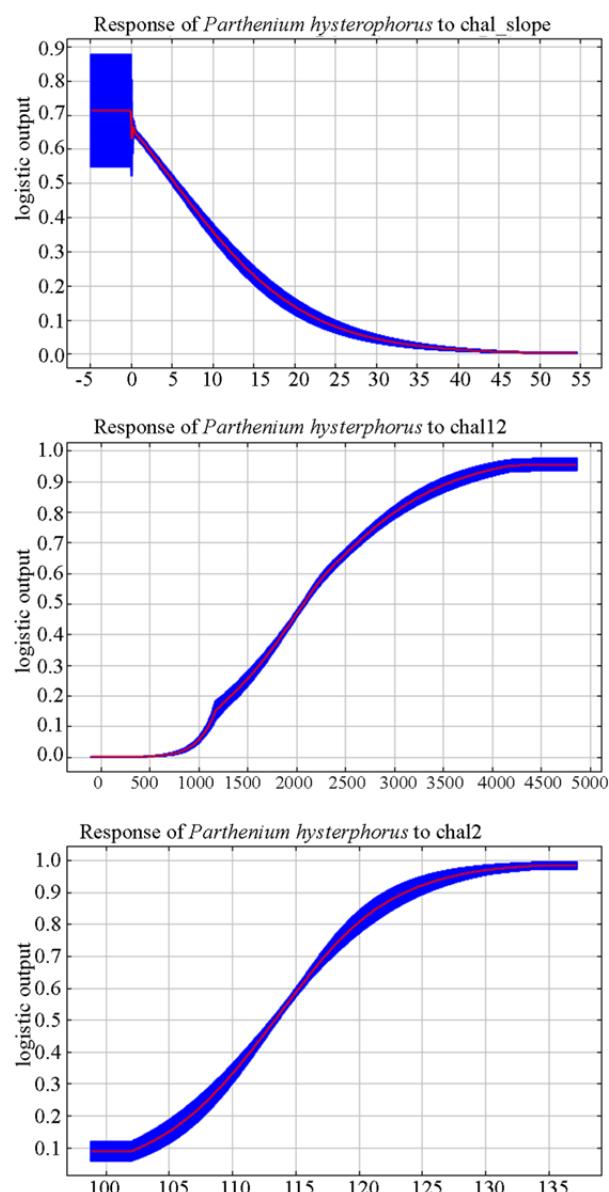


Figure 3 Response curve for probability of presence obtained for top three contributing variables slope, Bio12 (Annual precipitation) and Bio2 (Mean Diurnal Range) for *P. hysterophorus* for current condition.

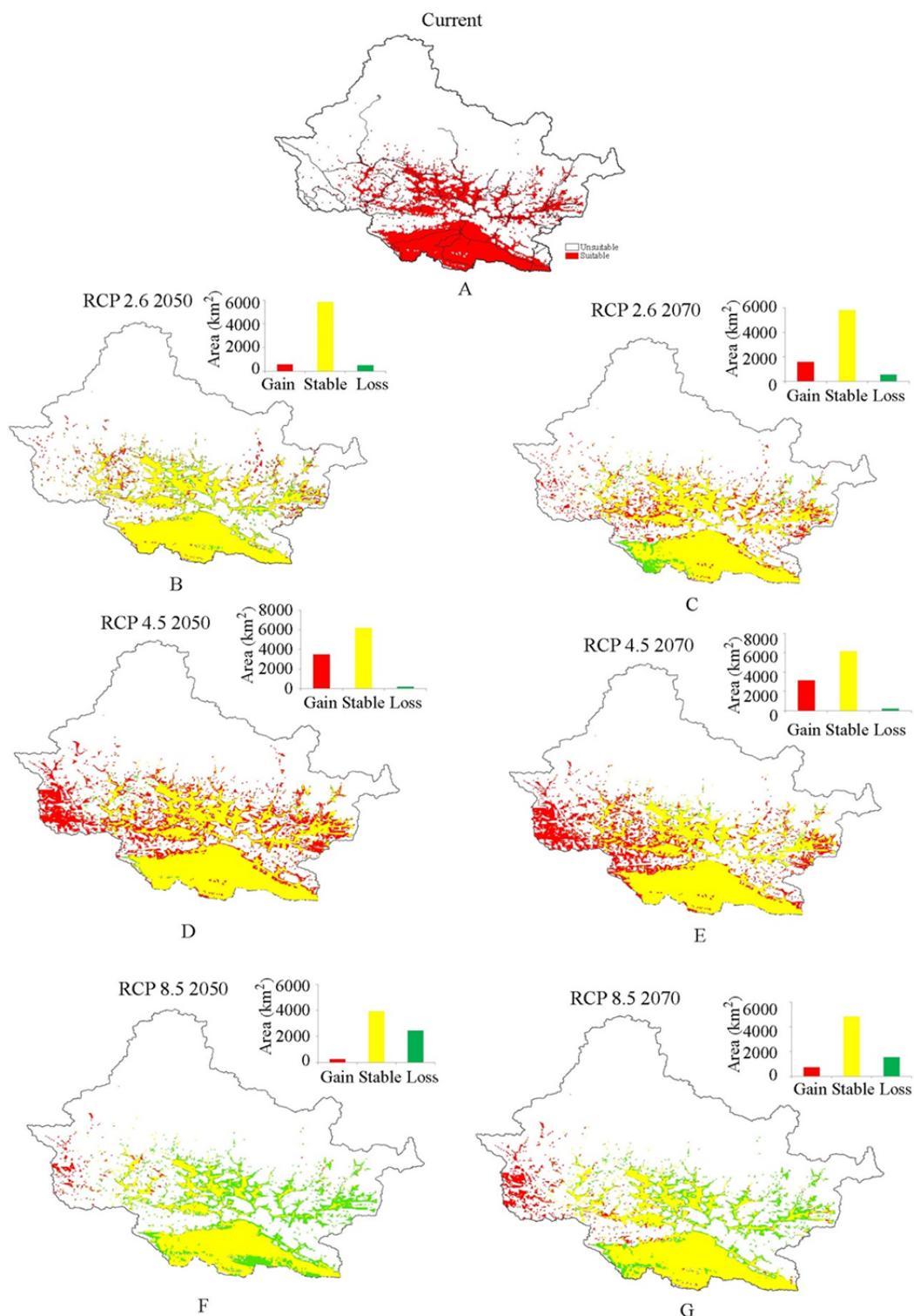


Figure 4 Current predicted suitable habitat of *P. hysterophorus* in CHAL and their linkage to road network (A), Range shifts for future climate change projected scenarios (map of the range shifts resulted from future projected scenarios – RCP 2.6, 4.5 and 8.5 for 2050 and 2070 (B - G). Area gained by *P. hysterophorus* is shown in red, stable areas are shown in yellow and areas lost are shown in green colour.

move upslope by 30 m in 2070 (RCP 2.6 and 4.5) compared to current uppermost elevation limit ([Figure 5](#)).

Out of four physiographic regions present in CHAL, three regions - Siwalik, Middle Mountain and High Mountain have suitable habitat for *P. hysterophorus* in current predicted scenario. The highest proportion of area in CHAL that is predicted to be suitable for *P. hysterophorus* under current climate lies in Siwalik region followed by Middle Mountain. A similar trend is seen in all future climate change scenarios except in RCP 2.6 for the year 2070 and 4.5 for both years ([Table 5](#)). However, with the changing climate when we compared the total suitable area in each physiographic region, suitable area will decline in Siwalik region in RCP 2.6 and 8.5 scenarios for both years, but it will increase in 2050 and 2070 under RCP 4.5 scenario. There will be an increase in suitable area of *P. hysterophorus* in mountainous region in two future climate change scenarios (RCP 2.6 and 4.5). Similar decreasing trend in suitable area is predicted for the year 2050 and 2070 in RCP 8.0 scenario. Middle Mountains seem to have more suitable habitat with changing climate in the future ([Figure 6](#)). Suitable habitat for *P. hysterophorus* is predicted in all protected areas of CHAL under current climatic condition. Almost all (97%) of the total area of Chitwan and Parsa National Park is found suitable for *P. hysterophorus*. In future climate change scenario, the mountainous protected areas (Langtang National Park, Annapurna Conservation Area and Manaslu Conservation Area) will have more suitable habitat for *P. hysterophorus* particularly in RCP 2.6 and 4.5 ([Table 6](#)). Manaslu conservation area is the only protected area in CHAL of which very negligible area (0.07% of total area) have suitable habitat for *P. hysterophorus* in current climatic condition. However, the change in climate in the future might develop more suitable habitat

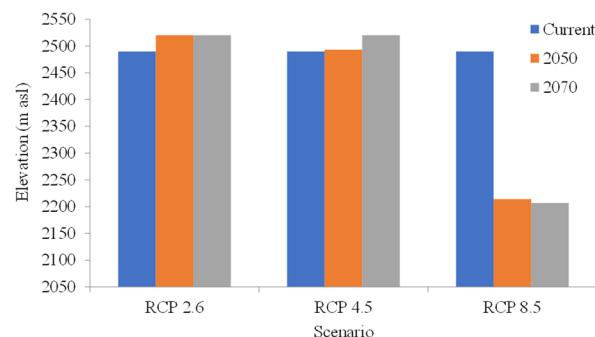


Figure 5 Shifts in uppermost elevation limit of *P. hysterophorus* with the future changing climate scenarios.

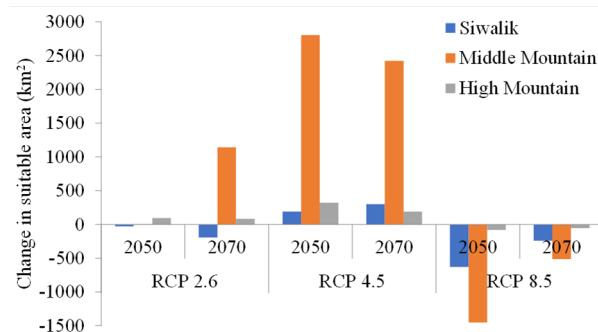


Figure 6 Change in suitable area in different physiographic regions of CHAL with changing climate in different Representative Concentration Pathways (RCP) scenarios.

for this weed ([Table 6](#)). Similarly, more areas in Langtang National Park will be affected.

3 Discussion

The model performance evaluation matrices indicated that all the habitat suitability models developed for *P. hysterophorus* performed better than random. Prevalence of all models was 2.6 % indicating 97.4% correct classification rate of the models ([Fielding and Bell 1997](#)). The model performance showing accuracy rate closer to 100%, TSS value closer to one and AUC value within the

Table 5 Total suitable area in terms of percentage in different physiographic regions of CHAL

Scenario	Total suitable area (km ²)			% Suitable area (of total suitable area in CHAL)		
	Siwalik	Middle Mountain	High Mountain	Siwalik	Middle Mountain	High Mountain
Current	3239.62	2950.86	162.98	50.82	46.29	2.56
RCP 2.6 2050	3209.46	2956.02	256.16	49.80	45.87	3.98
RCP 2.6 2070	3044.99	4091.94	246.95	41.13	55.28	3.34
RCP 4.5 2050	3430.31	5756.22	483.04	35.38	59.36	4.98
RCP 4.5 2070	3538.73	5373.75	353.74	38.05	57.79	3.80
RCP 8.5 2050	2609	1496.99	80.18	62.14	35.65	1.91
RCP 8.5 2070	2998.14	2438.66	107.68	53.88	43.82	1.93

Table 6 Change in the suitable area within protected areas of CHAL

Scenario	Suitable area for <i>P. hysterophorus</i> (km ²)			
	Langtang National Park and buffer zone (1225.56)	Annapurna Conservation Area (7754)	Manaslu Conservation Area (1645.99)	Chitwan-Parsa National Park and buffer zone (1941.29)
current	60.40	22.67	1.07	1878.03
RCP 2.6 2050	74.13	27.77	2.95	1874.22
RCP 2.6 2070	69.31	29.72	6.56	1707.59
RCP 4.5 2050	99.49	55.15	13.35	1905.82
RCP 4.5 2070	90.58	31.69	1.55	1926.49
RCP 8.5 2050	12.06	9.38	0	1524.78
RCP 8.5 2070	14.18	10.72	0	1755.41

Note: Total area of protected area is given in parenthesis. Increased areas with respect to current scenario are made bold.

range 0.8-0.9 is considered good (Swets 1988; Allouche et al. 2006). Hence, our model has shown a good performance based on the three performance matrices evaluated.

The impact of bioclimatic variables may be conditional on local topography because of the modification of local climate due to slope and aspect hence is significant for improved SDMs assessing the effect of climate change at local scale (Lassueur et al. 2006; Austin and Niel 2011). The slope of topography had the highest contribution in the predicting the suitable habitat of *P. hysterophorus* in current as well as in the future climate change scenario. This indicates that *P. hysterophorus* prefers to grow in plain land and the suitability decreases with increased slope of the land. The steeper slopes might act as buffer to some extent against the invasion probably due to higher radiation and drought events (Bennie et al. 2006). In addition, Nepal being a mountainous and agriculture dependent country where majority of people in the mountains rely on agriculture, more than 80% of land of central Nepal is under terrace cultivation by reducing the terrain slope (Neupane and Thapa 2001; Paudel and Thapa 2001). Terraced land which is no longer cropped is usually grazed and has the high probability of introduction of invasive species (Douglas et al. 1994; Lasanta et al. 2013). The seed bank record of *P. hysterophorus* (200,000 m⁻²) in the abandoned land in India by Joshi (1991) is probably the highest record for this species. Hence, its possible spread because of abandoned terraced land in mountainous countries like Nepal should be considered while preparing management plans.

With the changing climate, the projected temperature as well as precipitation in western

region of Nepal is comparatively higher (OECD 2003; NCVST 2009), leading to more favourable climatic conditions for the growth of invasive species than under current climate. Under laboratory conditions, elevated atmospheric CO₂ favours the growth of *P. hysterophorus* (Khan et al. 2014; Shabbir et al. 2014). These findings further suggest that global warming resulting from climate change is likely to facilitate invasion of *P. hysterophorus* into new areas. All of these observations stress the importance of implementation of an early detection and eradication plan of *P. hysterophorus* within the suitable areas where the weed has not yet spread widely. Adverse effects due to alien invasive species have been identified as one of the major biodiversity conservation issues in CHAL (WWF 2013). At present, *P. hysterophorus* is one of the problematic weeds in natural and agroecosystems of the Middle Mountain area in CHAL (Shrestha et al. 2019) and the situation may worsen in future due to climate change. The suitable habitat of *P. hysterophorus* will decrease by 6.79% in RCP 8.5 scenario. Lamsal et al. (2018) also found a decrease in suitable habitat by 7.44% in Himalayan range in RCP 8.5 scenario. Global mean surface temperature and annual mean precipitation will likely increase under the RCP 8.5 scenario (IPCC 2014). Hence, extreme high temperature (>40°C) and increased precipitation might be the limiting factor for the growth of *P. hysterophorus* due to compromised physiological activities and competition from other grass cover (Doley 1977; Dale 1981; Pandey et al. 2003) that may lead to the decrease in suitable habitat in extreme climate change scenarios. There is the difference in elevational range between occurrence records of *P.*

hysterophorus (142-2032 m asl) and current MaxEnt projection (119-2490m) indicating that the full extent of the distribution has probably not yet been achieved and there is also the report of this weed up to 2600 m asl in other parts of the world (McConnachie et al. 2011). Hence, there is still a chance of spread of this weed upslope in CHAL if there is not barrier for seed dispersal. The future changing climate combined with human interference may exacerbate the spread of this weed towards higher elevation.

The predicted current suitable map showed that the Siwalik region with elevation range from 200 to 1500 m asl (DHM 2017) has more suitable areas in CHAL and it is also the suitable elevational range (100-1600 m asl) for *P. hysterophorus* in its native range (Dale 1989). The scatter plot between the elevation and mean annual precipitation of the occurrence points used for modelling also showed that most of the *P. hysterophorus* presence records are confined to lower elevational range (below 700 m asl). A similar elevational range (below 500m asl) for the majority of *P. hysterophorus* has been reported from other parts of the world (McConnachie et al. 2011). *P. hysterophorus* spreads quickly along the roadsides due to vehicular movement (Bajwa et al. 2018). Invasive plants thrive along roadsides and spread quickly due to disturbance. So quick dispersal of seeds of these species with the help of vehicle over the long distances should be some of the earliest species to shift their ranges with the changing climate (Dukes and Mooney 1999). The protected areas, which are directly connected to the road network of Nepal (Chitwan National Park, Langtang National Park and Annapurna Conservation Area) are in the high risk of invasion by *P. hysterophorus*. Large area of Chitwan National Park being in the tropical (Siwalik) region have suitable habitat for *P. hysterophorus*. In the case of mountainous protected areas, Langtang National Park and Annapurna conservation area, increase in suitable habitat for *P. hysterophorus* in two RCP scenario might be further exacerbate due to their connectivity to the roads.

References

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true

4 Conclusion

Our study suggests that the suitable habitat of *P. hysterophorus* in CHAL will expand by 10% in the future, causing potential threats to the native vegetation. These findings further suggest that global warming resulting from climate change is likely to facilitate invasion of this weed into new areas. All of these observations stress the importance of implementation of an early detection and eradication plan of *P. hysterophorus* within the suitable areas. Therefore, our study could be helpful to assess the risk of spread of *P. hysterophorus* from present localities to non-invaded areas in the current and future climatic conditions and would help the scientific community and policy makers in planning and effective management to reduce the ecological and economic impacts of the weed in CHAL.

Acknowledgements

This study was made possible through support provided by the Feed the Future Innovation Lab for Integrated Pest Management of the U.S. Agency for International Development, under the terms of Cooperative Agreement No. AID-OAA-L-15-00001. Some of the occurrence data used in this manuscript were collected by BBS during field works supported by International Foundation for Science (Sweden), Nepal Academy of Science and Technology (Nepal), and National Trust for Nature Conservation (Nepal). We are thankful to Dr. DN Shah for his guidance during data preparation. Thanks to Ms. Sara Hendery for editing the language of this manuscript.

Electronic supplementary material: Supplementary material (Appendices 1-4) are available in the online version of this article at <https://doi.org/10.1007/s11629-019-5548-y>

skill statistic (TSS). Journal of Applied Ecology 43: 1223-1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>

- Austin MP, Niel KP Van (2011) Improving species distribution models for climate change studies: Variable selection and scale. *Journal of Biogeography* 38: 1-8.
<https://doi.org/10.1111/j.1365-2699.2010.02416.x>
- Bajwa AA, Nguyen T, Navie S, et al. (2018) Weed seed spread and its prevention: The role of roadside wash down. *Journal of Environmental Management* 208: 8-14.
<https://doi.org/10.1016/j.jenvman.2017.12.010>
- Bennie J, Hill MO, Baxter R, et al. (2006) Influence of slope and aspect on long-term vegetation change in British chalk grasslands. *Journal of Ecology* 94: 355-368.
<https://doi.org/10.1111/j.1365-2745.2006.01104.x>
- Bennie J, Huntley B, Wiltshire A, et al. (2008) Slope, aspect and climate: Spatially explicit and implicit models of topographic microclimate in chalk grassland. *Ecological Modelling* 216: 47-59. <https://doi.org/10.1016/j.ecolmodel.2008.04.010>
- Bhattarai KR, Måren IE, Subedi SC (2014) Biodiversity and invasibility: Distribution patterns of invasive plant species in the Himalayas, Nepal. *Journal of Mountain Science* 11: 688-696. <https://doi.org/10.1007/s11629-013-2821-3>
- Boria RA, Olson LE, Goodman SM, et al. (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling* 275: 73-77.
<https://doi.org/10.1016/j.ecolmodel.2013.12.012>
- Bradley BA, Wilcove DS, Oppenheimer M (2010) Climate change increases risk of plant invasion in the Eastern United States. *Biological Invasions* 12: 1855-1872.
<https://doi.org/10.1007/s10530-009-9597-y>
- Dale IJ (1989) Parthenium weed in the Americas: A report on the ecology of *Parthenium hysterophorus* in South, Central and North America, in: Australian Weeds, pp 8-14.
- DHM (2017) Observed climate trend analysis in the districts and physiographic zones of Nepal. Department of Hydrology and Meteorology, Government of Nepal, Kathmandu.
- Doley D (1977) Parthenium weed (*Parthenium hysterophorus* L.): Gas exchange characteristics as a basis for prediction of its geographical distribution. *Australian Journal of Agricultural Research* 28: 449-460.
<https://doi.org/10.1071/AR9770449>
- Douglas TD, Kirkby SJ, Critchley RW, et al. (1994) Agricultural terrace abandonment in the Alpujarra, Andalucia, Spain. *Land Degradation and Rehabilitation* 5: 281-291.
- Dukes JS, Mooney HA (1999) Does global change increase the success of biological invaders? *Tree* 14: 135-139.
- Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution* 1: 330-342. <https://doi.org/10.1111/j.2041-210X.2010.00036.x>
- Fielding AH, Bell JF (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24: 38-49.
- Franklin J (2009) Mapping species distributions: spatial inference and prediction. Cambridge University Press.
- Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecological Letters* 8: 993-1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>
- Haseler WH (1976) *Parthenium hysterophorus* L. in Australia. *PANS* 22: 515-517.
<https://doi.org/10.1080/09670877609414342>
- Hellmann JJ, Byers JE, Bierwagen BG, et al. (2008) Five potential consequences of climate change. *Conservation Biology* 22: 534-543.
<https://doi.org/10.1111/j.1523-1739.2008.00951.x>
- Hijmans RJ, Cameron SE, Parra JL, et al. (2005) Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* 25: 1965-1978.
<https://doi.org/10.1002/joc.1276>
- IPCC (2014) Climate Change 2014: Synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland.
- Jimenez-Valverde A, Peterson AT, Soberon J, et al. (2011) Use of niche models in invasive species risk assessments. *Biological Invasions* 13: 2785-2797.
<https://doi.org/10.1007/s10530-011-9963-4>
- Joshi S (1991) Biocontrol of *Parthenium hysterophorus* L. *Crop Protection* 10:429-431.
- Khan N, George D, Shabbir A, et al. (2014) Rising CO₂ can alter fodder-weed interactions and suppression of *Parthenium hysterophorus*. *Weed Research* 55: 113-117.
<https://doi.org/10.1111/wre.12127>
- Lamsal P, Kumar L, Aryal A, et al. (2018) Invasive alien plant species dynamics in the Himalayan region under climate change. *Ambio* 47: 697-710.
<https://doi.org/10.1007/s13280-018-1017-z>
- Lasanta T, Arnaez J, Flano PR, et al. (2013) Agricultural terraces in the Spanish mountains: An abandoned landscape and a potential resource. *Boletin de la Asociacion de Geografos Espanoles* 63: 487-492.
- Lassueur T, Joost S, Randin CF (2006) Very high resolution digital elevation models: Do they improve models of plant species distribution? *Ecological Modelling* 198: 139-153.
<https://doi.org/10.1016/j.ecolmodel.2006.04.004>
- Liu C, Berry PM, Dawson TP, et al. (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography (Cop.)* 28: 385-393.
- Liu Y, Oduor AMO, Zhang Z, et al. (2016) Do invasive alien plants benefit more from global environmental change than native plants? *Global Change Biology* 23: 3363-3370.
<https://doi.org/10.1111/gcb.13579>
- Lutz AF, ter Maat HW, Biemans H, et al. (2016) Selecting representative climate models for climate change impact studies: An advanced envelope-based selection approach. *International Journal of Climatology* 36: 3988-4005.
<https://doi.org/10.1002/joc.4608>
- Mackey BG, Lindenmayer DB (2001) Towards a hierarchical framework for modelling the spatial distribution of animals. *Journal of Biogeography* 28: 1147-1166.
<https://doi.org/10.1046/j.1365-2699.2001.00626.x>
- Mainali KP, Warren DL, Dhileepan K, et al. (2015) Projecting future expansion of invasive species: Comparing and improving methodologies for species distribution modeling. *Global Change Biology* 21: 4464-4480.
<https://doi.org/10.1111/gcb.13038>
- McConnachie AJ, Strathie LW, Mersie W, et al. (2011) Current and potential geographical distribution of the invasive plant *Parthenium hysterophorus* (Asteraceae) in eastern and southern Africa. *Weed Research* 51: 71-84.
<https://doi.org/10.1111/j.1365-3180.2010.00820.x>
- Merow C, Smith MJ, Silander JA (2013) A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. *Ecography* 36: 1058-1069.
<https://doi.org/10.1111/j.1600-0587.2013.07872.x>
- MoFSC (2015) Strategy and action plan 2016-2025, Chitwan-Annapurna Landscape, Nepal. Ministry of Forests and Soil Conservation, Singha Durbar, Kathmandu, Nepal.
- NCVST (2009) Vulnerability through the eyes of the vulnerable: Climate change induced uncertainties and Nepal's development predicaments. Institute for Social and Environmental Transition-Nepal (ISET-N), Nepal Climate Vulnerability Study Team (NCVST) Kathmandu National Disaster Report. UNDP-Nepal.
- Neupane RP, Thapa GB (2001) Impact of agroforestry intervention on soil fertility and farm income under the subsistence farming system of the middle hills, Nepal. *Agriculture, Ecosystems and Environment* 84: 157-167.
- OECD (2003) Development and climate change in Nepal: Focus on water resources and hydropower. Organization for Economic Cooperation and Development.
- Paini DR, Sheppard AW, Cook DC, et al. (2016) Global threat to agriculture from invasive species. *PNAS* 113: 7575-7579.
<https://doi.org/10.1073/pnas.1602205113>
- Pandey DK, Palni LMS, Joshi SC (2003) Growth, reproduction, and photosynthesis of ragweed parthenium (*Parthenium*

- hysterophorus*). Weed Science 52: 191-201.
- Paudel GS, Thapa GB (2001) Changing farmers' land management practices in the hills of Nepal. Environmental Management 28: 789-803.
<https://doi.org/10.1007/s002670010262>
- Pearson RG, Raxworthy CJ, Nakamura M, et al. (2007) Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. Journal of Biogeography 34: 102-117.
<https://doi.org/10.1111/j.1365-2699.2006.01594.x>
- Pejchar L, Mooney HA (2009) Invasive species, ecosystem services and human well-being. Trends in Ecology and Evolution 24: 497-504.
- Peterson AT, Viegas DA (2001) Predicting species invasions using ecological niche modeling: New approaches from bioinformatics attack a pressing problem. Bioscience 51: 363-371.
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling 190: 231-259.
<https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Phillips SJ, Dudik M (2008) Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. Ecography 31: 161-175.
<https://doi.org/10.1111/j.2007.0906-7590.05203.x>
- Qin Z, Ditomaso A, Wu RS, et al. (2014) Potential distribution of two *Ambrosia* species in China under projected climate change. Weed Research 54: 520-531.
<https://doi.org/10.1111/wre.12100>
- Ricciardi A, Neves RJ, Rasmussen JB (1998) Impending extinctions of North American freshwater mussels (Unionoida) following the zebra mussel (*Dreissena polymorpha*) invasion. Journal of Animal Ecology 67: 613-619.
<https://doi.org/10.1046/j.1365-2656.1998.00220.x>
- Sanford T, Frumhoff PC, Luers A, et al. (2014) The climate policy narrative for a dangerously warming world. Nature Climate Change 4: 164-166.
<https://doi.org/10.1038/nclimate2148>
- Shabbir A, Dhileepan K, Khan N, et al. (2014) Weed-pathogen interactions and elevated CO₂: Growth changes in favour of the biological control agent. Weed Research 54: 217-222.
<https://doi.org/10.1111/wre.12078>
- Shabbir A, Mcconnachie A, Adkins SW (2019) Spread. In: Adkins S, Shabbir A, Dhileepan K (ed), Parthenium Weed: Biology, Ecology and Management. CAB International. pp 40-56.
- Shrestha B, Shrestha U, Sharma K, et al. (2019) Community perception and prioritization of invasive alien plants in Chitwan Annapurna Landscape, Nepal. Journal of Environmental Management 229: 38-47.
<https://doi.org/10.1016/j.jenvman.2018.06.034>
- Shrestha BB (2016) Invasive alien plant species in Nepal. Frontiers of Botany 2016: 269-284.
- Shrestha BB (2014) Distribution of alien invasive weed *Parthenium hysterophorus* and its biological control agent in Nepal. A research report submitted to International Foundation for Science, Sweden (Unpublished).
- Shrestha BB (2012) Parthenium weed in Chitwan National Park, Nepal. International Parthenium News 5: 6-7.
- Shrestha BB, Kokh M, Karki JB (2016) Mapping of invasive alien plant species in Tarai Arc Landscape (TAL) and Chitwan - Annapurna Landscape (CHAL), Nepal. A report submitted to National Trust for Nature Conservation (NTNC), Kathmandu, Nepal (Unpublished).
- Shrestha BB, Shabbir A, Adkins SW (2015) *Parthenium hysterophorus* in Nepal: A review of its weed status and possibilities for management. Weed Research 55: 132-144.
<https://doi.org/10.1111/wre.12133>
- Shrestha UB, Sharma KP, Devkota A, et al. (2018) Potential impact of climate change on the distribution of six invasive alien plants in Nepal. Ecological Indicators 95: 99-107.
<https://doi.org/10.1016/j.ecolind.2018.07.009>
- Simberloff D (2000) Global climate change and introduced species in United States. Science of the Total Environment 262: 253-261.
- Siwakoti M, Shrestha BB, Devkota A, et al. (2016) Assessment of the effects of climate change on the distribution of invasive alien plant species in Nepal. A Report submitted to Nepal Academy of Science and Technology, Kathmandu (Unpublished).
- Sutherst RW, Maywald GF, Russell BL (2000) Estimating vulnerability under global change: modular modelling of pests. Agriculture, Ecosystems and Environment 82: 303-319.
- Swets JA (1988) Measuring the accuracy of diagnostic systems. Science 240: 1285-1293.
<https://doi.org/10.1126/science.3287615>
- Thapa S, Chitale V, Rijal SJ, et al. (2018) Understanding the dynamics in distribution of invasive alien plant species under predicted climate change in Western Himalaya. PLoS One: e0195752. <https://doi.org/10.1371/journal.pone.0195752>
- Thapa GJ, Wikramanayake E, Forrest J (2015) Climate-change impacts on the biodiversity of the Terai Arc Landscape and the Chitwan-Annapurna Landscape. Hariyo Ban, WWF Nepal, Kathmandu, Nepal.
- Thomas CD, Cameron A, Green RE, et al. (2004) Extinction risk from climate change. Nature 427: 145-148.
- Thuiller W, Richardson DM, Midgley GF (2007) Will climate change promote alien plant invasions? In: Nentwig W (ed.), Biological Invasions. pp 197-211.
<https://doi.org/10.1007/978-3-540-36920-2>
- Vitousek PM, D'Antonio CM, Loope LL, et al. (1997) Introduced species: A significant component of human-caused global change. New Zealand Journal of Ecology 21:1-16.
- WWF (2013) Chitwan-Annapurna Landscape biodiversity important areas and linkages. WWF Nepal, Hariyo Ban Program.

Assessing the Multi-pathway Threat from an Invasive Agricultural Pest: *Tuta absoluta* in Asia

Joseph McNitt¹, Young Yun Chungbaek², Henning Mortveit², Madhav Marathe², Mateus R Campos³, Nicolas Desneux³, Thierry Brévault^{4,5,6}, Rangaswamy Muniappan⁷, and Abhijin Adiga²

¹*Epic Systems Corporation, United States*, ²*Biocomplexity Institute & Initiative, University of Virginia*, ³*French National Institute for Agricultural Research*, ⁴*BIOPASS, CIRAD-IRD-ISRA-UCAD, Dakar, Senegal*, ⁵*CIRAD, UPR AIDA, F-34398 Montpellier, France*, ⁶*Université de Montpellier, CIRAD, Montpellier, France*, ⁷*Feed the Future Integrated Pest Management Innovation Lab, Virginia Tech*

Abstract

Modern food systems facilitate rapid dispersal of pests and pathogens through multiple pathways. The complexity of spread dynamics and data inadequacy make it challenging to model the phenomenon and also to prepare for emerging invasions. We present a generic framework to study the spatiotemporal spread of invasive species as a multi-scale propagation process over a time-varying network accounting for climate, biology, seasonal production, trade, and demographic information. Machine learning techniques are used in a novel manner to capture model variability and analyse parameter sensitivity. We applied the framework to understand the spread of a devastating pest of tomato, *Tuta absoluta*, in South and Southeast Asia, a region at the frontier of its current range. Analysis with respect to historical invasion records suggests that even with modest self-mediated spread capabilities, the pest can quickly expand its range through domestic city-to-city vegetable trade. Our models forecast that within five to seven years, *T. absoluta* will invade all major vegetable growing areas of Mainland Southeast Asia assuming unmitigated spread. Monitoring high consumption areas can help in early detection, and targeted interventions at major production areas can effectively reduce the rate of spread.

1 Introduction

As the intensity of trade and human mobility increase, so does the rate of exotic species invasions [15]. Climate change and the detrimental impact of intensive agriculture on natural resources further aggravate this problem [10]. Understanding the dynamics of invasive species spread is imperative for achieving zero hunger, no poverty, good health and well being, which are among the sustainable development goals drafted by the United Nations [20]. Models play an important role in predicting the spatiotemporal spread, identifying roles of different pathways, assessing efficacy of control strategies, and exposing gaps in the understanding of the phenomenon [7, 11]. However, impending invasions of agricultural pests present difficult challenges. Accounting for multiple drivers of dispersal invariably makes the model complex. At the same time, data inadequacy makes it nearly impossible to calibrate and validate these models. Despite these limitations, a natural goal for a modeller is to provide useful insights into the mechanisms of spread, and thus help design effective policies for its prevention and mitigation.

Network propagation models have been widely used to study large interacting biological, social and technical systems. Some examples include infectious disease spread, online social networks, cascading failures in infrastructure networks [2]. Increasingly, they are being applied to study invasive species dynamics [5, 9, 21]. Unlike pest risk maps generated by species distribution models [22], the resulting dynamics of such a validated model yields a causal description of the underlying complex system. Here, we present a multi-pathway propagation model to study the spread of invasive agricultural pests. The model accounts for both self-mediated and human-mediated spread and effectively encapsulate the spatial heterogeneity, temporal variations and multi-scale nature of the propagation mechanisms.

We applied this framework to study the spread of the South American tomato leafminer or *Tuta absoluta*, a pest of the tomato crop and representative of recent biological invasions that have significantly perturbed global food production. Indigenous to South America, *T. absoluta* was accidentally introduced to Spain in 2006, and since then has rapidly spread throughout Europe, Africa, Western and Central Asia, the Indian subcontinent, and parts of Central America [3, 8]. It is well accepted that trade played a critical role in *T. absoluta*'s rapid spread. On multiple occasions it has been discovered in packaging stations and its spread pattern is correlated with prime trade routes [16]. Our study region is South and Southeast Asia— a region at the frontier of its current range – comprising of 10 countries: members of the Association of Southeast Asian Nations (ASEAN) and Bangladesh. In recent years, there has been a thrust to improve vegetable production

in all the countries of this region. With the pest having already spread to major tomato producing areas in Bangladesh, there is a high chance that it will be introduced to the remaining countries in the near future. Such invasions can have devastating effect on the economy and livelihood of farmers. Moreover, the invasion in Mainland Southeast Asia in particular is a serious threat to China, the largest producer of tomato, and Australasian neighbours. To our knowledge, this is the first study that explicitly considers multiple pathways of introduction and spread of *T. absoluta*. Earlier modelling efforts have only accounted for ecological aspects and self-mediated spread [8, 13, 24].

Our contributions are multifold. We provide a generic data-driven approach to build complex multi-pathway models. We identified, analysed, and fused disparate datasets corresponding to biology, climate, production, and agricultural commodity flow. We estimated the spatiotemporal distribution of host crops by integrating information from research articles, annual reports, and models. Trade flows were estimated using a gravity model approach. Machine learning techniques such as clustering and decision-tree algorithms were used for parameterization, to capture model variability and study parameter sensitivity. The analysis provides valuable insights into the dynamics of *T. absoluta* spread and its control, particularly from the perspective of human-mediated spread.

2 Methods

Data. The global datasets used in the model and for analysis are described in Table S1 of the supplement. Country specific data on seasonal production, consumption, processing, and trade was obtained from websites of agriculture ministries, research articles and technical reports (Table S2 in the supplement). Almost all the datasets used are openly available. The details of *T. absoluta* biology can be found in the supplement.

Multi-pathway spread model. We developed a stochastic multi-scale propagation model to simulate the multi-pathway spread of *T. absoluta*. Key concepts are illustrated in Figure 1a. The study region is divided into cells – the smallest spatial units – by overlaying a grid ($0.25^\circ \times 0.25^\circ$). Each cell is in one of the three states: susceptible (S) denoting pest free state, exposed (E) denoting that the pest has been introduced but the population has not yet built up to influence other cells, and infectious (I) denoting that the pest has established and the cell can influence its neighbours. The cell states are updated in discrete time steps, where the interval between successive time steps corresponds to a month. The probability that a cell v transitions

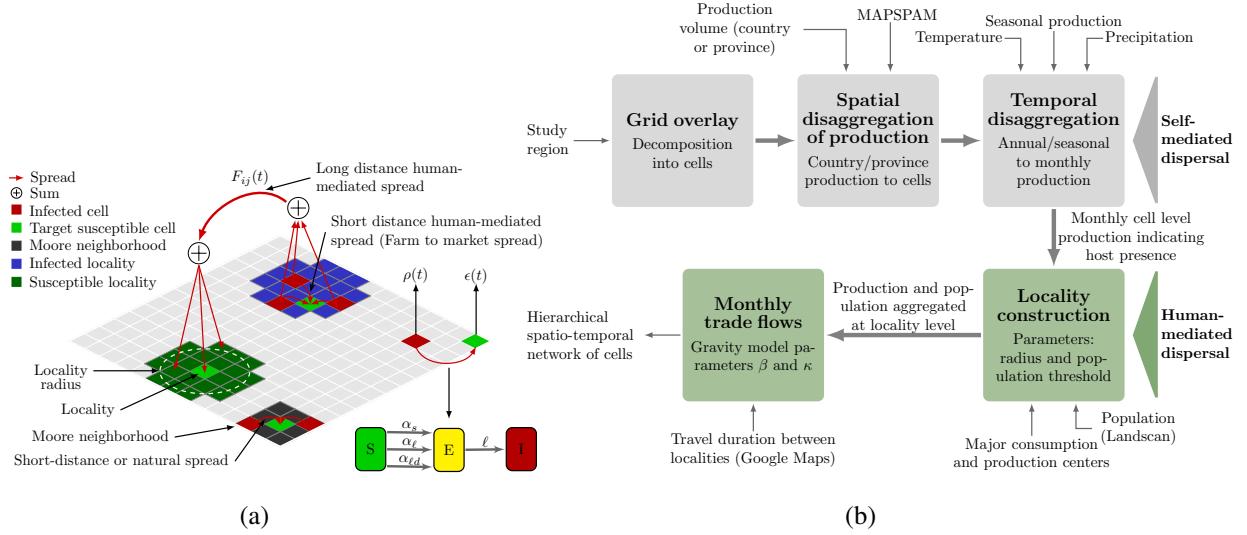


Figure 1: An illustration of the multi-pathway model (a) and the process of constructing the spatiotemporal network of cells.

from state S to E is determined by (i) suitability of the cell for *T. absoluta* to establish at that time step $\epsilon(v, t)$ and (ii) influence of “neighbouring” cells in state I depending on the pathway. An exposed cell transitions to state I after a latency period of ℓ time steps. This is the time required for the population to build up to infect other cells. Once the pest has established in a cell, the cell remains infected forever, a fair assumption considering that, historically, eradication of *T. absoluta* has not been successful (the only exception being United Kingdom). The infectiousness of a cell $\rho(v, t)$ is modelled as a linear function of host presence at time t , for which we use the weighted sum of production volume of tomato, eggplant, and potato in that cell at time t . The weights correspond to relative carrying capacity of *T. absoluta* on the three hosts [23].

There are three pathways by which a cell can become infected: short-distance dispersal, local human-mediated dispersal and long-distance dispersal. Short-distance dispersal captures the spread through natural means; from an infested cell to cells in its Moore neighbourhood of range r_M . The probability that a susceptible cell gets exposed (E) at time step t through short-distance spread is as follows:

$$p_s(v, t) = \epsilon(v, t) \left(1 - \exp \left(-\alpha_s \sum_{v' \in M_v(r_M)} \rho(v', t) \right) \right). \quad (1)$$

The probability depends on the suitability of the cell $\epsilon(v, t)$, infestation level of each neighbouring cell in the Moore neighbourhood with range r_M , $\rho(v', t)$, and the scaling factor, α_s , which is the transmission rate for this pathway. The function form is explained in Section S3.6.

For human-assisted spread we identified large urban areas in the region which we refer to as *localities* (Figure 1a) and considered interactions within and between localities. These areas have significant trade flows due to high consumption or production. Each *locality* consists of all grid cells which are within a certain distance (determined by *locality radius*) from its corresponding centre. Local human-mediated dispersal is modelled as the spread between cells belonging to a locality. Every cell v is influenced by cells in its locality L based on their infectiousness. The expression is similar to that in equation (1), but with cells in the locality instead of the Moore neighbourhood

$$p_\ell(v, t) = \epsilon(v, t) \left(1 - \exp \left(-\alpha_\ell \sum_{v' \in L} \rho(v', t) \right) \right), \quad (2)$$

where α_ℓ is the scaling factor. The details of locality construction are provided in Section S3.1 of the supplement.

Long-distance human-mediated dispersal corresponds to spread through trade between localities. For this purpose, we considered only tomato trade as there is not much evidence of *T. absoluta* spreading through trade of other hosts. We modelled domestic trade using a gravity model approach accounting for tomato production, processing, imports, and exports in each locality, and the travel time between localities. The probability of spread is directly proportional to the trade flow F_{ij} from one locality (i) to another (j). Suppose cell v belongs to locality i . Then, the probability of cell v transitioning from S to E due to long-distance human-mediated dispersal is given by:

$$p_{\ell d}(v, t) = \epsilon(v, t) \left(1 - \exp \left(-\alpha_{\ell d} \sum_{j \neq i} \sum_{v' \in L(j)} F_{ji} \rho(v', t) \right) \right), \quad (3)$$

where $\alpha_{\ell d}$ is the pathway scaling factor. The model parameters and their values are summarised in Table 1.

Network construction. Figure 1b provides a schematic of the network construction. The first step was to estimate monthly production volume of tomato, eggplant, and potato for each cell. We estimated annual production in each cell followed by disaggregation to monthly production. The annual production was estimated using the vegetable production available at the highest resolution for each country (at the level of province to just one value for the entire country) and a synthetic dataset called Spatial Production Allocation Model [28]. For monthly production, we used linear regression to model the production rate as a function of precipitation, temperature, and elevation. Seasonal tomato and eggplant production data for different

Table 1: Model parameters, their values and notes on parameter choices and ranges.

Parameter	Description	Value/range
r_M	Range of Moore neighbourhood	{1, 2, 3} corresponding to spread per month of approximately 25km, 50km and 75km respectively [13, 18].
ℓ	Latency period to transition from E to I	{1, 2, 3} months based on the time for the pest to complete life cycle (<i>T. absoluta</i> biology in Methods).
season	Disaggregation of annual production to monthly values	<i>Uniform</i> throughout the year or <i>seasonal</i> based on regression analysis (Methods).
β	Gravity model distance function exponent	{0, 1, 2}
κ	Gravity model distance function cut-off	Between 4 to 16 hours of travel time.
seed	Location and time of initial infestation	Scenarios based on countries (see Table S5)
locality radius	Determines cells assigned to a locality	100km (See Section S3.1 in the supplement for locality construction and analysis)
t_s	Time of initial infestation during parameterization	{3, 4, 5} corresponding to March, April and May respectively based on first report in Bangladesh [14].
$\alpha_s, \alpha_\ell, \alpha_\ell$	Pathway scaling factors	In the interval [0, 500].

regions of Philippines was used. For most of the other countries only qualitative information on seasonal production is available (Table S2). The regression function was applied to locations of these countries where this information is available and visually compared available data. The details are in Section S3.2 of the supplement.

To model locality-to-locality trade, we applied the approach of Venkatramanan et al. [26] with some modifications. We modelled the flow of fresh tomato crop between markets based on the following assumptions: (i) the total outflow from a city depends on the amount of produce in its surrounding regions and imports from countries outside the focus region at time t , and (ii) the total inflow depends on total consumption, processing demand, and exports from the city to countries outside the focus region. The details are in Section S3.3. Trade between countries of the focus region was not modelled as there is no adequate information on ports of entry or monthly flow volumes. But, it was accounted for while analyzing the possible routes of introduction.

Parameterization and experiment design. The goodness of fit of a parameter instance was determined by comparing the simulation output with *T. absoluta* incidence reports (Figure 2a and Table S3 for Bangladesh). The spread was simulated with infestation starting from the location of first report. For each cell, empirical probability that it is in state I at time t was computed (averaged over 100 repetitions). The output was compared to ground truth using a similarity function adapted from [5]. Let v be a reporting cell and t_v denote the month of the actual report of pest presence. To account for uncertainty in reporting, we consider a time window $U_\tau = [t_v - \tau, t_v + \tau]$ during comparison, where τ is the uncertainty parameter. We set $\tau = 2$, that is, error within ± 2 months is tolerated. Supposing \mathcal{C}_R is the set of cells corresponding to ground truth, and

$p(C, v, t)$ is the empirical probability that cell v is infected at time t for the input configuration C , then the similarity \mathcal{S} is given by

$$\mathcal{S}(C) = \frac{1}{|\mathcal{C}_R|} \sum_{v \in \mathcal{C}_R} \left(\sum_{t \in U_\tau} p(C, v, t) + \sum_{t \notin U_\tau} (1 - p(C, v, t)) \right). \quad (4)$$

Note that $\mathcal{S}(C)$ attains a maximum value of one when the simulation output is a perfect match (within the error tolerance limit). For parameter space exploration, we were motivated by a recent approach of using machine learning surrogates [17]. In our iterative “go with the winners” process [1] the subspace under consideration is sampled uniformly. Then, with model parameters as features and the similarity score as the dependent variable, we use Classification and Regression Trees (CART) approach to identify parts of the subspace for which the similarity score is high and reject the remaining. Simulations were performed for more than 500,000 parameter combinations using a high performance computing cluster. Configurations with similarity score $\mathcal{S}(C) \geq 0.75$ were chosen for further analysis.

Analysis of spread pattern. The objective here is to analyse the variability in the simulation outcomes within the set of best fit configurations. We leverage well-known machine learning techniques in a novel way to address this question. The methodology is outlined in Figure S5 in the supplement. First, we cluster the simulation outputs (time and cell indexed empirical probabilities) of selected configurations from the parameterization phase. This step captures the variability in outcomes; simulation outputs belonging to different clusters can be considered to be significantly different from one another. In the second step, we attempt to infer relationships between model parameters and cluster membership. To this end, our approach is to cast this as a classification problem using CART with model parameters as the features and cluster index as the label. The relationships are inferred from the decision tree that resulted from the algorithm. To avoid any bias introduced by the clustering algorithm, we also apply more than one method – hierarchical agglomerative clustering and the k -means algorithm. In both cases, we use the Euclidean distance as the distance measure to compare two simulation outputs. The analysis is repeated for different values of k , the number of clusters. More details are provided in Section S5 of the supplement.

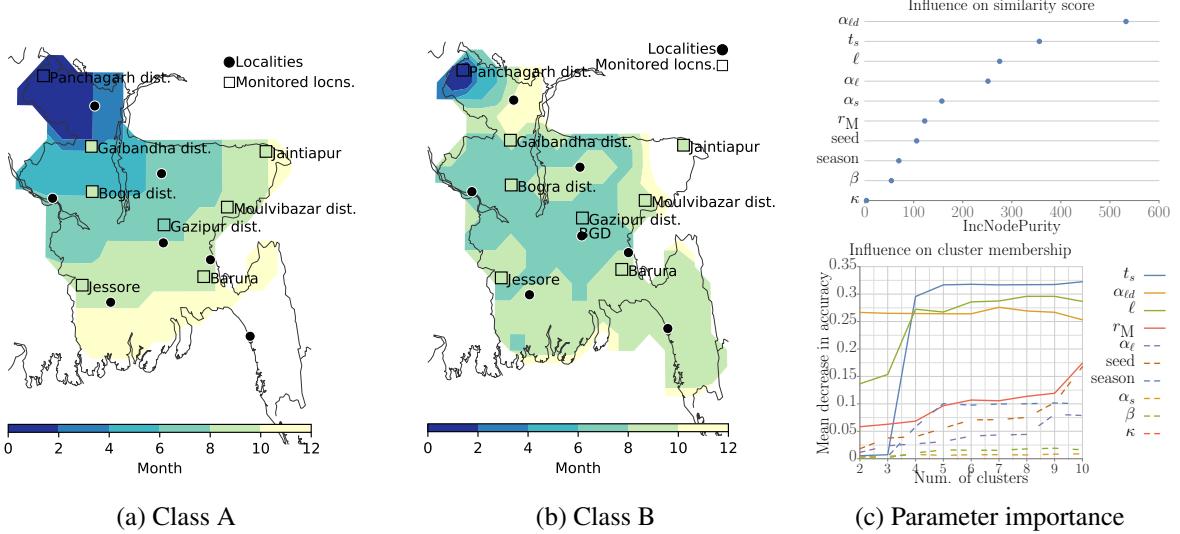


Figure 2: Possible spread patterns in Bangladesh. The contour plots show the spread starting from the location of first report in Panchagarh district for a simulation time of 12 months. Here, the time of infection for a cell is the minimum time step t such that the empirical probability that the cell is infected by time t is ≥ 0.8 . Also highlighted are the eight monitored locations and the localities applied in the model. The colours of the monitored locations correspond to the time of infection relative to the first report (Panchagarh). Two distinct spread patterns emerged from the cluster analysis. (a) and (b) show representative spreads observed for each class. The similarity in each case was $S > 0.8$. Importance of model parameters with respect to (i) similarity score S and (ii) cluster membership based on random forest method. The latter plot shows how the results vary with an increase in the number of clusters for hierarchical clustering algorithm. More results are presented in Figure S10 in the supplement. Videos depicting the spatiotemporal spread for each class are provided of the supplement.

3 Results

Variability in spread pattern. The clustering analysis of the configurations selected during the parameterization phase (approximately 8000 of them) reveals two distinct spread patterns primarily determined by the pathway parameters. The first class of models (Figure 2a), referred to as Class A, is characterised by the absence of long-distance human-mediated spread (α_{ld} negligible) and brisk spread between geographically adjacent cells, driven by the latency period ℓ , the Moore range r_M , and the short-distance scaling factor α_s . In contrast, for Class B models (Figure 2b), the long-distance pathway (α_{ld}) plays a significant role and there is relatively slow spread between geographically adjacent neighbours. Both hierarchical clustering and k -means clustering (Figures S6(b) and S7(a)) are consistent in this regard.

The Class A spread pattern does not capture the gap between the time of first report (Panchagarh) and the report in Gaibandha district (Figure 2a). Even though the distance between the two locations is only 185km, the latter reported the presence only after 10 months of first report, suggesting that self-mediated

spread might have been much slower. In the model output on the other hand, the corresponding cell gets infected between the second and fourth months. In Class B, this location is infected much later in comparison. However, the eastward spread towards the location Jaintiapur is slower than what was observed (Figure 2b). Even though Panchagarh is quite far from this location, pest presence was reported by February 2017, just nine months after the first report. As a baseline, we also simulated the spread using the cellular automata model developed by Guimapi [13] for Bangladesh. The spread pattern is similar to Class A as the model does not account for long-distance hops. However, the predicted rate of range expansion is much higher than our models (see Section S6.3 for model details and results).

The relative importance of model parameters was assessed using random forest algorithm with regard to their influence on (i) similarity score (\mathcal{S}) and (ii) spread pattern, which in our case, is akin to cluster membership. In the case of spread pattern, the importance was derived for each k (number of clusters) and clustering algorithm. Some results are presented in Figure 2c. We note that the long-distance scaling factor (α_{ld}) is among the top three important parameters. The start month (t_s) is also important for two reasons. Firstly, the distance between two time shifted simulation outputs can be large. Secondly, outputs are sensitive to seasonal variations or temporality of the network. Latency period (ℓ) and Moore range (r_M) together control the extent of radial spread in a time step. Typically for Class A models, r_M is high and ℓ is low and the other way round in the case of Class B models. Analysis of trade flows and seasonality is presented in Section S6.2 in the supplement.

Scenarios of pest introduction to countries in Southeast Asia. To identify routes of introduction to other countries in the region, we applied both Class A and Class B models. The starting point of the spread corresponds to the Panchagarh district (Figure 2a). Both model classes strongly indicate that *T. absoluta* is already present in parts of Myanmar (curves corresponding to time step 24, or two years from first report). Also, the pest is likely to enter Thailand from Myanmar, and subsequently move to Laos, Cambodia, and Vietnam as it spreads eastwards, and to China when it spreads northwards. From Thailand, spreading southward, it will enter Malaysia and subsequently enter Indonesia (Figure S12).

We also analysed the international tomato trade network (Section S6.1 of supplement) to assess the risk due to imports from *T. absoluta* infested countries outside this region. Malaysia and Singapore are important hubs with tomato imports from *T. absoluta* infested regions. There is a possibility that *T. absoluta* is directly introduced to these regions. However, in both cases, the import volume is very low. Also, the introduction risk

depends on the preventive measures taken by the exporting countries. With respect to both trade and natural pathways, there is a low chance that the pest will be introduced into Philippines from neighbouring countries, as there are no shared borders with any country in the region nor evidence of tomato trade. However, human mobility is a possible pathway. For example, the Middle East is the top destination for Filipino workers.

Predicted spread is model and region dependent. In the case of Class A models, the eastward spread is faster than southward spread (Figure S12a in the supplement). This is mainly because the Moore neighbourhood is smaller at the narrow region in the south of Myanmar and Thailand bordering Malaysia. However, in the case of Class B (Figure S12b), the spread is much faster in the same region aided by domestic trade flows from northern and central Thailand to the southern region. The Class A spread pattern predicts that within the next 4-5 years, much of the northern part of Mainland Southeast Asia will be invaded. The Class B spread pattern predicts that in the same period, *T. absoluta* will spread all over Malaysia and Singapore. However, the rate of spread observed is slower than that observed in Bangladesh for both classes. Also, even though the models exhibited a similar rate of spread for Bangladesh, we observed high variance in intensity of infestation as well as range expansion for the rest of the region. The results are in Figure S13. The reason for slow spread is as follows. Bangladesh has the highest tomato volume per country surface area ($\approx 2.5\text{tonnes}/\text{km}^2$). The next country is Vietnam ($\approx 1.5\text{tonnes}/\text{km}^2$). Therefore, in the case of Bangladesh, not only is the extent of infestation in a cell $\rho(\cdot)$ typically high, but also since it is a densely populated country, most cells have vegetable production. Hence, the rate of spread is much higher for relatively lower values of pathway parameters and Moore range. Also, we observed a strong dependence on Moore range (Figure S13b). In geographically larger countries, the production is scattered. Therefore, the lower the Moore range, the slower the spread.

Influence of domestic trade on spread pattern and rate. Here, we focus on long-distance dispersal and therefore restrict our discussion to Class B models. For the country-specific studies, the starting location was decided based on our analysis of possible entry points through different pathways (Section S4.1 in the supplement). We observed the following common spread pattern. When the invasive species is introduced to a country, dispersal is slow until the invasion front reaches a production source. Once it establishes at a source, the spread is very fast. Depending on the country, within 12 to 24 time steps (or 1-2 years), it spreads to almost all major localities of the country (see Figure 3b for example). Production areas which are very

close to high-consumption localities (large urban areas) are particularly vulnerable. Since local production typically does not satisfy demands of such localities, they have high inflows from other production areas and possibly from other countries. As a result, these localities are quickly infected. Once introduced to such localities, farmer–market interactions (local human-mediated dispersal) can facilitate the introduction of the pest to nearby production regions where it can establish and proliferate.

Given that monitoring and quarantining are both resource intensive and potentially disruptive, developing strategies that involve few locations yet provide near-optimal control is a goal for modellers. Market-level phytosanitary measures in terms of import restrictions have been undertaken by countries [25]. Here, we evaluated a simple strategy for containing the spread through the trade pathway. Localities associated with high annual outflows were identified (at most four in each country). As discussed earlier, pest establishment in these areas can potentially lead to rapid range expansion. The outflow from the targeted localities was cut off to mimic control at the trade/market level. In the strictest sense, this can be implemented by restricting trade of host crops. But, it is possible that phytosanitary measures have the same effect. Figure 3 shows results for two countries. More results are present in Figure S14 in the supplement. Consistently, across countries, we observed a significant reduction in range expansion as well as intensity of spread. Besides, as seen in Figure 3b, stifling these flows localises the spread that resembles those of Class A models, but with much less intensity.

4 Discussion

The variability in the spread patterns that explain the incidence reports exposes the lack of understanding of the pathways of spread. Nevertheless, the analysis does strongly indicate the role of human-assisted spread of *T. absoluta*. The pest was reported in May 2016 in the northwestern part of Bangladesh bordering India. The region is among the top three tomato producers in the country. By the beginning of the next production season, *T. absoluta* was found in almost every major urban region. Similar correlation between tomato trade and *T. absoluta* spread was observed in Nepal [26]. Studies on self-mediated spread (flying capability or by wind) can definitely help estimate more accurately the rate of self-mediated spread. It is also important to consider alternate scenarios of introduction. We recall that the far eastern part of Bangladesh (locality Jaintiapur in Figure 2b) reported pest presence nine months after the first report. This place happens to be close to an important trade route connecting Northeastern Bangladesh to Meghalaya in India, where *T. absoluta* was

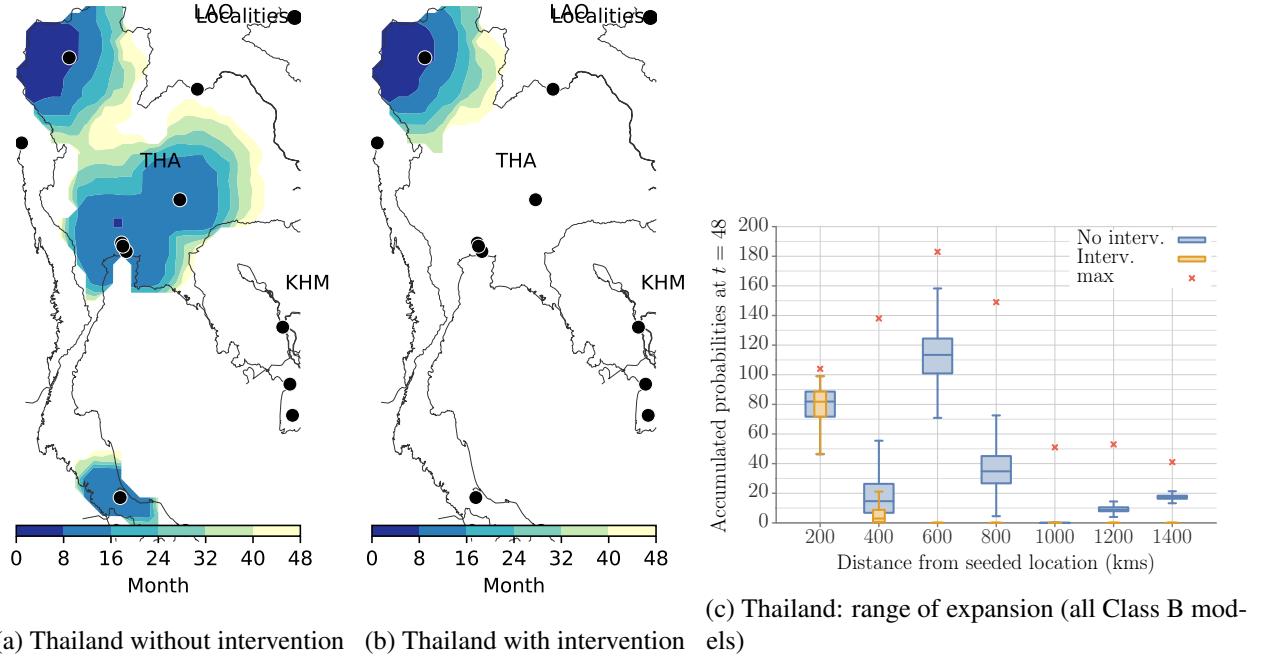


Figure 3: Rate and pattern of spread with and without intervention. Representative spread dynamics of Class B models ($r_M = 1, \ell = 3$) for two countries. More plots are shown in Figure S14. In each case, a cell close to a high production region was seeded. The first column corresponds to spread for 48 months after introduction. The colours indicate the time interval at which there is at least a 50% chance that a location will be infected. The second column corresponds to spread after cutting off flows from chosen localities. The third column shows average spread with respect to origin of infection for all Class B models. The cells are binned based on their distance from the origin of infection. Given time step t (48), let $\Pr_{\leq t}(v)$ be the probability that cell v is in state I by time t . For each configuration, we computed the “total infection” for every bin at time t by aggregating $\Pr_{\leq t}(v)$ for each v in the bin. The red points referred to as “max” correspond to the total number of cells in each bin, which is also the maximum possible accumulated probability for that bin.

officially reported in January 2017. Therefore, it is possible that multiple incursions took place.

Historically, international trade has played a strong role in the spread of *T. absoluta* between countries. For example, the pest was first reported by India in 2014. By early 2016, it was discovered in the Kathmandu area of Nepal and in the northern part of Bangladesh in May 2016. Both countries import significant volumes of tomatoes from India. However, there has been no report from Pakistan, another neighbour which does not import tomato from India. There are similar examples outside the region such as its slow advance from South America to Central America, or the fact that it is not reported in China despite being present in neighbouring Central Asian countries since 2015. We recall the discussion on slow predicted rate of spread in Mainland Southeast Asia compared to the observed rate in Bangladesh. One reason for this could be the unaccounted trade flows between countries. International trade within this region is not documented well. It is critical to address the data gaps concerning international trade, particularly considering that production and trade between countries in this region have been increasing over the years (details are in Section S6.1 of supplement).

While several integrated pest management strategies have been suggested for managing *T. absoluta*, hardly any work has been done in designing effective interventions at the trade level. Designing phytosanitary measures targeted towards markets and vehicles of transportation for preventing introductions (or reintroductions) is therefore a promising research direction. Some countries have already taken measures in this regard. Our results also indicate that monitoring markets with inflows from many regions is important. In the United States for example, the Animal and Plant Health Inspection Service of the Department of Agriculture (USDA-APHIS) has instituted quarantine regulations for imports from regions where the pest is present [25]. Identifying the optimal set of nodes in a network to reduce infectious disease spread is a widely studied topic. There are very few works that apply such techniques to invasive species spread (Nopsa et al. [21] for example). As the world moves towards concentrated and specialised agricultural production, focusing on this aspect becomes increasingly important.

Literature survey. Multi-pathway models are being increasingly used to study the role of invasive species dispersal. Douma et al. [9] survey the literature categorising various efforts into flow-based pathway models and agent-based models. Carrasco et al. [5] combine spatially explicit models of human-mediated spread with a phenology model to incorporate population dynamics of the western corn rootworm. Nopsa et al. [21] use a network science approach to study the role of transport and storage infrastructure in the spread of

pests and pathogens of wheat. Our model is in part motivated by the hybrid approaches used in the study of infectious diseases of humans and livestock [4, 27]. Although there is a general consensus that vegetable and seedling trade is a primary driver of *T. absoluta* spread, previous modelling efforts have exclusively focused on ecological aspects. Several studies [8, 24] provide risk maps using CLIMEX and take additional factors into account. Guimapi et al. [13] used a cellular automata approach to capture the global spread of the pest by factoring in temporal variations and spatial distribution of vegetation, temperature, and tomato production. A precursor to this work [26] modelled the seasonal production and trade of tomato in Nepal to study the role of trade in the spread of *T. absoluta* using a gravity model and network dynamics.

Emulators –based on Gaussian processes for example [12] – and machine learning surrogates [17] are emerging as solutions to overcome computational challenges, parameterization, and sensitivity analysis of complex agent-based models. The parameter importance study and parameterization was partly motivated by the work of Lamperti et al. [17]. We are not aware of any previous work that analyses the dynamics of simulation systems using unsupervised learning as presented in this paper. However, clustering has been considered in the context of multi-resolution simulation models as an interfacing component between simulators with different resolutions [6]. We cast the problem of deriving relationship between model parameters and cluster index as a classification problem. CART was our choice of algorithm since the learned model is a decision tree that can be interpreted. In principle, any such algorithm which provides such an explanation (such as multinomial regression for example) can be employed.

Challenges and limitations. Modelling emerging invasions is particularly challenging. Limited data on incidence and understanding of the underlying dynamics makes it nearly impossible to calibrate and validate the models. We have had to simplify or ignore some of the processes that might significantly influence the spread. For example, our model uses monthly production as a surrogate for infectiousness of a cell. Complex phenology models can be used instead (as in Carrasco et al. [5]), but would add to the complexity of the model. Since our focus region spans multiple countries, identifying and collecting data for each country was a lengthy process. For many countries, data had to be collected (or even inferred) from several publications and reports (Table S2). Further, these datasets were misaligned in time and spatial resolution. It is important to account for heterogeneity in production, consumption, awareness, cultural factors, etc. both within and between countries. Some countries are technologically more advanced than others, which manifests as differences in yield, crop loss, trade infrastructure, pest awareness, and preparation for invasion [10].

In particular, it is hard to model human-assisted spread owing to lack of seasonal trade data. To determine outflows and inflows for each locality, we had to identify major ports for import(s) and export(s) as well as estimate the fraction of production which was used for processing and was available only for a few countries. The farm–market-consumer interactions (local human-mediated spread) involves various actors such as farmers, wholesalers, retailers, wet markets, supermarkets, and so on. Modelling this is a challenge in itself. If data on actual flows of vegetables is provided, the gravity model can be improved or replaced by more sophisticated approaches. Also, the relationship between long-distance invasion risk and trade volume is hard to determine. While a direct relationship between volume and risk is plausible, whether the relation is linear (as assumed by our model) is unclear.

Conclusion. Traditionally, in developing countries, crops such as tomato are seasonal. However, over the past decade, due to rising demand and opportunities to export, there has been a thrust towards year-round production using protected cultivation methods and resilient varieties. An increase in urban population, short shelf life of vegetables, and the advantages of short marketing chains have encouraged urban agriculture in developing countries [19]. Our results indicate that such urban and peri-urban agriculture is particularly vulnerable to invasive species attacks. In particular, in Southeast Asia, vegetable production and internal trade have steadily increased. In comparison, the export of tomato outside of the focus region has risen steeply in recent years (after 2011), while the imports generally indicate a downward trend. Therefore, invasions from pests such as *T. absoluta* can have a huge negative impact on the socioeconomic fabric of this region. The modelling and analysis framework presented here is generic and applicable to other invasive species. The methodology is modular and leverages popular learning algorithms to analyse complex models under data scarcity. Other potential applications for this work include studies of natural or human-initiated disasters, climate change, and optimisation of food flows.

Data availability. The authors declare that the data supporting the findings of this study are available within the paper and its Supplementary Information file, or from the authors upon reasonable request.

Acknowledgements This work was supported in part by the United States Agency for International Development under the Cooperative Agreement NO. AID-OAA-L-15-00001 Feed the Future Innovation Lab for Integrated Pest Management, DTRA CNIMS Contract HDTRA1-11-D-0016-0001, NSF BIG DATA Grant IIS-1633028, NSF DIBBS Grant ACI-1443054, NIH Grant 1R01GM109718 and NSF NRT-DESE

Grant DGE-154362. We are grateful to Yousuf Mian, Nguyen Van Hoa, and Kimhian Seng for their help with obtaining country-specific information on production, trade, and pest incidence. We thank Richard Beckman, Irene Eckstrand and Erin Raymond for useful discussions on model design and paper organisation.

Author contributions. AA defined the scope of the research. AA, JM, TB, MRC collected and interpreted data. AA, MM conceived and designed the experiments. JM, AA and YYC performed the analysis. HM, ND, TB and RM provided assistance in interpreting the results. AA and JM wrote the paper with significant inputs from MRC and YYC. AA supervised the research. All authors discussed the results and commented on the manuscript.

References

- [1] D. Aldous and U. Vazirani. " go with the winners" algorithms. In *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pages 492–501. IEEE, 1994.
- [2] A. Barrat, M. Barthelemy, and A. Vespignani. *Dynamical processes on complex networks*. Cambridge university press, 2008.
- [3] A. Biondi, R. Guedes, and F. Wan. Ecology, Worldwide Spread, and Management of the Invasive South American Tomato Pinworm, *Tuta absoluta*: Past, Present, and Future. *Annual Review of Entomology*, (63):239–258, 2017.
- [4] R. A. Bradhurst, S. E. Roche, I. J. East, P. Kwan, and M. G. Garner. A hybrid modeling approach to simulating foot-and-mouth disease outbreaks in Australian livestock. *Frontiers in Environmental Science*, 3:17, 2015.
- [5] L. Carrasco, J. Mumford, A. MacLeod, T. Harwood, G. Grabenweger, A. Leach, J. Knight, and R. Baker. Unveiling human-assisted dispersal mechanisms in invasive alien insects: integration of spatial stochastic simulation and phenology models. *Ecological Modelling*, 221(17):2068–2075, 2010.
- [6] C. G. Cassandras, C. G. Panayiotou, G. Diehl, W. Gong, Z. Liu, and C. Zou. Clustering methods for multiresolution simulation modeling. In *Enabling Technology for Simulation Science IV*, volume 4026, pages 37–49. International Society for Optics and Photonics, 2000.

- [7] N. J. Cunniffe, B. Koskella, C. J. E. Metcalf, S. Parnell, T. R. Gottwald, and C. A. Gilligan. Thirteen challenges in modelling plant diseases. *Epidemics*, 10:6–10, 2015.
- [8] N. Desneux, E. Wajnberg, K. A. Wyckhuys, G. Burgio, S. Arpaia, C. A. Narváez-Vasquez, J. González-Cabrera, D. C. Ruescas, E. Tabone, J. Frandon, et al. Biological invasion of European tomato crops by *Tuta absoluta*, ecology, geographic expansion and prospects for biological control. *Journal of Pest Science*, 83(3):197–215, 2010.
- [9] J. Douma, M. Pautasso, R. Venette, C. Robinet, L. Hemerik, M. Mourits, J. Schans, and W. van der Werf. Pathway models for analysing and managing the introduction of alien plant pests an overview and categorization. *Ecological Modelling*, 339:58–67, 2016.
- [10] R. Early, B. A. Bradley, J. S. Dukes, J. J. Lawler, J. D. Olden, D. M. Blumenthal, P. Gonzalez, E. D. Grosholz, I. Ibañez, L. P. Miller, et al. Global threats from invasive alien species in the twenty-first century and national response capacities. *Nature Communications*, 7, 2016.
- [11] J. M. Epstein. Why model? *Journal of Artificial Societies and Social Simulation*, 11(4):12, 2008.
- [12] A. Fadikar, D. Higdon, J. Chen, B. Lewis, S. Venkatramanan, and M. Marathe. Calibrating a stochastic, agent-based model using quantile-based emulation. *SIAM/ASA Journal on Uncertainty Quantification*, 6(4):1685–1706, 2018.
- [13] R. Y. Guimapi, S. A. Mohamed, G. O. Okeyo, F. T. Ndjomatchoua, S. Ekesi, and H. E. Tonnang. Modeling the risk of invasion and spread of *Tuta absoluta* in Africa. *Ecological Complexity*, 28:77–93, 2016.
- [14] M. S. Hossain, M. Y. Mian, and R. Muniappan. First record of *Tuta absoluta* (Lepidoptera: Gelechiidae) from Bangladesh. *Journal of Agricultural and Urban Entomology*, 32(1):101–105, 2016.
- [15] P. E. Hulme. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46(1):10–18, 2009.
- [16] O. Karadjova, Z. Ilieva, V. Krumov, E. Petrova, V. Ventsislavov, et al. *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae): Potential for entry, establishment and spread in Bulgaria. *Bulgarian Journal of Agricultural Science*, 19(3):563–571, 2013.

- [17] F. Lamperti, A. Roventini, and A. Sani. Agent-based model calibration using machine learning surrogates. *Journal of Economic Dynamics and Control*, 90:366–389, 2018.
- [18] J. C. Martins, M. C. Picanço, R. S. Silva, A. H. Gonring, T. V. Galdino, and R. N. Guedes. Assessing the spatial distribution of *Tuta absoluta* (Lepidoptera: Gelechiidae) eggs in open-field tomato cultivation through geostatistical analysis. *Pest management science*, 74(1):30–36, 2018.
- [19] P. Moustier and H. Renting. Urban agriculture and short chain food marketing in developing countries. *Cities and Agriculture—Developing Resilient Urban Food Systems; de Zeeuw, H., Drechsel, P., Eds*, pages 121–138, 2015.
- [20] U. Nations. Sustainable development goals, Apr. 2019.
- [21] J. F. H. Nopsa, G. J. Daglish, D. W. Hagstrum, J. F. Leslie, T. W. Phillips, C. Scoglio, S. Thomas-Sharma, G. H. Walter, and K. A. Garrett. Ecological networks in stored grain: Key postharvest nodes for emerging pests, pathogens, and mycotoxins. *BioScience*, page biv122, 2015.
- [22] R. G. Pearson. Species’ distribution modeling for conservation educators and practitioners. *Synthesis. American Museum of Natural History*, 50:54–89, 2007.
- [23] S. E. Sylla, T. Brévault, L. Monticelli, K. Diarra, and N. Desneux. Geographic variation of host preference by the invasive tomato leafminer *Tuta absoluta*: implications for host range expansion. *Journal of Pest Science, Accepted*, 2018.
- [24] H. E. Tonnang, S. F. Mohamed, F. Khamis, and S. Ekesi. Identification and risk assessment for worldwide invasion and spread of *Tuta absoluta* with a focus on Sub-Saharan Africa: implications for phytosanitary measures and management. *PloS one*, 10(8):e0135283, 2015.
- [25] USDA. New Pest Response Guidelines: Tomato Leafminer (*Tuta absoluta*). *Animal and Plant Health Inspection Service, Plant Protection and Quarantine*, 2012.
- [26] S. Venkatramanan, S. Wu, B. Shi, A. Marathe, M. Marathe, S. Eubank, L. Sah, A. Giri, L. Colavito, K. Nitin, et al. Modeling commodity flow in the context of invasive species spread: Study of *tuta absoluta* in Nepal. *Crop Protection*, 2019.

- [27] W. Yang, D. R. Olson, and J. Shaman. Forecasting Influenza Outbreaks in Boroughs and Neighborhoods of New York City. *PLoS Computational Biology*, 12(11):1–19, 2016.
- [28] L. You, U. Wood-Sichra, S. Fritz, Z. Guo, L. See, and J. Koo. Spatial Production Allocation Model (SPAM) 2005 v3.2. <http://mapspam.info>, 2017.

PAC Learnability of Node Functions in Networked Dynamical Systems

Abhijin Adiga¹ Chris J. Kuhlman¹ Madhav V. Marathe^{1,2} S. S. Ravi^{1,3} Anil K. Vullikanti^{1,2}

Abstract

We consider the PAC learnability of the functions at the nodes of a discrete networked dynamical system, assuming that the underlying network is known. We provide tight bounds on the sample complexity of learning threshold functions. We establish a computational intractability result for efficient PAC learning of such functions. We develop efficient consistent learners when the number of negative examples is small. Using synthetic and real-world networks, we experimentally study how the network structure and sample complexity influence the quality of inference.

1. Introduction

Many real world phenomena (disease, influence and social behavior, neuronal activity, magnetic systems) can be formally represented as networked discrete dynamical systems (Valente, 1996; Schelling, 1978; Amini & Fountoulakis, 2014). Many such systems that include popular models such as SEIR (Newman, 2002) and Linear Threshold (Kempe et al., 2003) fall into the generic class of **Synchronous dynamical systems** (SyDSs). A SyDS consists of a graph whose nodes represent entities and whose edges represent interactions among the entities. Nodes have states and a local function at each node determines the next state of the node using the current states of the node and its neighbors. In a SyDS, the propagation of a contagion evolves in discrete time steps. In practice, some components of a SyDS are unknown; learning them is an active area of research. Some of this research is based on observing the system (Adiga et al., 2017; González-Bailón et al., 2011; Narasimhan et al., 2015; Kempe et al., 2003; Lokhov, 2016) while others rely on active interactions with the system in the form of queries

¹Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, VA, USA. ²Also with the Department of Computer Science, University of Virginia, Charlottesville, VA, USA. ³Also with the Department of Computer Science, University at Albany – SUNY, Albany, NY, USA. Correspondence to: Abhijin Adiga <abhijin@virginia.edu>.

(Bei et al., 2016; Kleinberg et al., 2017; Adiga et al., 2018).

We consider the problem of learning the node functions of a SyDS with a known graph and binary node states under the probably approximately correct (PAC) learning model pioneered by Valiant (1984). The PAC model has been applied to many different learning contexts such as learning classes of Boolean functions, half spaces, automata, etc. (Kearns & Vazirani, 1994). While PAC learnability has been studied for classes of individual Boolean functions (e.g., (Hellerstein & Servedio, 2007)), there has been limited work on dynamical systems, which can be viewed as Boolean functions that are connected through a network. Only recently, it has been applied in the context of learning influence functions of nodes in stochastic networked dynamical systems (Narasimhan et al., 2015; He et al., 2016).

Threshold functions are a commonly used class of local functions in dynamical system models (Granovetter, 1978; Watts, 2002). A t -threshold function is a Boolean function whose value is 1 iff at least t of its inputs have value 1. Variants of this model include bithreshold (Kuhlman et al., 2011) and progressive threshold (Amini & Fountoulakis, 2014). Our focus is on learning this class of functions (i.e., the threshold value of each node in the network) under the PAC model. Let n be the number of nodes in the graph. A **configuration** (s_1, s_2, \dots, s_n) specifies the state of each node at a certain time. Given a configuration C_1 at time τ , the configuration C_2 at time $\tau + 1$ is called the **successor** of C_1 . Each example given to the learner is a pair of configurations (C_1, C_2) . An example is labeled ‘positive’ if C_2 is the successor of C_1 ; otherwise, it is labeled ‘negative’. Under the PAC model, these examples are drawn from an unknown distribution.

Summary of results. Our focus is on SyDSs, where the local function at each node is a threshold function; that is, the hypothesis space consists of n -tuples of the form (t_1, t_2, \dots, t_n) , where each t_i , a non-negative integer, is the threshold value assigned to node v_i , $1 \leq i \leq n$. Following Mitchell (1997), we use the phrase “sample complexity”, represented by $\mathcal{M}(\epsilon, \delta)$, to denote the number of examples needed under the PAC model to learn a hypothesis space for given pair of error and probability values (ϵ, δ) . We assume that $\epsilon, \delta \in (0, \frac{1}{2})$. Our results are summarized below.

1. **Upper bounds:** For learning the hypothesis space of

threshold functions for SyDSs, we show that $\mathcal{M}(\epsilon, \delta)$ is at most $\frac{1}{\epsilon}(n \log(d_{\text{avg}} + 3) + \log(1/\delta))$ where d_{avg} is the average node degree of the underlying graph. We also extend this to other classes of threshold functions.

2. Lower bounds: We prove that the Vapnik-Chervonenkis (VC) dimension of the hypothesis space of threshold functions is at least $\lfloor n/4 \rfloor$ for *any* graph. From well known results in computational learning theory (Kearns & Vazirani, 1994), these lower bounds on the VC dimension imply that the sample complexity for learning threshold functions is $\Omega(n/\epsilon)$. Since $d_{\text{avg}} \leq n - 1$, for any fixed δ , the upper bound on the sample complexity mentioned in Item 1 above is within a factor $O(\log n)$ of the lower bound. Further, for graphs with constant average degree, this factor is $\Theta(1)$. As a special case, when the underlying graph is a clique on n nodes, we show that the VC dimension of the hypothesis space is at most $n + 1$. In contrast, most prior work on PAC learnability of graph dynamical system properties has been restricted to showing polynomial upper bounds on the sample complexity, e.g., (Narasimhan et al., 2015).

3. Hardness of learning: Even though there is a polynomial upper bound on the sample complexity, we show that when there are positive and negative examples, the hypothesis class of threshold functions is not *efficiently* PAC learnable, unless the complexity classes **NP** and **RP** (Randomized Polynomial time) coincide. Such complexity results are known in the learning theory literature for several other problems such as learning k -term DNF, neural networks, etc. (Kearns & Vazirani, 1994).

4. Efficient learning algorithms: For the case when there are only positive examples, we present an algorithm which learns the thresholds in time $O(|\mathcal{E}|n)$, where \mathcal{E} is the set of examples and n is the number of nodes. Further, when a set \mathcal{E}_N of negative examples is also given, we present a dynamic programming algorithm that learns in time $O(2^{|\mathcal{E}_N|} \text{poly}(n))$, which is polynomial when $|\mathcal{E}_N| = O(\log n)$. Also, using submodular function maximization under matroid constraints, we present an efficient learner which is consistent with all the positive examples and at least $(1 - 1/e)$ fraction of the negative examples. These results show that computational intractability arises in this case when the number of negative examples is large.

5. Experiments: We present experimental results using both synthetic and real-world networks to demonstrate how network structure and sample complexity influence the quality of the inferred system. We also provide experimental results that interpolate between theoretical results of limiting cases: as graph density decreases from a fully connected graph to a sparse graph, differences between the true and inferred systems first increase, and then decrease, thus exhibiting non-linear and non-monotonic behavior. Another interesting finding is how, in learning the same dynamical

system, differences in the distributions of configurations can lead to widely different qualities of inferred systems.

Due to limited space, many proofs and additional experimental results appear in (Adiga et al., 2019).

Related work. Inferring properties of networked dynamical systems from time-series data of node activations is a popular topic. Brugere et al. (2018) and Guille et al. (2013) provide comprehensive surveys of the literature on inferring networks and propagation model parameters from information or infection cascades. González-Bailón et al. (2011) present techniques for learning thresholds of nodes in a Twitter network using data from retweets. In contrast, PAC learnability of networked dynamical system is an emerging area of research. Recently, Narasimhan et al. (2015) and He et al. (2016) studied the PAC learnability of the influence function of popular stochastic propagation models – independent cascade and linear threshold from complete and partial observations. Lokhov (2016) uses a dynamic message-passing algorithm to reconstruct parameters of a spreading model given infection cascades.

There has been extensive research on PAC learning threshold functions and in general, Boolean functions. Hellerstein & Servedio (2007) provide a survey covering learnability of halfspaces, polynomial threshold functions, decision trees and disjunctive normal form (DNF) formulas. Learning the local function of a single vertex of a threshold SyDS is a special case of learning halfspaces (Blumer et al., 1989) with all weights equal to 1. Our results on learning threshold functions of SyDSs under the PAC model show that the network structure plays an important role in determining the sample complexity.

Recently, Adiga et al. (2017; 2018) considered the problem of inferring threshold SyDS, but for different models of observation. While Adiga et al. (2017) devise algorithms for learning thresholds of a dynamical system using information about the system’s trajectories, Adiga et al. (2018) study inference under active querying, where the user may ask for the successor of an arbitrary configuration.

2. Preliminaries

The dynamical system model. We use \mathbb{B} to denote the Boolean domain $\{0,1\}$. A Synchronous Dynamical System (SyDS) \mathcal{S} over \mathbb{B} is a pair $\mathcal{S} = (G, \mathcal{F})$, where (i) $G(V, E)$, an undirected graph with $|V| = n$, represents the underlying graph of the SyDS, with node set V and edge set E , and (ii) $\mathcal{F} = \{f_1, f_2, \dots, f_n\}$ is a collection of functions in the system, with f_i denoting the **local function** associated with node v_i , $1 \leq i \leq n$. At any time, each node of G has a state value from \mathbb{B} . The inputs to function f_i are the state of v_i and those of the neighbors of v_i in G ; for each input, the function f_i outputs a value in \mathbb{B} , and this value is the

next state of v_i . In a SyDS, all nodes compute and update their next state *synchronously*. Other update disciplines (e.g., sequential updates) have also been considered in the literature (Mortveit & Reidys, 2007). At any time τ , if $s_i^\tau \in \mathbb{B}$ is the state of node v_i ($1 \leq i \leq n$), the **configuration** C of the SyDS is the n -vector $(s_1^\tau, s_2^\tau, \dots, s_n^\tau)$. The system evolves in discrete time steps by the repeated application of \mathcal{F} . If a SyDS has a one step transition from configuration C_1 to C_2 , then C_2 is the **successor** of C_1 . In this paper, the local functions considered are deterministic, and therefore, the successor of each configuration is *unique*.

Threshold functions. The local function f_v associated with node v of a SyDS \mathcal{S} is a t_v -**threshold** function for some integer $t_v \geq 0$ if the following condition holds: the value of f_v is 1 if the number of 1's in the input to f_v is *at least* t_v ; otherwise, the value is 0. Using $d(v)$ to denote the degree of node v , the number of inputs to the function f_v is $d(v) + 1$. We assume that $0 \leq t_v \leq d(v) + 2$. The threshold values 0 and $d(v) + 2$ give rise to local functions that always output 1 and 0 respectively. Given a configuration C and a node v , the **score** of configuration C with respect to v , $\text{score}(C, v)$, is the number of 1's in the input provided by C to the local function at v .

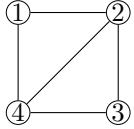


Figure 1. An example of a SyDS. The thresholds are $t_1 = 1$, $t_2 = 1$, $t_3 = 2$ and $t_4 = 3$.

Example: We present the graph of a threshold SyDS in Figure 1. The initial configuration is $(0, 0, 0, 1)$. It can be verified that the system goes through the following sequence of configurations during the next three time steps: $(0, 0, 0, 1) \rightarrow (1, 1, 0, 0) \rightarrow (1, 1, 1, 0) \rightarrow (1, 1, 1, 1)$.

Positive and negative examples. The goal is to learn the concept class of threshold functions for SyDSs whose underlying graphs are known. Thus, given an n -node SyDS, the hypothesis space H for the concept class contains n -tuples of the form (t_1, t_2, \dots, t_n) , where each t_i is the threshold value assigned to node v_i . Each example η given to learner is a pair of configurations (C_1, C_2) of the SyDS. We use \mathcal{E} to denote the set of examples given to a learner. An example is labeled ‘positive’ if C_2 is the successor of C_1 and ‘negative’ otherwise. Thus, the instance space from which examples are drawn consists of pairs of configurations, with an appropriate label for each example. Positive examples can be generated by observing or querying a system (Adiga et al., 2017; 2018). A negative example can be constructed from a positive example (C_1, C_2) by randomly complementing one or more bits of C_2 . The positive and negative examples are similar to membership queries used to learn Boolean functions (e.g., (Angluin & Slonim, 1994)).

Some PAC model definitions. We will assume that the

reader is familiar with the basic concepts of the PAC model such as efficient PAC learnability covered in many texts (e.g., (Kearns & Vazirani, 1994; Mitchell, 1997)). The **true error** (denoted commonly by $\text{error}_{\mathcal{D}}(h)$) of a hypothesis h is the probability that h will misclassify an example chosen at random using a distribution \mathcal{D} ; that is, $\text{error}_{\mathcal{D}}(h) = \Pr_{x \in \mathcal{D}}[c(x) \neq h(x)]$, where c is the target concept. For given values of ϵ and δ , the **sample complexity** of a learner is the number of examples needed by the learner to output an appropriate hypothesis h . We denote this quantity by $\mathcal{M}(\epsilon, \delta)$. Below is a well-known upper bound on $\mathcal{M}(\epsilon, \delta)$ established in (Haussler, 1988) based on the size of the hypothesis space H .

$$\mathcal{M}(\epsilon, \delta) \leq \frac{1}{\epsilon} (\log |H| + \log(1/\delta)). \quad (1)$$

We also use the concept of Vapnik-Chervonenkis (VC) dimension (Kearns & Vazirani, 1994). Given a set \mathcal{E} of labeled examples, we say that a hypothesis h (i.e., a threshold assignment to the nodes of the SyDS) from the hypothesis space H is **consistent** with \mathcal{E} , if h correctly classifies each example in \mathcal{E} . Given a set \mathcal{E} of unlabeled examples from the instance space, a **dichotomy** of \mathcal{E} partitions \mathcal{E} into two subsets \mathcal{E}_P and \mathcal{E}_N of positive and negative examples respectively. A set of examples \mathcal{E} is **shattered** by the hypothesis space H if for every dichotomy of \mathcal{E} , there is a hypothesis $h \in H$ that is consistent with the labeled examples generated by the dichotomy. The **VC dimension** Δ of H defined over the instance space is the largest finite subset of the instance space that can be shattered by H . Since our hypothesis space H is finite, the VC dimension of H is also finite. Below is a result from (Hanneke, 2016) on the bounds for sample complexity in terms of the VC dimension Δ .

Lemma 1. For any $\epsilon \in (1, 1/8]$ and $\delta \in (0, 1/100]$, sample complexity $\mathcal{M}(\epsilon, \delta) = O\left(\frac{1}{\epsilon}(\Delta \log(1/\epsilon) + \log(1/\delta))\right)$ and $\mathcal{M}(\epsilon, \delta) = \Omega\left(\frac{1}{\epsilon}(\Delta + \log(1/\delta))\right)$, where $\log(z) = \log(\max(z, e))$ and e is the base of common logarithm. \square

When ϵ and δ are fixed, the above lemma points out that for a hypothesis space with VC dimension d , the sample complexity is $\Theta(\Delta)$.

3. Bounds on the Sample Complexity

Upper bound. Here, we bound the size of the hypothesis space; the bound on sample complexity follows by a direct application of Inequality (1).

Theorem 1. Let $G(V, E)$ be a graph with n nodes and average degree d_{avg} . Let $\epsilon, \delta > 0$ be given. The sample complexity $\mathcal{M}(\epsilon, \delta)$ of threshold SyDS satisfies $\mathcal{M}(\epsilon, \delta) \leq \frac{1}{\epsilon}(n \log(d_{\text{avg}} + 3) + \log(1/\delta))$.

Proof: Each node v , with degree $d(v)$, can be assigned a threshold in the range 0 to $d(v) + 2$. So, for each

node, there are $d(v) + 3$ possible thresholds. Therefore, $|H| = \prod_{v \in V} (d(v) + 3) \leq \left(\frac{1}{n} \sum_{v \in V} (d(v) + 3) \right)^n = (d_{\text{avg}} + 3)^n$. The inequality follows from arithmetic mean–geometric mean inequality. The upper bound on $\mathcal{M}(\epsilon, \delta)$ follows by applying Equation (1). \square

In (Adiga et al., 2019), we show similar results for SyDSs with bithreshold and progressive threshold functions.

Theorem 1 corresponds to an extreme case where no additional information on the threshold functions is known. There are at least two factors that influence the size of the hypothesis space. Firstly, if local functions of different vertices are correlated and the relationship is known, then determining one threshold value would lead to estimates of others. Such models have been considered in disease dynamics (Miller, 2009). Secondly, any additional information about the local function restricts the number of hypotheses possible (e.g. (Romero et al., 2011)). To elaborate on this point, we consider the following setting. Let $\{V_1, V_2, \dots, V_k\}$ be a partition of the node set V . Every node in V_i has the same threshold $t_i \in T_i$, where T_i is a subset of threshold values. Then, the number of possible threshold assignments is at most $\prod_{i=1}^k |T_i|$ and therefore, from Equation (1), $\mathcal{M}(\epsilon, \delta) \leq \frac{1}{\epsilon} (\sum_{i=1}^k \log |T_i| + \log(1/\delta))$. Theorem 1 is the limiting case with $k = n$ and $T_i = \{0, \dots, d(v_i) + 2\}$.

Lower bound. We will establish a lower bound on the VC dimension of the hypothesis class and then apply Lemma 1 to obtain a lower bound on the sample complexity. We note that these results hold for the class of bithreshold SyDSs since threshold SyDSs are contained in this class.

Theorem 2. For a threshold SyDS defined on a graph $G(V, E)$ with n nodes, the VC dimension of the hypothesis space is $\geq \lceil n/4 \rceil$. Hence, $\mathcal{M}(\epsilon, \delta) = \Omega(n)$ for constant ϵ and δ .

We need two lemmas to prove the above theorem.

Lemma 2. Let I be an independent set in graph $G(V, E)$. There exists a set of configuration–successor pairs of size $|I|$ that can be shattered.

Proof: Let $v_1, v_2, \dots, v_{|I|}$ denote an arbitrary ordering of the vertices of I . The remaining vertices are ordered arbitrarily following $v_{|I|}$. We construct a set T with $|I|$ configuration–successor pairs as follows. Let C_i be the configuration in which the state of v_i is 1 and the states of all other nodes is 0, $1 \leq i \leq |I|$. Let $T = \{(C_i, C_i) : 1 \leq i \leq |I|\}$. We now show that T can be shattered. Let $A \subseteq T$. Vertex thresholds are assigned as follows: If $(C_i, C_i) \in A$, then set $t(v_i) = 1$; else set $t(v_i) = 2$. If $(C_i, C_i) \notin A$, then the successor of C_i for the assigned thresholds is C_i since $\text{score}(v_i, C_i) = t(v_i) = 1$ while for

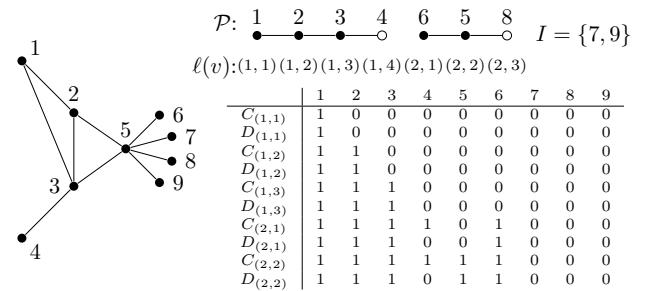
every other vertex v_j , $\text{score}(v_j, C_i) \leq 1 < t(v_j) = 2$. Suppose $(C_i, C_i) \notin A$; then, $\text{score}(v_i, C_i) = 1 < t(v_i) = 2$. Thus, the state of v_i is 0 and therefore, the successor is not C_i . \square

Definition 1. An **ordered path** in $G(V, E)$ is an ordered subset of vertices $P = (v_1, v_2, \dots, v_k)$, where $k \geq 2$, such that edge $\{v_i, v_{i+1}\} \in E$ for $1 \leq i \leq k - 1$.

Lemma 3. Let \mathcal{P} be a collection of node-disjoint ordered paths in $G(V, E)$. Let n' be the number of participating nodes in \mathcal{P} . Then, there exists a set of configuration–successor pairs of size at least $\lceil n'/2 \rceil$ that can be shattered.

Proof sketch: We will describe the construction of the configuration–successor pairs. The proof that the set can be shattered is in (Adiga et al., 2019). We will arbitrarily order the paths in \mathcal{P} . Each node v in \mathcal{P} is associated with a tuple $\ell(v) = (o_v, i_v)$ referred to as its *label*, where o_v (outer position) is the position of its path in the ordering of the paths and i_v (inner position) is the node's position in the path. For any two distinct vertices v and v' , $\ell(v') < \ell(v)$ if and only if (i) $o_{v'} < o_v$ or (ii) $o_{v'} = o_v$ and $i_{v'} < i_v$. Let $N(v)$ be the set of neighbors of v (open neighborhood). For any i , define $d_{(o,i)}(v) = |N(v) \cap \{v' \mid \ell(v') \leq (o, i)\}|$.

Let L be the set of labels $\ell = (o, i)$ such that the i th vertex is not the last node in the corresponding ordered path. For each $\ell \in L$, the configuration–successor pair (C_ℓ, D_ℓ) is as follows: In C_ℓ , all labeled nodes v with $\ell(v) \leq \ell$ are set to state 1 and the rest are set to 0. In D_ℓ , all v such that $\ell(v) \in L$ and $\ell(v) \leq \ell$ are set to state 1 and the rest to state 0. All nodes not in \mathcal{P} are also set to 0. The proof that the set $T = \{(C_\ell, D_\ell) \mid \ell \in L\}$ can be shattered is in (Adiga et al., 2019). See Figure 2 for an example. \square



For $A = \{(C_{(1,1)}, D_{(1,1)}), (C_{(2,1)}, D_{(2,1)})\}$, threshold assignments are $t(1) = d_{(1,1)} + 1 = 1$; $t(2) = d_{(1,2)} + 2 = 3$; $t(3) = d_{(1,3)} + 2 = 4$; $t(6) = d_{(2,1)} + 1 = 1$; $t(5) = d_{(2,2)} + 2 = 4$; $t(4) = t(8) = t(7) = t(9) = n + 1 = 10$

Figure 2. Example of configuration–successor pair construction using ordered paths. Here, T is the set of $(C_{(i,j)}, D_{(i,j)})$ pairs defined in the table and A is a subset of T .

Proof of Theorem 2: We use a greedy strategy to obtain \mathcal{P} . Let $V' = V$. At the start of each iteration, remove a vertex

from $v \in V'$ and set $P = \{v\}$. The current path has only v , which is an end point. The path is “grown” as follows. For an end point of the current path in P , pick a neighbor v' from $V' \setminus P$ if one exists. Move v' from V' to P and update the path with v' as an end point and v as an internal point if $|P| > 2$. The process terminates when no new neighbor can be added from V' for the current path’s end points. The path is ordered starting and ending with the current end points. P is added to $\mathcal{P}_{\geq 3}$ (i.e., the subset of paths with ≥ 3 nodes) and the process continues until V' has no paths of length ≥ 2 . In that case, we set $V_{<3} = V'$. What remains of V' induces an independent set of size at least $|V'|/2$. This is set to I . If $|I| \geq n/4$, then, the proof follows from Lemma 2. Otherwise, the number of vertices in \mathcal{P} is at least $n/2$. In that case, the proof follows from Lemma 3. \square

A tighter bound for complete graphs. Here, we observe that the relationship between sample complexity and edge density is not straightforward. When the underlying network G is a complete graph, from Theorem 1, the sample complexity for the threshold SyDSs is $O(n \log n)$ for constant ϵ and δ . Our next result shows that the sample complexity in this case is actually $O(n)$.

Theorem 3. *For a threshold SyDS defined on a complete graph G of size n , the VC-dimension of the hypothesis space is at most $n + 1$.*

Proof (idea): We establish an upper bound on the VC dimension. The proof is based on pigeon-hole principle and appears in (Adiga et al., 2019). \square

4. Efficiency of Learning Algorithms

We will show that the complexity of learning the thresholds hinges on the number of negative examples. We will also develop efficient exact and approximation algorithms. Given a collection \mathcal{E} of labeled examples, a learner is **consistent** with respect to \mathcal{E} if the hypothesis (a threshold assignment for all the nodes) produced by the learner satisfies *all* the examples in \mathcal{E} ; that is, (a) for each positive example (C_1, C_2) , C_2 is the successor of C_1 and (b) for each negative example (C_1, C_2) , C_2 is *not* the successor of C_1 .

Complexity of learning from positive and negative examples. Using a known approach from the literature (see e.g., (Kearns & Vazirani, 1994)), we show that if there is an efficient PAC learning algorithm with positive and negative examples, then there is a randomized polynomial time (**RP**-time) algorithm for every problem in **NP**, thus implying that **NP** = **RP**. To prove this result, we use an **NP**-complete problem defined in (Adiga et al., 2017). Recall that a configuration C of a SyDS is a *fixed point* if the successor of C is C itself; otherwise, it is a non-fixed point.

Problem 1. TA-FNF (Adiga et al., 2017): Given the graph $G(V, E)$ of a SyDS \mathcal{S} and two disjoint sets of con-

figurations F_1 and F_2 , is there a threshold assignment for the nodes of \mathcal{S} such that each configuration in F_1 is a fixed point and each configuration in F_2 is *not* a fixed point?

Now, we state our complexity result.

Theorem 4. *Given a graph $G(V, E)$ and a set \mathcal{E} of positive and negative examples, the concept class of threshold functions is not efficiently PAC learnable, unless **NP** = **RP**.*

Proof sketch: Suppose the concept class of threshold functions is efficiently PAC learnable from a set of positive and negative examples. Let \mathcal{A} be an efficient learning algorithm whose running time is polynomial in the size of the problem instance and the values $1/\delta$ and $1/\epsilon$. We show that \mathcal{A} can be used to devise an **RP** algorithm for TA-FNF. For brevity, we denote the learning problem by POS-NEG-LEARN. Given an instance I of the TA-FNF problem, we construct an instance I' of POS-NEG-LEARN as follows.

- (a) The underlying graph $G(V, E)$ of the SyDS \mathcal{S} for the POS-NEG-LEARN problem is the same as that of the TA-FNF problem.
- (b) For each configuration $C \in F_1 \cup F_2$, we construct the example (C, C) . If $C \in F_1$, the example is labeled positive, else, it is labeled negative. This is the set of examples \mathcal{E} .
- (c) Let $\epsilon = 1/(2|\mathcal{E}|)$ and $\delta = 0.1$.

Using the assumed efficient learning algorithm \mathcal{A} , we can now construct an algorithm \mathcal{A}_1 for TA-FNF as follows.

1. Run the learning algorithm \mathcal{A} on instance I' . Whenever \mathcal{A} requests an example, it is given one chosen uniformly randomly from \mathcal{E} . (Thus, the chosen distribution does not produce any examples that are not in \mathcal{E} .)
2. If \mathcal{A} produces a threshold assignment that is consistent with all the examples in \mathcal{E} , then \mathcal{A}_1 outputs the message “Yes”; otherwise (i.e., either \mathcal{A} does not produce a hypothesis or the produced hypothesis h is not consistent with \mathcal{E}), \mathcal{A}_1 outputs the message “No”.

Since \mathcal{A} runs in polynomial time, \mathcal{A}_1 also runs in polynomial time. We have the following lemma which shows that \mathcal{A}_1 is indeed an **RP** algorithm for TA-FNF (proof in (Adiga et al., 2019)).

Lemma 4. *(i) If the TA-FNF instance has a solution, then Algorithm \mathcal{A}_1 produces the message “Yes” with probability at least 0.9. (ii) If TA-FNF instance does not have a solution, Algorithm \mathcal{A}_1 produces the message “No”.*

Thus, starting with the assumption of an efficient learning algorithm \mathcal{A} , Lemma 4 shows that the TA-FNF problem is in **RP**, contradicting the assumption that **NP** ≠ **RP**. This completes our proof of Theorem 4. \square

We note that Theorem 4 holds for the case of proper learning

where the hypothesis class and the concept class coincide; here, they are both assumed to be n -tuples of threshold values. It is of interest to investigate whether the result can be extended to PAC learning under the representation-independent setting (see, e.g., (Warmuth, 1989)).

Learning from positive examples. Here, we show that there is an efficient consistent learner when all the examples in \mathcal{E} are positive. The consistent learner is described in Algorithm 1. This is similar to an algorithm in (Adiga et al., 2017) for learning thresholds from observations. The idea behind the algorithm is the following. Suppose $(C_1, C_2) \in \mathcal{E}$. Consider any node v . Recall that $\text{score}(C_1, v)$ is the number of 1's provided by C_1 to the local function at v . Let $C_2(v)$ denote the state of v in C_2 . If $C_2(v) = 0$, then we can conclude that $t(v) > \text{score}(C_1, v)$; otherwise, $t(v) \leq \text{score}(C_1, v)$. For each node v , these inequalities can be used to find non-negative integers $\ell(v)$ and $u(v)$ such that $\ell(v) \leq t(v) \leq u(v)$. It is easy to see that there is a consistent assignment of threshold values to nodes only if for each node v , the range $R(v) = [\ell(v) \dots u(v)]$ is non-empty. Since there are at most $|\mathcal{E}|$ inequalities for each node v , the time needed to construct the range of threshold values for each node and check the feasibility is $O(|\mathcal{E}||V|)$, which is a polynomial function of the input size. Therefore, we have the following result.

Algorithm 1 A consistent learner for positive examples.

Require: The underlying graph $G(V, E)$ of a SyDS and a collection \mathcal{E} of positive examples.
Ensure: If a solution exists, then a consistent assignment of threshold values to the nodes in V .

```

1:  $X = \emptyset$ .
2: for all  $(C_1, C_2) \in \mathcal{E}$  do
3:   for all  $v \in V$  do
4:     if  $C_2(v) = 0$  then
5:       Add the inequality " $t(v) > \text{score}(C_1, v)$ " to  $X$ .
6:     else
7:       Add the inequality " $t(v) < \text{score}(C_1, v)$ " to  $X$ .
8:     end if
9:   end for
10: end for
11: If there is a solution to the set of inequalities in  $X$ , output a solution. Otherwise, output "no consistent assignment".

```

Theorem 5. *When all the examples are positive, there is an efficient algorithm to determine whether there is a consistent threshold assignment; if so, the algorithm finds one such assignment.* \square

Learning with a small number of negative examples. Theorem 4 points out the difficulty of developing an efficient consistent learner for threshold SyDSs when there are both positive and negative examples. Since there is an efficient learner for the situation when the input has only positive examples (Theorem 5), it is seen that the computational intractability result arises due to the negative examples. Our

next result shows that when the number of negative examples is *small* (i.e., $O(\log n)$ where n is the number of nodes), an efficient consistent learner can be developed.

Let \mathcal{E}_N be the set of negative examples. Given a negative example $x = (C_1, C_2) \in \mathcal{E}_N$, we say that a pair (v, t) “handles” x if setting the threshold of v to t makes the value of v in the successor of C_1 to be *different from* $C_2(v)$. There might be multiple pairs $(v, t), (v', t')$, which handle an example $x \in \mathcal{E}_N$. Note that v may or may not be same as v' , and t may or may not be the same as t' . Also, t should be within the range $R(v)$ determined from examining the positive examples \mathcal{E}_P . Let $S(v, t) = \{x \in \mathcal{E}_N : (v, t) \text{ handles } x\}$. Let $B = \{(v, t) : S(v, t) \neq \emptyset\}$. We have the following lemma (proof in (Adiga et al., 2019)).

Lemma 5. *There exists a consistent threshold assignment for \mathcal{E} if and only if there exists a subset $B' = \{(v_1, t_1), \dots, (v_r, t_r)\} \subseteq B$ where v_1, \dots, v_r are distinct nodes, and $\bigcup_{i=1}^r S(v_i, t_i) = \mathcal{E}_N$, such that threshold values t_i for node v_i are consistent with the ranges $R(v)$ inferred from the positive examples \mathcal{E}_P .*

We develop a dynamic programming algorithm to check whether there is a consistent learner when there are positive and negative examples. Let v_1, \dots, v_n be an ordering of the nodes. The algorithm uses the following steps.

- (1) We maintain information in a table $M[j, S]$ where $j = 1, \dots, n$ and S ranges over all the subsets of \mathcal{E}_N . We define $M[j, S] = 1$ if there exists an assignment $\{(v_1, t_1), \dots, (v_{j'}, t_{j'})\}$ for $j' \leq j$, such that all $x \in S$ are handled by this assignment; otherwise $M[j, S] = 0$. The entries of $M[\cdot, \cdot]$ are computed as follows.
- (2) The base case is $j = 1$. For each $(v_1, t) \in B$, we set $M[1, S'] = 1$ for all $S' \subseteq S(v_1, t)$. We set $M[1, S] = 0$ for all remaining S .
- (3) We set $M[j+1, S] = 1$ if $M[j, S] = 1$ or if $M[j, S - S(v_{j+1}, t)] = 1$ for some $(v_{j+1}, t) \in B$; otherwise, we set $M[j+1, S] = 0$.

The lemma below (proved in (Adiga et al., 2019)) establishes the correctness and the running time of the algorithm.

Lemma 6. *The above dynamic program has space and time complexity $O(2^{|\mathcal{E}_N|} \text{poly}(n))$, which is polynomial if $|\mathcal{E}_N| = O(\log n)$. Further, $M[n, \mathcal{E}_N] = 1$ if and only if there exists a consistent threshold labeling for \mathcal{E} .*

Approximately satisfying constraints in \mathcal{E}_N . We use the notation and a result from (Călinescu et al., 2011). A partition matroid $\mathcal{M} = (X, \mathcal{I})$ is defined in the following manner: X is partitioned into ℓ sets X_1, \dots, X_ℓ , with associated integers k_1, \dots, k_ℓ . A set $A \subseteq X$ is independent iff $|A \cap X_i| \leq k_i$ for all i . A function $f : 2^X \rightarrow \mathbb{R}_+$ is submodular if for all $P \subseteq Q \subset X$, $p \notin Q$, we have $f(P \cup \{p\}) - f(P) \geq f(Q \cup \{p\}) - f(Q)$. Finally, f

is monotone if $f(P) \leq f(Q)$ for all $P \subseteq Q$. Călinescu et al. (2011) design an algorithm that gives a $(1 - 1/e)$ -approximation for a monotone submodular function with matroid constraints.

Using our earlier notation, define $B(v) = \{(v, t) \in B : S(v, t) \neq \emptyset\}$ and $X = B$. Define the function $g : 2^X \rightarrow \mathbb{R}_+$ as $g(P) = |\bigcup_{(v,t) \in P} S(v, t)|$. Our algorithm involves the following steps.

- (1) $B(v_1), \dots, B(v_n)$ is a partition of X . We fix $k_i = 1$ for $i = 1, \dots, n$. The matroid $\mathcal{M} = (X, \mathcal{I})$ is as above.
- (2) Use the algorithm of (Călinescu et al., 2011) to find a subset $P \in \mathcal{I}$ that has the maximum value $g(P)$.
- (3) The set $P = \{(v_{j_1}, t_{j_1}), \dots, (v_{j_r}, t_{j_r})\}$ gives the threshold assignment.

The following lemma gives the performance guarantee provided by the above algorithm.

Lemma 7. *The assignment returned by the above algorithm ensures that at least $(1 - 1/e)$ fraction of the examples in \mathcal{E}_N are handled.*

The lemma follows by noting that function $g(\cdot)$ is monotone submodular and then applying Theorem 1.1 from (Călinescu et al., 2011).

5. Experimental Results

Analysis procedures and networks. Figure 3 provides an overview of the major steps in the experiments. Mined graphs are obtained from the web and synthetic networks of different classes are generated; see Table 1. There are five graph instances for each of the random regular (RR) and scale-free (SF) synthetic networks for a specified n (number of nodes). For each graph instance, five sets of true threshold assignments are made, where the threshold for each node v is chosen uniformly at random from the range $[0, d_v + 2]$ (the two limiting cases: 0 means the node will always change to state 1; $d_v + 2$ means the node will not transition to state 1 and will transition to state 0 if it is 1). For each true assignment of thresholds to nodes, Algorithm 1 is used to estimate ten inferred threshold assignments. This is done because the outcome of Algorithm 1 is stochastic due to the fact that the positive examples are chosen randomly. These sets of threshold estimates are also generated for different numbers m of queries (or configurations), ranging from 10 to 10^5 . Each true threshold assignment, represented by \mathcal{F} , is compared to each of the estimated threshold sets (for each particular value of m), represented by \mathcal{F}' , by counting the fraction f_{ne} of times that $\mathcal{F}(C) \neq \mathcal{F}'(C)$. Based on preliminary analyses, we take the number n_t of configurations C (i.e., comparisons or trials) to be $n_t = 10n$. In Algorithm 1 for estimating thresholds and in the evaluation of the dynamics between \mathcal{F} and \mathcal{F}' , configurations are re-

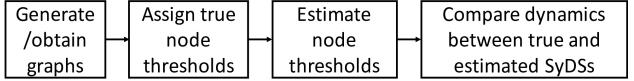


Figure 3. Overview of pipeline work items for experiments. The third box uses Algorithm 1.

quired. For consistency, configurations must be sampled from the same distribution for both operations. We evaluate two distributions. One configuration distribution \mathcal{D}_u is a uniform distribution where each node is set to state 1 with probability p (and to state 0 with probability $(1 - p)$). This distribution has been analyzed in the context of learning halfspaces (Long, 1995). The second configuration distribution \mathcal{D}_{PL} generates a probability in $(0,1)$ uniformly at random, and using a power law relationship, determines the number n_1 of nodes whose state will be 1 in the configuration (He et al., 2016). Then, uniformly at random n_1 nodes are selected from V to be in state 1, with the remaining nodes in state 0. Since our experiments use positive examples only and the underlying system is deterministic, once a random configuration C is chosen, its successor C' is determined. Thus, in our experiments, we don't sample pairs of configurations. We perform the steps of Figure 3 using a full factorial design, and results are presented next.

Table 1. Mined and synthetic networks, and their attributes.

Network	Properties			
	n	$ E $	d_{ave}	d_{max}
Jazz	198	2742	27.70	100
NRV	769	4551	11.84	20
euEmail	986	16064	32.58	345
Ran Reg ^{a,1}	11–1000	$n d_{avg}/2$	10	10
Scl free ^{a,2}	20–1000	$\sim n d_{avg}/2$	9.5–9.9	13–149
Cliques ³	400	$n_q n_c (n_c - 1)/2$	$n_c - 1$	$n_c - 1$

^a 5 replicates per n value.

¹ n values are 11, 20, 40, 60, 80, 100, 200, 400, 600, 800, 1000.

² n values are 20, 40, 60, 80, 100, 200, 400, 600, 800, 1000.

³ number n_c of nodes per clique are 10, 20, 40, 80, 100, 200, 400.

We use the above methods to investigate the behavior of the PAC model for different (i) network classes, sizes of networks, d_{ave} , and graph density; (ii) threshold estimation parameters such as configuration distributions \mathcal{D} , node probability p , and numbers m of queries for Algorithm 1; and (iii) (ϵ, δ) pairs of the PAC model. Selected results are provided here; others are in (Adiga et al., 2019).

Effects of graph size and number of queries. Figure 4 shows f_{ne} vs. m for the mined networks. The first plot shows boxplots for the Jazz network. Each box represents 50 values of f_{ne} . (This is because there are 5 threshold assignments, and for each assignment, we generated 10 estimated threshold assignments.) As expected, as the number m of queries increases, the fraction of configurations C for which $\mathcal{F}(C) \neq \mathcal{F}'(C)$ decreases. In the second plot, the av-

verage values of f_{ne} are plotted against m for the three mined networks. The curves, highest to lowest, correspond to decreasing numbers of nodes in the graphs. That is, the larger the graph in terms of numbers of nodes, all other things being equal, the larger is the number of queries required to get accurate thresholds from Algorithm 1 because distance-1 neighbors in a graph typically do not make progress in narrowing threshold estimates in one step.

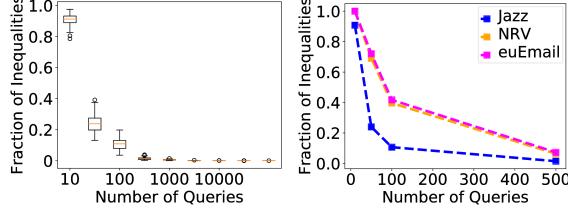


Figure 4. Fraction of inequalities f_{ne} versus m for mined networks (a) Jazz and (b) for the three mined networks. In (a), the number of queries ranges from 10 to 100,000. As m increases from 10 to 100,000, $f_{ne} \rightarrow 0$. In (b) the focus is smaller values of m ; errors are lesser in the Jazz network and greatest in euEmail as these networks have the minimum and maximum numbers of nodes of the mined graphs. There are 50 data points per m value on the x-axis (1 graph instance \times 5 true threshold assignments per graph \times 10 estimated instances per true instance). In both plots, we use \mathcal{D}_u with $p = 0.25$.

Effect of graph density on f_{ne} . For cliques (densest graphs) and independent sets of nodes (least dense graphs), theoretical results show that f_{ne} is the same in both extreme cases. Hence, a natural question is what happens at intermediate values of graph density. To study this, we take a series of RR graphs and SF graphs with $d_{ave} = k = 10$ and increase n from 11 (a clique) to 1000 (a less dense graph) to determine the fraction of errors in comparing \mathcal{F} and \mathcal{F}' . Figure 5 shows f_{ne} vs. n for all of the random regular (RR) and scale-free (SF) networks. In this plot “ n increasing” is synonymous with “graph density decreasing.” Each box represents 250 values of f_{ne} . As expected, in the first plot (RR) median f_{ne} values are comparable at extreme ends. However, at intermediate values of n (and density), the behavior is nonlinear and non-monotonic, showing increased f_{ne} . In the second plot, the effect exists in that data are shown for the smallest $n = 20$. For $n = 11$, the SF network would become a clique (keeping d_{ave} the same) and the data from the plot at $n = 11$ shows smaller f_{ne} values. The two different graph structures suggests that this effect may be robust across different network structures.

Effect of distributions for sampling configurations. We investigate the effect of different distributions— \mathcal{D}_u and \mathcal{D}_{PL} —for sampling configurations for generating estimated thresholds and for comparing \mathcal{F} and \mathcal{F}' . Results are shown in Figure 6 for RR networks. For \mathcal{D}_u in the first plot, we take $p = 0.25$, and comparatively small f_{ne} are observed.

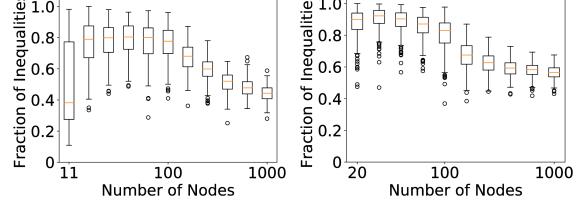


Figure 5. Fraction of inequalities f_{ne} as a function of n (and as a function of graph density, because graph density decreases as n increases, for fixed value of d_{ave}) for (a) random regular (RR) and (b) scale-free (SF) networks. In both plots, $q = 0.1$, so the number of queries scales with n according to $m = qn$. There are 250 data points per box plot. These plots illustrate that f_{ne} increases at intermediate values of graph density, and decreases with very high and very low density. We use \mathcal{D}_u with $p = 0.25$.

In contrast, for \mathcal{D}_{PL} in the second plot, $f_{ne} = 1$ for almost all n . A fundamental difference between the two models is that for \mathcal{D}_u , the expected number n_1 of nodes in state 1 in each sampled configuration is the same ($n_1 = pn$), while for \mathcal{D}_{PL} , n_1 can vary as $0 \leq n_1 \leq n$. As a result, there may be less consistency in the configurations produced by \mathcal{D}_{PL} , leading to greater inconsistency in configurations for estimating thresholds and for comparing \mathcal{F} and \mathcal{F}' .

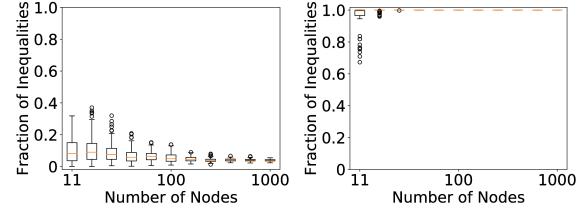


Figure 6. Fraction of inequalities f_{ne} as a function of n for different distributions of configurations: (a) \mathcal{D}_u with $p = 0.25$ and (b) \mathcal{D}_{PL} . These results, on RR networks, show that widely different f_{ne} can result from different models for generating configurations. Here, $m = qn$ with $q = 0.1$.

6. Summary and Future Work

We examined the learnability of local functions of SyDSs under the PAC model. Our focus was on threshold functions with emphasis on sample complexity, efficiency of learning and effect of network structure. There are several possible directions for future work. For example, one may consider learning other classes of local functions including stochastic functions. More general observation models can be explored. For example, observations can span multiple time steps. The difficulty here is that a single error in one step can lead to errors in the ensuing steps. Another and challenging research direction is to investigate learning both the graph topology (which we currently assume is known) and the local functions under the PAC model.

Acknowledgments

We thank the reviewers for providing valuable feedback. This work is supported by DTRA CNIMS (Contract HDTRA1-11-D-0016-0001), NSF DIBBS Grant ACI-1443054, NSF EAGER Grant CMMI-1745207, NSF BIG DATA Grant IIS-1633028 and NSF CRISP 2.0 Grant 1832587. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

References

- Adiga, A., Kuhlman, C. J., Marathe, M. V., Ravi, S. S., Rosenkrantz, D. J., and Stearns, R. E. Inferring local transition functions of discrete dynamical systems from observations of system behavior. *Theor. CS.*, 679:126–144, 2017.
- Adiga, A., Kuhlman, C. J., Marathe, M. V., Ravi, S. S., Rosenkrantz, D. J., and Stearns, R. E. Learning the behavior of a dynamical system via a “20 questions” approach. In *Thirty second AAAI Conference on Artificial Intelligence*, pp. 4630–4637, 2018.
- Adiga, A., Kuhlman, C. J., Marathe, M. V., Ravi, S. S., and Vullikanti, A. K. PAC learnability of node functions in networked dynamical systems. Technical Report, Network Systems Science and Advanced Computing Division, Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, VA, 2019.
- Amini, H. and Fountoulakis, N. Bootstrap percolation in power-law random graphs. *Journal of Statistical Physics*, 155(1):72–92, 2014.
- Angluin, D. and Slonim, D. K. Randomly fallible teachers: Learning monotone DNF with an incomplete membership oracle. *Machine Learning*, 14(1):7–26, 1994.
- Bei, X., Chen, W., Garg, J., Hoefer, M., and Sun, X. Learning market parameters using aggregate demand queries. In *Proc. AAAI*, pp. 411–417, 2016.
- Blumer, A., Ehrenfeucht, A., Haussler, D., and Warmuth, M. K. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM (JACM)*, 36(4):929–965, 1989.
- Brugere, I., Gallagher, B., and Berger-Wolf, T. Y. Network structure inference, a survey: Motivations, methods, and applications. *ACM Computing Surveys (CSUR)*, 51(2), 2018. (24 pages).
- Călinescu, G., Chekuri, C., Pál, M., and Vondrák, J. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM J. Comput.*, 40(6):1740–1766, 2011.
- González-Bailón, S., Borge-Holthoefer, J., Rivero, A., and Moreno, Y. The dynamics of protest recruitment through an online network. *Scientific Reports*, 1, 2011. (7 pages).
- Granovetter, M. Threshold models of collective behavior. *American Journal of Sociology*, pp. 1420–1443, 1978.
- Guille, A., Hadid, H., Favre, C., and Zighed, D. A. Information diffusion in online social networks: A survey. *ACM SIGMOD Record*, 42(2):17–28, 2013.
- Hanneke, S. The Optimal Sample Complexity of PAC Learning. *J. Machine Learning Research*, 17:1–15, 2016. ISSN 15337928. URL <http://arxiv.org/abs/1507.00473>.
- Haussler, D. Quantifying inductive bias: AI learning algorithms and Valiant’s learning framework. *Artificial intelligence*, 36(2):177–221, 1988.
- He, X., Xu, K., Kempe, D., and Liu, Y. Learning influence functions from incomplete observations. In *Advances in Neural Information Processing Systems*, pp. 2073–2081, 2016.
- Hellerstein, L. and Servedio, R. A. On PAC learning algorithms for rich boolean function classes. *Theoretical Computer Science*, 384(1):66–76, 2007.
- Kearns, M. J. and Vazirani, V. V. *An Introduction to Computational Learning Theory*. MIT Press, Cambridge, MA, 1994.
- Kempe, D., Kleinberg, J., and Tardos, E. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137–146, 2003.
- Kleinberg, J., Mullainathan, S., and Ugander, J. Comparison-based choices. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pp. 127–144. ACM, 2017.
- Kuhlman, C. J., Kumar, V. A., Marathe, M. V., Swarup, S., Tuli, G., Ravi, S., and Rosenkrantz, D. J. Inhibiting the diffusion of contagions in bi-threshold systems: Analytical and experimental results. In *AAAI Fall Symposium: Complex Adaptive Systems*, 2011.
- Lokhov, A. Reconstructing parameters of spreading models from partial observations. In *Advances in Neural Information Processing Systems*, pp. 3467–3475, 2016.
- Long, P. M. On the sample complexity of PAC learning half-spaces against the uniform distribution. *IEEE Transactions on Neural Networks*, 6(6):1556–1559, 1995.

- Miller, J. C. Spread of infectious disease through clustered populations. *Journal of the Royal Society Interface*, 6 (41):1121–1134, 2009.
- Mitchell, T. M. *Machine Learning*. McGraw Hill, Boston, MA, 1997.
- Mortveit, H. and Reidys, C. *An Introduction to Sequential Dynamical Systems*. Springer Science & Business Media, New York, NY, 2007.
- Narasimhan, H., Parkes, D. C., and Singer, Y. Learnability of influence in networks. In *Advances in Neural Information Processing Systems*, pp. 3186–3194, 2015.
- Newman, M. E. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128:1–016128:12, 2002.
- Romero, D. M., Meeder, B., and Kleinberg, J. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th International Conference on World Wide Web*, pp. 695–704. ACM, 2011.
- Schelling, T. C. *Micromotives and Macrobbehavior*. Norton, New York, NY, 1978.
- Valente, T. W. Social network thresholds in the diffusion of innovations. *Social Networks*, 18:69–89, 1996.
- Valiant, L. G. A theory of the learnable. *Communications of the ACM*, 18(11):1134–1142, 1984.
- Warmuth, M. K. Towards representation independence in PAC learning. In *Proc. International Workshop on Analogical and Inductive Inference (AII'89)*, pp. 78–103, Oct. 1989.
- Watts, D. J. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99:5766–5771, 2002.

Modeling Commodity Flow in the Context of Invasive Species Spread: Study of *Tuta absoluta* in Nepal

S. Venkatramanan¹, S. Wu², B. Shi⁴, A. Marathe^{1,5}, M. Marathe^{1,3}, S. Eubank^{1,5}, L. P. Sah^{7,8,9}, A. P. Giri^{7,8,9}, L. A. Colavito^{7,8,9}, K. S. Nitin¹⁰, V. Sridhar¹⁰, R. Asokan¹⁰, R. Muniappan⁷, G. Norton⁶, A. Adiga¹

¹Biocomplexity Institute & Initiative, University of Virginia

²Department of Computer Science, Virginia Tech

³Department of Computer Science, University of Virginia

⁴Department of Economics, Virginia Tech

⁵Department of Public Health Sciences, University of Virginia

⁶Department of Agriculture and Applied Economics, Virginia Tech

⁷Feed the Future Integrated Pest Management Innovation Lab

⁸Feed the Future Asian Vegetable and Mango Innovation Lab

⁹International Development Enterprises, Nepal

¹⁰Indian Institute of Horticultural Research

Trade and transport of goods is widely accepted as a primary pathway for the introduction and dispersal of invasive species. However, understanding commodity flows remains a challenge owing to its complex nature, unavailability of quality data, and lack of systematic modeling methods. A robust network-based approach is proposed to model seasonal flow of agricultural produce and examine its role in pest spread. It is applied to study the spread of *Tuta absoluta*, a devastating pest of tomato in Nepal. Further, the long-term establishment potential of the pest and its economic impact on the country are assessed. Our analysis strongly indicates that trade plays an important role in the spread of *T. absoluta*. The economic impact of this invasion could range from USD 17-25 million. The proposed approach is generic and particularly suited for data-poor scenarios.

1 Introduction

Invasive alien species spread is a complex phenomenon driven by various natural and anthropogenic factors. While the knowledge of biology and climate is essential to assess establishment risk and devise sustainable management strategies ([Sutherst 2000](#)), it is critical to understand human-mediated pathways to prevent introduction and mitigate immediate impact ([Hulme 2009](#); [Banks et al. 2015](#)). Increasingly, the need for a comprehensive outlook to tackle this problem is being highlighted ([Cunniffe et al. 2015](#); [Robinet et al. 2012](#); [Carrasco et al. 2010](#)); yet studies that analyze trade and travel in the context of food safety are only beginning to emerge ([Colunga-Garcia and Haack 2015](#); [Tatem 2009](#); [Ercsey-Ravasz et al. 2012](#); [Nopsa et al. 2015](#); [Early et al. 2016](#)).

As is the case with many built infrastructures, trade of goods naturally yields to network representations. Typically, nodes of the network represent locations—ranging from markets to continents depending on the context—connected by transportation infrastructure. The influence of one node on another is captured by an edge with weight that is a function of the transaction volume across that edge. A major challenge in constructing such networks is estimating the temporal flows. The intricate web of supply chain logistics makes it hard to document transactions, and even in economically developed countries, obtaining commodity specific flow data is a challenge (Magarey et al. 2011). On the other hand, it is comparatively easy to obtain datasets on production, population, and economic indicators at finer spatial resolution, thus allowing the use of spatial interaction models such as the gravity model (Haynes et al. 1984). Such an approach is also better suited for data-poor regions.

A representative flow network can yield valuable insights into the phenomenon. The network structure helps identify possible entry points and hubs (Nopsa et al. 2015; Sutrave et al. 2012). Network dynamical processes such as the SEIR (Susceptible → Exposed → Infected → Recovered) model from epidemiology (Pastor-Satorras et al. 2015) are applied to capture the spatio-temporal evolution of the invasion. Model selection and validation is challenging due to the lack of accurate ground surveillance, since very few countries have the capacity to effectively react to impending and emerging invasions (Early et al. 2016). Therefore, there is a need for a robust modeling framework that can operate with limited observational data.

This work addresses three key issues that current modelers encounter: (*i*) lack of a systematic approach to investigate the role of human-mediated pathways, (*ii*) the difficulty in evaluating models in the absence of accurate validation datasets, and (*iii*) the need to synthesize disparate datasets and modeling methodologies in order to provide a comprehensive assessment of the situation. Our framework is outlined in Figure 1. We apply it to study the spread dynamics of *Tuta absoluta* (*Gelechiidae, Lepidoptera*) (Meyrick, 1917), a devastating pest of the tomato crop. The region of interest is Nepal, a biodiversity hotspot (Kindlmann 2011) and largely agrarian economy, which recently reported *T. absoluta* invasion in 2016 (Bajracharya et al. 2016).

Indigenous to South America, *Tuta absoluta* or the South American tomato leaf miner was accidentally introduced to Spain in 2006 (Desneux et al. 2010). It has rapidly spread throughout Europe, Africa, Western Asia, the Indian subcontinent, and parts of Central America (Biondi et al. 2018) over the past decade. Since tomato is among the top two traded vegetables worldwide (<http://www.fao.org>), it is strongly suspected that trade played a critical role in *T. absoluta*'s rapid spread. Indeed, on multiple occasions it has been discovered in packaging stations, e.g., in Netherlands (NPPO 2009). Karadjova et al. (2013) observed that the spread pattern in Bulgaria was correlated with prime trade routes. With tomato being a commercially important crop (Grousset et al. 2015), *T. absoluta* has had a significant global impact. For example, in Turkey, the annual estimated intervention cost is €167M per year (Oztemiz 2014). Due to extensive insecticide treatment in Europe (Gontijo et al. 2013), insecticidal resistance has been recently observed in populations of *T. absoluta* (Guedes and Siqueira 2013). Lack of effective natural enemies has made integrated pest management challenging.

We propose a network-based model to represent seasonal flow of agricultural produce among major markets accounting for production, population, per capita income, and transportation infrastructure. Under the hypothesis that trade is the primary driver of spread, pest dispersal is modeled as a stochastic dynamical process over the seasonal flow networks. Applying techniques from an emerging field in network science, namely *epidemic origin inference* (see for example (Shah and Zaman 2011)), we establish a strong correlation between the expected spread over the flow network and observational data. This is the first application of this method in the context of invasive species spread.

Our analysis strongly indicates that given its dependency on vegetable imports and the pattern of agricultural produce flow unraveled by our model, Nepal is extremely vulnerable to attacks from pests and pathogens that can spread through trade. Our assessment of suitability for *T. absoluta* establishment based on growth and stress factors and observational data is that except for the mountainous regions in the north, *T. absoluta* can survive year-round in Nepal. However, the highest damage potential is in the southeastern parts. Our assessment of the economic impact based on a partial equilibrium approach predicts a social welfare loss of \$17.5-24.7 million due to this invasion.

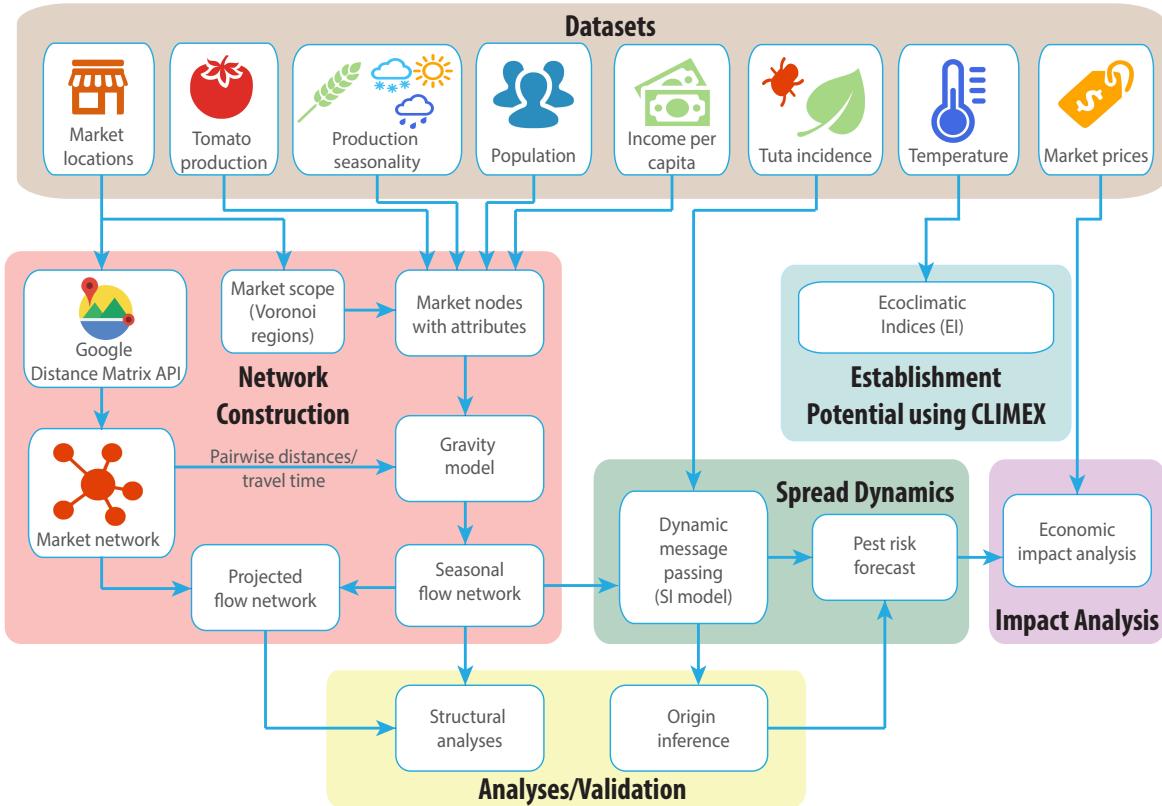


Figure 1: A schematic of the modeling framework showing the diverse datasets used and the different modules.

2 Methods

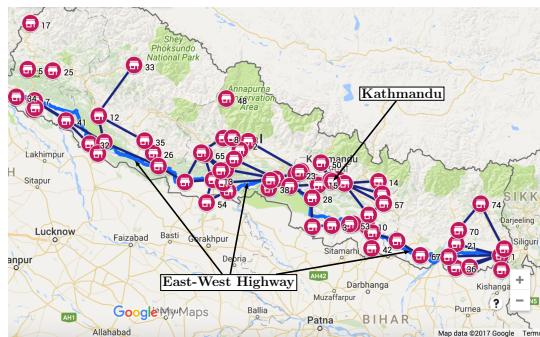
2.1 Data

Datasets used in our framework are listed in Table ??, Supplementary Information. We link several open-source datasets along with qualitative inputs from local experts in order to model the seasonal trade of the tomato crop as well as pest dynamics. Some of the challenges arise from the fact that the datasets vary in their spatial and temporal resolution and their year of release. Owing to the unique geography of Nepal, the vegetable production cycle varies with altitude (see Figure 2b). The annual production data was combined with the knowledge of production cycle to model the spatio-temporal variation in production across seasons. Major vegetable markets were geolocated using Google Maps, and Google Distance Matrix API was used to construct the road network and compute travel times. Several organizations have been involved in the monitoring of *T. absoluta* spread in Nepal: Nepal Agricultural Research Council (NARC), United States Agency for International Development (USAID), iDE Nepal, ENBAITA, and Agricare Pvt. Ltd. The pest was monitored using pheromone traps that were installed in several Village Development Committees (VDC). In May 2016, *T. absoluta* was officially reported by NARC's entomology division in Lalitpur (near Kathmandu). More details are provided in Section ?? of Supplementary Information. For the purpose of our analysis, we used surveillance data until May 2017, a year after the first official report (Figure 2c).

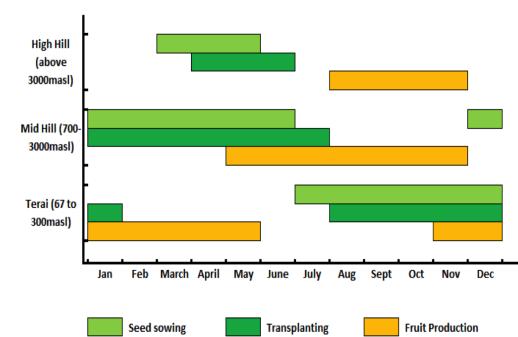
2.2 Nepal's physiography, tomato production, trade and demographics

Despite being a small country (800km along the Himalayan axis, 150-250km across), Nepal has high geographic diversity owing to its altitude. Along the north-south axis, Nepal is divided into three belts,

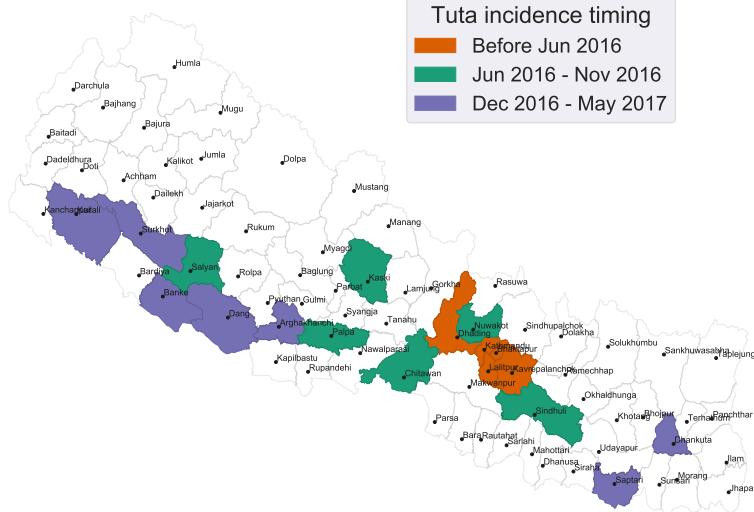
namely Terai, Mid Hills and High Hills (or Mountains) (Figure 3a). Terai is a densely populated region (with more than 400 people/sq.km. (Ali 2000)), while the High Hills are the least populated (with the exception of Kathmandu valley, which has the highest population density in Nepal). The altitude variations also impose different seasonal cycles for agricultural production (see Figure 2b). Production in the Mid Hills and High Hills is largely restricted to the summer months of May to November, while the Terai region produces during the winter months of November to May. Even though Nepal imports tomatoes from India all year round, the volume is not significant enough to influence domestic trade. For instance, in 2014 Nepal exported only 1% of its tomatoes and imported about 6-7% of its total consumption (<http://www.fao.org/faostat>). Also, the tomato processing industry in Nepal is not well developed, and fresh tomatoes are mostly traded for consumption (Nepal Economic, Agriculture, and Trade 2011). This motivates the use of population and per capita income as indicators of tomato consumption in a given district. Nepal is divided into development regions from east to west as shown in Figure 3e. The Central Development Region (CDR), which includes Kathmandu, generates 80% of the country's revenue, and 60% of government expenditure is allotted to it (Vandernoot and Van Hove 2014). In comparison, the revenue generated by the Far Western Development Region was less than 5% of that of CDR.



(a) Market network



(b) Production seasonality



(c) *T. absoluta* incidence timings

Figure 2: Market locations, production cycle and *T. absoluta* monitoring and distribution in Nepal. Data shown represents surveillance until May 2017.

2.3 Modeling the domestic tomato trade network

The domestic trade of tomato is modeled as a network of flows between major vegetable markets in Nepal based on the following premise: the total outflow from a market is a function of the amount of produce in its surrounding regions, and the total inflow is a function of consumption, which in turn is a function of the population to which it caters and the corresponding per capita income. The main assumptions in this model are based on the discussion in the previous section. They are: (i) imports and exports are not significant enough to influence domestic trade; (ii) fresh tomatoes are mainly traded for consumption as there is negligible processing; and (iii) the higher the per capita income, the greater the consumption.

The nodes of the flow network are the major markets, 69 in all (Figure 2a). The seasonal flows are estimated using a doubly constrained gravity model (Kaluza et al. 2010; Anderson 2011). The flow F_{ij} from location i to j is given by $F_{ij} = a_i b_j O_i I_j f(d_{ij})$ where, O_i is the total outflow from i , I_j is the total inflow to j , d_{ij} is the travel duration from i to j , $f(\cdot)$ is the *distance deterrence function*, and coefficients a_i and b_j are computed through an iterative process to ensure flow balance. The reader should note that the resulting network can contain self loops representing volume consumed locally. The total outflow O_i is a function of the seasonal production assigned to that market. The total inflow I_j is a function of consumption at the market. Based on Figure 2b, we identified two seasons – S1: June to November and S2: December to May. We partitioned the districts into two groups: Mid Hills and High Hills belong to group 1, while the Terai districts belong to group 2. All districts belonging to group i were assigned their respective annual production for season S_i and zero for the other season. Thus, we derived two distinct flow networks, one for each season by assigning for every district, its production to one season. The details of how these quantities were estimated is in Section ??, Supplementary Information. Also provided are the methods used for validation and sensitivity analysis.

2.4 Spread dynamics

We apply a discrete-time SI (Susceptible-Infected) epidemic model (Pastor-Satorras et al. 2015) on the domestic trade networks to simulate the spread of *T. absoluta* by the domestic trade pathway. Each node is either susceptible S (free from pest) or infected I (pest is present). Henceforth, we use the term “infected” for a node or a region frequently to imply *T. absoluta* infestation at that location. A node i in state I infects each of its out-neighbors j in the network with probability proportional to the flow F_{ij} at each time step t . The infection probabilities are obtained by normalizing flows globally: $\lambda_{ij} = \frac{F_{ij}}{\max_{i,j} F_{ij}}$. The model is based on two assumptions: (i) an infected node remains infected and continues to infect its neighbors, and (ii) the chance of infection is directly proportional to the volume traded. Considering the fact that Nepal was ill-prepared for this invasion and the lack of effective intervention methods, (i) is a fair assumption. Historically, *T. absoluta* has spread rapidly in regions where tomato trade has been the highest (parts of Europe and Middle-East for example) thus motivating assumption (ii). The quantity of interest is the probability that node i is infected by time t given the initial condition f_0 (which assigns probability of infection at time step $t = 0$ to each node). We denote it by $P_I(i, t, f_0)$. In general, computing it is difficult. We adopt the *dynamic message passing algorithm* by Lokhov et al. (Lokhov et al. 2014) to estimate this. This is described in (Venkatraman et al. 2017).

The initial configuration f_0 is chosen to mimic a spatially dispersed seeding scenario. We first select a *central* seed node, and then use a Gaussian kernel with parameter σ around the seed node to assign initial infection probabilities for neighboring markets. A market at a geodesic distance d from the seed is assigned the infection probability $e^{-\frac{d^2}{2\sigma^2}}$. The kernel accounts for factors such as uncertainty in determining the pest location, the possibility of spread of the pest through natural means, as well as interactions between these markets.

A Bayesian inference approach for model evaluation. Given that Nepal had limited resources to monitor the *T. absoluta* invasion, the pest incidence reports are not adequate. While it is hard to determine month of invasion at any location, some confidence can be placed on whether introduction happened by the end of a particular season. We evaluated our model based on the following backward inference problem: for an observation of node states at time t , what is the most likely origin of invasion (also known as the source detection problem (Shah and Zaman 2011))? We examined the likelihood of

markets or regions being the source nodes, and in particular, we compared this with the likelihood of the region around Kathmandu being the source. Suppose \mathcal{O} is the observation criterion; it consists of pairs (v, X) where v is a node and $X \in \{S, I\}$ is a state. For each candidate initial condition f_0 , we estimated the joint probability of \mathcal{O} at a time step t as a product of the marginal probability estimates from the message passing algorithm and define an *energy function* for each tuple (f_0, t) as

$$\phi(\mathcal{O}|f_0, t) = -\log \left(\prod_{(v, X) \in \mathcal{O}} P_X(i, t, f_0) \right)$$

The lower the value of ϕ , the higher the likelihood of f_0 being the initial condition. Secondly, recalling the uncertainty in interpreting time step t , we examined the relative likelihoods of each f_0 and the stability of the ranking across a range of model parameters. This is again discussed in ([Venkatramanan et al. 2017](#)).

2.5 Long-term establishment

In order to evaluate the long term establishment potential of *T. absoluta* in Nepal, various growth and stress factors regulating the development of *T. absoluta* were computed with particula reference to Kathmandu District. The parameter fitting values used in our study for CLIMEX modeling were iteratively changed after Tonnang et al. ([Tonnang et al. 2015](#)). This was done particularly for lower temperature threshold (DV0) and upper temperature threshold (DV3) to 7°C and to 40°C, respectively, in order to accommodate areas with *T. absoluta* infestation reported in Nepal. We used EI scaled up to 100 in order to get a clear gradient picture in terms of suitability of the pest in different geographical localities within Nepal, with 0 EI indicating non-suitable areas for *T. absoluta* establishment and areas with higher EIs indicating proportionately more favorable climatic conditions for the pest. Meteorological data of Nepal over the years indicates an increase in the average temperature at a rate of 0.05°C/year between 1975 to 2005 ([Marahatta et al. 2009](#)). Rising temperatures may a have a direct impact on the spread and establishment of the insect pests. Keeping this in view, an impact of 1°C rise in temperature scenario was also modeled to determine the near future potential areas for the establishment of *T. absoluta*. The resulting differences in EI values were mapped to find the influence of a 1°C rise in temperature for different regions in Nepal for *T. absoluta* establishment.

2.6 Economic impact

Direct economic impact. The direct impact measures the direct revenue loss from the tomato crop as determined by $P \sum_i h y_i l_i z_i$ which is the sum of loss encountered by each district. For each of Nepal's 75 districts, the spread dynamics module provides the probability of invasion and hence the proportion of area affected (z_i). The proportional loss in the affected area (h) is assumed to be 0.25. Here y_i is the yield per unit of land in district i before being affected and l_i is the tomato production area in the district.

Total economic impact. We use the partial equilibrium approach ([Alston et al. 1995](#); [Soliman et al. 2012](#)), which focuses on the dynamics of the tomato market and assumes that the price for substitute and complementary goods are given, i.e. no changes occur in the substitute and complementary goods market due to changes in tomato prices. Due to *T. absoluta* infestation, the supply curve of tomato shifts leftwards in each district based on the probability of invasion. Assuming no changes occur in the downward sloping demand curve, shift in supply results in a higher equilibrium price and lower equilibrium quantity. This further impacts the consumers' and producers' welfare as measured by the consumers' surplus and producers' surplus (profits).

The total economic impact, or the change in social welfare, is the sum of change in consumers' and producers' surplus. To measure this, we first estimate the demand parameters given the initial price and then use it to estimate the new equilibrium price. Once both prices are known, we calculate the changes in consumers' and producers' surplus and hence the social welfare. For details on the economic impact calculation, please see the Supplementary Information.

3 Results

3.1 Structural properties of the trade network

Recall that there are two networks, one for each season S1 and S2, each comprising of 69 nodes. Analysis of the networks with respect to the gravity model parameters is discussed in (Venkatramanan et al. 2017). In Figure 3, we present representative observations of tomato trade between markets resulting from the gravity model. It is helpful to view these results in terms of the physiographic divisions (Figure 3a) and development regions (Figure 3e). Our model accounts for the fact that the High Hills/Mid Hills and the Terai are the primary sources of tomato during seasons S1 and S2 respectively. This is clearly reflected in the net flow diagram between geographic regions: north (High Hills/Mid Hills) to south (Terai) in S1 and south to north in S2. However, an interesting pattern to be noted is the significant flow from east to west during S1 as observed in the net flow diagram between the Development Regions. The East and Central Development Regions produce significantly more tomatoes than the other regions. Also, the presence of the arterial East-West highway that almost covers the entire breadth of the country (Figure 2a) helps in the east to west trade.

Comparison with the annual flow network. To evaluate the importance of seasons, we constructed the annual flow network by using the gravity model with annual production for each district. The resulting flows are shown in Figures 3h and 3d. Compared to the seasonal flows we see that annual flows are of shorter distance and thus there is not much flow between regions (either between east and west or south and north). This comparison clearly emphasizes the need to account for seasonal trade.

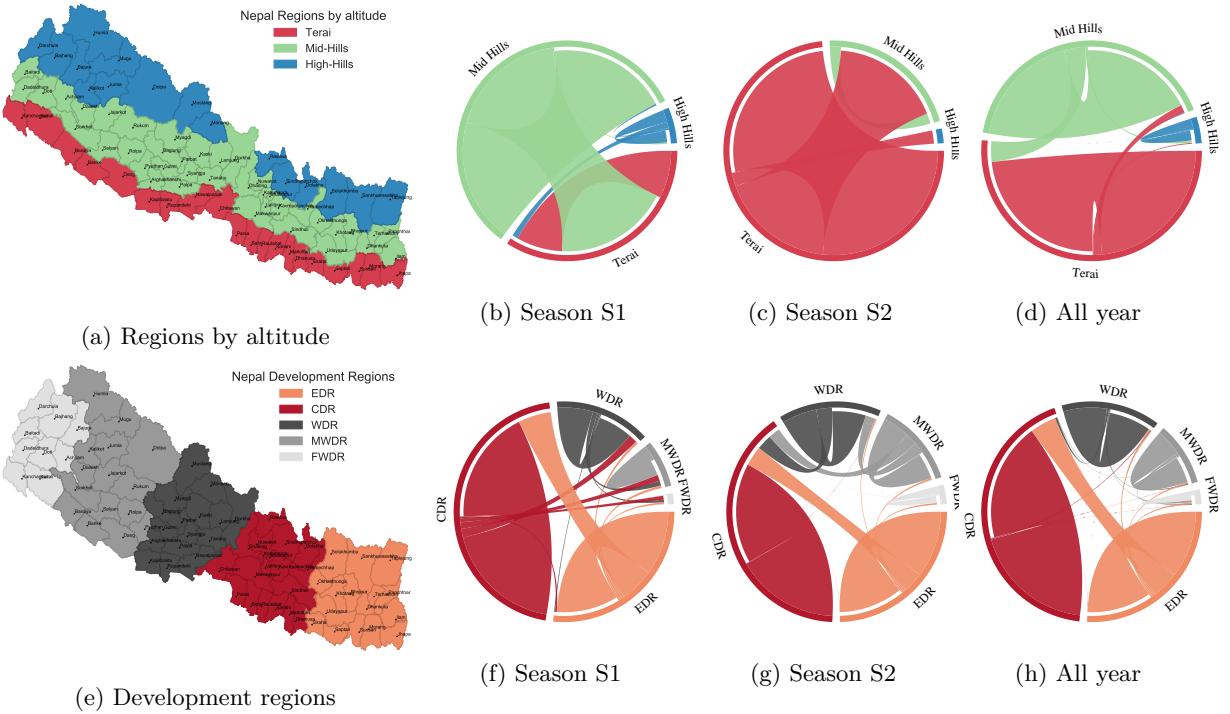


Figure 3: **The spatio-temporal structure of the flow network.** S1 ranges from May to November, while S2 ranges from December to April. The first row shows the flow from east to west between development regions of Nepal. The second row depicts the flow from north to south between regions of different altitudes. While the second and third columns correspond to seasonal flow, the last column corresponds to the flows generated from annual data.

3.2 Role of trade network in pest spread

The experiment was setup under the premise that *T. absoluta* was first introduced to Nepal in the Kathmandu valley. Ground experts have high confidence in this assumption since the pest was not discovered in the previous growing season in other parts of Nepal. Given the pest reports till December 2016 (Figure 2c), we evaluate our model based on the following backward inference problem: for an observation of node states at time t , what is the most likely origin of invasion? This is also known as the source detection problem ([Shah and Zaman 2011](#)). Using the energy function defined in Methods, we examine the likelihood of markets or regions being the source nodes, and in particular, we compare this with the likelihood of the region around Kathmandu being the source (see Figure 4). Secondly, recalling the uncertainty in interpreting time step t , we examined the relative likelihoods of each f_0 and the stability of the ranking across a range of model parameters.

We considered the spread during June-November (season S1) for model evaluation. Using the S1 flow network, our objective was to rank various starting configurations f_0 based on $\phi(\mathcal{O}|f_0, t)$ given \mathcal{O}, t . For a given σ , we evaluated the likelihood of each node being the central node. We considered two criteria based on which the likelihood of each f_0 as the starting configuration was computed: (i) \mathcal{O}_G : this is the set of all pairs (v, I) where v is a market node that belongs to a district that reported pest presence by December 2016. (ii) \mathcal{O}_B : this is the set of (v, I) for all nodes v , setting up the baseline which assumes no observational data.

The results are shown in Figure 4. Firstly, we observed that for both criteria \mathcal{O}_G and \mathcal{O}_B , the top few ranks are relatively robust to varying network and model parameters. Also, for both criteria markets from the Central Development Region (CDR) that belong to Kathmandu and its adjacent districts are among the top ranked nodes. Interestingly, for the criterion \mathcal{O}_G , Dhankuta (EDR), with the highest assigned production has a very low rank (Figure 4a) and a low ϕ value compared to the top market in \mathcal{O}_G . However, for \mathcal{O}_B , it is ranked second (Figure 4b). This clearly shows that while Dhankuta has the potential to infect a large number of areas, given what has been observed it is very unlikely that it was the source of infection. Dhankuta reported presence of the pest only towards the end of 2016 (see Figure 2c).

Spread in season S2 To study the spread from November 2016 to May 2017, we considered the dynamics on season S2 network. To set the initial conditions, we used the results of our inference study and chose Kathmandu with $\sigma = 10$ as the seed distribution. For this initial condition, we obtained the probability of infection for all nodes in S1 for T1 time steps. This distribution is used as initial condition for the S2 network spread. Figure 4c shows the infection probabilities for a particular combination of (T_1, T_2) . As seen in Figure 4c, our model suggests that most of Terai and Mid Hills regions of CDR and WDR were affected by the end of May 2017, and in the subsequent seasons, the pest should affect most of the Terai and Mid Hills region. From Figure 2c, we see that regions belonging to Terai in CDR and Mid Hills of WDR and MWDR had already reported pest presence (marked in Figure 4c). The latest reports indeed corroborate our findings. With the exception of 3-4 districts, the pest is present in the entire Terai and Mid Hills regions.

While the intended usage of the origin inference formulation is to determine the source of infection, we have adapted it to compare expected spread in the model with observed data. Our results demonstrate that this framework is in general very useful in finding the likely pathways of introduction of the pest.

3.3 Establishment potential

Figure 5a shows the plot of Ecoclimatic Indices (EI). The Terai and Mid Hills show relatively favorable climatic conditions for *T. absoluta* establishment. The locations that already report pest incidence have EI ranging from 30-75. High Hills does not favor its establishment, which may be attributed to severe cold stress. Eastern parts of Nepal (districts such as Udaypur, Saptari, Sindhuli and parts of Sansari) are more favorable to *T. absoluta* establishment than the central and western parts. With a rise of 1°C in temperature, most of the Terai regions of Nepal may experience a reduction in the EI value (ranging from 15-20), though the pest can survive and establish itself in these regions (see Figure 5b). But some parts of Mid Hills may experience an increase (15-20) in EI values with a 1°C rise. However, considering the increasing adoption of protected cultivation methods ([Nepal Ministry of](#)

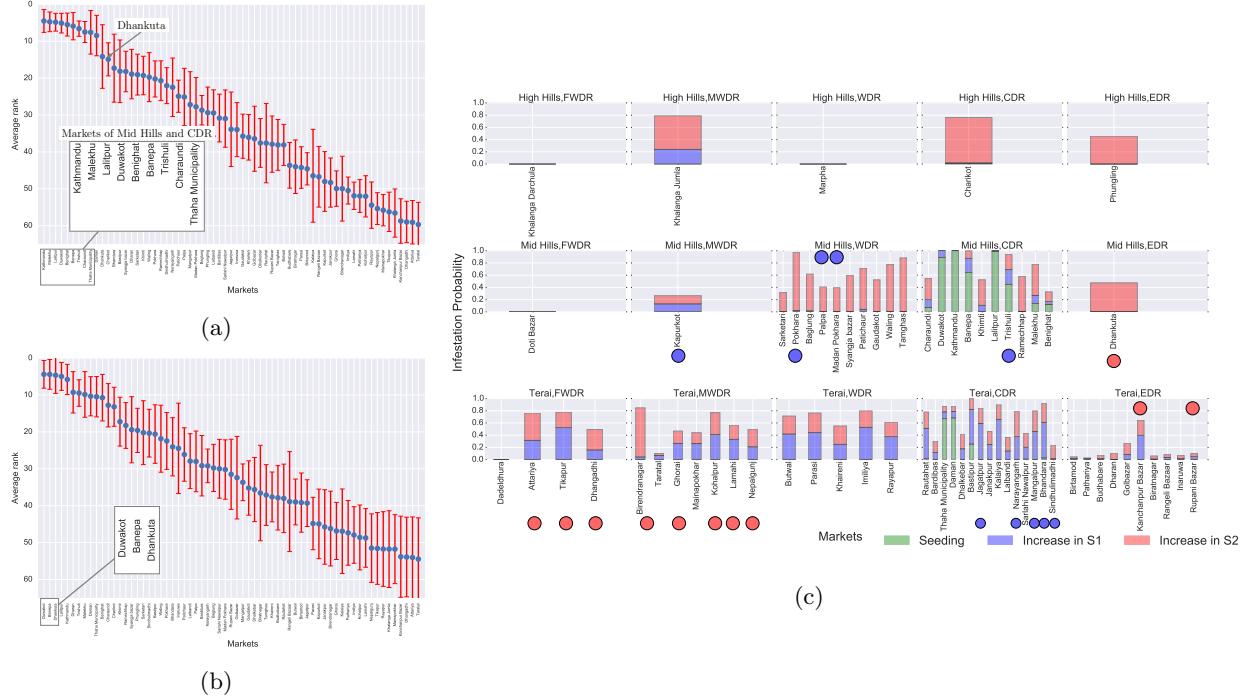


Figure 4: **Evaluating the spread model using epidemic source inference framework.** (a) The average rank of each market based on the likelihood for the criterion O_G for a range of model parameters. (b) Same as (a), but for criterion O_B . (c) **Spread in S2.** The parameters used were $\beta = 2$, $\kappa = 500$, $\sigma = 15$, $\gamma = 1$, $T_1 = T_2 = 10$ with Kathmandu as the seed node. The blue dots correspond to markets whose districts reported *T. absoluta* presence before December 2016 (season S1), while the red dots correspond to markets which reported later. More details on sensitivity analysis are presented in (Venkatramanan et al. 2017).

Agriculture Development 2012), this is in some sense a conservative estimate. As observed in Europe and Mediterranean regions (Karadjova et al. 2013), *T. absoluta* has been known to survive unfavorable weather conditions in greenhouses.

In particular, Kathmandu and its surrounding locations have an EI in the range of 15–45, corresponding to low to moderate level of infestation. Growth index analysis for Kathmandu region (see Figure S4 in Supplementary Information) indicates that infestation can be expected throughout the year. The weekly growth index peaks during February–March and October–November, indicating the possibility of higher incidence of *T. absoluta*; the lower weekly growth index during April–May indicates lower incidence of the pest. Another growth promoting factor for *T. absoluta* identified for Kathmandu is the moisture index which directly influences the weekly growth index of the pest.

3.4 Economic impact analysis

We evaluated the economic impact of *T. absoluta* in Nepal based on the projection of its spread from the initial infestation in Kathmandu. Since Nepal's tomato production is less than 0.2% of the global tomato production, most of which is used to meet the domestic demand for tomatoes, and it exports only 1% of its production, so we treated it as a small closed economy (<http://www.fao.org/faostat>). A 25% crop loss (Bajracharya et al. 2016) in cultivated areas infected by *T. absoluta* results in a direct economic impact of \$19.7M (Supplementary Information). The direct impact, however, does not account for the change in the market price of tomatoes due to loss in the crop or the impact of price change on consumers' and producers' welfare.

To calculate a more comprehensive total economic impact, we used the partial equilibrium approach (Alston et al. 1995; Soliman et al. 2012). The comprehensive economic impact analysis shows a social

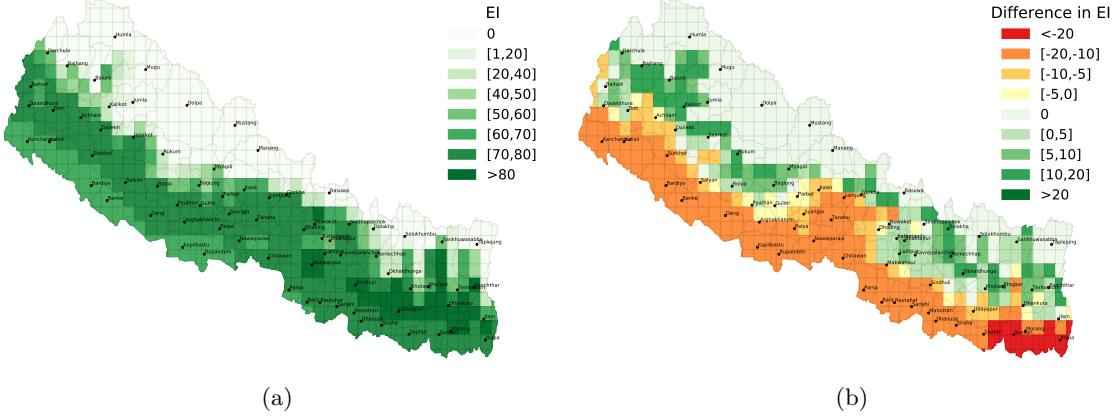


Figure 5: **Long-term establishment potential.** (a) Plot of the Ecoclimatic Index (EI) for current climate conditions. (b) The difference between EI's of current scenario and 1°C rise in temperature.

welfare loss of \$22.4M and a price increase of 32%. The analysis assumed ‘demand elasticity to price’ to be -0.7, ‘supply elasticity to price’ to be 0.5 (<https://data.ers.usda.gov>), base price at \$400 per ton (www.kalimatinmarket.com.np), crop loss due to pest invasion to be 25% (Bajracharya et al. 2016), and reduced net price due to increased cost of production from pest control to be 80% (Khidr et al. 2013). The pest risk probabilities were obtained after running the spread model for 10 time steps.

As mentioned in the model evaluation, due to inaccurate and inadequate incidence reports, it is not clear what time duration each simulation step t corresponds to. So we performed a sensitivity analysis of the economic impact for two other time steps; one where the spread model is run for only 5 time steps and the other where it is run for 20 time steps. The longer run forecasts result in higher pest risk probabilities and hence higher economic impact, and vice versa. Pest risk probabilities based on 5 time steps, ceteris paribus, result in a direct economic impact of \$16M and a total economic impact of \$17.5M, whereas after 20 time steps these numbers increase to \$21.5M and \$24.7M respectively.

4 Discussion

Even though Nepal is mainly an agrarian society, its agriculture has been predominantly characterized by subsistence farming, poor marketing infrastructure, and dependency on imports from neighboring India. However, in the past decade, there has been a surge of efforts to improve this situation (Nepal Ministry of Agriculture Development 2012; Nepal Economic, Agriculture, and Trade 2011). Increasingly, farmers have been adapting protected cultivation methods such as tunnel farming (Nepal Ministry of Agriculture Development 2012) to increase yield, and reverse the trend of importing. Also, given its unique geography, Nepal is among the most susceptible regions to climate change (Kindlmann 2011, Chapter 2). Under such conditions, major invasion events such as *T. absoluta* are detrimental to its biodiversity, economy, and societal well-being in general. Therefore, there is a desperate need for concerted efforts to understand the complex food system of this country and increase its reactive capacity and resilience to attacks from invasive species. We believe that our work has taken the first steps in this direction.

Although there is general consensus that vegetable and seedling trade is a primary driver of *T. absoluta* spread (Desneux et al. 2010; Karadjova et al. 2013; USDA 2011; Campos et al. 2017), previous modeling efforts have only focused on establishment potential (Desneux et al. 2010; Tonnang et al. 2015) and spatial dispersion (Guimapi et al. 2016). This is the first work that analyzes human-mediated pathways in the context of *T. absoluta*. Some methodological aspects pertaining to this work appear in Venkatramanan et al. (2017).

Related work. In recent years, there has been a lot of interest in studying the role of international trade and travel in invasive species spread. Nopsa et al. (2015) studied the structure of a stored grain network

induced by storage facilities and an underlying rail network. Carrasco et al. (2010) use phenology models along with gravity models for human-assisted dispersal. Nopsa et al. (2015) evaluated the structure of rail networks in the US and Australia for pest and mycotoxin dispersal. Colunga-Garcia et al. (2009) used the regional freight transport information to characterize risk of urban and periurban areas to exotic forest insect pests in the US. Robinet et al. (2009, 2016) developed a long-distance dispersal model to study the spread of pests that accounts for heterogeneous human population densities in the study region. International (Early et al. 2016; Ercsey-Ravasz et al. 2012; Kaluza et al. 2010) and domestic (Magarey et al. 2011; Colunga-Garcia and Haack 2015) trade datasets have been analyzed to assess the susceptibility of countries to invasive alien species and contaminants. Tatem (2009) showed that the world-wide airline network increases the risks of establishment by providing busy transport links between spatially distant, but climatically similar regions of the world.

Challenges. Agro-trade networks are a complex system problem. The networks depend on varied factors, including seasonal production, population distribution, cultural factors, economic activity, storage, and transport infrastructure. Furthermore, data needed to develop agro-trade networks is often sparse, noisy, misaligned in reporting time and is not openly available. Apart from quantitative datasets, there is also need for qualitative information pertaining to the study region such as cultural practices, seasonal production cycles, etc. Interpreting this data and integrating it into model design requires local knowledge. Challenges exist in obtaining high-resolution pest distribution data. Monitoring is a resource intensive task. Also, the pest might not have been detected during the off season due to host unavailability. Its reporting might be delayed by farmers due to lack of awareness or fear of quarantining. Given these constraints, there might be several months of delay in reporting pest presence. Another challenge is validating the network representations. While international trade data is available at the commodity level, domestic data is hard to come by. Even in data-rich regions such as the US, the available sample data (e.g. Freight Analysis Framework¹) is aggregated at the commodity level. The role of these networks in the study of invasive species requires one to understand the ecological contagion processes.

Limitations and possible improvements. Since our study is one of the first to consider regional commodity flow analysis in the context of pest spread, especially *T. absoluta*, there are several avenues for improvement. While some of the limitations arise from lack of refined data, others are due to the limited understanding of the underlying complexity of pest invasions. The former may be the norm for emerging contagions in a data-poor region, whereas the latter will need several iterations of model development and validation by the scientific community. Our model predominantly focuses on commodity flow, and does not explicitly account for natural or other modes of spread (infected seedlings from nurseries for example). A more comprehensive model will need to integrate ecological suitability and biology directly in the diffusion process. Further studies will be needed to understand the pest's flying capacity, influence of wind direction, alternate hosts, etc. to realistically capture spread dynamics. Transportation of infected seedlings from nurseries to production sites and dispersal through packaging material are other possible causes of long-distance spread, thus hinting at multiple pathways that need to be accounted for within the umbrella term of "commodity flow". While gravity models have been applied to study a diverse set of phenomena that concern interaction between entities (Erlander and Stewart 1990; Kaluza et al. 2010; Bossenbroek et al. 2001; Thiemann et al. 2010; Jongejans et al. 2015; Krings et al. 2009) they do have known shortcomings (Simini et al. 2012; Rothlisberger and Lodge 2011). Further, the model could be refined by accounting for price variations and commodity varieties. The assignment of demand and supply attributes was done assuming homogeneous distribution of district level production and population information. This can be improved with (i) population and production estimates at a higher spatial resolution, (ii) ground survey data on each market's scope. The same holds for temporal resolution in assigning seasonal production.

We have described a first-principles based commodity modeling framework that integrates easily available datasets on population, production, etc. to model the flow of agricultural produce. We have demonstrated the validity of the constructed networks, and have used it to understand the impact of commodity flow on pest spread. Despite being limited by the availability of quality validation datasets, a

¹https://ops.fhwa.dot.gov/freight/freight_analysis/faf/

bare bones framework such as ours can be quickly extended to other vegetables, pests and regions with minimal effort.

Acknowledgments

This work was supported in part by the United States Agency for International Development under the Cooperative Agreement NO. AID-OAA-L-15-00001 Feed the Future Innovation Lab for Integrated Pest Management, DTRA CNIMS Contract HDTRA1-11-D-0016-0001, NSF BIG DATA Grant IIS-1633028, NSF DIBBS Grant ACI-1443054, NIH Grant 1R01GM109718 and NSF NRT-DESE Grant DGE-154362. G.N. was also partly supported by Virginia Agricultural Experiment Station project VA-136324.

References

- Ali, M. (2000). *Dynamics of vegetable production, distribution and consumption in Asia*. Shanhua (Taiwan) AVRDC.
- Alston, J. M., Norton, G. W., and Pardey, P. G. (1995). Science under scarcity. *CAB International. Wallingford, Oxon, UK*.
- Anderson, J. E. (2011). The gravity model. *Annual Review of Economics*, 3(1):133–160.
- Bajracharya, A. S. R., Mainali, R. P., Bhat, B., Bista, S., Shashank, P., and Meshram, N. (2016). The first record of South American tomato leaf miner, *Tuta absoluta* (Meyrick 1917)(Lepidoptera: Gelechiidae) in Nepal. *J. Entomol. Zool. Stud*, 4:1359–1363.
- Banks, N. C., Paini, D. R., Bayliss, K. L., and Hodda, M. (2015). The role of global trade and transport network topology in the human-mediated dispersal of alien species. *Ecology letters*, 18(2):188–199.
- Biondi, A., Guedes, R. N. C., Wan, F.-H., and Desneux, N. (2018). Ecology, worldwide spread, and management of the invasive south american tomato pinworm, *tuta absoluta*: past, present, and future. *Annual review of entomology*, 63:239–258.
- Bossenbroek, J. M., Kraft, C. E., and Nekola, J. C. (2001). Prediction of long-distance dispersal using gravity models: zebra mussel invasion of inland lakes. *Ecological Applications*, 11(6):1778–1788.
- Campos, M. R., Biondi, A., Adiga, A., Guedes, R. N., and Desneux, N. (2017). From the Western Palaearctic region to beyond: *Tuta absoluta* ten years after invading Europe. *Journal of Pest Science*.
- Carrasco, L., Mumford, J., MacLeod, A., Harwood, T., Grabenweger, G., Leach, A., Knight, J., and Baker, R. (2010). Unveiling human-assisted dispersal mechanisms in invasive alien insects: integration of spatial stochastic simulation and phenology models. *Ecological Modelling*, 221(17):2068–2075.
- Colunga-Garcia, M. and Haack, R. A. (2015). Following the transportation trail to anticipate human-mediated invasions in terrestrial ecosystems. *Pest Risk Modelling and Mapping for Invasive Alien Species. CAB International, Wallingford, UK*, pages 35–48.
- Colunga-Garcia, M., Haack, R. A., and Adelaja, A. O. (2009). Freight transportation and the potential for invasions of exotic insects in urban and periurban forests of the united states. *Journal of Economic Entomology*, 102(1):237–246.
- Cunniffe, N. J., Koskella, B., Metcalf, C. J. E., Parnell, S., Gottwald, T. R., and Gilligan, C. A. (2015). Thirteen challenges in modelling plant diseases. *Epidemics*, 10:6–10.
- Desneux, N., Wajnberg, E., Wyckhuys, K. A., Burgio, G., Arpaia, S., Narváez-Vasquez, C. A., González-Cabrera, J., Ruescas, D. C., Tabone, E., Frandon, J., et al. (2010). Biological invasion of European tomato crops by *Tuta absoluta*, ecology, geographic expansion and prospects for biological control. *Journal of Pest Science*, 83(3):197–215.

- Early, R., Bradley, B. A., Dukes, J. S., Lawler, J. J., Olden, J. D., Blumenthal, D. M., Gonzalez, P., Grosholz, E. D., Ibañez, I., Miller, L. P., Sorte, C. J. B., and Tatem, A. J. (2016). Global threats from invasive alien species in the twenty-first century and national response capacities. *Nature Communications*, 7.
- Ercsey-Ravasz, M., Toroczkai, Z., Lakner, Z., and Baranyi, J. (2012). Complexity of the international agro-food trade network and its impact on food safety. *PloS one*, 7(5):e37810.
- Erlander, S. and Stewart, N. (1990). *The Gravity Model in Transportation Analysis: Theory and Extensions*. VSP, Utrecht.
- Gontijo, P., Picanço, M., Pereira, E., Martins, J., Chediak, M., and Guedes, R. (2013). Spatial and temporal variation in the control failure likelihood of the tomato leaf miner, *Tuta absoluta*. *Annals of Applied Biology*, 162(1):50–59.
- Grousset, F., Suffert, M., and Petter, F. (2015). EPPO Study on pest risks associated with the import of tomato fruit. *EPPO Bulletin*, 45(1):153–156.
- Guedes, R. and Siqueira, H. (2013). The tomato borer *Tuta absoluta*: insecticide resistance and control failure. *Plant Sciences Reviews 2012*, page 245.
- Guimapi, R. Y., Mohamed, S. A., Okeyo, G. O., Ndjomatchoua, F. T., Ekesi, S., and Tonnang, H. E. (2016). Modeling the risk of invasion and spread of *Tuta absoluta* in Africa. *Ecological Complexity*, 28:77–93.
- Haynes, K. E., Fotheringham, A. S., et al. (1984). *Gravity and spatial interaction models*, volume 2. Sage Beverly Hills, CA.
- Hulme, P. E. (2009). Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46(1):10–18.
- Jongejans, E., Skarpaas, O., Ferrari, M. J., Long, E. S., Dauer, J. T., Schwarz, C. M., Rauschert, E. S., Jabbour, R., Mortensen, D. A., Isard, S. A., et al. (2015). A unifying gravity framework for dispersal. *Theoretical Ecology*, 8(2):207–223.
- Kaluza, P., Kölzsch, A., Gastner, M. T., and Blasius, B. (2010). The complex network of global cargo ship movements. *Journal of the Royal Society Interface*, 7(48):1093–1103.
- Karadjova, O., Ilieva, Z., Krumov, V., Petrova, E., Ventsislavov, V., et al. (2013). *Tuta absoluta* (Meyrick)(Lepidoptera: Gelechiidae): Potential for entry, establishment and spread in Bulgaria. *Bulgarian Journal of Agricultural Science*, 19(3):563–571.
- Khidr, A., Gaffar, S., Maha, S., Nada, A., Taman, A., Fathia, A., and Salem, A. (2013). New approaches for controlling tomato leafminer, *Tuta absoluta* (Meyrick) in tomato fields in Egypt. *Egyptian Journal of Agricultural Research*, 91(1):335–348.
- Kindlmann, P. (2011). *Himalayan biodiversity in the changing world*. Springer Science & Business Media.
- Krings, G., Calabrese, F., Ratti, C., and Blondel, V. D. (2009). Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(07):L07003.
- Lokhov, A. Y., Mézard, M., Ohta, H., and Zdeborová, L. (2014). Inferring the origin of an epidemic with a dynamic message-passing algorithm. *Physical Review E*, 90(1):012801.
- Magarey, R. D., Borchert, D., Engle, J., Colunga-Garcia, M., Koch, F. H., and Yemshanov, D. (2011). Risk maps for targeting exotic plant pest detection programs in the United States. *EPPO Bulletin*, 41(1):46–56.
- Marahatta, S., Dangol, B. S., and Gurung, G. B. (2009). *Temporal and Spatial Variability of Climate Change Over Nepal, 1976-2005*. Practical Action Nepal Office.

- Nepal Economic, Agriculture, and Trade (2011). Value Chain/Market Analysis of the vegetable Sub-Sector in Nepal. Technical report, USAID/Nepal.
- Nepal Ministry of Agriculture Development (2012). Value chain development plan for tomato. <http://pact.gov.np/docs/publication/Value%20Chain%20Dev%20for%20Tomato%20book.pdf>.
- Nopsa, J. F. H., Daglish, G. J., Hagstrum, D. W., Leslie, J. F., Phillips, T. W., Scoglio, C., Thomas-Sharma, S., Walter, G. H., and Garrett, K. A. (2015). Ecological networks in stored grain: Key postharvest nodes for emerging pests, pathogens, and mycotoxins. *BioScience*, page biv122.
- NPPO (2009). *Tuta absoluta* Povolny (Gelechiidae) - tomato leaf miner - in tomato packaging facility in The Netherlands, National Plant Protection Organization (NPPO), Wageningen.
- Oztemiz, S. (2014). *Tuta absoluta* povolny (Lepidoptera: Gelechiidae), the exotic pest in Turkey. *Romanian Journal of Biology*.
- Pastor-Satorras, R., Castellano, C., Van Mieghem, P., and Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925.
- Robinet, C., Kehlenbeck, H., Kriticos, D. J., Baker, R. H., Battisti, A., Brunel, S., Dupin, M., Eyre, D., Faccoli, M., Ilieva, Z., et al. (2012). A suite of models to support the quantitative assessment of spread in pest risk analysis. *PLoS One*, 7(10):e43366.
- Robinet, C., Roques, A., Pan, H., Fang, G., Ye, J., Zhang, Y., and Sun, J. (2009). Role of human-mediated dispersal in the spread of the pinewood nematode in China. *PLoS One*, 4(2):e4646.
- Robinet, C., Suppo, C., and Darrouzet, E. (2016). Rapid spread of the invasive yellow-legged hornet in France: the role of human-mediated dispersal and the effects of control measures. *Journal of Applied Ecology*.
- Rothlisberger, J. D. and Lodge, D. M. (2011). Limitations of gravity models in predicting the spread of Eurasian watermilfoil. *Conservation Biology*, 25(1):64–72.
- Shah, D. and Zaman, T. (2011). Rumors in a network: Who's the culprit? *IEEE Transactions on information theory*, 57(8):5163–5181.
- Simini, F., González, M. C., Maritan, A., and Barabási, A.-L. (2012). A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100.
- Soliman, T., Mourits, M. C., Van Der Werf, W., Hengeveld, G. M., Robinet, C., and Lansink, A. G. O. (2012). Framework for modelling economic impacts of invasive species, applied to pine wood nematode in Europe. *PLoS One*, 7(9):e45505.
- Sutherst, R. W. (2000). Climate change and invasive species: a conceptual framework. *Invasive species in a changing world*, pages 211–240.
- Suttrave, S., Scoglio, C., Isard, S. A., Hutchinson, J. S., and Garrett, K. A. (2012). Identifying highly connected counties compensates for resource limitations when evaluating national spread of an invasive pathogen. *PLoS One*, 7(6):e37793.
- Tatem, A. J. (2009). The worldwide airline network and the dispersal of exotic species: 2007–2010. *Ecography*, 32(1):94–102.
- Thiemann, C., Theis, F., Grady, D., Brune, R., and Brockmann, D. (2010). The structure of borders in a small world. *PloS one*, 5(11):e15422.
- Tonnang, H. E., Mohamed, S. F., Khamis, F., and Ekesi, S. (2015). Identification and risk assessment for worldwide invasion and spread of *Tuta absoluta* with a focus on Sub-Saharan Africa: implications for phytosanitary measures and management. *PloS one*, 10(8):e0135283.
- USDA (2011). New Pest Response Guidelines: Tomato Leafminer (*Tuta absoluta*). *Animal and Plant Health Inspection Service, Plant Protection and Quarantine*.

Vandernoot, J. and Van Hove, C. (2014). Disparities between development regions and district development committees in Nepal. *International Advances in Economic Research*, 20(3):353–355.

Venkatramanan, S., Wu, S., Shi, B., Marathe, A., Marathe, M., Eubank, S., Sah, L. P., Giri, A., Colavito, L. A., Nitin, K., et al. (2017). Towards robust models of food flows and their role in invasive species spread. In *Big Data (Big Data), 2017 IEEE International Conference on*, pages 435–444. IEEE.

Mechanistic and Data-Driven Agent-Based Models to Explain Human Behavior in Online Networked Group Anagram Games

Vanessa Cedeno-Mieles^{†§}, Zhihao Hu[‡], Xinwei Deng[‡], Yihui Ren[#],

Abhijin Adiga^{*}, Christopher Barrett^{*}, Saliya Ekanayake^{*}, Gizem Korkmaz^{*}, Chris J. Kuhlman^{*}, Dustin Machi^{*}, Madhav V. Marathe^{*}, S. S. Ravi^{*}, Brian J. Goode[&], Naren Ramakrishnan^{¶†}, Parang Saraf^{¶†}, Nathan Self^{¶†}

Departments of Computer Science[†] and Statistics[‡], Discovery Analytics Center[¶], Biocomplexity Institute[&]

Virginia Tech - Blacksburg, VA

Escuela Superior Politécnica del Litoral, ESPOL - Guayaquil, Ecuador[§]

Brookhaven National Laboratory - Upton, NY[#]

Biocomplexity Institute, University of Virginia - Charlottesville, VA^{*}

Noshir Contractor

*Department of Industrial Engineering
and Management Sciences
Northwestern University
Evanston, IL*

Joshua M. Epstein

*Department of Epidemiology, NYU
New York, NY
Santa Fe Institute
Santa Fe, NM*

Michael W. Macy

*Department of Sociology
Cornell University
Ithaca, NY*

Abstract—In anagram games, players are provided with letters for forming as many words as possible over a specified time duration. Anagram games have been used in controlled experiments to study problems such as collective identity, effects of goal-setting, internal-external attributions, test anxiety, and others. The majority of work on anagram games involves individual players. Recently, work has expanded to group anagram games where players cooperate by sharing letters. In this work, we analyze experimental data from online social networked experiments of group anagram games. We develop mechanistic and data-driven models of human decision-making to predict detailed game player actions (e.g., what word to form next). With these results, we develop a composite agent-based modeling and simulation platform that incorporates the models from data analysis. We compare model predictions against experimental data, which enables us to provide explanations of human decision-making and behavior. Finally, we provide illustrative case studies using agent-based simulations to demonstrate the efficacy of models to provide insights that are beyond those from experiments alone.

I. INTRODUCTION

A. Background and Motivation

In one form of an **individual anagram game**, a player is provided with a set of alphabetical letters to form as

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ASONAM '19, August 27-30, 2019, Vancouver, Canada

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6868-1/19/08

<http://dx.doi.org/10.1145/3341161.3342965>

many words as possible in a prescribed time duration. The performance of a player is often quantified based on the number of words formed.

In a **group anagram game** (GrAG), multiple players collaborate. Each player is given letters and forms words with her own letters, and can share letters with her neighbors to enable everyone to form more words. Figure 1 provides a schematic of a 3-player GrAG. Each player (v_1 , v_2 , and v_3) is initially provided with $n_l = 3$ letters as shown. A player may form words, and through the communication channels in gray, may request letters and reply to letter requests.

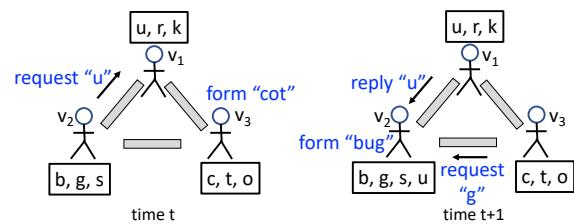


Fig. 1. Simplified view of a networked group anagram game (GrAG), with illustrative actions among $n = 3$ players that communicate and share letters through the gray channels. Each player is initially given $n_l = 3$ letters. Letters that a player has “in-hand” to form words are shown in boxes. Player actions are shown in blue. At time t , v_2 requests a “u” from v_1 and v_3 forms the word “cot.” At the next time, v_2 receives a “u” from v_1 , forms the word “bug,” and receives a request from v_3 .

Overwhelmingly, research on anagram games considers the individual setting. It has been extensively studied (over 20 published works) for more than 60 years to analyze phe-

nomena such as goal-setting, compensation types, internal-external attributions, and test anxiety (e.g., [1], [2]). Other names for anagram game are *word formation game* and *word construction game*.

There are several reasons to study GrAGs. The research in [3] used them to study experimentally the formation of collective identity (CI), defined in social psychology as an individual's cognitive, moral, and emotional connection with a broader community, category, practice, or institution [4]. A second motivation is their relevance to other types of group dynamics, notably intergroup and intragroup cooperation and competition (e.g., [5]). A third motivation is that many of the phenomena listed above for the individual anagram game (e.g., goal-setting) could be studied in group settings with models of group behavior.

Overall, researches involving anagram games encompass a broad range of disciplines like sociology, economics, management science, and (social) psychology [1], [2], [6]. It is clear that using anagram games is valuable in various fields of research. The first and only work on modeling GrAGs was recently completed [7]. We enumerate the differences between our work and [7] in Section I-B immediately below.

B. Our Work Scope and Differentiators from Previous Work

Work scope. Our work starts with data from online social network GrAGs. (The game platform and online experiments are *not* the focus in this work.) With these data: (i) data analytics are performed to support model development; (ii) different models for different player actions in the GrAG are developed; (iii) the models are evaluated against experimental data; and (iv) these models are then recast as agent-based models and executed within an agent-based modeling and simulation (ABMS) platform to produce computational results that go beyond the experiments.

Based on this work scope, all of the following are completely different in this work, compared to that in [7]: data analytics, the aspects of the game that are being modeled, the types of modeling techniques used, the models themselves, and the quantities that the models predict. We address particular differences between [7] and our work now.

Work in Ref. [7]. The subject of [7] is the *action type and time (ATAT) model*, which uses multinomial logistic regression to build the model. In that work, the goal was to develop models to predict the *type* of action taken in time, e.g., predictions of the form: player v_i takes action type “form word” at time t . Also, if a player action is form word, and the player has letters that cannot form a word (e.g., letters q , z , and r) then that model will nevertheless form an unspecified (unrealistic) word from these letters. Moreover, the models of [7] do not consider the particular letters assigned to players in a game. Consequently, all player behaviors will tend toward the same mean behavior in agent-based simulations (ABSs).

Our work. In contrast, our work focuses on three *component models*. Different models are developed for the actions “form word,” “request letter,” and “reply to (letter) request.” Our models account for network structure, letter assignments and

letters in-hand (i.e., letters that a player has to form words), and particular player parameter assignments—all of which can vary among players—so results will remain distinct across agents. That is, we capture heterogeneity in several ways.

Our ABMS framework uses a **composite model**: a combination of the ATAT model (to determine what action types players take in time) and the three component models developed herein (to predict the specifics of each action). The composite model is our agent-based model (ABM). This ABMS system simulates GrAG scenarios beyond those of the experiments.

C. Novelty of Our Work

First, our work is an exemplar of a detailed procedure for combining mechanistic and data-driven models to form single models of human *decision-making* that output human *actions* in a game. *Mechanistic models*, for our purposes, have the following characteristics: (i) the models are based on first principles and are not tied to any particular domain; and (ii) the models are specified, implemented, and executed without any experimental data. To augment mechanistic models by accounting for variability in player behaviors, *data-driven models* are constructed from analyses of experimental data. Second, because we prove that the mechanistic models capture player behavior, these models *explain* behaviors, as described in our contributions below. Third, our mechanistic models are novel: Levenshtein Distance (LD) [9] (see Section IV-A) and a greedy optimization procedure describe human decision-making and have not been used in anagrams contexts (we could not find LD used in any modeling of human behavior, as we do here). As called for in the social sciences, our focus is on model construction and predictions, and explanations of human behavior [10], [11].

D. Contributions

1. A process for combining mechanistic and data-driven approaches to build models of human decision-making. We provide the details of our process in Section IV. See Figure 2. First, mechanistic models are conjectured and evaluated by comparing their predictions to experimental data. This does three things: (i) enables comparisons of model predictions with experimental data, and if these comparisons are favorable (which they are), then (ii) the structures of the models, and the mechanisms embedded in them, provide *explanations* for human decision-making [12], [13], and (iii) the mechanistic models form the basis of the ABMs. Second, because the mechanistic models can be improved by including data from experiments, we use data-driven modeling approaches to introduce stochasticity to account for variability across human subject game players. Hence we utilize these two modeling approaches in a well-defined process.

2. Mechanistic models. We use concepts such as LD, word corpora, word proximity networks (WPNs), and a greedy optimization algorithm (all defined in Section IV) to develop mechanistic models for two of the three player actions (see Figure 2). The LD model, used for word formation, could be used within any agent that is required to form words, and

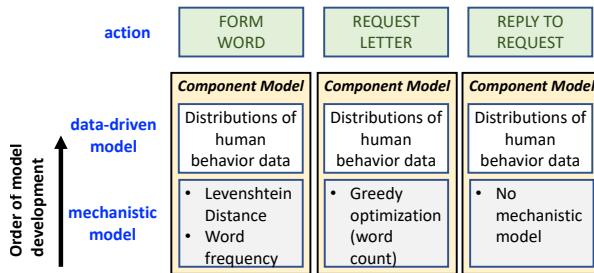


Fig. 2. **Component models** (i.e., combined *mechanistic* and *data-driven* models) for the three player actions in the GrAG. These are models of human decision-making, which output specific player actions in the game. The particulars of the mechanistic and data-driven models are given in the respective boxes under the actions and are detailed in Section IV. Mechanistic models are built first, and then augmented with data-driven models.

the greedy optimization algorithm, used for requesting letters, could be used by agents to make a choice from among a finite set of options. That is, these models are not tied to our GrAG. But the next contribution presents their utility within the GrAG.

3. New experimental findings and explanations of player behaviors based on cognitive and economic theories. The analyses focus on data for three types of player actions: (1) form a word; (2) request a letter; and (3) reply to a letter request. See Figure 2. A summary of some explanations follows. A word w_2 that a player forms is explained by considering (i) the letters that the player has in-hand (i.e., in her possession) and (ii) LD [9] between the most recently formed word w_1 and the next word to be formed w_2 from a candidate set of words (Section IV-B). This is motivated by, and consistent with, cognitive load theory [14] in that people try to reduce cognitive load during learning. Here, the closer the next word formed is to the previously formed word—as measured by LD—the lesser the cognitive load in forming a new word. For letter requests, we use the idea that player action is based on rational choice theory [15]. Our analyses (Section IV-C) demonstrate that the letter that a player requests from her neighbors is explained by identifying the letter that maximally increases the number of words that the player can form, when also considering the letters that the player has in-hand (greedy optimization algorithm). This behavior is consistent with rational choice theory. This is because players’ earnings in games are proportional to the number of words formed, so it is rational for a player to choose a letter to maximize the size of their candidate word set. It is interesting that our explanation means that players are reasoning beyond more naive approaches, such as simply requesting some “most frequently” used letter (e.g., preferring *e* over *z*). (We have modeled this naive approach—results not shown here—and this model’s results are not consistent with the data.) Also, the experimental data clearly show that players do not request all the available letters at the outset of a game. Rather, they request letters throughout the game as they identify use for them. Finally, we also show that there are four types of

behavior in replying to letter requests (Section IV-D).

4. Agent-based models and results. A family of ABMs are developed, yielding a composite model, where each ABM is comprised of a distinct model for each of the three actions, with user-specified parameter values for player/agent characteristics, such as the agent’s vocabulary and their aptitude, i.e., the degree to which they perform optimally. The multi-logit regression model based on [7] is adopted to determine which action type each agent selects at each discrete time in a simulation (time granularity is seconds). The selected action type then determines the appropriate model developed herein to predict details of the action. Note that there is a fourth action, a no-operation (no-op), where the agent does nothing at particular times, which represents agent thinking and requires no model. We also provide new insights from exercising the ABMs (see Section V), such as demonstrating how player performance decreases with decreasing player aptitude and the effects of heterogeneous initial letter assignments to players.

II. RELATED WORK

By far, the most relevant study to our work is the modeling in [7], which is **agent-based modeling of anagram games**. To the best of our knowledge that is the only work prior to ours that models the GrAG [7]. That work was discussed in detail in relation to our work in Sections I-A and I-B. We now address other topics related to our work.

Anagram experiments. Over 20 experiment works (e.g., [1], [2]) use *single player* anagram games. The only cooperative GrAG, which is *face-to-face*, is reported in [3]. The game is used to foster CI among teammates.

Networked experiments and modeling. There are several other online (e.g., [16]) and in-person (e.g., [3]) experiments with interacting participants that can be represented as networks, and analyses of network populations (e.g., [17], [18]), where edges represent interaction channels.

Mechanistic and data-driven modeling. Several works use AI methods and data to model behavior (e.g., tutoring and learning [19]). Also, neuroscientists are using neuro-imaging to understand human decision-making; [20] discusses optimization methods, such as the one we use in the model for requesting letters.

Explanatory modeling. There are many works (e.g., [12], [13]) that describe different definitions of *explanations*, different types of explanations that models provide, and procedures for arriving at explanations. We follow ideas from [12], [13]: that the structure of mechanistic models that adequately predict human behavior can be used to explain behavior.

III. ONLINE SOCIAL-NETWORKED GROUP ANAGRAM GAME

We built a customized web application (web app) for an online GrAG. Players recruited through Amazon Mechanical Turk (MTurk), are provided game instructions, participate in the GrAG through their web browsers, and are paid based on their performance. A total of 48 experiments were performed using a total of 367 players, with numbers of players per game

ranging from 3 to 17. The game duration is 5 minutes. In the following, we describe the GrAG/experiment.

Figure 1 provides a description of the game setup and actions. A game begins with n players, v_1 through v_n . Each player has a degree d that specifies the number of connections to other players. A connection (edge) between two players denotes a communication channel where a letter ℓ can be requested and sent (sending a letter is a reply). Thus, an experiment configuration is a graph $G(V, E)$ with player set V and communication channels E . In experiments, G is a k -regular random graph ($k \equiv d$), with uniform degree $2 \leq k \leq 8$. Each player starts the game with n_ℓ initial letters, which they can use to form words or share among their neighbors, when requested. At the beginning of a game, a word corpus C^W is defined with a list of words a player can form during the game. For this we use a list of the top 5000 words from the 450 million word Corpus of Contemporary American English, the only large and balanced corpus of American English [22]. The three major player actions in a game are now described.

Player action: forming a word. At any point during a game, a player v_i can form a word w_i . All letters in the word w_i must come from the set of letters v_i has in-hand L_i^{ih} (superscript ih). A single letter ℓ in L_i^{ih} can appear any number of times in a word. For a word submission to be accepted in the game, the word has to be in the word corpus C^W . The C^W is specified, but it is not provided to players. Rather, players have to recognize possible words that can be formed from the letters they have. The C^W is the same for all players in all games. A player can submit a word only once; multiple players can form the same word.

Player action: requesting a letter. At any point during a game, a player v_i can request a letter ℓ_{ij}^{req} from a neighbor v_j 's set of n_ℓ initial letters L_j^{init} . The anagram game screen shows all neighbors' initial letters as available for request. A letter received by v_i is put into the set L_i^{ih} .

Player action: replying with a letter. At any point during a game, a player v_i can reply with a letter ℓ_{ij}^{rep} to a neighbor v_j 's request (ℓ_{ij}^{rep} must be in L_i^{init}). The anagram game screen for v_i shows all of the letters requested of v_i .

To encourage cooperation, any letter in L_i^{ih} can be used any number of times in forming words, and the letter is not lost; the letter bestows an infinite supply of use. Similarly, if v_i requests a letter ℓ from v_j , and v_j replies with it, v_j still retains a copy of the letter and can use it. Also, earnings for the team are based on the total number of words formed, and all players receive $(1/n)$ of the total earnings. Typical player earnings are \$7 to \$10 per game.

IV. DATA ANALYSIS AND MODEL DEVELOPMENT

Figure 2 provides the roadmap for building the models for the three player actions, which is the focus of this section. Ultimately, our goal is to use these models as ABMs in an ABMS framework to study GrAGs well beyond those of experiments.

For each action—which is a component model of the ABM—we provide: (i) our premise for understanding player

behavior and the key concepts for this premise, (ii) experimental analyses and results for these key ideas that construct and justify (i.e., give evidence for) the component model of the composite ABM, and (iii) a formal algorithm for the component model for the action in Figure 2. Note that the steps of algorithms that we specify below are not focused on efficient implementation, but rather on conveying the steps of the algorithms as they relate to the data analyses. First, we address preliminaries.

A. Preliminaries

We introduce two concepts used in data analysis and modeling. **Levenshtein distance** (d^L) [9], an edit distance, is prominent in our work and the work's novelty, and is motivated by work in linguistics and bioinformatics [21]. It quantifies the difference in letters between two words. In starting with one word to obtain a second word, a letter substitution counts as one, as does each of letter insertion and letter deletion. Hence, going from *had* to *hats* requires $d^L = 2$: one to substitute *t* for *d* and one for inserting an *s*.

A **word proximity network** (WPN) is a clique graph $H(V_H, E_H)$ where vertices V_H are words that can be formed, according to a word corpus C^W , with the letters that a player currently has in-hand and E_H is the set of edges between pairs of words, labeled with the d^L between the two words.

B. Player Action: Form Word

Basic premise, assumptions, and key concepts. We seek to identify a method that explains the process of players selecting words to form. Our premise is that given the last word w_1 that v_i has formed, the next word w_2 that v_i will form will be one with minimal d^L from w_1 because this requires a minimal number of letter manipulations (i.e., lesser cognitive load [14]). For the first word, v_i selects the most frequent word from the corpus that can be formed with its letters in-hand L_i^{ih} . (The word corpus provides the frequency of occurrence of each word.) We note that for each player v_i , there is a set L_i^{ih} of letters that she has in-hand and a corresponding set $W_i^{ih} \subseteq C^W$ of words that v_i can form from the entire corpus C^W of words, based on the letters in L_i^{ih} . As v_i requests and receives more letters from her neighbors, the cardinalities of L_i^{ih} and W_i^{ih} will (typically) increase. Also note that for a given word w_1 formed by v_i in a game, W_i^{ih} can be partitioned based on $d^L(w_1, w_2)$ for fixed w_1 and for each $w_2 \in W_i^{ih}$ using the WPN. Let $W_i^{ih}(w_1, d^L) \subseteq W_i^{ih}$ be the set of words at d^L from w_1 that v_i can form.

Our data analysis is based on two central ideas, for each player v_i . First, we compare d^L values between two consecutive words formed (w_1 and then w_2), both the actual value $d_{i,act}^L(w_1, w_2)$ measured from experiments and the optimal (i.e., minimal) value of d^L , denoted $d_{min}^L(w_1, w^*)$, for some w^* in W_i^{ih} that is at a minimum LD from w_1 . Both d^L values are based on v_i 's set L_i^{ih} . (We drop the arguments when they are obvious from context.) Second, for a given set of words at some d^L from w_1 , denoted $W_i^{ih}(w_1, d^L)$, we select w_2 based on the popularity of words as provided by the rank

(frequency of use) from [22]. All of these parameters are either inputs (e.g., C^W), measured in experiments, or computed from experimental data. These high-level steps enable us to understand players' behavior in forming words, as described next.

Data analysis. Analysis step 1. For each player v_i in a game, we consider pairs of consecutive words formed, (w_1, w_2) . From this, we compute $d_{i,act}^L(w_1, w_2)$, the actual d^L . Also from these data and from L_i^{ih} at the time w_2 was formed, we can compute d_{min}^L and the word set $W_i^{ih}(w_1, d_{min}^L)$. We compute $\Delta d^L = d_{i,act}^L - d_{min}^L$. A value of zero means that the player is performing optimally according to our premise; a value > 0 means that v_i is performing suboptimally— v_i is making more letter edits (expending greater effort) than is required by the data.

We rank the players by their average Δd^L , Δd_{ave}^L , over all pairs of words (w_1, w_2) that they form in a game. We partition the ranking of players into five equi-sized bins, P_1 through P_5 , such that players in P_1 (resp., P_5) have the smallest (resp., largest) values of Δd_{ave}^L . That is, the players in P_1 perform closest to optimal. A player v_i 's aptitude b_i^{wf} in forming words takes a value from P_1 through P_5 . We take this player-centric approach because we want to produce agent models based on individual player and groups of players' behaviors.

Analysis step 2. For each of the five groups of players P_j ($1 \leq j \leq 5$), we plot all data points $(x, y) = (d_{min}^L, d_{i,act}^L(w_1, w_2))$ for each person in that group, in Figure 3. In each plot, for each d_{min}^L on the x-axis (the *mechanistic model prediction*), there is a range of $d_{i,act}^L(w_1, w_2)$ (from the data) for all v_i in a particular 20% bin. If we break the players down into 10% bins (instead of the 20% bins), the top 30% of players perform such that the median value of $d_{i,act}^L(w_1, w_2)$ equals d_{min}^L . That is, in a median sense, these top 30% of players form words w_2 such that $d_{i,act}^L(w_1, w_2) = d_{min}^L$, and hence w_2 is formed optimally (i.e., according to the mechanistic model). Moreover, if we look at the top 80% of players, then $d_{min}^L \leq d_{i,act}^L(w_1, w_2) \leq d_{min}^L + 1$. That is, for 80% of all players, the pairs of consecutive words, (w_1, w_2) , produce $d_{i,act}^L$ values that differ by at most 1 from the d_{min}^L . These data for $|C^W| = 5000$ substantiate our premise that players form word w_2 based on d^L . Although not shown, similar results are generated for $|C^W| = 1000, 2000, 3000$, and 4000, if we take these sets as the 1000, 2000, 3000, and 4000 most frequently used words in the original corpus of 5000 words.

Analysis step 3. For each box plot in Figure 3, we form a frequency distribution \mathcal{D}^{d^L} as a function of the triple $(C_i^W, b_i^{wf}, d_{min}^L)$. Figure 4 provides one such distribution. So, given a C_i^W , an aptitude b_i^{wf} for forming words, and a d_{min}^L , one can sample an actual LD, $d_{i,act}^L$, in forming w_2 from w_1 .

Analysis step 4. For a given w_1 and $d_{i,act}^L$, $W_i^{ih}(w_1, d_{i,act}^L) \subseteq W_i^{ih}$ is the candidate set of words that v_i can form as w_2 . The issue is how players extract a particular word from $W_i^{ih}(w_1, d_{i,act}^L)$ as w_2 . Figure 5 provides the answer. For each v_i , we rank the words in $W_i^{ih}(w_1, d_{i,act}^L)$

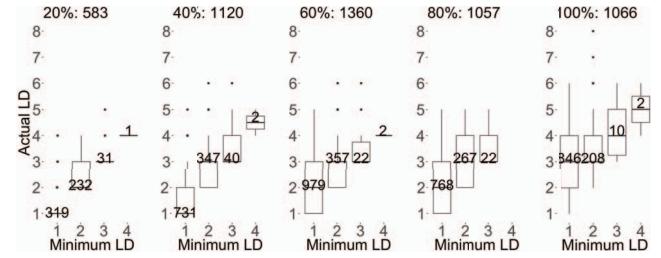


Fig. 3. Comparison of *mechanistic* model predictions against data for the *form word* model. Mechanistic predictions are the values on the x-axis (d_{min}^L); data are on the y-axis ($d_{i,act}^L$). We use the $|C^W| = 5000$ word corpus. Each plot corresponds to a grouping of players by d^L , and represents, in turn, P_j , $j \in \{1, 2, 3, 4, 5\}$, moving left to right. Numbers are numbers of observations in the data. If $d_{i,act}^L(w_1, w_2) = d_{min}^L$, then the experimental data correspond exactly with the mechanistic model.

$$C_i^W = 5000, b_i^{wf} = P_1, d_{min}^L = 1$$

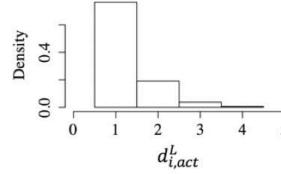


Fig. 4. For $(C_i^W, b_i^{wf}, d_{min}^L) = (5000 \text{ words}, P_1, 1)$, the distribution \mathcal{D}^{d^L} of $d_{i,act}^L$ from experiments is shown. For a given d_{min}^L computed for optimal behavior, the appropriate distribution is sampled to obtain $d_{i,act}^L$ for v_i . These distributions are formed from the data in Figure 3 and they are part of the *data-driven* model of form word.

in decreasing order of frequency of occurrence (which is obtained from the word corpus itself), such that the first ranked word is the most frequently used word. This plot shows the number of times the chosen word w_2 is of a particular rank. It is clear that players select w_2 based on the frequency of the word's use, e.g., the top-ranked word is selected almost 700 times from the corpus. This result also holds over different corpus sizes from 1000 to 5000 words. These data support our use of a mechanistic model of selecting the word with highest frequency of use in a word corpus from the candidate set of words.

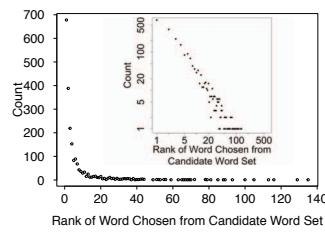


Fig. 5. Experimental data for $|C_i^W| = 5000$. Log-log scale plot (inset) of the distribution of ranks of words formed by players from the word set $W_i^{ih}(w_1, d_{i,act}^L)$. Lesser rank means higher word frequency from corpus. Players most often choose words with lesser rank (i.e., greater frequency).

Remark: These data analyses substantiate our claim that our models are explanatory. The data are consistent with the explanation that humans reason about what word to form using LD and word frequency (familiarity), consistent with cognitive load theory [14].

Remark: It is emphasized that players in the experiments are not given a word corpus, frequency of letter use, d^L concepts and values, etc. Our construction and procedures presented here are our representation of the mental decision-making processes that players engage in, resulting in human behavior in the form of detailed actions. In experiments, players are

Input: Agent $v_i \in V$. Agent word-forming aptitude b_i^{wf} . Word corpus or vocabulary C_i^W for v_i . Letters in-hand L_i^{ih} . Most recent word formed by v_i , w_1 . Words W_i^f formed up to now by v_i . Distribution \mathcal{D}^{wr} of word frequencies from C_i^W and distribution \mathcal{D}^{d^L} of $d_{i,act}^L$ frequency as a function of tuple $(C_i^W, b_i^{wf}, d_{min}^L)$.

Output: Next word w_2 that v_i forms, if any.

Steps:

- 1) From letters in-hand L_i^{ih} , construct the set W_i^{ih} of words that v_i can form (and that v_i has not yet formed). Set $V_H = W_i^{ih}$ and let H be the WPN network induced by V_H . Let the edge set be E_H , with edge labels of d^L .
- 2) If V_H is empty, terminate algorithm and return no word.
- 3) From the values of the edge labels $d^L(w_1, w_j)$, for all edges $\{w_1, w^*\} \in E_H$ of WPN H , where $w_1, w^* \in V_H$, determine the minimum LD, d_{min}^L .
- 4) For the triple $(C_i^W, b_i^{wf}, d_{min}^L)$, sample from the distribution \mathcal{D}^{wr} to obtain the actual LD, $d_{i,act}^L$, that v_i will use to form the next word. (Example provided in Figure 4.)
- 5) From the set $W_i^{d_{i,act}^L} \subseteq V_H$ of words at $d_{i,act}^L$ from w_1 , order the words from most frequently used word to least (C^W provides this ranking).
- 6) From the frequency distribution \mathcal{D}^{wr} of words in $W_i^{d_{i,act}^L}$, draw a rank r_i of a word. Select the unique word w_2 that corresponds to rank r_i . Return w_2 .

Fig. 6. Steps of the Algorithm FORM WORD. This algorithm returns a word that an agent forms.

only given letters and the ability to share them. This remark holds for the next two models, too.

Algorithm for form word. The algorithm is in Figure 6, and follows directly from the above data analysis. This is cast as the *agent* model in the ABMS.

C. Player Action: Request Letter

Basic premise, assumptions, and key concepts. Our goal is to uncover a process that explains how players select the next letter to request from their neighbors. Our premise is that player v_i will select the next letter to request as the letter from the set of candidate neighboring letters L'_i that produces the greatest increase in the number of words that v_i can form. The key idea is to examine each candidate letter ℓ and determine the number of new words $|W_i^{ih\ell}|$ that can be formed with existing letters in L_i^{ih} and the requested letter combined (this word set is $W_i^{ih\ell}$), rank these letters in decreasing order of $|W_i^{ih\ell}|$, and select the letter to request based on this ranking. This is a greedy process—in the sense of selecting the best letter (i.e., the letter that ranks first), one at a time—and is our mechanistic model. This is a rational choice approach [15] because players are incentivized to form as many words as possible, so it is rational to select a letter that maximally increases the number of words that can be formed. Note that as more letters have been requested and received, the number of letters to request, $|L'_i|$, decreases because once a player has a letter, she can use it any number of times. We now provide the evidence for behavior that is aligned with this premise.

Data analysis. Analysis step 1. We rank all players by their performance in requesting letters in the GrAG, as follows. For each v_i , and for each actual letter request, we rank the candidate letters to request in L'_i according to our greedy

model (given immediately above), and then identify the rank $r_{i,act}$ of the letter $\ell_{i,act}$ actually requested. Then we compute an average rank of letter requests $r_{i,ave}$ for each v_i , over the first 1/2 of all v_i 's requests. We use only the first 1/2 of requests in computing $r_{i,ave}$ because as $|L'_i|$ decreases, the selected rank and the top-ranked letters will be more closely aligned because there are so few letters left. Hence, in order to not bias the results, we use only the first 1/2 of letter requests. The players v_i are ranked by $r_{i,ave}$, smallest to largest value, and the players are partitioned into five equi-sized bins Q_1 through Q_5 , where players in Q_1 (resp., Q_5) select letters to request that are most (resp., least) conformant to our mechanistic model. A player v_i 's aptitude b_i^{req} in requesting letters takes a value from Q_1 through Q_5 . This partitioning is to ensure a sufficient number of observations for each bin. Again, we partition based on players because we want to develop agent behaviors based on player behavior.

Analysis step 2. We analyze each Q_j , $j \in \{1, 2, 3, 4, 5\}$, separately, as follows. We take each $v_i \in Q_j$, note each rank $r_{i,act}$ corresponding to each letter request in the first 1/2 of requests, count the number of occurrences of the ranks of each requested letter, and sum the counts over all players. Results are shown in the left-most plot of Figure 7 for $b_i^{req} = Q_1 = 20\%$. (Note that player v_i 's aptitude b_i^{req} in requesting letters may take values Q_1 through Q_5 .) These data are compared against our mechanistic model (in green), which predicts all letter requests will be of rank 1 in this plot. Note that for the $b_i^{req} = Q_1 = 20\%$ data, the number of occurrences of a selected rank generally increases as the rank decreases, though the effect is sometimes less pronounced for some cases. We claim that the data support our premise, i.e., our model explains the data. Players request letters that generate the greatest increase in the words that they can form.

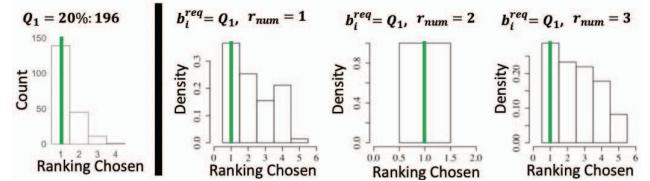


Fig. 7. Comparison of *mechanistic* model predictions (in green) against data (the distributions) for the *request letter* model. Our mechanistic model predicts all letter requests will be of rank-1 in each of the four plots. (LEFT) Experimental data are for the 5000-word corpus, aptitude $b_i^{req} = Q_1 = 20\%$ for letter requests (plots for Q_j , $j \in \{2, 3, 4, 5\}$ are not shown). For aptitude Q_1 , the frequency of the rank of the chosen letter is plotted. These data show that players most often choose letters with lower rank, meaning that they choose letters that can form relatively more words. (RIGHT) These three plots break down the left-most plot by showing distributions for different request numbers r_{num} of 1, 2, and 3, by v_i . These distributions $\mathcal{D}^{\ell r}$ are used to sample $r_{i,act}$ based on $(C_i^W, b_i^{req}, r_{num})$.

Analysis step 3. We break down each plot of the type in Figure 7, at the left, to account for C_i^W , b_i^{req} , and the number r_{num} of the letter request in the three right-most plots of the figure. By sampling from frequency distributions $\mathcal{D}^{\ell r}$ based on $(C_i^W, b_i^{req}, r_{num})$ for v_i , we obtain the rank of the actual letter

Input: Agent $v_i \in V$. Agent letter requesting aptitude b_i^{req} . Word corpus C_i^W . Letters in-hand L_i^{ih} . The set L'_i of letters that v_i 's neighbors were initially assigned that v_i has not yet requested; this is the candidate set of letters to request. The request number r_{num} . Distributions D^{ler} of letter ranks for triples $(C_i^W, b_i^{req}, r_{num})$.

Output: Next letter ℓ^* that v_i requests, if any.

Steps:

- 1) If L'_i is empty, terminate and return no letter.
- 2) For each candidate letter to request $\ell \in L'_i$ that has yet to be requested, determine the new words $W_i^{ih\ell}$ that can be formed from C_i^W with the letters in set $L_i^{ih} \cup \{\ell\}$ (include only words that have not yet been formed).
- 3) If every word set $W_i^{ih\ell}$ for all ℓ is empty, remove an arbitrary letter ℓ^* from L'_i , terminate this algorithm and return ℓ^* .
- 4) Rank the letters $\ell \in L'_i$ in decreasing values of $|W_i^{ih\ell}|$. Let $r(\ell)$ be the rank of ℓ .
- 5) Determine the rank $r_{i,act}$ of the letter to select for requesting by sampling from distribution D^{ler} using as input $(C_i^W, b_i^{req}, r_{num})$. (See Figure 7 for three examples.)
- 6) Select the letter ℓ^* such that $r(\ell^*) = r_{i,act}$. Break ties arbitrarily. Remove ℓ^* from L'_i . Return ℓ^* .

Fig. 8. Steps of the Algorithm REQUEST LETTER. This algorithm returns a letter that an agent requests.

requested $r_{i,act}$ in the model. This provides finer modeling granularity by accounting for the number of the letter request.

Algorithm for request letter. The algorithm is in Figure 8 and follows directly from the data analysis just presented. This algorithm is presented in the form of an *agent* model.

Remark: These analyses and data provide evidence for our claim that this model is explanatory. Players generally request letters by (roughly) maximizing the increase in number of words that they can form, which follows rational choice theory [15].

D. Player Action: Reply to Letter Requests

Unlike the previous two models, this model is purely data-driven. For space reasons, we provide an abbreviated description here.

Basic premise, assumptions, key ideas, and data analysis. The goal is to produce a model that explains how players respond to letter requests from their neighbors. The basic premise is that players can be partitioned into categories of behavior. We determined from the data these four categories: (1) those players that respond to all queued (pending) letter requests in their buffer (called FB for full buffer); (2) those that respond to some fraction of all pending letter requests in their buffer (called LTFB for less than full buffer); (3) those that sometime behave as FB and sometimes as LTFB (called Mixed); and (4) those that never reply to letter requests (called NR). The key ideas are that for each category, we need to determine: (i) how many replies to letter requests are made uninterrupted (i.e., contiguously) for categories LTFB and Mixed, and (ii) for each number of letter replies, the time duration over which these letter replies are made (for categories FB, LTFB, and Mixed). These are the four values for a player v_i 's aptitude b_i^{rpl} in replying to letter requests.

Algorithm for reply to (letter) request. Owing to space limitations, the algorithm is not provided here, but will be

in an extended version of this work.

Remark: In these various algorithms, elements of sets are returned, or a distribution corresponding to particular inputs is sampled. In some cases, there are no data for specified conditions. For these types of situations, we implement a recursive search technique to sample from the distribution or set with the closest set of inputs.

V. AGENT-BASED SIMULATIONS AND RESULTS

Remark: *Model evaluation* is an important step and has been performed. Figures 3, 4, and 7 are part of this process.

Simulation model. We conduct discrete time agent-based simulations (ABSS) of the GrAG. Each time unit is one second of the 300-second GrAG. At each time and for each agent, an action is selected. Based on the action chosen, the corresponding model for that action, developed herein, is executed (Figures 6 and 8 for “form word” and “request letter,” respectively, and the thinking action is a no-op). Thus, all agents behavior in the simulations follow these models. We run $n_{runs} = 100$ runs or simulation instances and average the results. We use the 5000-word corpus C^W . These are purely simulation studies and are not tied to the experiments. The goal is to demonstrate that the models alone provide insights into human behavior.

Study 1: Effects of model aptitude properties. We use a game configuration $G(V, E)$ consisting of six players that form a circle, with each player having two neighbors. The initial letter assignments are given in Table I. We systematically vary the aptitudes of players in forming words b_i^{wf} , in requesting letters b_i^{req} , and in replying to letter requests b_i^{rpl} . See Table II. Recall that these aptitudes correspond to the skill levels of players.

TABLE I
STUDY 1 INITIAL LETTER ASSIGNMENTS TO PLAYERS IN SIMULATIONS
FOR SIX PLAYERS ARRANGED AS 2-REGULAR GRAPH.

Player #:	1	2	3	4	5	6
Init. Ltrs:	b, a, t	m, e, n	l, u, t	s, o, p	h, u, g	r, i, e

TABLE II
PARAMETERS THAT ARE SYSTEMATICALLY VARIED IN THE SIMULATIONS
OF STUDY 1. THESE APTITUDE (b_i^{wf} , b_i^{req} , b_i^{rpl}) SETTINGS ARE THE
SAME FOR ALL AGENTS IN A SIMULATION.

Sim. No.	b_i^{wf}	b_i^{req}	b_i^{rpl}	Sim. No.	b_i^{wf}	b_i^{req}	b_i^{rpl}
1	P_1	Q_1	FB	5	P_5	Q_5	FB
2	P_2	Q_2	FB	6	P_5	Q_5	LTFB
3	P_3	Q_3	FB	7	P_5	Q_5	NR
4	P_4	Q_4	FB	—	—	—	—

Figure 9 (left) shows the average number of interactions (requests sent, replies received, requests received, replies sent) and the average number of words formed per player for the first five simulation numbers (sim. no.) of Table II. There is a drop-off in performance in going from $b_i^{wf} = P_1$ to P_5 , $b_i^{req} = Q_1$ to Q_5 , for fixed $b_i^{rpl} = \text{FB}$. We observe that decreasing the

letter request aptitude b_i^{req} and the word formation aptitude b_i^{wf} decreases the quality of letters requested and hence the number of words that can be formed.

To determine how b_i^{rpl} affects performance, we plot in Figure 9 (right) results from simulation numbers 5, 6, and 7 of Table II. Using $b_i^{wf} = P_5$ and $b_i^{req} = Q_5$ as a reference, there is a large decrease in numbers of reply interactions in going from $b_i^{rpl} = \text{LTFB}$ to $b_i^{rpl} = \text{NR}$, as expected, since NR means that agents do not reply to letter requests.

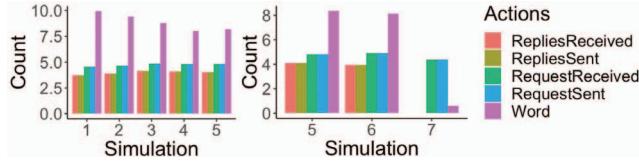


Fig. 9. (Left) Simulation results for Sim. nos. 1 through 5 of Table II. The average number of words formed per player drops in going from $b_i^{wf} = P_1$ to P_5 , $b_i^{req} = Q_1$ to Q_5 , for fixed $b_i^{rpl} = \text{FB}$. (Right) Simulation results for Sim. nos. 5, 6, and 7 of Table II. Using $b_i^{wf} = P_5$ and $b_i^{req} = Q_5$ as a baseline, these results show a precipitous drop-off in replies to letter requests, and to words formed, in going from $b_i^{rpl} = \text{LTFB}$ to $b_i^{rpl} = \text{NR}$. Results in counts for $b_i^{rpl} = \text{LTFB}$ are slightly less than those for $b_i^{rpl} = \text{FB}$.

Study 2: Effects of heterogeneity: network connectivity and quality of letter assignments to players. We use a game configuration $G(V, E)$ consisting of four players v_i ($1 \leq i \leq 4$) that form a star. The initial letter assignments are given in Figure 10. All players have the following conditions $b_i^{wf} = P_1$, $b_i^{req} = Q_1$, and $b_i^{rpl} = \text{FB}$. Players are assigned heterogeneous numbers and qualities of letters; see the figure caption. The numbers of requests received and replies sent are greatest for player v_1 owing to its centrality; this affects the number of words player v_1 forms, which is less than those for v_2 and v_3 . Players v_2 and v_3 have more requests received from v_1 (compared to v_4) because their letters (i.e., popular consonants) create larger sets of possible words to form. The number of words formed is least for player v_4 because of the poorer quality of assigned letters.

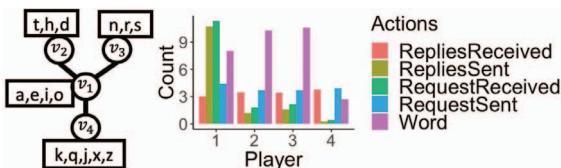


Fig. 10. Simulation for players v_i ($1 \leq i \leq 4$), arranged in a star. All players have the following conditions $b_i^{wf} = P_1$, $b_i^{req} = Q_1$, and $b_i^{rpl} = \text{FB}$. Player v_1 is at the center with three neighbors. v_1 is assigned the four most popular vowels in the alphabet; v_2, v_3 are assigned the six most popular consonants, and v_4 is assigned the five least popular consonants. See text for discussion of results.

VI. SUMMARY AND FUTURE WORK

We have developed mechanistic and data-driven models for representing the decision-making and actions of humans in online networked GrAGs. Our contributions are in Section I-D.

We would like to conduct more experiments with more network structures. This would also (ideally) produce sufficient data to more finely partition aptitudes—player behavior—into ten 10% bins (currently, we have five 20% bins). These experiments would be used to further evaluate the models and improve them.

ACKNOWLEDGMENT

We thank the anonymous reviewers for their helpful feedback. We thank our colleagues at NSSAC and computer system administrators: Dominik Borkowski, Jason Decker, Miles Gentry, Jeremy Johnson, William Marmagis, Douglas McMaster, and Kevin Shinpaugh. This work has been partially supported by NSF CRISP 2.0 (Grant 1832587), DARPA Cooperative Agreement D17AC00003 (NGS2), DTRA CNIMS (Contract HDTRA1-11-D-0016-0001), DTRA Comprehensive National Incident Management System Contract HDTRA1-17-D-0023. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

REFERENCES

- [1] W. L. Davis and D. E. Davis, “Internal-external control and attribution of responsibility for success and failure,” *J. of Personality*, 1972.
- [2] G. Charness, P. Kuhn, and M. C. Villeval, “Competition and the ratchet effect,” *Journal of Labor Economics*, vol. 29, pp. 513–547, 2011.
- [3] G. Charness, R. Cobo-Reyes *et al.*, “Identities, selection, and contributions in a public-goods game,” *Games and Economic Behavior*, 2014.
- [4] F. Polletta and J. M. Jasper, “Collective identity and social movements,” *Annual Review of Sociology*, vol. 27, pp. 283–305, 2001.
- [5] M. Goldman, J. W. Stockbauer *et al.*, “Intergroup and intragroup competition and cooperation,” *J. of Exper. Soc. Psych.*, 1977.
- [6] C. B. Cadsby *et al.*, “Sorting and incentive effects of pay for performance: An experimental investigation,” *Acad. of Mgmt. Jour.*, 2007.
- [7] Y. Ren *et al.*, “Generative modeling of human behavior and social interactions using abductive analysis,” in *ASONAM*, 2018, pp. 413–420.
- [8] V. Cedeno-Mieles *et al.*, “Modeling anagram games,” Tech. Rep., 2019, <https://github.com/vcedeno/asonam19/blob/master/tr.pdf>.
- [9] V. I. Levenshtein, “Binary codes capable of correcting deletions, insertions and reversals,” *Soviet Physics Doklady*, 1966.
- [10] J. M. Hofman, A. Sharma, and D. J. Watts, “Prediction and explanation in social systems,” *Science*, vol. 355, pp. 486–488, 2017.
- [11] D. J. Watts, “Should social science be more solution-oriented?” *Nat. Hum. Behav.*, pp. 1–5, 2017.
- [12] A. Bokulich, “How scientific models can explain,” *Synthese*, vol. 180, pp. 33–45, 2011.
- [13] J. Jebeile and A. G. Kennedy, “Explaining with models: The role of idealization,” *Int. Studies in the Philos. of Science*, pp. 383–392, 2015.
- [14] J. Sweller, “Cognitive load during problem solving: Effects on learning,” *Cognitive Science*, vol. 12, pp. 257–285, 1988.
- [15] G. S. Becker, *The economic approach to human behavior*. University of Chicago Press Chicago, 1976.
- [16] W. Mason and D. J. Watts, “Collaborative learning in networks,” *PNAS*, vol. 109, pp. 764–769, 2012.
- [17] N. P. Nguyen, G. Yan *et al.*, “Containment of misinformation spread in online social networks,” in *Web Science Conference*, 2012, pp. 213–222.
- [18] D. O’Callaghan *et al.*, “Uncovering the wider structure of extreme right communities spanning popular online networks,” in *Web Sci*, 2013.
- [19] Z. Zhao, M. Madaio *et al.*, “Socially-conditioned task reasoning for a virtual tutoring agent,” in *AAMAS*, 2018, pp. 2265–2267.
- [20] J. O’Doherty and P. Bossaerts, “Toward a mechanistic understanding of human decision making: Contributions of functional neuroimaging,” *Current Directions in Psychological Science*, vol. 17, pp. 119–123, 2008.
- [21] S. Spaulding *et al.*, “A social robot system for modeling children’s word pronunciation: Socially interactive agents track,” in *AAMAS*, 2018.
- [22] “Word frequency data: Corpus of contemporary american english,” <https://www.wordfrequency.info/free.asp>, accessed: 2018-11-16.



Graphical dynamical systems and their applications to bio-social systems

Abhijin Adiga¹ · Chris J. Kuhlman¹ · Madhav V. Marathe^{1,2} · Henning S. Mortveit¹ · S. S. Ravi^{1,3} · Anil Vullikanti^{1,2}

Published online: 15 December 2018
© Indian Institute of Technology Madras 2018

Abstract In this review paper, we discuss graphical dynamical systems (GDSs) and their applications to biological and social systems (bio-social systems). Traditionally, differential equation-based models have been central in modeling bio-social systems. GDSs provide an alternate modeling framework. This framework explicitly represents individual components of the system and captures the interactions among them via a network. The purpose of this review is to enable modelers to obtain an understanding of this basic mathematical and computational framework so that it can be used to study specific bio-social applications. The work covers the range from computational theory to simulation-based analysis. We also provide some directions for future work.

Keywords Graphical dynamical systems · Mathematical modeling · Simulation science · Complexity theory · Biological and social systems

1 Introduction

This review article focuses on the foundations of large interacting biological and social (bio-social) systems. Such systems are comprised of a large number of interacting entities/agents. The global properties of such systems are among the outcomes of the interactions of individual agents/entities with a relatively small number of other entities. In general, the underlying interaction structure (often called a network) is heterogeneous and time varying. Examples of such systems include: (1) the mammalian immune system, (2) epidemiology of infectious diseases, (3) gene interaction networks and (4) spatial ecologies. These systems are highly nonlinear and complex; the heterogeneity in such systems exists among individual agents as well as the interactions among them. This heterogeneity is a hallmark of such systems and is often responsible for the rich dynamics they exhibit. Furthermore, the global dynamics and the interaction structure coevolve. This makes studying such systems challenging. Simple aggregate models, although good starting points, are not adequate to capture the complexity of the systems and lead to results that ignore the effects of heterogeneities. What is needed is a mathematical formalism that can capture the interaction structures and diversity of local agents and their interactions.

In this article, we describe a formal theory based on *graphical dynamical systems* to study such massive bio-social systems. Abstractly, a graphical dynamical system

✉ Madhav V. Marathe
mvm7hz@virginia.edu

Abhijin Adiga
abhijin@virginia.edu

Chris J. Kuhlman
cjk8gx@virginia.edu

Henning S. Mortveit
henning.mortveit@vt.edu

S. S. Ravi
ssravi@virginia.edu

Anil Vullikanti
vsakumar@virginia.edu

¹ Biocomplexity Institute and Initiative, University of Virginia, 1001 Research Park Boulevard, Charlottesville, VA 22911, USA

² Present Address: Computer Science Department, University of Virginia, Charlottesville, VA, USA

³ University at Albany – State University of New York, Albany, NY, USA

(GDS) captures the following situation: There is a set of entities (e.g., genes in a gene regulatory network) represented by a set $\{1, 2, \dots, n\}$ of nodes, where each node i has a state s_i from a set K such as $\{0, 1\}$. Here $s_i = 1$ could encode that gene i is expressed, while $s_i = 0$ could encode the non-expressed state. A set of *local transition functions* $\mathcal{F} = \{f_1, \dots, f_n\}$ governs the local dynamics in the sense that f_i is used to determine how the state of node i evolves from time t to $t + 1$ for some suitable time scale. To specify how the system evolves globally, an update scheme \mathcal{U} is used to determine how the functions \mathcal{F} assemble to a map $F : K^n \rightarrow K^n$ of the form

$$F = (F_1, F_2, \dots, F_n). \quad (1)$$

For example, if the update scheme is to “apply the functions f_i in parallel,” we have $F_i = f_i$ ($1 \leq i \leq n$) in Eq. (1). We note that the set \mathcal{F} has an associated graph G called the *dependency graph* or *wiring diagram* that captures the dependence among variables: This graph has nodes $\{1, 2, \dots, n\}$ (e.g., the genes), and there is a directed edge $\{i, j\}$ whenever a function f_i depends on the state of node j . Generally, however, the graph G is given by the system or application (e.g., the graph of the gene regulatory network).

Abstracting the above, a graphical dynamical system (GDS) is a triple

$$(G, \mathcal{F}, \mathcal{U}) \quad (2)$$

with an associated GDS map $F : K^n \rightarrow K^n$ as in Eq. (1) assembled from $\{f_1, \dots, f_n\}$ via the update scheme \mathcal{U} . A large part of GDS theory seeks to infer properties about dynamics from its defining constituents G , \mathcal{F} and \mathcal{U} , an approach often referred to as *structure-to-function* analysis.

1.1 Agent-based models and GDS

GDSs are closely related to agent-based models (ABMs); in fact, they are a natural language for formally specifying ABMs. In other words, given the description of an ABM under the GDS framework, formal statements about the dynamics of the ABM can be readily translated into formal properties of the corresponding GDS. Also, formal properties determined from analysis of a GDS can guide or assist in verification of an ABM. As discussed in recent papers [1–4], ABMs are thought of as a modeling framework for generative sciences. We will discuss this aspect further in Sect. 5. For many classes of GDSs, computational theory can be used to demonstrate that certain analysis questions are computationally intractable; that is, unless some widely believed hypotheses in computational complexity turn out to be false, exhaustive enumeration is provably the most efficient way to solve those analysis problems. To cope with this intractability, systems are

usually analyzed through computation using a *simulation model*, that is, by judiciously running an implementation of the mathematical model describing the system on a computer (see Sect. 5). In fact, the interplay between analysis and carefully executed simulations can be a powerful driver of the theory.

In Sect. 4, we present some theory that assesses the *computational cost* associated with GDS analysis questions over its phase space, such as determining the number of *fixed points*.¹

1.2 Organization

We review the motivation behind the construction of the GDS framework in Sect. 2 and also discuss related work. Terminology and definitions are given in Sect. 3 before presenting computational theory and applications in Sect. 4. Simulation approaches are presented in Sect. 5. A brief discussion on a game theoretic view of GDSs appears in Sect. 6. We close with some directions for future work in Sect. 7.

2 GDS background and related work

2.1 Background

The framework of graphical dynamical systems was introduced by the authors in the context of *massively interacting systems*. Examples of systems motivating the GDS model include the transportation analysis system TRANSIMS [5], networked epidemiology systems Epi-Sims, EpiSimdemics, Epifast and Indemics [6–9], socio-technical systems where humans interact with critical infrastructure in crisis scenarios [10, 11] and the mammalian immune system [12, 13]. See [2–4, 14] for additional examples in the context of energy systems, computational social sciences and ecological systems.

A mathematical and computational theory based on GDS aims to (1) accurately model large interacting bio-social systems such as the above, (2) express analysis questions involving policy formulation, interventions and control in a precise formal manner, (3) support analytic reasoning for model validation, verification, sensitivity analysis, uncertainty quantification, optimization, etc. and (4) facilitate mapping of these models efficiently onto high-performance computing architectures [9, 15]. GDSs provide a natural framework to abstract the essence of analysis problems and therefore develop theory across application domains.

¹ Concepts such as phase space and fixed points are defined in Sect. 3.

2.2 Related work

Several subclasses of graphical dynamical systems have been analyzed in the literature. *Cellular automata* (CA) were introduced by von Neumann in [16] as a model for computation and were later studied by many others [17–20]. Generally, a CA is defined over a regular graph such as a k -dimensional lattice.

Boolean networks were introduced by Kauffman as models for gene regulatory networks; see [21, 22]. In the original setting, this class used $\{0, 1\}$ as the state space, typically involved regular graphs, and employed a synchronous update scheme (see Sect. 3). Some of these conditions have been relaxed in later work. *Automata networks* were introduced by Goles and, in many ways, overlap with synchronous and sequential GDSs [23–26]. *Probabilistic (or random) Boolean networks* (PBNs) are constructed as Boolean networks, but the vertex functions are chosen in a stochastic manner; in this case, one may view the evolution as that of an ensemble of Boolean networks, one for each function configuration. PBNs have been studied as a modeling framework for gene regulatory networks [27–30]. Synchronous and sequential graphical dynamical systems are *finite dynamical systems* and may therefore be regarded as *polynomial dynamical systems* [31, 32]. This approach is a starting point for the use of approaches based on computational algebra to analyze the dynamics as well as to reconstruct the system from appropriate time series data.

In our work, the focus is on the discrete time/discrete state setting. We remark that there is active research on networked ordinary differential equation (ODE) models (see [33–35]); however, these have yet to scale to the levels of GDS-based simulation models. It may in fact be interesting to see which results under the GDS framework extend to the analysis of networked ODE models. Examples of Boolean network models as alternatives to ODE-based models of biological systems are given in [36, 37].

Computationally, GDSs are closely related to other well-studied models, including recurrent neural networks (including Hopfield networks), concurrent transition systems and graph automata; see [2–4, 38–40] for further discussion on this topic.

3 Graphical dynamical system terminology and definitions

3.1 Synchronous and sequential graphical dynamical systems

This section presents definitions and terminology used throughout the paper. Let $G(V, E)$ be a graph on n nodes. Here we assume that the graph is undirected; extending

definitions to the directed case is straightforward, but adds bookkeeping. Each node is assigned a state from a finite domain K , frequently taken to be the two-valued Boolean domain $\mathbb{B} = \{0, 1\}$. We denote the state assigned to node i by s_i , or by s_i^t , when we want to explicitly reference the time t . A *configuration* at time t is an n -tuple $s^t = (s_1^t, s_2^t, \dots, s_n^t) \in K^n$. To each node i , we assign a *G-local transition function* f_i . This function is local in the sense that it only depends on the state of node i and those of the neighbors of i in G . We write $n[i]$ for the sorted sequence of these nodes and denote by $s[i]$ the corresponding sub-configuration of s . We also introduce the functions $F_i : K^n \rightarrow K^n$ defined by

$$F_i(s_1, \dots, s_n) = (s_1, \dots, s_{i-1}, f_i(s[i]), s_{i+1}, \dots, s_n). \quad (3)$$

Definition 1 (*Synchronous/sequential graphical dynamical systems*) [38, 41] Let $G(V, E)$ be a graph, $\mathcal{F} = \{f_1, f_2, \dots, f_n\}$ local transition functions and $\pi = (\pi_1, \dots, \pi_n)$ a permutation of V . The synchronous graph dynamical system (SyGDS) over graph G with local transition functions \mathcal{F} is the triple $\mathcal{S} = (G, \mathcal{F}, \mathcal{U} = \text{synch})$ with associated map $F : \mathbb{B}^n \rightarrow \mathbb{B}^n$ defined by

$$F(s_1, s_2, \dots, s_n) = (f_1(s[1]), f_2(s[2]), \dots, f_n(s[n])). \quad (4)$$

The triple $\mathcal{S} = (G, \mathcal{F}, \pi)$ is the *sequential graphical dynamical system* (SeGDS) over G with local transition functions \mathcal{F} , update sequence π and associated map $F_\pi : K^n \rightarrow K^n$ defined by

$$F_\pi = F_{\pi_n} \circ F_{\pi_{n-1}} \circ \dots \circ F_{\pi_1}. \quad (5)$$

A key difference between a SyGDS and an SeGDS is that for an SeGDS, all states $s_{\pi_1}^{t+1}$ through $s_{\pi_i}^{t+1}$ are used in the update of vertex states $s_{\pi_j > i}^t$. More general update schemes such as block synchronous updates [42] and stochastic rate-based schemes [43] have also been studied.

3.1.1 Classes of local transition functions

Here we present several classes of Boolean local transition functions referenced throughout the paper. One such class is symmetric Boolean functions [38]. A Boolean function $f : \mathbb{B}^n \rightarrow \mathbb{B}$ is symmetric if there exists a function $\eta : \{0, 1, \dots, n\} \rightarrow \mathbb{B}$ such that $f(s_1, \dots, s_n) = \eta(\sum_i s_i)$ for all $s \in \mathbb{B}^n$. Clearly, a symmetric Boolean function can be specified by giving the $n + 1$ values for its matching map η . *Threshold functions* (e.g., [40, 44]) are symmetric Boolean functions specified by an integer $h \geq 0$. The h -threshold function has value 1 precisely when h or more of its inputs are 1; see [45, 46] for applications.

Example 1 Consider the SyGDS whose underlying graph is shown in Fig. 1. The local transition functions at the

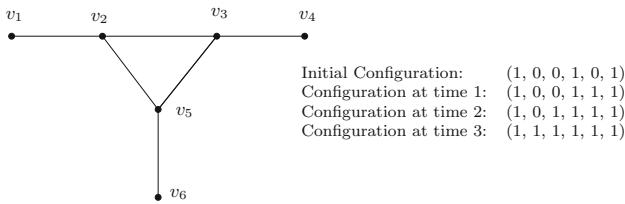


Fig. 1 SyGDS of Example 1. Each configuration has the form $(s_1, s_2, s_3, s_4, s_5, s_6)$, where s_i is the state of node v_i with $1 \leq i \leq 6$. The configuration at time 3 is a fixed point

nodes v_1, v_4, v_5, v_6 are one-threshold functions; the function at v_3 is the two-threshold function, while the function at v_2 is the three-threshold function. Assume that initially v_1, v_4 and v_6 are in state 1 and all other nodes are in state 0. During the first time step, the state of node v_5 changes to 1 since its neighbor v_6 is in state 1; the states of other nodes do not change. The configurations at subsequent time steps are shown in the figure. The system reaches the configuration $(1, 1, 1, 1, 1, 1)$ at time step 3. Subsequently, none of the nodes changes its state. Thus, the configuration $(1, 1, 1, 1, 1, 1)$ is a *fixed point* for this system. \square

A more general class of Boolean functions is that of r -symmetric functions where $r \geq 1$ is an integer. A Boolean function f is r -symmetric function if its input arguments can be partitioned into r subsets A_1 through A_r such that f factors as $f(s) = \eta'(\sum_{i \in A_1} s_i, \dots, \sum_{i \in A_r} s_i)$. This class generalizes threshold functions ($r = 1$) as well as bi-threshold functions ($r = 2$) [47–50], that is, functions for which the transitions from 0 to 1 and from 1 to 0 are governed by separate threshold functions with thresholds t_1 and t_0 , respectively.

Let $F : K^n \rightarrow K^n$ be a GDS map. If $F(\mathcal{C}) = \mathcal{C}'$, we say that \mathcal{C}' is the successor of \mathcal{C} and that \mathcal{C} is a predecessor of \mathcal{C}' under F . A configuration \mathcal{C} is called a fixed point if the successor of \mathcal{C} is \mathcal{C} itself. A configuration \mathcal{C} with no predecessors is called a Garden of Eden (GE) configuration. A configuration \mathcal{C} for which there exists an integer $r > 0$ such that $F^r(\mathcal{C}) = \mathcal{C}$ is a periodic point, and the sequence $(\mathcal{C}, F(\mathcal{C}), F^2(\mathcal{C}), \dots, F^{r-1}(\mathcal{C}))$ is a limit cycle. A configuration that is not a periodic point is called a transient point.

The phase space of a GDS map $F : K^n \rightarrow K^n$, denoted by \mathcal{P}_F , is the directed graph defined as follows: Its node set is K^n and its edge set is $\{(\mathcal{C}, F(\mathcal{C})) \mid \mathcal{C} \in K^n\}$. If $K = \mathbb{B}$, the number of nodes in the phase space is 2^n ; thus, the size of phase space is *exponential* in the size of the description of the graph G . Clearly, each node in the phase space has outdegree 1, fixed points are those nodes having a self-loop, and GE configurations are the nodes with indegree zero. A forward trajectory of F is a sequence of the form $(s, F(s), F^2(s), \dots, F^k(s))$ for some integer $k \geq 0$.

Example 2 A small SyGDS with 3 nodes and its phase space (which has 8 nodes) are shown in Fig. 2. In this example, $(0, 0, 1)$ is the successor of $(1, 1, 0)$. Both $(0, 0, 1)$ and $(1, 0, 0)$ are predecessors of $(0, 1, 1)$. Further, both $(0, 1, 1)$ and $(0, 1, 0)$ are fixed points and both $(1, 0, 1)$ and $(1, 0, 0)$ are Garden of Eden configurations. The sequence $(0, 0, 0) \rightarrow (1, 1, 0) \rightarrow (0, 0, 1) \rightarrow (0, 1, 1)$ is a forward trajectory.

3.2 Stochastic SyGDS

SyGDSs with stochastic local transition functions, denoted by SSyGDSs, are useful in many applications; see [51, 52]. Using states from a domain \mathbb{D} and denoting the neighbors of node i by $n[i] = (i_1, \dots, i_r)$, the local transition functions are specified in this case as a collection of probabilities

$$f_i(\theta', \theta_{i_1}^1, \dots, \theta_{i_r}^1, \theta) = \Pr\{s_i^t = \theta | s_i^{t-1} = \theta', s_{i_j}^{t-1} = \theta_{i_j}^1, 1 \leq j \leq r\}, \quad (6)$$

where for each combination of inputs, the sum of the probabilities assigned by f_i over all the values $\theta \in \mathbb{D}$ must be 1. All nodes update their states synchronously and independently of each other. Thus, for any pair of configurations $\mathcal{C}_1 = (b_1, b_2, \dots, b_n)$ and $\mathcal{C}_2 = (b'_1, b'_2, \dots, b'_n)$ we have

$$\Pr\{\mathcal{C}_1 \rightarrow \mathcal{C}_2\} = \prod_{i=1}^n f_i(b_i, b_{i_1}, \dots, b_{i_r}, b'_i). \quad (7)$$

The generalized phase space \mathcal{P}_S of an SSyGDS S is the directed graph where the nodes are the set of all possible configurations. There is a directed edge from \mathcal{C}_1 to \mathcal{C}_2 labeled by $p = \Pr\{\mathcal{C}_1 \rightarrow \mathcal{C}_2\}$ whenever $p > 0$. A SSyGDS is thus a particular type of Markov chain on the configuration space; the size of the Markov chain is *exponential* in the size of the underlying graph G .

Example 3 A SSyGDS whose graph G has the node set $V = \{a, b, c, d\}$ is shown in the left panel of Fig. 3. For this system, the transition function f_a is the three-input stochastic OR function: When all its inputs are 0, $\Pr\{f_a = 1\} = 0$; for all other input combinations, we specify that $\Pr\{f_a = 1\} = 1/2$. The function f_b is a deterministic three-input OR function: When all its inputs are 0, $\Pr\{f_b = 1\} = 0$; when at least one of its inputs is 1, $\Pr\{f_b = 1\} = 1$. The transition function f_c is a four-input stochastic AND function where $\Pr\{f_c = 1\} = 3/4$ for the input configuration $(1, 1, 1, 1)$ and $\Pr\{f_c = 1\} = 0$ in all other cases. Finally, f_d is a two-input deterministic OR function similar to f_b . From the specifications of the stochastic local transition functions, the reader can verify that for $\mathcal{C}_1 = (1, 1, 1, 1)$ and $\mathcal{C}_2 = (0, 1, 0, 0)$, \Pr

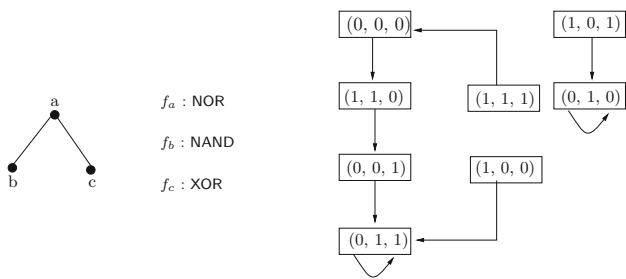


Fig. 2 A small SyGDS and its phase space (see Example 2). Note: The domain of state values for each node is $\{0, 1\}$. As indicated above, the Boolean functions at nodes a , b and c are NOR, NAND and XOR, respectively. (The configuration $(1, 0, 1)$ indicates that $s_a = 1$, $s_b = 0$ and $s_c = 1$.)

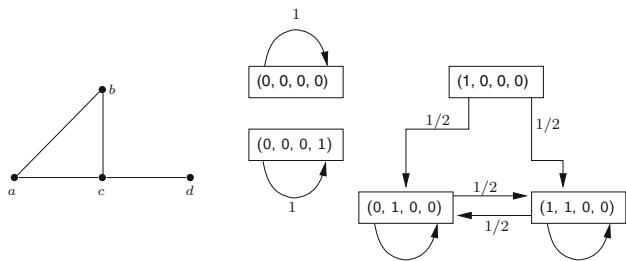


Fig. 3 A stochastic SyGDS (see Example 3) and a portion of its generalized phase space. The complete generalized phase space, which has 16 nodes and many edges, is not shown to avoid clutter

$\{\mathcal{C}_1 \rightarrow \mathcal{C}_2\} = 1/8$. A portion of the generalized phase space of the SSyGDS in the left panel of Fig. 3 is shown in the right panel of that figure.

3.2.1 Notational convention

For the rest of the paper, we use GDS to mean graphical dynamical systems without any specific constraints on update order, location functions or the type of network. We will use SeGDS and SyGDS to refer to a sequential graphical dynamical system and a synchronous graphical dynamical system, respectively. Finally, we use SSyGDS to refer to a stochastic synchronous graphical dynamical system. In some of our earlier papers, we have used SDS and SyDS to refer to sequential graphical dynamical systems and synchronous graphical dynamical systems, respectively. The current notation makes the graphical aspect more explicit.

3.2.2 Additional notes

Stochastic sequential graphical dynamical systems are not discussed in this paper. Due to space reasons, most of the results discussed here pertain to SyGDSs; in many cases, analogous results hold for SeGDSs as well. Furthermore,

we assume that the underlying graph G is given. Synthesizing the network that models a bio-social system is an important area of research. Finally, again for space reasons, our focus on formal GDS results in the next section is on *decision* problems. *Counting* problems (i.e., find the number of solutions), *ambiguous* problems (i.e., given a solution, decide whether there is a different solution) and *uniqueness* problems (i.e., decide whether the instance has exactly one solution) are not discussed here; nevertheless, they are interesting and have been studied in the literature (for example, see [53–55]).

4 Application of computational and complexity theory to GDSs

4.1 Overview

We now discuss some computational problems (called analysis problems) for dynamical systems that are useful in addressing questions that arise in the study of bio-social systems. We first discuss these problems for deterministic SyGDSs and then mention the extensions to stochastic SyGDSs. Our focus is on the computational complexity results for these problems and the significance of such results.

4.2 Analysis problems under the SyGDS model

Analysis problems arise in the context of determining the *dynamic behavior* of a system, given its static description. We consider systems whose static description is specified as a SyGDS; that is, it is assumed that the underlying graph and the local transition functions are given. Since the phase space of a SyGDS captures the dynamic behavior of the corresponding system, analysis questions correspond to structural properties of the directed graph representing the phase space of the given SyGDS. In this section, we define several such questions for SyGDSs and summarize known results about those questions.

Of the many applications that can be studied using the SyGDS model, we focus on diffusion processes, which arise in diverse kinds of phenomena, such as the spread of epidemics, fads and beliefs in social networks (see, e.g., [46, 56–59]). Here, the underlying graph of the SyGDS represents a social contact network, with each node representing a person and an edge between two nodes indicating that the corresponding people came into contact with each other; thus, each edge represents an opportunity for one of the end points to influence the other. Each node takes on state values from $\{0, 1\}$. We say that a node in state 1(0) is “infected” (“uninfected”). Here, infection is a general term denoting that the node is infected due to a

disease or has adopted a new belief, etc., depending on the phenomenon being modeled. The local transition function at a node v determines whether node v gets infected when some of its neighbors are infected. For example, suppose the local transition function at v is the k -threshold function for some integer $k \geq 1$. In such a case, node v gets infected at a time step t if k or more of its neighbors were infected prior to t . Starting from an initial configuration (where some nodes are in state 1), the system goes through a sequence of successive configurations. The analysis problems discussed below capture several aspects of this form of dynamical behavior.

4.2.1 Configuration reachability

This form of analysis problem concerns the *reachability* of certain configurations. Formally, given a SyGDS \mathcal{S} and two configurations \mathcal{C} and \mathcal{C}' , the goal of the reachability problem is to determine whether \mathcal{S} starting from \mathcal{C} can reach \mathcal{C}' . One can also formulate a timed version of the reachability problem where an additional time parameter t is specified as input to the problem, and the goal is to decide whether \mathcal{C}' can be reached in t or fewer time steps. In the context of diffusion processes, one application of the reachability problem is in deciding whether an undesirable situation (such as a large infected population) can occur within t time steps, given the current conditions.

Many papers in the literature have addressed the reachability problem for discrete dynamical systems. Barrett et al. [38, 60] showed that, in general, the problem is computationally intractable (technically, PSPACE-complete²) even when all nodes have the same symmetric local transition function. When each local transition function is a threshold function, it was shown in [38] that the reachability problem can be solved efficiently. However, when both threshold functions and negative threshold functions³ are permitted as local transition functions, the reachability problem becomes PSPACE-complete [40]. It is also known that the reachability problem for SyGDSs with bi-threshold functions can be solved efficiently when all the threshold values are 1 [48]. Kuhlman et al. [50] show that the reachability problem remains efficiently solvable for SyGDSs with bi-threshold local transition functions when the underlying graph is directed and acyclic. For some classes of local transition functions, upper and lower bounds on the number of time units to reach fixed points were established in [62].

² For definitions concerning complexity classes, we refer the reader to [61].

³ A negative threshold function is the negation of a threshold function. For example, a negative three-threshold function has the value 1 if and only if two or fewer of its inputs have the value 1.

4.2.2 Predecessor existence

The reachability problem mentioned above addresses how a contagion process progresses through a network over time. Another problem addresses the question of finding a configuration that can occur immediately before the current configuration. A formulation of this predecessor existence problem is the following: Given a SyGDS \mathcal{S} and a configuration \mathcal{C} , find a configuration \mathcal{C}' such that \mathcal{C}' is a (one-step) predecessor of \mathcal{C} (if one exists). A generalization of this one-step predecessor problem is the k -step predecessor problem, where the goal is to find a configuration \mathcal{C}' such that the SyGDS reaches \mathcal{C} from \mathcal{C}' in exactly k time steps, for some $k \geq 2$. In diffusion processes, solutions to k -predecessor problems for appropriate values of k help in understanding how the process may have spread through a population.

Barrett et al. [54] present a comprehensive study of the predecessor existence problem and its variants (e.g., determining whether a configuration has a unique predecessor, counting the number of predecessors) for discrete dynamical systems. They consider the problem for various classes of graphs and local transition functions. For example, they show that the predecessor existence problem is NP-complete even when the underlying graph of the SyGDS is a grid and each local transition function is symmetric. Further, they show that when the underlying graph is treewidth bounded⁴ and each local transition function is r -symmetric for some fixed integer r , the predecessor existence problem and its counting version can be solved efficiently. They also present an extension of this algorithm to the case where predecessor configurations must satisfy additional constraints (e.g., a configuration in which at least α nodes are in state 1, for a given integer α). Kuhlman et al. [50] show that the predecessor existence problem remains NP-complete for SyGDSs with bi-threshold functions even when the maximum node degree of the underlying graph is 3; they also show that the problem can be solved efficiently when each node has a degree of at most 2.

4.2.3 Fixed-point existence

Recall that a fixed point of a SyGDS is a configuration which is its own successor. When a SyGDS is used to model a diffusion process, fixed points represent situations in which no additional infections can occur. This motivates the problem of determining whether a given SyGDS has a fixed point. Formally, given a SyGDS \mathcal{S} , the goal of the fixed-point existence problem is to determine whether \mathcal{S} has a fixed point. Note that given a SyGDS \mathcal{S} and a

⁴ For definitions related to treewidth, we refer the reader to [63].

configuration \mathcal{C} , determining if \mathcal{C} is a fixed point is trivial. The fixed-point existence problem has been studied by several groups of researchers [64–66]. Barrett et al. [64] showed that the problem is NP-complete in general, but efficiently solvable for several classes of local transition functions (e.g., monotone functions⁵). Other researchers (e.g., [65, 66]) have established complexity results for the problem of counting the number of fixed points for various classes of local functions.

4.2.4 Analysis problems for more general phase space properties

As mentioned earlier, each fixed point of a SyGDS \mathcal{S} is a self-loop (i.e., a directed cycle with one edge) in the phase space of \mathcal{S} . In some biological applications [67], it is useful to determine whether the phase space contains longer cycles. An approach for studying the complexity aspects of this and more general properties of SyGDSs (e.g., the phase space has two node-disjoint simple paths each with seven nodes) has been proposed in [53]. This approach encodes phase space properties as appropriate graph predicates and develops a theoretic framework for analyzing these predicates. A number of known complexity results and efficient algorithms for testing phase space properties of SyGDSs follow as special cases from this general framework.

4.2.5 Dynamical systems with nested canalyzing local functions

Two special classes of Boolean functions, called canalyzing and nested canalyzing functions, were proposed and studied in [21, 68, 69] to model stability in gene regulatory networks. The complexity of analysis problems for SyGDSs in which each local function is a nested canalyzing function (NCF) has been studied in [39]. This reference shows that the reachability problem for SyGDSs with NCF local functions remains PSPACE-complete, while predecessor existence and fixed-point existence remain NP-complete. Thus, restricting the local functions to NCFs does not make these analysis problems computationally easier for SyGDSs.

For the reader's convenience, definitions of the analysis problems discussed above and some references which mention applications of the problems to bio-social systems are given in Table 1. For simplicity, we define only the basic versions of the problems in the table. The references cited in this section discuss several variants of the basic analysis problems.

⁵ A Boolean function is monotone if it does not change from 1 to 0 when one more of the inputs is changed from 0 to 1. For example, every k -threshold function (for any integer $k \geq 0$) is monotone.

4.3 Analysis problems under the SSyGDS model

Algorithmic aspects of analysis problems under the SSyGDS model (i.e., the SyGDS model with stochastic local transition functions) have not received as much attention in the literature as the deterministic SyGDS model. We will now summarize the known results for analysis problems under the SSyGDS model.

4.3.1 Reachability and predecessor existence

The work reported in [51, 52] is motivated by applications in epidemiology, and the focus is on problems related to reachability and predecessor existence for SSyGDSs. Since transitions in SSyGDS are stochastic, the reachability (decision) problem is reformulated as follows: Given a SSyGDSS, two configurations \mathcal{C} and \mathcal{C}' , and a probability value p , determine whether \mathcal{S} starting from \mathcal{C} reaches \mathcal{C}' with a probability of at least p . Likewise, the definition of the predecessor existence problem is as follows: given a SSyGDS \mathcal{S} , a configuration \mathcal{C} and a probability value p , is there a configuration \mathcal{C}' such that the one-step transition probability from \mathcal{C}' to \mathcal{C} is at least p ?

When \mathcal{C} represents the initial onset of a disease and \mathcal{C}' represents an undesirable situation (e.g., with a large percentage of infected population), the goal of the reachability problem is to determine whether an undesirable state is likely to occur in the absence of any effort to contain the disease. As mentioned earlier, the predecessor existence and its k -step generalization are helpful in understanding the progression of an epidemic.

It is shown in [52] that the reachability problem for SSyGDSs is hard for the complexity class RSPACE(n), which is the class of problems that can be solved by a probabilistic Turing machine which uses space bounded by a linear function of the input size. As pointed out in [52], under standard assumptions in complexity theory, reachability problems for SSyGDSs are likely to be computationally more difficult than those for (deterministic) SyGDSs. It is also observed that the NP-hardness of the predecessor problem for SSyGDSs follows from the corresponding result for SyGDSs. However, the problem is shown to be efficiently solvable when the following conditions hold: (1) the underlying graph is treewidth bounded, (2) the local stochastic functions are r -symmetric for some fixed integer r and (3) the number of distinct probability values used in all the local transition functions is bounded. Barrett et al. [52] also address other problems related to reachability. They point out that the problem of determining the most likely successor of a given configuration of a SSyGDS can be solved efficiently, while computing the two-step transition probability from a configuration \mathcal{C}_1

Table 1 A list of analysis problems for GDS that arise in practice while studying bio-social systems

Problem	Definition	References
Reachability	Does a given SyGDS starting from \mathcal{C} reach \mathcal{C}' ?	[38, 39, 70, 71]
Predecessor existence	Does the configuration \mathcal{C} of a given SyGDS have a predecessor? How many predecessors does \mathcal{C} have? Does \mathcal{C} have two distinct predecessors?	[54, 72]
Fixed-point existence	Does a given SyGDS have a fixed point? How many fixed points does a given SyGDS have?	[64]
Other properties	Ex: Does the phase space of a given SyGDS have a cycle of length ≥ 3 ?	[53, 67]

to a configuration \mathcal{C}_2 is computationally intractable (technically, #P-complete).

5 Scalable simulations of bio-social systems based on the GDS formalism

In this section, we first relate the GDS framework to simulation. Then we provide a motivation for simulation in terms of the analysis problems of Sect. 4. Thereafter, we present three illustrative examples of simulation that use the GDS framework. We provide additional examples of simulation in the bio-social realm and note parallel developments in discrete dynamical system theory and simulation. We conclude with a brief mention of other modeling approaches.

5.1 Aspects of GDS-based simulation

Simulation is defined here as the process of executing software to mimic or replicate the behavior of a system. Our focus is on systems of interacting agents or entities, where entities can be any combination of humans, animals, insects, plants and inanimate objects: any object that can act or be acted on. We represent this system of agents as a (time-varying) graph, where nodes represent agents and edges represent pairwise interactions. We note, by comparison, that there can be many entities (e.g., humans) in a system, but they do not have to interact (see, e.g., [73]).

Here, we confine ourselves to the sub-topic of simulation that uses implementations of algorithms that are based (at least in part) on the theoretic foundations presented in Sects. 1 and 3. The central steps in building a simulation system, then, are as follows: (1) formally specify the system components, namely the agents and their interactions; (2) using the GDS framework, specify mathematically the local functions (i.e., behaviors) of the agents in the system; (3) formalize the dynamics (e.g., behaviors and interactions) using algorithms; and (4) build and verify the software system that implements the algorithms used to simulate the system under study. Many agent-based models (ABMs) and agent-based modeling and simulation

(ABMS) systems can be represented within the GDS framework; see, e.g., [74, 75]. Several reviews of ABMs and software modeling systems have appeared in the literature [74–77].

5.2 Simulations as approximate representations of dynamics of systems

There is a close connection between each of the GDS formalisms of Sect. 3 and the analysis problems of Sect. 4, and simulation. First, the ideal outcome from evaluating a GDS is that the complete dynamical behavior of the system is computed. From Sect. 3, this means computing the phase space of the dynamical system. But recall that the phase space is exponential in the number of nodes of the graph (i.e., the population). This means that for dynamical systems of hundreds, thousands or billions of nodes (i.e., interacting entities), computing the phase space is not possible given the current state of computing technology. Second, the great majority of the analysis problems in Sect. 4 are computationally intractable. Thus, solving these problems for large GDSs is also not currently possible. This is why there are many works on specializing computationally intractable problems along one or more dimensions to make them computationally efficient (e.g., [38, 48, 53, 54, 62, 64, 65]) and why there are many works that devise heuristics to solve intractable dynamical system problems (e.g., [46, 78–81]).

Both of these issues are addressed, at least in part, by computing forward trajectories, and this is where simulations play a vital role. Rather than generating the entire phase space (i.e., all system transitions), a single simulation starts with a prescribed state of the system (that is part of the “initial conditions”) and runs the dynamics of the system forward for a prescribed number of steps or for some specified duration. This sequence of computed successive system configurations—a forward trajectory—is part of the phase space. By running multiple simulations with different initial conditions (or the same initial conditions to evaluate the effects of stochasticity), one seeks to generate a portion of the phase space that is relevant to the particular problem of interest.

5.3 Three illustrative examples

We provide three examples of simulation in biological and social systems. The first example explores dynamics on a gene regulatory network in a plant [26, 82]. While the simulation and analysis are very close to the mathematical setup in Sect. 3, there is a direct connection to experimental observations: The model provides the first explanation of the importance of a particular hormone for plant development. The second example is a simulation system for inflammatory and regulatory immune pathways governed by cell interactions in the human gut [13, 83, 84]—this utilizes a much more complex set of local functions than does the first example. Our final example is a stochastic model of epidemic spread which is used extensively in public health analysis.

These examples have been chosen to demonstrate the range in utility of a GDS-based simulation approach for systems of various sizes. In the first example, there are 12 genes (agents) in the model. The second example has 10^7 cells (agents), a six orders of magnitude increase in problem size. The third example can have 10^9 agents, e.g., when used to model epidemic spread in the continental US.

5.3.1 Example 1: gene regulatory networks

Boolean networks were first used to study regulatory networks in [21, 85]. Since that time, a host of works on regulatory networks has been published, evolving from networks represented as regular graphs to current considerations of wirings in actual networks, to various forms of local transition functions, to different update schemes [22, 28, 82, 86, 87]. Here we highlight the work of [82].

In [82], the regulatory system is modeled as a GDS, described as follows. The regulatory network $G(V, E)$ represents a set V of genes and a set E of interactions. See Fig. 4. A gene is either activated (state 1) or inactivated (state 0), so that the state set—the possible states of a node/gene—is $K = \{0, 1\}$. Each of the G -local transition functions f_i is a generalized threshold function given by

$$f_i(x) = H\left(\left(\sum_{j \in n[i]} (w_{j,i}x_j)\right) - k_i\right), \quad (8)$$

where H is the Heaviside function $H : \mathbb{R} \rightarrow \mathbb{R}$ given by $H(x) = 1$ if $x > 0$ and $H(x) = 0$ if $x \leq 0$. In Eq. (8), k_i is the threshold of node i , and $w_{j,i}$ is the weight assigned to the directed edge from j to i , the latter capturing the influence of node j on i . The edge weights and vertex thresholds may be positive or negative. A negative edge weight $w_{j,i}$ means that j tries to inhibit the transition of i .

There are multiple dimensions to this work. First, observing that the states of several nodes become constant

(i.e., 0) after a few time steps, Demongeot et al. [82] show that the original GDS can be transformed into a system \mathcal{F}' for which the matrix of weights W is symmetric. For the symmetric case, they apply theorems guaranteeing that all limit cycles must be either fixed points or two-cycle. Second, they highlight six fixed points and seven two-cycle as attractors that are significant in terms of the genes of a particular plant species. Third, robustness is characterized numerically by attractors (i.e., limit cycles) and the attractor basins. (Note that the basin of an attractor is the set of system states that are forward asymptotic to the attractor.) Viewing phase space as a directed graph, the attractor basin is the set of states in the weakly connected component of the attractor; the attractor states are those in the strongly connected component. Clearly, phase space is partitioned into attractor basins. We denote the set of attractors by $\mathcal{A} = \{A_1, \dots, A_m\}$ and the corresponding set of basins by $\mathcal{B} = \{B_1, \dots, B_m\}$.

Robustness is assessed in terms of perturbations of the vertex states in regulatory networks (that is, changing a gene's state from 0 to 1 or from 1 to 0). At issue is whether a change of a single vertex's state will cause the newly generated state to move to a different attractor basin. That is, for a configuration x in basin B , will the perturbation x' by a single vertex state of x belong to $B' \neq B$? Depending on the system, the states of two or more vertices may have to be perturbed in order for a system state to map to a different attractor basin. Then, a question is to determine, for a particular system state and target attractor basin, the minimum number of vertices whose state must be changed to transform the system state into a new system state in the target attractor basin. This number is called the Hamming distance. An overarching theme of these questions is the number of perturbations required to alter the long-term dynamics (i.e., limit cycles) of a system. The robustness of a system increases as the number of perturbations required

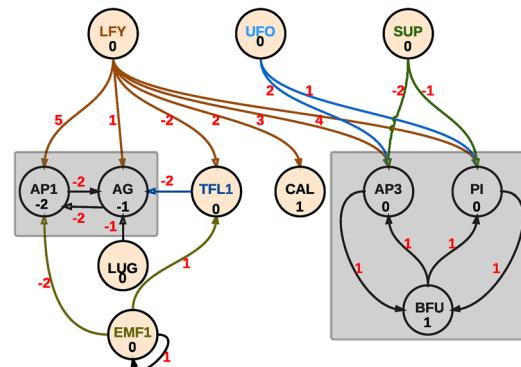


Fig. 4 Gene regulatory network of Example 1, reproduced from Figure 3 of [82]. The two strongly connected components are highlighted in gray. Genes are shown as circles. A gene's threshold is depicted within its circle. Edge weights for the GDS are shown in red (color figure online)

to change the attractor basin increases. These types of issues are explored in [82], where not only the number of gene state changes is studied, but also the particular genes whose states are changed are studied in the context of a particular plant species. Other studies of robustness include [22, 88, 89].

In [82], the authors also consider stability properties with respect to the update mode. One may choose to apply the vertex functions in Eq. (8) synchronously, sequentially, block sequentially or possibly in some random order. Understanding attractor structure as a function of update mode gives insight into model stability and quite possibly the underlying gene regulatory network.

Here we give an illustration of how one may assess this type of stability under all possible sequential update orders. In this case, an order $\pi = (\pi_1, \pi_2, \dots, \pi_n)$ of the nodes/genes specifies the order in which the functions f_i are applied. Using different orders, one would possibly expect different attractor structures. For the graph G in Fig. 4 with $n = 12$ vertices, there are $12! = 479,001,600$ different update orders, and potentially this many different attractor structures. Using the theory of κ -equivalence from [90, 91], one may show that this number is in fact a lot smaller. For this, we will say that two systems F_π and F'_π are cycle equivalent if their multi-sets of limit cycle sizes coincide. (For a GDS map F with one three-cycle, two two-cycles and three fixed points, its multi-set of limit cycle sizes is $\{1(3), 2(2), 3(1)\}$.) The measure $\kappa(G)$ can be evaluated as $T_G(1, 0)$ where T_G is the Tutte polynomial of G [92]. Applying this to the network in Fig. 4 we obtain $\kappa(G) = 210$. In other words, without looking at the form of the functions \mathcal{F} , we can immediately conclude that one may at most observe 210 distinct attractor structures when varying the update order. However, and as shown in [82], the long-term behavior of the system is effectively modeled by the subgraph contained in the gray boxes of Fig. 4, and the fact that the state of the gene labeled as UFO is fixed at its initial value. Thus, for each choice of $x_{\text{UFO}} \in \{0, 1\}$, the limit cycle structure of the gene regulatory network is governed by the disjoint sub-networks indicated in the gray boxes. For this graph, one can show that the κ -value is 1. In other words, for each choice of x_{UFO} at $t = 0$, there is only one possible limit cycle structure, and *one may observe at most two distinct limit cycle structures under sequential update modes*. Moreover, from the analysis in [82], we know that fixed points and two-cycle are the only possibilities for limit cycles in these two cases.

5.3.2 Example 2: mammalian immune networks

Graphical dynamics systems and associated agent-based models have recently been used to model numerous

processes that underlie host-pathogen interactions and the immune system. A pathogen here can be thought of as any infectious agent that can lead to the illness of a host (pathogenesis). The mammalian intestinal tract constitutes the largest component of the immune system. The mammalian gut comprises non-hemopoietic cells (epithelia, Paneth cells, goblet cells), hemopoietic cells (macrophages, dendritic cells, T cells) and a large community of microbes (now called the microbiota). Enteric pathogenesis in the gastrointestinal (GI) tract is often caused by ingestion of microbes in food and water. Upon microbe entry, immune cells in the GI tract mount an immuno-inflammatory response that eliminates the microbe, but may also cause tissue damage. This collateral damage is often the basis of disease pathogenesis.

As the GI tract is constantly exposed to foreign antigens, most of which are innocuous, this inherent inflammatory response must be regulated so that the system does not remain in a constant state of tissue-damaging hyper-inflammation. Immune regulation is carried out by the *regulatory* (or anti-inflammatory) immune response triggered by factors such as host tissue damage or commensal gut microflora. The prevailing understanding of the gut mucosa is one in which immune cells are in a dynamic balance (homeostasis) between regulatory and inflammatory responses, with regulatory phenotypes generally predominating [93, 94].

An understanding of the role of individual components of these immune pathways is needed to elucidate the mechanisms that lead to microbial persistence and severity of symptoms. This in turn can lead to improved treatment and prevention against pathogenic strains of *Escherichia coli*, *Clostridium difficile* and *Helicobacter pylori* for example.

Over the last two decades, several ABMs have been developed to study various immune processes. This includes: Agent-based Artificial IS (AbAIS) [95], CAFISS [96], ImmSim [97], ImmSim3 [98], C-ImmSim [99], Par-Imm [99], ImmunoGrid [100], Rhapsody [101, 102], SIS [103], SIMMUNE [104], NFSim [105], BIS [106] and the work reported in [107]. The GDS framework and its extensions provide a natural underlying mathematical abstraction for these ABMs. Comprehensive surveys of ABMs for immune processes can be found in [108–110]. See [12, 83] for a detailed account of these ABMs and their relative strengths.

As part of ongoing work, we have developed *ENteric Immunity Simulator* (ENISI), a scalable high-performance agent-based modeling environment to study the inflammatory and regulatory immune pathways initiated by microbe-immune cell interactions in the gut [12, 13]. ENISI is an interaction-based model where individual cells are modeled, along with their movement through different

tissues, and the probabilistic outcomes of cell–cell interaction.

The ENISI modeling environment is unique in its scope and approach. The modeling environment is designed to specifically represent regulatory mechanisms of both adaptive and innate immunity, multi-location migration of cells, and cross talk between antigen presenting cells and T cells. This is done by explicitly representing each participating cell of the immune pathway. This facilitates mapping of model parameter specifications and predictions to laboratory techniques which manipulate specific cell populations. In other words, the model is mechanistic and represents cells individually.

ENISI has the ability to simulate 10^7 or more individual cells. With ENISI, mucosal immunologists can test and generate hypotheses for enteric disease pathology and propose interventions through experimental infection of an *in silico* gut. This information can then be used to better understand immunological mechanisms and to generate novel treatment strategies that can be tested in the laboratory using mouse and pig models of infection as well as human clinical experimentation.

The ENISI modeling environment has already been illustrated by developing (1) a *in silico* model and dynamic simulation of *H. pylori* and (2) a simulation of dysentery resulting from *B. hyodysenteriae* infection so as to identify aspects of the host immune pathways that lead to continued inflammation-induced tissue damage even after pathogen elimination.

GDSs (and their extension to time-varying systems) provide a natural mathematical abstraction of the immune system as well as the computer simulation. The cells and the microbial community can be thought of as nodes. The network is dynamic in such a setting; vertices as well as edges come and go during the course of the system evolution. The cell/bacterial movement, their birth/death and their phenotypic change form the basis of the time-varying system. Cells and bacteria interact with each other via signaling pathways; see [12, 83] for further details.

A state chart-like formalism was used to model each cell type, including each cell's individualized state transitions, the other cell types that are required in interactions that cause the state transitions, and transition probabilities. The cell types are: epithelial cell, inflammatory bacteria, tolerogenic bacteria, commensal bacteria, sampling dendritic cell, dendritic cell, macrophage, conventional CD4+ T cell and natural T-regulatory cell (nTreg). As an example, the state chart-like formal specification for tolerogenic bacteria cell type is provided in Fig. 5. In this figure, ovals represent states of the cell type, the solid arrow represents a transition that occurs in the next time step after transition to the state represented by the arrow's tail, and dashed arrows represent single-cell interaction-initiated state transitions

where the edge labels characterize the initiator of the transition.

Of particular note here is that the cells and their interactions can be represented by a time-varying graph G . Thus, descriptions such as the one in Fig. 5 can be converted to local transition functions that capture the interactions of different cells i . Consequently, an expression can be written for the local function of each node of the form $s_i^{(t+1)} = f_i(s_{j_1}^t, s_{j_2}^t, \dots, s_{j_k}^t)$, where $s_i^{(t+1)}$ is the next state of node (cell, agent) i , and $j_\ell \in J$ are the k elements of the closed neighborhood J of node i . (This form of local function was described in Sect. 3.) In ENISI, the synchronous update method is used in simulations.

5.3.3 Example 3: networked epidemiology

Controlling epidemics caused by infectious diseases remains an important societal challenge, despite significant medical advances. Understanding and forecasting the dynamics of the disease, inferring the source and developing interventions to control the spread are some of the fundamental questions studied by epidemiologists and public health policy planners. Since there are limited data available for any outbreak, especially in the early stages, mathematical and computational models play a key role in studying epidemic spread. These are typically considered at the level of an individual, with a simple model capturing the spread of the disease from one individual to those coming in contact with him/her. Many researchers have used ABMs to model the spread of epidemics; see, e.g., [7, 9, 111–114]. Details incorporated in different models vary widely. We refer the reader to [56, 115] for a discussion on different models of epidemic spread.

One of the most commonly used SSyGDS models is the SIR model of epidemic spread. In the basic SIR model, each node is in one of the following states: Susceptible (\mathbb{S}), Infected (\mathbb{I}) or Recovered/Removed (\mathbb{R}) states, i.e., $K = \{\mathbb{S}, \mathbb{I}, \mathbb{R}\}$. A node j in state \mathbb{I} spreads the infection to each neighbor i in state \mathbb{S} with probability p , independent

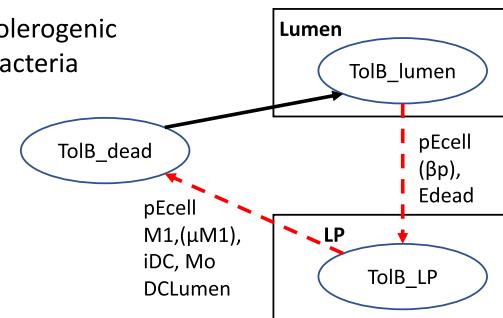


Fig. 5 One of nine state chart-like formal specifications used to represent cell types (cells are agents) in ABMs of immunological processes in the gut. Adapted from [12]

of the other neighbors of i . A node in state \mathbb{I} recovers after a prescribed duration t_{inf} in the infectious state, i.e., switches to the R state. Formally, for the \mathbb{S} to \mathbb{I} transition, the local function for node i is given by:

$$f_i(\mathbb{S}, \mathbb{I}) = 1 - (1 - p)^k,$$

where k is the number of infected neighbors of i , and $f_i(\mathbb{I}, \mathbb{R}) = 1$ after $t_{\text{inf}} = 1$ units of infectivity. Most often, the update method \mathcal{U} is synchronous. Many other variants of this basic model are used commonly, e.g., SI, SIS and SIRS, where the letters indicate the state transitions. For example, in the SI model, an infected node stays infected, whereas in the SIS model an infected node switches to state \mathbb{S} after a period of time.

If the network $G(V, E)$ does not change over time and there are no new externally induced infections, the GDSs corresponding to all these models have fixed points in which there are no infected nodes. Computing the values of parameters related to epidemic analysis can be mapped to computations that use properties of the phase space of the corresponding GDS (similar to the analysis questions discussed in Sect. 4.2). Examples include the following: (1) the average outbreak size is equal to the average number of nodes in state \mathbb{R} in the fixed points reachable from a given initial configuration and (2) the (average) duration of an epidemic is the (average) length of a transient to a fixed point with no infected nodes. An important topic in epidemiology is *forecasting*: Here, the objective is to determine the future dynamics, given the current configuration. An example of a forecasting problem is the following: Given the current configuration, will the system reach a fixed point with a large number of infected nodes? The dynamics of epidemic spread are very complex in heterogeneous real-world networks and cannot be computed analytically. This makes efficient simulations necessary. Designing highly scalable epidemic simulation tools which can handle SIR models at a national or global scale (with billions of agents) is a fundamental research problem [15, 116].

5.3.4 Additional examples of bio-social systems

Beyond the applications described above, a number of other biological and social systems can be fruitfully analyzed using computer simulations that implement ABMs based on GDSs. These include: depression [117], incarceration [118], segregation [119], invasive species [120, 121], drug trials and predictions of drug effects [76], reproductive growth of bacteria [122], forest management and fire protection [123], cell biology [124], computer viruses/malware [125], finance (works on housing, mortgages, banking systems and credit risk appear in [126]),

disasters [127, 128] and ecology [129, 130]. See [70, 71, 74, 131, 132] for additional applications.

Social science modeling continues to make ever greater use of ABM. Reviews of social modeling include [133–135]. In particular, Bianchi and Squazzoni [135] break down ABM work by different topics such as cooperation and social norms, reciprocity, reputation, punishment, conventions, trust, social influence, culture and social inequality.

These and other references illustrate a development in ABM that parallels the one in dynamical systems. That is, as explained in Sect. 2.2, dynamical system formulations have progressed from cellular automata, to Boolean networks, and to GDS, among others. The interaction structures have evolved from lattice (regular) patterns, to random patterns, to structured patterns of interactions based on data. The local functions have also grown in sophistication within each of these dynamical system constructs, over time. We observe the same types of evolution in ABM (in social systems). Early ABM works of segregation [119] and user choices [136], for example, are cellular automata formulations wherein agents are located in a grid (or lattice) and interact with their geo-spatial nearest neighbors, and possibly change state, based on some rule (i.e., local function). Later, in [137], a grid system is still employed to measure civil disobedience, but the local functions are more complicated. Current modeling efforts, in the three examples above, and in the other references cited in this subsection, illustrate increasing complexity in terms of population representation, environment, state update scheme and agent local functions. Clearly, this evolution is not without exceptions, but there appears to be a general trend of increasing complexity in discrete dynamical systems and ABM, along similar dimensions.

5.3.5 Other modeling approaches

We discussed ABM in the context of GDSs above. In contrast, analytical approaches have also been developed (see, e.g., [138, 139]). Intermediate solutions in terms of granularity—between analytical solutions and fine-grained ABMs—include ODE and compartmental models [140–146]. Various kinds of modeling of biological systems are presented in [147]. Multi-scale aspects of biological modeling are discussed in [148].

6 Decentralized dynamics: game theoretic models

The GDS model, as defined in Sect. 3, works with a fixed local function at each node. Furthermore, the analysis problems discussed earlier consider GDSs as computing

devices; thus, we are interested in estimating the computational resources needed to determine phase space properties. In many other settings, GDSs can be viewed as graphical models of games. From this point of view, we can think of each node as an agent that interacts with its neighbors to optimize individual utility. The local function at a node v can then be viewed as a strategy adopted by the player represented by v . We illustrate this in the specific context of vaccination decisions for epidemic control. Recall the SI/SIS/SIR models of epidemic spread discussed in Sect. 5. An individual v can protect oneself by getting vaccinated (which can be modeled by the function $f_v = 0$); however, this involves a certain cost C_v , which might be the cost of the vaccine, or the side effects. On the other hand, if enough nodes in the network are vaccinated, the probability that v gets infected would be low, in which case node v has no incentive to get vaccinated. This is a natural game theoretic setting and has been studied extensively using noncooperative game models.

Vaccination games were first introduced by Bauch and Earn [149], who model vaccination decisions as a strategic noncooperative game. This was first extended to a network setting by Aspnes et al. [150] for the SI model, and we use their terminology to describe the basic formulation. We assume that the infection starts from a random source $s \in V$ (which may be vaccinated, in which case the disease dies out immediately). We assume the vaccine has 100% efficacy, so that if a node is vaccinated, it is protected from the infection (and also does not participate in the spread of the infection). Therefore, the vaccination by node v can be modeled by removing v from G . We also restrict ourselves to *pure* decisions, where the nodes only have two choices: (1) get vaccinated (denoted by $a_v = 1$), in which case it incurs a fixed cost C_v ; (2) not get vaccinated (denoted by $a_v = 0$): In this case, the node incurs zero cost if it does not get infected, but incurs cost L_v if it gets infected. The expected cost of infection is $\Pr[v \text{ gets infected}] \cdot L_v$. Note that the probability v gets infected depends on the strategies of all the nodes. We denote $\mathbf{a} = (a_v : v \in V)$ as the strategy vector, and $G_{\mathbf{a}}$ as the graph obtained by removing the subset $S = \{v : a_v = 1\}$. Putting these together, the expected cost for node v , given a strategy vector \mathbf{a} , can be expressed as

$$\text{cost}_v(\mathbf{a}) = a_v C_v + (1 - a_v) L_v \Pr[v \text{ gets infected in } G_{\mathbf{a}}].$$

The main solution concept in noncooperative game theory is a *Nash equilibrium* (NE): A strategy \mathbf{a} is said to be a NE if no node v can decrease its cost by unilaterally switching its strategy a_v . See [56, 151] for an introduction to game theory. Aspnes et al. [150] characterize the structure of Nash equilibria (NE) in epidemic games on networks and the complexity of computing them. They show that a pure NE always exists (in which $a_v \in \{0, 1\}$), and can be

computed using an iterative *best response* method: Each node v updates its strategy a_v if that results in lower cost. Let $\text{cost}(\mathbf{a}) = \sum_v \text{cost}_v(\mathbf{a})$; a strategy vector \mathbf{a} that minimizes the cost is referred to as a social optimum. A quantity of interest in noncooperative games is the Price of Anarchy (PoA), which is the maximum ratio of the cost of a NE to the cost of a social optimum. Aspnes et al. show that for a game involving n players, the PoA can be $\Theta(n)$ in general, which implies that decentralized decisions can lead to highly sub-optimal outcomes for the system as a whole.

7 Final remarks

We discussed various models of graphical dynamical systems and pointed out how they are useful in studying several problems that arise in the context of bio-social systems. Due to space reasons, our focus was on a few basic analysis problems. Many other kinds of phase space analysis problems arise in various bio-social systems in the literature; their mathematical and computational study is an important topic of research. We also considered only a few classes of local functions. Many other classes of functions are likely to be useful in modeling various aspects of bio-social systems. We close by discussing the limitations of the GDS framework, highlighting some open questions and providing a brief discussion on other research topics which have been studied in the GDS context.

7.1 Limitations

Despite being very useful as a modeling framework, the GDS approach has some limitations. As stated in Sect. 5.2, one limitation is that the size of the phase space of a GDS is exponential in the number of nodes, and hence, methods such as simulation must be used for systems of large size to approximate the dynamics. Also, under standard assumptions in computational complexity [152], many analysis problems cannot be solved efficiently. However, this difficulty also opens up a research direction, namely the development of efficient heuristics for the analysis problems. For biological and social systems, determining the exact local functions can be difficult and may require significant numbers of data points. However, this difficulty is not unique to the GDS formalism; producing accurate models of agent behavior is difficult in many contexts.

7.2 Some open questions and research directions

For many analysis problems which are known to be computationally intractable in general, it is of interest to study whether special cases can be solved efficiently. Some

results along this direction were mentioned in Sect. 4, such as for the reachability, predecessor existence, the fixed-point existence problems and analysis of more general phase space properties. As discussed, many of these problems are computationally very hard, in general, but can be solved efficiently in some cases if the graphs have bounded treewidth. The applicability of these results to real bio-social systems is of considerable interest. In particular, an important open direction is to determine whether such problems can be solved efficiently for SyGDS instances whose underlying networks arise in real bio-social systems. Such networks have many special properties, e.g., high clustering and low diameter (see, e.g., [56]); exploiting these properties to develop efficient algorithms for the various analysis problems is an important research topic. A first step in this direction might be to develop better characterizations of phase space properties and their sensitivity in SyGDS instances on such networks. Another important direction is to develop “approximate” notions of these problems, as a way to address the computational hardness. For instance, in the predecessor existence problem, one might ask whether there is a configuration C' such that its successor is “close” to a given configuration C . Such notions are likely to make more bio-social instances amenable to analysis; they can also help in incorporating uncertainty.

We also discussed the use of simulations to understand the behavior of bio-social systems modeled as graph dynamical systems. This topic also offers many interesting research issues. For example, Adiga et al. [153] discuss over a dozen future research directions for the modeling of epidemics in tropical climates; each of them can be cast as a specific analysis problem involving one or more components (e.g., network, local functions) of a GDS. Exploiting the features of practical bio-social systems so that simulations can be carried out efficiently on commonly available computing platforms is an additional direction for future work.

7.3 Other research topics

There are many additional topics beyond those covered here—this includes inference, control and optimization problems. For example, in epidemiology and more broadly the spread of contagions over networks, one is often interested in accentuating [78, 154–156] or thwarting [81, 143, 157–159] the spread of contagions. Maximizing the spread of information (e.g., through judicious seeding) is an example of the former, while efficient allocation of pharmaceutical interventions (e.g., antivirals or vaccines) is an example of the latter. A very important open topic is to design effective interventions (e.g., determining whom to vaccinate), given a limited set of resources. These are

challenging optimization problems, even when the intervention is implemented in a centralized manner. In practice, game theoretic issues need to be taken into consideration, as discussed in Sect. 6.

The GDS formalism allows researchers to use established mathematical and computational machinery in addressing these issues for bio-social systems. Inference problems cut across all the areas just mentioned. As an example, one may want to determine a system’s GDS representation (assuming that one exists) through observations of the system’s transitions (e.g., a time series) [160]. Other examples arise in epidemiology, wherein one wants to determine the index node (also called “patient zero”), given some observations regarding the state of the epidemic after a certain time.

Acknowledgements We thank members of the Network Dynamics and Simulation Science Laboratory (NDSSL) for their comments and input. Specifically, we thank Chris Barrett, Christian Reidys, Daniel Rosenkrantz and Richard Stearns for their collaboration on several papers discussed in this article. We thank the computer systems administrators and managers at the Biocomplexity Institute of Virginia Tech for their help in this and many other works: Dominik Borkowski, William Miles Gentry, Jeremy Johnson, William Marangas, Douglas McMaster, Kevin Shinpaugh and Robert Wills. This work has been partially supported by DTRA CNIMS (Contract HDTRA1-11-D-0016-0001), NSF BIG DATA Grant IIS-1633028, NSF DIBBS Grant ACI-1443054 and NSF EAGER Grant CMMI-1745207. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

Disclaimer The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the US Government.

References

- Epstein, J.M.: Generative Social Science: Studies in Agent-Based Computational Modeling. Princeton University Press, Princeton (2007)
- Barrett, C., Bisset, K., Eubank, S., Marathe, M., Kumar, V., Mortveit, H.: Modeling and simulation of large biological, information and socio-technical systems: an interaction based approach. In: Laubenbacher, R. (ed.) Modeling and Simulation of Biological Networks, pp. 101–147. American Mathematical Society (2007)
- Barrett, C., Eubank, S., Marathe, M.: Modeling and simulation of large biological, information and socio-technical systems: an interaction based approach. In: Goldin, D., Smolka, S.A., Wegner, P. (eds.) Interactive Computation, pp. 353–392. Springer, Berlin (2006)
- Kuhlman, C.J., Kumar, V.A., Marathe, M.V., Mortveit, H.S., Swarup, S., Tuli, G., Ravi, S., Rosenkrantz, D.J.: A general-purpose graph dynamical system modeling framework. In: Simulation Conference (WSC), Proceedings of the 2011 Winter, pp. 296–308. IEEE (2011)

5. Barrett, C., Beckman, R., Berkbigler, K., Bisset, K., Bush, B., Campbell, K., Eubank, S., Henson, K., Hurford, J., Kubicek, D., Marathe, M., Romero, P., Smith, J., Smith, L., Speckman, P., Stretz, P., Thayer, G., Eeckhout, E., Williams, M.: TRANSIMS: transportation analysis simulation system. Technical Report LA-UR-00-1725, LANL (2001)
6. Eubank, S.: Scalable, efficient epidemiological simulation. In: Proceedings of Symposium on Applied Computing, pp. 139–145 (2002)
7. Eubank, S., Guclu, H., Kumar, V.S.A., Marathe, M., Srinivasan, A., Toroczkai, Z., Wang, N.: Modelling disease outbreaks in realistic urban social networks. *Nature* **429**, 180–184 (2004)
8. Bisset, K.R., Chen, J., Deodhar, S., Feng, X., Ma, Y., Marathe, M.V.: Indemics: an interactive high-performance computing framework for data intensive epidemic modeling. *ACM Trans. Model. Comput. Simul. (TOMACS)* **24**(1), 4:1–4:32 (2014). (**Special Issue on Simulation in Complex Service Systems**)
9. Barrett, C.L., Bisset, K.R., Eubank, S.G., Feng, X., Marathe, M.V.: Episimdemics: an efficient algorithm for simulating the spread of infectious disease over large realistic social networks. In: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing, pp. 37:1–37:12 (2008)
10. Parikh, N., Swarup, S., Stretz, P.E., Rivers, C.M., Lewis, B.L., Marathe, M.V., Eubank, S.G., Barrett, C.L., Lum, K., Chungbaek, Y.: Modeling human behavior in the aftermath of a hypothetical improvised nuclear detonation. In: Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS), pp. 949–956. Saint Paul, MN, USA (2013)
11. Barrett, C., Bisset, K., Chandan, S., Chen, J., Chungbaek, Y., Eubank, S., Evrenosoğlu, Y., Lewis, B., Lum, K., Marathe, A., Marathe, M., Mortveit, H., Parikh, N., Phadke, A., Reed, J., Rivers, C., Saha, S., Stretz, P., Swarup, S., Thorp, J., Vullikanti, A., Xie, D.: Planning and response in the aftermath of a large crisis: an agent-based informatics framework. In: Pasupathy, R., Kim, S.H., Tolk, A., Hill, R., Kuhl, M.E. (eds.) Proceedings of the 2013 Winter Simulation Conference, pp. 1515–1526 (2013)
12. Bisset, K., Alam, M., Bassaganya-Riera, J., Carbo, A., Eubank, S., Hontecillas, R., Hoops, S., Mei, Y., Wendelsdorf, K., Xie, D., Yeom, J., Marathe, M.: High-performance interaction-based simulation of gut immunopathologies with enteric immunity simulator (ENISI). In: Proceedings of International Parallel and Distributed Processing Symposium (IPDPS), pp. 48–59 (2012)
13. Alam, M., Deng, X., Philipson, C., Bassaganya-Riera, J., Bisset, K., Carbo, A., Eubank, S., Hontecillas, R., Hoops, S., Mei, Y., Abedi, V., Marathe, M.: Sensitivity analysis of an ENteric Immunity Simulator (ENISI)-based model of immune responses to *Helicobacter pylori* infection. *PLoS ONE* **10**(9), e0136139 (2015)
14. Barrett, C., Eubank, S., Kumar, V.S.A., Marathe, M.: Understanding large-scale social and infrastructure networks: a simulation-based approach. *SIAM News* **37**(4), 1–5 (2004)
15. Yeom, J.S., Bhatale, A., Bisset, K.R., Bohm, E., Gupta, A., Kale, L.V., Marathe, M., Nikolopoulos, D.S., Schulz, M., Wesolowski, L.: Overcoming the scalability challenges of epidemic simulations on Blue Waters. In: Proceedings of the IEEE 28th International Parallel and Distributed Processing Symposium, pp. 755–764 (2014)
16. von Neumann, J.: Theory of Self-Reproducing Automata. University of Illinois Press, Champaign (1966). (**Edited and completed by Arthur W. Burks**)
17. Kari, J.: Theory of cellular automata: a survey. *Theor. Comput. Sci.* **334**, 3–33 (2005)
18. Ilachinski, A.: Cellular Automata: A Discrete Universe. World Scientific Publishing Company, Cambridge (2001)
19. Delorme, M., Mazoyer, J. (eds.): Cellular Automata—A Parallel Model, Mathematics and Its Applications, vol. 460. Kluwer Academic Publishers, Alphen aan den Rijn (1999)
20. Wolfram, S.: Theory and Applications of Cellular Automata, Advanced Series on Complex Systems, vol. 1. World Scientific Publishing Company, Singapore (1986)
21. Kauffman, S.: Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* **22**(3), 437–467 (1969)
22. Ribeiro, A.S., Kauffman, S.A.: Noisy attractors and ergodic sets in models of gene regulatory networks. *J. Theor. Biol.* **247**(4), 743–755 (2007)
23. Goles, E., Martínez, S.: Neural and Automata Networks: Dynamical Behaviour and Applications. Kluwer Academic Publishers, Alphen aan den Rijn (1990)
24. Goles-Chacc, E., Fogelman-Soulie, F., Pellegrin, D.: Decreasing energy functions as a tool for studying threshold networks. *Discrete Appl. Math.* **12**, 261–277 (1985)
25. Goles, E., Olivos, J.: Periodic behavior in generalized threshold functions. *Discrete Math.* **30**, 187–189 (1980)
26. Ruz, G.A., Goles, E.: Reconstruction and update robustness of the mammalian cell cycle network. In: Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), pp. 397–403 (2012)
27. Gershenson, C.: Introduction to random Boolean networks (2004). [arXiv:nlin/040806v3-12Aug2004](https://arxiv.org/abs/nlin/040806v3). Accessed Aug 2005
28. Shmulevich, I., Dougherty, E.R., Zhang, W.: From Boolean to probabilistic Boolean networks as models of genetic regulatory networks. *Proc. IEEE* **90**(11), 1778–1792 (2002)
29. Shmulevich, I., Dougherty, E.R., Kim, S., Zhang, W.: Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* **18**(2), 261–274 (2002)
30. Shmulevich, I., Kauffman, S.A.: Activities and sensitivities in Boolean network models. *Phys. Rev. Lett.* **93**(4), 048701:1–048701:4 (2004)
31. Jarrah, A.S., Laubenbacher, R.: On the Algebraic Geometry of Polynomial Dynamical Systems, The IMA Volumes in Mathematics and Its Applications, vol. 149, pp. 109–123. Springer, New York (2009)
32. Jarrah, A.S., Raposa, B., Laubenbacher, R.: Nested canalyzing, unate cascade, and polynomial functions. *Physica D* **233**, 167–174 (2007)
33. Kaneko, K.: Pattern dynamics in spatiotemporal chaos. *Physica D* **34**, 1–41 (1989)
34. Golubitsky, M., Pivato, M., Stewart, I.: Interior symmetry and local bifurcations in coupled cell networks. *Dyn. Syst.* **19**(4), 389–407 (2004)
35. Nishikawa, T., Sun, J., Motter, A.E.: Sensitive dependence of optimal network dynamics on network structure. *Phys. Rev. X* **7**, 041044:1–041044:21 (2017)
36. Davidich, M., Bornholdt, S.: The transition from differential equations to Boolean networks: a case study in simplifying a regulatory network model. *J. Theor. Biol.* **255**(3), 269–277 (2008)
37. Veliz-Cuba, A., Stigler, B.: Boolean models can explain bistability in the lac Operon. *J. Comput. Biol.* **18**, 783–794 (2011)
38. Barrett, C.L., Hunt III, H.B., Marathe, M.V., Ravi, S.S., Rosenkrantz, D.J., Stearns, R.E.: Complexity of reachability problems for finite discrete dynamical systems. *J. Comput. Syst. Sci.* **72**(8), 1317–1345 (2006)
39. Rosenkrantz, D.J., Marathe, M.V., Ravi, S.S., Stearns, R.E.: Testing phase space properties of synchronous dynamical systems with nested canalyzing local functions. In: Proceedings of 17th International Conference on Autonomous Agents and

- Multiagent Systems (AAMAS), Stockholm, Sweden, pp. 1585–1594 (2018)
40. Barrett, C.L., Hunt III, H.B., Marathe, M.V., Ravi, S.S., Rosenkrantz, D.J., Stearns, R.E.: Reachability problems for sequential dynamical systems with threshold functions. *Theor. Comput. Sci.* **295**(1–3), 41–64 (2003)
 41. Mortveit, H.S., Reidys, C.M.: An Introduction to Sequential Dynamical Systems. Springer, Berlin (2007)
 42. Abdelhamid, S.H.E., Kuhlman, C.J., Marathe, M.V., Mortveit, H.S., Ravi, S.S.: GDSCalc: a web-based application for discrete graph dynamical systems. *PLoS ONE* **10**(8), 24 (2015)
 43. El Samad, H., Khammash, M., Petzold, L., Gillespie, D.: Stochastic modeling of gene regulatory networks. *Int. J. Robust Nonlinear Control* **15**, 691–711 (2005)
 44. Eubank, S., Kumar, V.S.A., Marathe, M.V., Srinivasan, A., Wang, N.: Structure of social contact networks and their impact on epidemics. In: Abello, J.M., Cormode, G. (eds.) Discrete Methods in Epidemiology, vol. 70, pp. 179–200. American Mathematical Society, Providence, RI (2006)
 45. Granovetter, M.: Threshold models of collective behavior. *Am. J. Sociol.* **83**(6), 1420–1443 (1978)
 46. Kuhlman, C.J., Kumar, V.S.A., Marathe, M.V., Ravi, S.S., Rosenkrantz, D.J.: Inhibiting diffusion of complex contagions in social networks: theoretical and experimental results. *Data Min. Knowl. Discov.* **29**(2), 423–465 (2015)
 47. Kuhlman, C.J., Kumar, V.S.A., Marathe, M.V., Swarup, S., Tuli, G., Ravi, S.S., Rosenkrantz, D.J.: Inhibiting the diffusion of contagions in bi-threshold systems: Analytical and experimental results. In: Complex Adaptive Systems: Energy, Information, and Intelligence, Papers from the 2011 AAAI Fall Symposium, Arlington, Virginia, pp. 91–100 (2011)
 48. Kuhlman, C.J.: High performance computational social science modeling of networked populations. Ph.D. thesis, Computer Science Department, Virginia Tech, Blacksburg, VA, USA (2013)
 49. Kuhlman, C.J., Kumar, V.S.A., Marathe, M.V., Ravi, S.S., Rosenkrantz, D.J., Swarup, S., Tuli, G.: A bi-threshold model of complex contagion and its application to the spread of smoking behavior. In: Proceedings of SNA-KDD Workshop, pp. 18.1–18.10 (2011)
 50. Kuhlman, C.J., Marathe, M.V., Kumar, V.S.A., Ravi, S.S., Rosenkrantz, D.J., Stearns, R.E.: Analysis problems for special classes of bi-threshold dynamical systems. In: Proceedings of Workshop on Multiagent Interaction Networks (MAIN 2013), held in conjunction with AAMAS, pp. 26–33 (2013)
 51. Barrett, C.L., Hunt III, H.B., Marathe, M.V., Ravi, S.S., Rosenkrantz, D.J., Stearns, R.E., Thakur, M.: Computational aspects of analyzing social network dynamics. In: IJCAI 2007, Proceedings of 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, 6–12 Jan 2007, pp. 2268–2273 (2007)
 52. Barrett, C.L., Hunt III, H.B., Marathe, M.V., Ravi, S.S., Rosenkrantz, D.J., Stearns, R.E.: Modeling and analyzing social network dynamics using stochastic discrete graphical dynamical systems. *Theor. Comput. Sci.* **412**(30), 3932–3946 (2011)
 53. Rosenkrantz, D.J., Marathe, M.V., Hunt III, H.B., Ravi, S.S., Stearns, R.E.: Analysis problems for graphical dynamical systems: a unified approach through graph predicates. In: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2015, Istanbul, Turkey, 4–8 May 2015, pp. 1501–1509 (2015)
 54. Barrett, C.L., Hunt III, H.B., Marathe, M.V., Ravi, S.S., Rosenkrantz, D.J., Stearns, R.E., Thakur, M.: Predecessor existence problems for finite discrete dynamical systems. *Theor. Comput. Sci.* **386**(1–2), 3–37 (2007)
 55. Barrett, C.L., Hunt III, H.B., Marathe, M.V., Ravi, S.S., Rosenkrantz, D.J., Stearns, R.E.: Predecessor and permutation existence problems for sequential dynamical systems. In: DMCS, pp. 69–80 (2003)
 56. Easley, D., Kleinberg, J.: Networks, Crowds and Markets: Reasoning About a Highly Connected World. Cambridge University Press, New York, NY (2010)
 57. Dodds, P., Watts, D.: A generalized model of social and biological contagion. *J. Theor. Biol.* **232**(4), 587–604 (2005)
 58. Centola, D., Macy, M.: Complex contagions and the weakness of long ties. *Am. J. Sociol.* **113**(3), 702–734 (2007)
 59. Kleinberg, J.: Cascading behavior in networks: algorithmic and economic issues. In: Nissan, N., Roughgarden, T., Tardos, E., Vazirani, V. (eds.) Algorithmic Game Theory, pp. 613–632. Cambridge University Press, New York, NY (2007)
 60. Barrett, C.L., Hunt III, H.B., Marathe, M.V., Ravi, S.S., Rosenkrantz, D.J., Stearns, R.E.: Analysis problems for sequential dynamical systems and communicating state machines. In: Proceedings of MFCS, pp. 159–172 (2001)
 61. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-completeness. W. H. Freeman and Co., San Francisco, CA (1979)
 62. Barrett, C.L., Hunt III, H.B., Marathe, M.V., Ravi, S.S., Rosenkrantz, D.J., Stearns, R.E.: On special classes of sequential dynamical systems. *Ann. Comb.* **7**, 381–408 (2003)
 63. Bodlaender, H.: Treewidth: algorithmic techniques and results. In: Proceedings of 22nd Symposium on Mathematical Foundations of Computer Science, pp. 29–36 (1997)
 64. Barrett, C.L., Hunt III, H.B., Marathe, M.V., Ravi, S.S., Rosenkrantz, D.J., Stearns, R.E., Tomic, P.T.: Gardens of Eden and fixed points in sequential dynamical systems. In: Proceedings of International Conference on Discrete Models Combinatorics, Computation and Geometry (DM-CCG), pp. 95–110 (2001)
 65. Kosub, S., Homan, C.M.: Dichotomy results for fixed point counting in Boolean dynamical systems. In: Proceedings of ICTCS, pp. 163–174 (2007)
 66. Tomic, P.T.: On the complexity of enumerating possible dynamics of sparsely connected Boolean network automata with simple update rules. In: Automata 2010—16th International Workshop on CA and DCS, pp. 125–144 (2010)
 67. Akutsu, T., Kosub, S., Melkman, A., Tamura, T.: Finding a periodic attractor of a Boolean network. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **9**(5), 1410–1421 (2012)
 68. Kauffman, S.: The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press, New York, NY (1993)
 69. Kauffman, S., Peterson, C., Samuelsson, B., Troein, C.: Random Boolean network models and the yeast transcriptional network. *Proc. Natl. Acad. Sci. (PNAS)* **100**(25), 14796–14799 (2003)
 70. Bornholdt, S.: Boolean network models of cellular regulation: prospects and limitations. *J. R. Soc. Interface* **5**(Suppl 1), S85–S94 (2008)
 71. Wang, R.S., Saadatpour, A., Albert, R.: Boolean modeling in systems biology: an overview of methodology and applications. *Phys. Biol.* **9**(5), 055001 (2012)
 72. Laschov, D., Margaliot, M., Even, G.: Observability of Boolean networks: a graph-theoretic approach. *Automatica* **49**(8), 2351–2362 (2013)
 73. Nguyen, C., Schlesinger, K.J., Carlson, J.M.: Data-driven models for individual and group decision making. In: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 852–859 (2017)
 74. Macal, C., North, M.: Introductory tutorial: agent-based modeling and simulation. In: Proceedings of the 2014 Winter Simulation Conference, pp. 6–20 (2014)

75. Weimer, C.W., Miller, J.O., Hill, R.R.: Agent-based modeling: an introduction and primer. In: Proceedings of the 2016 Winter Simulation Conference, pp. 65–79 (2016)
76. An, G., Mi, Q., Dutta-Moscato, J., Vodovotz, Y.: Agent-based models in translational systems biology. Wiley Interdiscip. Rev. Syst. Biol. Med. **1**, 159–171 (2009)
77. Nikolai, C., Madey, G.: Tools of the trade: a survey of various agent based modeling platforms. J. Artif. Soc. Soc. Simul. **12**, 1–37 (2009)
78. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137–146 (2003)
79. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 199–208 (2009)
80. Nguyen, N.P., Yan, G., Thai, M.T., Eidenbenz, S.: Containment of misinformation spread in online social networks. In: Proceedings of the 4th Annual ACM Web Science Conference, pp. 213–222 (2012)
81. Kuhlman, C.J., Tuli, G., Swarup, S., Marathe, M.V., Ravi, S.S.: Blocking simple and complex contagion by edge removal. In: 2013 IEEE 13th International Conference on Data Mining, Dallas, TX, USA, 7–10 Dec 2013, pp. 399–408 (2013)
82. Demongeot, J., Goles, E., Morvan, M., Noual, M., Sene, S.: Attraction basins as gauges of robustness against boundary conditions in biological complex systems. PLoS ONE **5**, e11793-1–e11793-18 (2010)
83. Wendelsdorf, K., Alam, M., Bassaganya-Riera, J., Bisset, K., Eubank, S., Hontecillas, R., Hoops, S., Marathe, M.: Enteric immunity simulator: a tool for *in silico* study of gastroenteric infections. IEEE Trans. NanoBiosci. **11**(3), 273–288 (2012)
84. Wendelsdorf, K., Bassaganya-Riera, J., Bisset, K., Eubank, S., Hontecillas, R., Marathe, M.: Enteric immunity simulator: a tool for *in silico* study of gut immunopathologies. In: Proceedings of the IEEE International Conference Bioinformatics and Biomedicine, pp. 462–469 (2011)
85. Kauffman, S.: Homeostasis and differentiation in random genetic control networks. Nature **224**, 177–178 (1969)
86. Thomas, R.: Boolean formalisation of genetic control circuits. J. Theor. Biol. **42**(3), 563–585 (1973)
87. Goles, E., Salinas, L.: Comparison between parallel and serial dynamics of Boolean networks. Theor. Comput. Sci. **396**, 247–253 (2008)
88. Serra, R., Villani, M., Barbieri, A., Kauffman, S., Colacci, A.: On the dynamics of random Boolean networks subject to noise: attractors, ergodic sets and cell types. J. Theor. Biol. **265**, 185–193 (2010)
89. Luo, J.X., Turner, M.S.: Evolving sensitivity balances Boolean networks. PLoS ONE **7**, e36010-1–e36010-8 (2012)
90. Macauley, M., Mortveit, H.S.: On enumeration of conjugacy classes of Coxeter elements. Proc. Am. Math. Soc. **136**(12), 4157–4165 (2008)
91. Macauley, M., Mortveit, H.: Cycle equivalence of graph dynamical systems. Nonlinearity **22**(2), 421–436 (2009)
92. Welsh, D.: The Tutte polynomial. Random Struct. Algorithms **15**, 210–228 (1999)
93. Gordon, S., Taylor, P.R.: Monocyte and macrophage heterogeneity. Nat. Rev. Immunol. **5**(12), 953–64 (2005)
94. Iwasaki, A.: Mucosal dendritic cells. Ann. Rev. Immunol. **25**, 381–418 (2007)
95. Grilo, A., Caetano, A., Rosa, A.: Agent based artificial immune system. In: Proceedings of GECCO-01, vol. LBP pp. 145–151 (2001)
96. Tay, J.C., Jhavar, A.: CAFISS: A complex adaptive framework for immune system simulation. In: Proceedings of the 2005 ACM Symposium on Applied Computing, SAC '05, pp. 158–164. ACM, New York, NY (2005)
97. Celada, F., Seiden, P.E.: A computer model of cellular interactions in the immune system. Immunol. Today **13**(2), 56–62 (1992)
98. Castiglione, F., Duca, K., Jarrah, A., Laubenbacher, R., Hochberg, D., Thorley-Lawson, D.: Simulating Epstein-Barr virus infection with C-ImmSim. Bioinformatics **23**, 1371–1377 (2007)
99. Bernaschi, M., Castiglione, F.: Design and implementation of an immune system simulator. Comput. Biol. Med. **31**(5), 303–31 (2001)
100. Emerson, A., Rossi, E.: Immunogrid—the virtual human immune system project. Stud. Health Technol. Inform. **126**, 87–92 (2007)
101. Efroni, S., Harel, D., Cohen, I.: Reactive animation: realistic modeling of complex dynamic systems. IEEE Comput. **38**(1), 38–47 (2005)
102. Swerdliv, N., Cohen, I.R., Harel, D.: The lymph node B cell immune response: dynamic analysis *in-silico*. Proc. IEEE **96**(8), 1421–1443 (2008)
103. Mata, J., Cohn, M.: Cellular automata-based modeling program: synthetic immune system. Immunol. Rev. **216**(1), 198–212 (2007)
104. Meier-Schellersheim, M., Mack, G.: SIMMUNE, a tool for simulating and analyzing immune system behavior. CoRR cs.MA/9903017 (1999)
105. Sneddon, M.W., Faeder, J.R., Emonet, T.: Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. Nat. Methods **8**(2), 177–83 (2011)
106. Folcik, V.A., An, G.C., Orosz, C.G.: The basic immune simulator: an agent-based model to study the interactions between innate and adaptive immunity. Theor. Biol. Med. Model. **4**, 39 (2007)
107. Sutterlin, T., Huber, S., Dickhaus, H., Grabe, N.: Modeling multi-cellular behavior in epidermal tissue homeostasis via finite state machines in multi-agent systems. Bioinformatics **25**(16), 2057–2063 (2009)
108. Bauer, A.L., Beauchemin, C.A.A., Perelson, A.S.: Agent-based modeling of host-pathogen systems: the successes and challenges. Inf. Sci. **179**(10), 1379–1389 (2009)
109. Fachada, N., Lopes, V.V., Rosa, A.: Agent based modelling and simulation of the immune system: a review. Technical report, Systems and Robotics Institute, Instituto Superior Tecnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal (2000)
110. Pappalardo, F., Zhang, P., Halling-Brown, M., Basford, K., Scalia, A., Shepherd, A.J., Moss, D., Motta, S., Brusic, V.: Computational simulations of the immune system for personalized medicine: state of the art and challenges. Curr. Pharmaco-geneom. Personal. Med. **6**(4), 260–271 (2008)
111. Ferguson, N., Cummings, D., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S., Burke, D.: Strategies for containing an emerging Influenza pandemic in Southeast Asia. Nature **437**, 209–214 (2005)
112. Barrett, C., Bissett, K., Chen, J., Feng, X., Kumar, V.S.A., Marathe, M.: Epifast: a fast algorithm for large scale realistic epidemic simulations on distributed memory systems. In: International Conference on Supercomputing (ICS), pp. 430–439 (2009)
113. Kumar, S., Piper, K., Galloway, D.D., Hadler, J.L., Grefenstette, J.J.: Is population structure sufficient to generate area-level inequalities in influenza rates? An examination using agent-based models. BMC Public Health **15**, 947 (2015)

114. Fox, S.J., Miller, J.C., Meyers, L.A.: Seasonality in risk of pandemic influenza emergence. *PLoS Comput. Biol.* **13**(10), e1005749-1–e1005749-23 (2017)
115. Marathe, M., Vullikanti, A.: Computational epidemiology. *Commun. ACM* **56**(7), 88–96 (2013)
116. Bhatele, A., Yeom, J.S., Jain, N., Kuhlman, C.J., Livnat, Y., Bisset, K.R., Kale, L.V., Marathe, M.V.: Massively parallel simulations of spread of infectious diseases over realistic social networks. In: Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, pp. 689–694 (2017)
117. Abdelhamid, S., Kuhlman, C.J., Marathe, M.V., Ravi, S.S., Reid, K.: Agent-based modeling and simulation of depression and its impact on students' success and academic retention. In: American Society for Engineering Education (ASEE) (2016)
118. Lum, K., Swarup, S., Eubank, S., Hawdon, J.: The contagious nature of imprisonment: an agent-based model to explain racial disparities in incarceration rates. *J. R. Soc. Interface* **11**(98), 12 (2014)
119. Schelling, T.C.: Dynamic models of segregation. *J. Math. Sociol.* **1**, 143–186 (1971)
120. Ameden, H.A., Boxall, P.C., Cash, S.B., Vickers, D.A.: An agent-based model of border enforcement for invasive species management. *Can. J. Agric. Econ.* **57**, 481–496 (2009)
121. Tonnang, H.E., Herve, B.D., Biber-Freudenberger, L., Salifu, D., Subramanian, S., Ngowi, V.B., Guimapi, R.Y., et al.: Advances in crop insect modelling method—towards a whole system approach. *Ecol. Model.* **354**, 88–103 (2017)
122. Merkey, B.V., Lardon, L.A., Seoane, J.M., Kreft, J.U., Smets, B.F.: Growth dependence of conjugation explains limited plasmid invasion in biofilms: an individual-based modelling study. *Environ. Microbiol.* **13**, 2435–2452 (2011)
123. Spies, T.A., White, E., Ager, A., Kline, J.D., Bolte, J.P., Platt, E.K., Olsen, K.A., et al.: Using an agent-based model to examine forest management outcomes in a fire-prone landscape in Oregon, USA. *Ecol. Soc.* **22**, 20 (2017)
124. Thorne, B.C., Bailey, A.M., Peirce, S.M.: Combining experiments with multi-cell agent-based modeling to study biological tissue patterning. *Brief. Bioinform.* **8**(4), 245–257 (2007)
125. Channakeshava, K., Bisset, K., Kumar, V.A., Marathe, M., Yardi, S.: High performance scalable and expressive modeling environment to study mobile malware in large dynamic networks. In: 25th IEEE International Parallel & Distributed Processing Symposium (IPDPS), pp. 770–781 (2011)
126. Bookstaber, R.: Using agent-based models for analyzing threats to financial stability. Technical Report Working Paper No. 0003, U.S. Dept. of Treasury (2012)
127. Paul, M., Dredze, M.: A model for mining public health topics from Twitter. *Health* **11**, 16–6 (2012)
128. Zhang, B., Chan, W.K.V., Ukkusuri, S.V.: Agent-based modeling for household level hurricane evacuation. In: Proceedings of the 2009 Winter Simulation Conference, pp. 2778–2784 (2009)
129. Grimm, V.: Ten years of individual-based modelling in ecology: what have we learned and what could we learn in the future? *Ecol. Model.* **115**, 129–148 (1999)
130. Zhu, Y., Xie, K., Ozbay, K., Yang, H.: Hurricane evacuation modeling using behavior models and scenario-driven agent-based simulations. *Procedia Comput. Sci.* **130**, 836–843 (2018)
131. Shmulevich, I., Lähdesmäki, H., Dougherty, E.R., Astola, J., Zhang, W.: The role of certain post classes in Boolean network models of genetic networks. *Proc. Natl. Acad. Sci.* **100**(19), 10734–10739 (2003)
132. Heckbert, S., Baynes, T., Reeson, A.: Agent-based modeling in ecological economics. *Ann. N. Y. Acad. Sci.* **1185**, 39–53 (2010)
133. Squazzoni, F.: The impact of agent-based models in the social sciences after 15 years of incursions. *Hist. Econ. Ideas* **18**, 197–233 (2010)
134. Bruch, E., Atwell, J.: Agent-based models in empirical social research. *Sociol. Methods Res.* **44**, 186–221 (2013)
135. Bianchi, F., Squazzoni, F.: Agent-based models in sociology. *WIREs Comput. Stat.* **7**, 284–306 (2015)
136. Axelrod, R.: The dissemination of culture. *J. Confl. Resolut.* **41**, 203–226 (1997)
137. Epstein, J.M.: Modeling civil violence: an agent-based computational approach. *Proc. Natl. Acad. Sci. (PNAS)* **99**, 7243–7250 (2002)
138. Hethcote, H.: The mathematics of infectious diseases. *SIAM Rev.* **42**(4), 599–653 (2000)
139. Newman, M.: The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003)
140. Longini, I.M., Nizam, A., Xu, S., Ungchusak, K., Hanshaoworakul, W., Cummings, D.A., Halloran, E.M.: Containing pandemic influenza at the source. *Science* **309**(5737), 1083–1087 (2005)
141. Sander, B., Nizam, A., Garrison, L.P., Postma, M.J., Halloran, M.E., Ira, M., Longini, J.: Economic evaluation of influenza pandemic mitigation strategies in the United States using a stochastic microsimulation transmission model. *Value Health* **12**, 226–233 (2009)
142. Yang, Y., Sugimoto, J., Halloran, M., Basta, N., Chao, D., Matrajt, L., Potter, G., Kenah, E., Longini, I.M.: The transmissibility and control of pandemic influenza a (H1N1) virus. *Science* **326**(5953), 729–733 (2009)
143. Halloran, M., Ferguson, N., Eubank, S., Longini, I., Cummings, D., Lewis, B., Xu, S., Fraser, C., Vullikanti, A., Germann, T., Wagener, D., Beckman, R., Kadau, K., Barrett, C., Macken, C., Burke, D., Cooley, P.: Modeling targeted layered containment of an influenza pandemic in the United States. *Proc. Natl. Acad. Sci. (PNAS)* **105**(12), 4639–4644 (2008)
144. Pandey, A., Atkins, K.E., Medlock, J., Wenzel, N., Townsend, J.P., Childs, J.E., Nyenswah, T.G., Ndeffo-Mbah, M.L., Galvani, A.P.: Strategies for containing Ebola in West Africa. *Science* **346**(6212), 991–995 (2014)
145. Rivers, C., Lofgren, E., Marathe, M., Eubank, S., Lewis, B.: Modeling the impact of interventions on an epidemic of Ebola in Sierra Leone and Liberia. *PLoS Curr.* (2014). <https://doi.org/10.1371/currents.outbreaks.4d41fe5d6c05e9df30ddce33c66d084c>
146. Venkatraman, S., Chen, J., Gupta, S., Lewis, B.L., Marathe, M., Mortveit, H.S., Vullikanti, A.: Spatio-temporal optimization of seasonal vaccination using a metapopulation model of Influenza. In: 2017 IEEE International Conference on Healthcare Informatics, ICHI, pp. 134–143 (2017)
147. Ji, Z., Yan, K., Li, W., Hu, H., Zhu, X.: Mathematical and computational modeling in complex biological systems. *Hindawi BioMed Res. Int.* **2017**, 1–16 (2017)
148. Walpole, J., Papin, J.A., Peirce, S.M.: Mathematical and computational modeling in complex biological systems. *Ann. Rev. Biomed. Eng.* **15**, 137–154 (2013)
149. Bauch, C., Earn, D.: Vaccination and the theory of games. *Proc. Natl. Acad. Sci. (PNAS)* **101**(36), 13391–13394 (2004)
150. Aspnes, J., Rustagi, N., Saia, J.: Worm versus alert: who wins in a battle for control of a large-scale network? In: Proceedings of Principles of Distributed Systems, 11th International Conference, OPODIS 2007, pp. 443–456 (2007)
151. Narahari, Y.: Game Theory and Mechanism Design, IISc Lecture Notes, vol. 4. World Scientific, Singapore (2014)
152. Papadimitriou, C.H.: Computational Complexity. Pearson Publishing, Reading, MA (1993)

153. Adiga, A., Chu, S., Eubank, S., Kuhlman, C.J., Lewis, B., Marathe, A., Marathe, M., Nordberg, E.K., Swarup, S., Vulikanti, A., Wilson, M.L.: Disparities in spread and control of Influenza in slums of Delhi: findings from an agent-based modelling study. *BMJ Open* **8**(1), 12 (2018)
154. Domingos, P., Richardson, M.: Mining the network value of customers. In: KDD '01: Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 57–66 (2001)
155. Domingos, P., Richardson, M.: Mining knowledge-sharing sites for viral marketing. In: Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining, pp. 61–70 (2002)
156. Kitsak, M., Gallos, L., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H., Makse, H.: Identifying influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893 (2010)
157. Halloran, M., Longini Jr., I., Nizam, A., Yang, Y.: Possible containment of bio-terrorist smallpox. *Science* **298**, 1428–1432 (2002)
158. Barrett, C., Chen, J., Eubank, S., Kumar, V., Lewis, B., Marathe, A., Marathe, M.: Role of vulnerable and critical nodes in controlling epidemics in social networks. In: Proceedings of Epidemics (2008)
159. Barrett, C., Bisset, K., Leidig, J., Marathe, A., Marathe, M.: Economic and social impact of influenza mitigation strategies by demographic class. *Epidemics* **3**(1), 19–31 (2011)
160. Adiga, A., Kuhlman, C., Marathe, M.V., Ravi, S.S., Rosenkrantz, D.J., Stearns, R.E.: Inferring local transition functions of discrete dynamical systems from observations of system behavior. *Theor. Comput. Sci.* **679**, 126–144 (2017)

Infering Probabilistic Contagion Models Over Networks Using Active Queries

Abhijin Adiga
Virginia Tech
abhijin@vt.edu

Madhav V. Marathe
Virginia Tech
mmarathe@vt.edu

Vanessa Cedeno-Mieles*
Virginia Tech
vcedeno@vt.edu

S. S. Ravi†
Virginia Tech
ssravi0@gmail.com

Chris J. Kuhlman
Virginia Tech
ckuhlman@vt.edu

Daniel J. Rosenkrantz
University at Albany – SUNY
drosenkrantz@gmail.com

Richard E. Stearns
University at Albany – SUNY
thestearns2@gmail.com

ABSTRACT

The problem of inferring unknown parameters of a networked social system is of considerable practical importance. We consider this problem for the independent cascade model using an active query framework. More specifically, given a network whose edge probabilities are unknown, the goal is to infer the probability value on each edge by querying the system. The optimization objective is to use as few queries as possible in carrying out the inference. We present approximation algorithms that provide provably good estimates of edge probabilities. We also present results from an experimental evaluation of our algorithms on several real-world networks.

CCS CONCEPTS

• Computing methodologies → Machine learning algorithms;

KEYWORDS

Independent cascade model; Active inference; Edge coloring; Approximation algorithms

ACM Reference Format:

Abhijin Adiga, Vanessa Cedeno-Mieles, Chris J. Kuhlman, Madhav V. Marathe, S. S. Ravi, Daniel J. Rosenkrantz, and Richard E. Stearns. 2018. Infering Probabilistic Contagion Models Over Networks Using Active Queries. In *The 27th ACM International Conference on Information and Knowledge Management (CIKM '18), October 22–26, 2018, Torino, Italy*. ACM, New York, NY, USA, Article 4, 10 pages. <https://doi.org/10.1145/3269206.3271790>

*Also with Escuela Superior Politécnica del Litoral (ESPOL), Guayaquil, ECUADOR.

†Also with University at Albany – SUNY.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CIKM '18, October 22–26, 2018, Torino, Italy

© 2018 Association for Computing Machinery.

ACM ISBN 978-1-4503-6014-2/18/10...\$15.00

<https://doi.org/10.1145/3269206.3271790>

1 INTRODUCTION

Background and Motivation. Due to the tremendous increase in the use of networked social systems (e.g., Facebook, Twitter, LinkedIn), researchers are actively studying various aspects of such systems. In particular, the study of contagion propagation in networked systems is an active area of research in many disciplines (e.g., computer science, social science, business, economics) since contagions can be used to model many different phenomena including disease spread, propagation of influence and social trends and flow of information (see e.g., [10, 29]). Many classes of diffusion phenomena over networks have been studied in the literature when the network and the associated parameters are known (e.g., [9, 10]). In actual social systems, many parameters of the network (e.g., behavior characteristics of nodes, transmission probabilities associated with edges) are not generally known. To understand diffusion phenomena over such networks, and subsequently apply the understanding to forecast, maximize influence, control the spread, etc., it is essential to have good estimates of model parameters. For example, for systems where the node behaviors can be captured by appropriate threshold functions, techniques for inferring those functions from media or other observational data have appeared in [2, 13, 26].

In this paper, our focus is on obtaining provably good estimates of the edge (or influence) probabilities of a given directed social network. Researchers have studied this problem under a model where observational data about the dynamics of the system (e.g., log of user activities, a time-ordered trace specifying the set of nodes influenced at each time step) is available (see e.g., [14, 27]). We consider the problem under an *active* query model, where a query specifies state values for the nodes and the response to the query provides the state of each node at the next time step. This active query model is appropriate for networked systems that arise in the context of online social experiments carried out under controlled settings (see e.g., [6, 18, 22]). We develop a precise formulation of the edge probability inference problem under the *independent cascade* (IC) model of diffusion. This diffusion model, which was first considered in the context of interacting particle systems [9, 19], has been widely used in the study of influence and disease propagation (see e.g., [12, 15]). To make our algorithms scale to large social

networks (with millions of nodes and about 200 million edges), we also formulate a problem whose goal is to find an appropriate subgraph of a large network so that the inference algorithm can be applied to the smaller subgraph. This formulation exploits the fact that in several application contexts, edges of a social network are partitioned into classes such that edges within the same class have the same (transmission) probability.

Summary of Results. Our results, summarized below, include provably good approximations of edge probabilities as well as experimental evaluations using several real-world and synthetic networks.

(1) For the IC model, we develop a precise formulation of the edge probability inference problem for a directed network under the active query framework.

(2) Given a directed network and values ϵ and δ , where $0 < \epsilon, \delta < 1$, we present an (ϵ, δ) -approximation algorithm to infer the edge probabilities of G for the IC model. Formally, our algorithm ensures that for every edge e , the probability that the estimated probability \hat{p}_e differs from the actual probability p_e by more than ϵp_e is at most δ . This approximation relies on two algorithmic ideas. First, it uses a stopping criterion for Monte Carlo sampling developed in [7]. Second, to minimize the number of queries used, it uses a novel edge coloring formulation (which we call fan-out edge coloring) for directed graphs.

(3) In practice, edge sets of large social networks are partitioned into classes such that all the edges in the same class have the same transmission probability. We rely on this idea to make our algorithms scale to very large social networks (with millions of nodes and hundreds of millions of edges). In particular, we formulate a combinatorial problem (called the **Minimum Cost Covering Subgraph** or MCCS problem) to identify a subgraph which has a small number of nodes and which contains at least one edge from each class. We show that this problem is **NP**-complete but present an approximation algorithm which provides a performance guarantee of $O(\sqrt{k})$, where k is the number of classes. This allows us to work with the smaller subgraphs generated by the approximation algorithm. It should be noted that our focus is on learning edge probabilities. We assume that a partition of the edges into subsets, where all edges in the same subset have the same probability, is available to our inference algorithms.

(4) We evaluate our algorithm for estimating edge probabilities on many real-world and synthetic networks. In a first set of experiments, we exploit edge labeling and use MCCS to reduce the sizes of large networks (in terms of numbers of nodes and edges) by several orders of magnitude, and infer edge probabilities. We evaluate a second set of intermediate sized-networks to address the case where there are no edge labels. Our assessments of these methods consider accuracy of the probability estimates, number of queries required to obtain these estimates and errors in contagion dynamics on networks under the IC model when using true and estimated probabilities.

Related Work. Many researchers have proposed formal models for contagion propagation in social networks (see e.g., [5, 10, 25]). This line of research generally assumes that all the parameters of the underlying network are known. Recently, there has been a considerable amount of interest in learning the parameters of

networked systems. For example, for systems where state changes of nodes are determined by threshold values of nodes (i.e., a node changes to state 1 only when at least a specified number of its neighbors are in state 1), many papers have addressed the problem of learning the node thresholds (see e.g., [2, 13, 26]). The problem of learning influence probabilities in networks has also received attention in the literature. For example, Goyal et al. [14] study the problem assuming that a log of users' actions is available. They develop algorithms for learning the edge probabilities under a variety of influence models. Saito et al. [27] consider the problem of estimating the edge probabilities for the IC model of diffusion. They assume that data in the form of a system trace which gives for each time instant t , the set of nodes which changed to state 1 at t is available. They use an algorithm based on expectation maximization to obtain estimates of edge probabilities. Liu et al. [20] address the probability inference problem for heterogeneous networks. Our work differs from the previous work in that we use an *active* query model (explained in Section 2) instead of observational data. An active query model in a different context (namely, determining users' choices) has been studied recently in [16]. Our query model enables us to obtain provably good estimates of edge probabilities. **Organization.** The remainder of this paper is organized as follows. In Section 2, we provide precise specifications of our active query model and the problem of inferring the edge probabilities under the IC model. In Section 3, we present our algorithm for the inference problem. To enable our algorithm to scale to very large social networks, we formulate the minimum covering subgraph problem, establish its complexity and present an efficient approximation algorithm for the problem in Section 4. We report results from our experiments (with and without finding a subgraph) in Section 5. Conclusions and directions for future work appear in Section 6. For space reasons, proofs of many results mentioned in the paper are omitted; they are available in [1].

2 MODEL DESCRIPTION AND PROBLEM FORMULATION

We assume that the underlying social network $G(V, E)$ is *directed*, with V and E denoting the vertex and edge sets respectively. An edge $e = (u, v)$, where u and v are the end points of e , is directed from u to v . In this case, u is an *in-neighbor* of v and v is an *out-neighbor* of u . For a node u , the indegree of u (denoted by $\deg_{\text{in}}(u)$) and outdegree of u (denoted by $\deg_{\text{out}}(u)$) are respectively the number of incoming and outgoing edges. The maximum indegree and outdegree of a graph are denoted by Δ_{in} and Δ_{out} respectively. Each directed edge is associated with a probability value; however, these probability values are not given and the goal is to obtain provably good estimates of those values.

We consider the independent cascade (IC) model of diffusion (defined below). We assume that every node has a state value from $\{0, 1\}$, where the state value 0 indicates that the node is "not infected" (or "not influenced") and 1 indicates that the node is "infected" (or "influenced"). In each time step and for each node v whose state value is 0, certain in-neighbors of v which are in state 1 try to influence v . This is a stochastic process whose nature is governed by the definition of the IC model (given below). A **configuration** of the network at time t is the binary tuple comprising

the state of every node in the network at time t . Given the configuration at time t , the diffusion process specified by the IC model determines a **successor** configuration at time $t + 1$. As the system is stochastic, the successor of a configuration need not be unique.

Independent Cascade (IC) model. Let $G(V, E)$ be a directed network where every edge $e \in E$ is associated with a (transmission or influence) probability $p_e > 0$. As mentioned earlier, each node may be in state 1 (influenced/infected) or state 0. At time t , a node v in state 0 is influenced independently by each in-neighbor u which changed to state 1 at time $t - 1$ with influence probability $p_{(u,v)}$. Subsequently, if the state of v changes to 1, then v influences its state 0 out-neighbors for exactly one time step, and never changes its state to 0.

Active query model. To estimate the edge probability values, our query model assumes an active form of interaction with the system. This type of query model has been studied in the literature in several contexts, including determining node behaviors [2] and inferring users' choices [16]. In our model, the user issues a query q that specifies a system configuration (i.e., the tuple of state values of all the nodes of the system) at a certain time instant; the response to the query is a successor configuration of q determined by the underlying stochastic process of the IC model. Formally, given a network G , the active query model corresponds to a *query function* $Q_G^{\text{IC}}(\cdot)$, which takes a configuration or *query* q as input and returns a successor $Q_G^{\text{IC}}(q)$ configuration of q as output. As the successor of a given configuration q is, in general, not unique, the system may return any successor of q that is consistent with the underlying stochastic process. Since generating responses to queries can be expensive, it is important to minimize the number of queries used. For a query q and a node v , we use $q(v)$ to denote the 0 or 1 value assigned to v by q . We can now present a precise definition of the problem of inferring edge probabilities under the IC model.

PROBLEM 2.1 (INFERIC). *Given a directed network $G(V, E)$ and a query function Q_G^{IC} corresponding to an IC model over G , infer the influence probability p_e for every edge $e \in E$ using a minimum number of queries.*

3 RESULTS FOR THE IC MODEL

Overview. We first propose an (ϵ, δ) -approximation algorithm for the INFERIC problem. The algorithm gives the following guarantee: with probability at least $1 - \delta$, the estimated influence probability \hat{p}_e is at most ϵp_e away from the actual probability p_e for every edge e . Next, we discuss how our algorithm compares with an optimal solution with respect to the number of queries used. We then consider the special case of a *homogeneous* IC model where all edges have the same influence probability.

Our approach. Consider a directed edge $e = (u, v)$. If u is in state 1, then it may influence v with probability p_e . To estimate p_e , we a design query q in such a manner that $q(u) = 1$, $q(v) = 0$ and u is the only in-neighbor of v that is in state 1. Suppose $q' = Q_G^{\text{IC}}(q)$; that is, q' is the successor of q returned by the system. Then, $q'(v)$ is a 0 – 1 random variable with $\Pr(q'(v) = 1) = p_e$. When query q is repeated N times, we obtain in N independent samples of $q'(v)$

which can be used to estimate p_e . To determine a suitable value of N , we employ the stopping criterion discussed in [7].

To minimize the number of queries used, we utilize a specific edge coloring scheme which enables us to simultaneously estimate the influence probabilities of a group of edges. Suppose each edge in E is assigned a color from $\{1, 2, \dots, \tau\}$ for some positive integer τ . (The coloring must satisfy certain conditions which will be discussed later in this section.) This coloring induces a partition $\{E_1, E_2, \dots, E_\tau\}$ of E , where E_i is the subset of edges assigned color i , $1 \leq i \leq \tau$. Then, query q_i is constructed as follows: for every $(u, v) \in E_i$, set $q(u) = 1$ and the rest of the vertices of the graph to 0. We repeatedly try these τ queries until the stopping criterion is satisfied for each edge in that query, thus ensuring the accuracy of the estimate for every edge.

Stopping rule. Suppose Z_1, Z_2, \dots are i.i.d. random variables distributed according to Z in the interval $[0, 1]$ with mean p . Dagum et al. [7] proposed an (ϵ, δ) -approximation algorithm, which we refer to as STOPPINGRULE (Algorithm 1), to estimate the value p with a near-optimal number of samples. Steps 2 and 9 of this Algorithm use the function $T(x, y)$ defined by

$$T(x, y) = 4(e - 2) \log(2/y)/x^2. \quad (1)$$

The following theorem from [7] shows the approximation quality of the estimate and the number of samples used.

Algorithm 1: STOPPINGRULE($\epsilon, \delta, \{Z_1, Z_2, \dots\}$) of [7]

Data: Values $\epsilon, \delta \in (0, 1)$ and i.i.d. random variables Z_1, Z_2, \dots according to Z in the interval $[0, 1]$ with mean p
Result: Estimate \hat{p} of p

```

1 /*Step 1*/
2 Let  $\epsilon' = \min(1/2, \sqrt{\epsilon})$ ,  $T_1 = 1 + (1 + \epsilon')T(\epsilon', \delta/3)$ ;
3 Initialize  $N = 0, S = 0$ ;
4 while  $S < T_1$  do
5   |  $N = N + 1, S = S + Z_N$ ;
6 end
7  $p' = T_1/N$ ;
8 /*Step 2*/
9 Let  $T_2 = 2(1 + \sqrt{\epsilon})(1 + 2\sqrt{\epsilon})(1 + \log(3/2)/\log(2/\delta))T(\epsilon, \delta)$ ;
10 Let  $N' = T_2\epsilon'/p'$  and  $S = 0$ ;
11 for  $i = 1, \dots, N'$  do
12   |  $S = S + (Z_{N+2i-1} - Z_{N+2i})^2/2$ ;
13 end
14  $\rho' = \max(S/N', \epsilon p')$ ;
15 /*Step 3*/
16 Let  $N'' = T_2\rho'^2/p'^2$  and  $S = 0$ ;
17 for  $i = 1, \dots, N''$  do
18   |  $S = S + Z_{N+N'+i}$ ;
19 end
20  $\hat{p} = S/N''$ ;

```

THEOREM 3.1. *Let Z be any random variable distributed in $[0, 1]$. Let $\mu_Z = E[Z] > 0$ be the mean of Z , σ_Z^2 be the variance of Z*

and $\rho_Z = \max\{\sigma_Z^2, \epsilon\mu_Z\}$. Let $\tilde{\mu}_Z$ be the approximation produced by Algorithm 1 and let N_Z be the number of experiments run by the algorithm with respect to Z for input parameters ϵ and δ . Then (i) $\Pr[\mu_Z(1 - \epsilon) \leq \tilde{\mu}_Z \leq \mu_Z(1 + \epsilon)] \geq 1 - \delta$ and (ii) there is a universal constant c' such that $E[N_Z] \leq c' T(\epsilon, \delta) \rho_Z / \mu_Z^2$. \square

From the above theorem, it can be seen that the estimate \hat{p} produced by Algorithm 1 satisfies $\Pr[|\hat{p} - p| > \epsilon p] \leq \delta$; further, the number of samples used is within a constant factor of the minimum number of samples required in expectation.

Fan-out edge coloring. The goal is to obtain a partition of the edge set E into $\{E_1, E_2, \dots, E_\tau\}$ (for some τ), where E_i has all the edges with color i ($1 \leq i \leq \tau$). Further, each subset E_i in the partition must satisfy the following condition: for any $e = (u, v) \in E_i$, v does not have any other incoming edge with color i and v does not have any outgoing edge with color i . In such a coloring, any color class induces a graph which is a collection of stars with edges “fanning out” from a central vertex. Hence, we refer to this as a “fan-out edge coloring”. Once we have such a coloring, the key idea is that one query is sufficient to estimate the probability for all the edges in the same color class. For each color class E_i ($1 \leq i \leq \tau$), such a query q_i can be constructed as follows. Let $Y_i \subseteq V$ be the subset of nodes such that for each node $u \in Y_i$, there is an outgoing edge $(u, v) \in E_i$. For each node $u \in Y_i$, we set $q_i(u) = 1$; for all other nodes $w \in V - Y_i$, we set $q(w) = 0$. An example of fan-out edge coloring and the construction of queries from the coloring are presented in Figure 1.

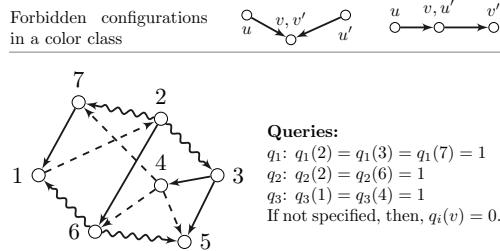


Figure 1: The constraints for fan-out edge coloring, an example of valid coloring and the resulting queries.

We now describe a method to realize a fanout edge coloring. We first construct an undirected graph \mathcal{E}_G from G as follows. The vertex set of \mathcal{E}_G is in one-to-one correspondence with E , the edge set of G . For two edges $e = (u, v)$ and $e' = (u', v')$ in E , their corresponding vertices in \mathcal{E}_G are adjacent iff either (i) $v = v'$ or (ii) $v = u'$ (see Figure 1). Thus, in \mathcal{E}_G , two nodes are adjacent iff the corresponding edges in E cannot be assigned the same color. Hence, any proper vertex coloring of \mathcal{E}_G corresponds to an edge coloring of G that satisfies the conditions for a valid fan-out edge coloring mentioned above. We note that \mathcal{E}_G is a variant of the well-known *line graph* defined for an undirected graph. Further, a simple greedy coloring strategy based on Brooks's theorem [31] can be used to color \mathcal{E}_G with at most $\Delta(\mathcal{E}_G) + 1$ colors, where $\Delta(\mathcal{E}_G) = \max_{e=(u,v)\in E} \deg_{in}(u) + \deg_{in}(v) + \deg_{out}(v) - 1$.

Our algorithm APPROXINFERIC, which uses Algorithm 1, is described in Algorithm 2. The algorithm first constructs a fan-out

edge coloring of E as discussed above. It then constructs a query q_i for each color class i and repeatedly obtains a successor of q_i . The number of repetitions is determined using Algorithm 1. We can now show that Algorithm 2 indeed provides a provably good estimate of the true probability for each edge.

Algorithm 2: APPROXINFERIC($G, Q_G^{IC}, \epsilon, \delta$)

Data: Directed graph $G(V, E)$, query function Q_G^{IC} which returns a successor of the input query q and values $\epsilon, \delta \in (0, 1)$.

Result: For every $e \in E$, an estimate \hat{p}_e of the influence probability p_e

- 1 Construct a fan-out edge coloring $\{E_1, \dots, E_\tau\}$ of E ;
- 2 **for** $i = 1$ to τ **do**
- 3 /*Construct query q_i */
- 4 $q_i = 0$; /*first set every vertex state to 0 in q_i */
- 5 **for** $e = (u, v) \in E_i$ **do**
- 6 $| q_i(u) = 1$;
- 7 **end**
- 8 /*Query until stopping criterion is satisfied*/
- 9 Initialize $N = 0$, $\forall e \in E_i, S_e = \emptyset$;
- 10 **while** $\forall e \in E_i, STOPPINGRULE(\epsilon, \delta / |E|, S_e)$ has not terminated **do**
- 11 $q'_N = Q_G^{IC}(q_i)$;
- 12 $\forall e = (u, v) \in E_i, S_e = S_e \cup q'_N(v)$;
- 13 $N = N + 1$;
- 14 **end**
- 15 **end**
- 16 $\forall e \in E, \hat{p}_e$ is the output of $STOPPINGRULE(\epsilon, \delta / |E|, S_e)$.

THEOREM 3.2. Let $G(V, E)$ be a directed graph, \mathcal{E} be a fan-out edge coloring of G with τ colors and $\epsilon, \delta \in (0, 1)$. For any independent cascade model defined over G , under the active query model, APPROXINFERIC estimates the influence probabilities with the following guarantees: (i) $\Pr(\forall e \in E, |\hat{p}_e - p_e| < \epsilon p_e) \geq 1 - \delta$, where for an edge e , p_e and \hat{p}_e are the actual and estimated influence probabilities respectively. (ii) The expected number of queries is at most $c\tau T(\epsilon, \delta / |E|) \rho / p_{min}$ where c is a positive constant, $p_{min} = \min_{e \in E} \{p_e\}$ and $\rho = \max(p_{min}, \epsilon)$.

PROOF. We recall that each $p_e > 0$. First, we show that the algorithm satisfies the following condition for each edge e .

$$\Pr(|\hat{p}_e - p_e| > \epsilon p_e) < \delta / |E|. \quad (2)$$

Let $e = (u, v)$, and let $e \in E_i$. By definition, $q_i(u) = 1$ and $q_i(v) = 0$. Let $q' = Q_G^{IC}(q_i)$. By the definition of the fan-out coloring, e is the only incoming edge of v in E_i , and therefore, u is the only neighbor of v whose state is 1 in q_i . Hence, $q'(v) = 1$ with probability p_e and 0 with probability $1 - p_e$. Note that the repetitions of q_i are independent of each other. Therefore, by Theorem 3.1, the estimate of $STOPPINGRULE(\epsilon, \delta / |E|, S_e)$ satisfies Equation (2). We now use

the union bound [23] to prove Part (i) of Theorem 3.2.

$$\begin{aligned} \Pr(\exists e \in E, |\hat{p}_e - p_e| > \epsilon p_e) &\leq \sum_{e \in E} \Pr(|\hat{p}_e - p_e| > \epsilon p_e) \\ &\leq \sum_{e \in E} \delta / |E| = \delta. \end{aligned}$$

Now we prove Part (ii) of the theorem. From Part (ii) of Theorem 3.1, for a random variable Z taking values in $[0, 1]$, with mean p and variance σ^2 , the expected number of queries required by the STOPPING-RULE for (ϵ', δ') is at most $f(p, \epsilon', \delta') = cT(\epsilon', \delta') \max(\sigma^2, \epsilon p)/p^2$. Since, in our case, Z is a $0 - 1$ random variable, it follows that $\sigma^2 = p(1 - p)$. Also, it can be verified that when $p_1 > p_2$, for any ϵ and δ satisfying $0 < \epsilon, \delta \leq 1$, $f(p_1, \epsilon, \delta) < f(p_2, \epsilon, \delta)$; that is, the edge that requires the most number of repetitions is the one with the minimum influence probability p_{\min} . Therefore, the number of times a query must be repeated is at most $cT(\epsilon, \delta/|E|)\rho/p_{\min}$ in expectation, for some constant $c > 0$. Since there are at most τ distinct queries, the expected total number of queries used is at most $c\tau T(\epsilon, \delta/|E|)\rho/p_{\min}$. \square

Bounds on the optimal number of queries. Dagum et al. [7] show that for a single random variable, the expected number of samples used by the stopping rule algorithm (Algorithm 1) is within a constant factor of the optimum expected value. For the INFERIC problem, there are two factors that influence the number of queries required: (i) network structure which influences the total number of distinct configurations that can be used for querying, and (ii) the distribution of edge probabilities that determines how many times each query is repeated. The interplay between these two factors makes it challenging to obtain good bounds on the number of queries. Theorem 3.2 gives an upper bound for the optimal number of queries.

For the lower bound, under the assumption that an algorithm must infer the probability of each edge independently, the number of distinct queries required is at least Δ_{in} , the maximum in-degree of the network. This is because, a vertex with in-degree $= \Delta_{\text{in}}$ has that many incoming edges which must be evaluated in distinct queries. Suppose p_{\max} is the maximum edge probability in the IC model. Then, every query must be repeated at least $c \Delta_{\text{in}} T(\epsilon, \delta/|E|)\rho/p_{\max}$ times for some constant c , where T is the function defined in Equation (1).

Uniform probability. When all the edges have the same probability p , the number of queries can be substantially reduced. Firstly, we need not consider all the color classes. It is enough to just use a color class with the maximum number of edges. Secondly, in every query, we obtain s samples of the same random variable, where s is the number of edges in the corresponding color class. Therefore, using a proof similar to that of Theorem 3.2, the expected number of queries used can be seen to be at most $\frac{cT(\epsilon, \delta/|E|)\rho}{p|E_{\max}|}$, where E_{\max} is a color class with the maximum number of edges. If the number of colors is τ , note that $|E_{\max}| \geq |E|/\tau$. Hence, the number of queries is at most $\frac{c\tau T(\epsilon, \delta/|E|)\rho}{p|E|}$, which is $1/|E|$ times the upper bound for the non-uniform probabilities case.

4 MINIMUM COVERING SUBGRAPH PROBLEM

4.1 Motivation for Considering Subgraphs

While the algorithm discussed in the previous section provides a good performance guarantee, it is difficult to use it directly with very large social networks with several million nodes and about 200 million edges. The reason is that with a large number of edges and the number of times a query must be repeated (for each edge) to obtain the desired performance guarantees (in terms of the parameters ϵ and δ), the required number of queries to be tried is prohibitively large. (One can get an idea of this from the experimental results in Section 5.3 for smaller graphs for which the number of edges varies from about 84,000 to about 940,000.) To ensure scalability of our algorithm, we exploit a feature of social networks that is commonly present in practical epidemic and social simulations. In these simulations, the edges of the social network are partitioned into a certain number of classes, and all the edges within the same class have the same (transmission) probability value. Such partitioning schemes rely on the fact that each edge in the social network represents a type of interaction among the two individuals corresponding to the end points of the edge, and each interaction type can be modeled by a single probability value. A good discussion on why interactions and degrees of influence can be grouped and modeled in this manner appears in [8, 24, 28]. The interaction types are generally based on node and edge attributes (e.g., ages of the individuals, the duration of interactions), and the number of types of interactions is much less than the number of edges in the network. Thus, instead of considering a very large social network, one can consider a subgraph which has at least one edge for each type of interaction. The inference algorithm under the IC model from the previous section can be applied on the subgraph and the results can be translated to the larger social network. In Section 5, we describe a scheme that we used in our experiments to partition the edge set into classes. In the next subsection, we provide a precise formulation of the subgraph problem, establish its complexity and present a provably good approximation algorithm for the problem.

4.2 Problem Definition and Results

As mentioned above, we want to find a subgraph G' of a given social network G so that G' contains all the edge types that are in the larger network. To make the problem formulation and analysis easy to understand, we will define this problem for undirected graphs. (The directions of the edges don't play a role in deciding which edges are chosen.) In our experiments, however, we used directed graphs. A general version this problem can be formulated as follows.

Minimum Cost Covering Subgraph (MCCS)

Instance: An undirected graph $G(V, E)$, a nonnegative cost $c(v)$ for each node $v \in V$, a partition of the edge set E into k classes E_1, E_2, \dots, E_k and a budget $B \leq \sum_{v \in V} c(v)$.

Question: Is there a subset $V' \subseteq V$ such that the total cost of the nodes in V' is at most B and the subgraph $G'(V', E')$ induced on V' contains at least one edge from each class?

The following result points out that MCCS is unlikely to be solvable efficiently.

THEOREM 4.1. *MCCS is NP-complete even when the underlying graph is bipartite and the cost of each node is 1.*

PROOF. (Idea) We prove the NP-hardness through a reduction from Minimum Set Cover (MSC) problem which is known to be NP-complete [11]. The details appear in [1]. \square

Theorem 4.1 shows that the MCCS problem is NP-hard even when all the nodes have a cost of 1; that is, the cost of the subgraph G' is the number of nodes in G' . For this version, we now present an approximation algorithm which provides a performance guarantee of $O(\sqrt{k})$, where k is the number of edge classes. In other words, the number of nodes in the subgraph chosen by APPROX-MCCS is always within the factor $O(\sqrt{k})$ of the number of nodes in an optimal subgraph. The details of this approximation algorithm, which we refer to as APPROX-MCCS, appear in Figure 3. The following theorem, whose proof appears in [1], shows the worst-case performance guarantee provided by APPROX-MCCS.

Algorithm 3: Approximation Algorithm for MCCS

```

Input : An undirected graph  $G(V, E)$ , a partition of the
         edge set  $E$  into  $k$  classes  $E_1, E_2, \dots, E_k$ .
Output: A subgraph  $G'(V', E')$  of  $G$  such that  $E'$  contains
         at least one edge from each class  $E_i$  ( $1 \leq i \leq k$ ) and
         the number of nodes in  $V'$  is as small as possible.

1 Let  $C = V' = \emptyset$ . (Note:  $C$  is the set of classes of edges from
   which at least one edge has been chosen and  $V'$  is the
   subset of nodes in the resulting subgraph  $G'$ .)
2 while ( $|C| < k$ ) do
3   Find a node  $v$  such that among all the nodes in  $V$ , the
      set of edges incident on  $v$  has the largest number of
      new classes which are not in  $C$ .
4   For each new class of edges, choose one edge from that
      class incident on  $v$  and add it to  $E'$ .
5   Update  $C$  by adding the new classes of edges chosen in
      Step 4.
6   Add  $v$  and the other end points of the edges chosen in
      Step 4 to  $V'$ .
7 end
8 Output the subgraph  $G'$  of  $G$  induced on  $V'$ .

```

THEOREM 4.2. *For any instance of MCCS given by a graph $G(V, E)$ where the number of classes into which the edge set E has partitioned is k and the cost of each node is 1, Algorithm APPROX-MCCS provides a performance guarantee of $O(\sqrt{k})$.* \square

5 EXPERIMENTAL RESULTS

5.1 Overview

Set 1 experiments. We report on two sets of experiments. In the first set, we start with larger synthetic populations, and generate labeled *directed* networks from them with 10^5 to 10^6 nodes and 10^6 to 10^8 edges (see Table 1). We then specify an edge labeling process and use Algorithm 3 to produce much smaller subgraphs that preserve the edge labels from the original graphs. These operations

are summarized in the top row of Figure 2, and will be explained below. Each edge label corresponds to an edge probability. We use Algorithm 2 to compute fanout edge colorings and estimated edge probabilities on these subgraphs, and map the results back to the larger graphs. These operations are summarized in the second and third rows of Figure 2, and will be explained below. This entire process demonstrates how to take large labeled networks, generate smaller networks, analyze them, and map the results back to the original large network. That is, this process demonstrates a *scalable* approach to inferring edge probabilities under the IC model.

Table 1: Networks used in our experiments and their properties. Set 1 networks are in rows 1 through 3. Set 2 networks are in rows 4 through 8. To conserve space, we have provided ranges of values for some network families in Set 2. Fan-out coloring is not applicable (N/A) to Set 1 networks because we perform the coloring on subgraphs.

Network	Properties			
	n	$ E $	$\Delta_{\text{in}}, \Delta_{\text{out}}$	Fanout(τ)
NRV	152,661	8,301,322	772	N/A
Miami	2,165,398	108,618,252	846	N/A
Seattle	3,405,279	197,374,320	892	N/A
Enron	33,696	361,622	1,383	1,393
Epinions	75,877	811,478	3,044	3,050
Slashdot0811	77,360	938,360	2,539	2,540
Erdős-Rényi (5)	70,000	$\approx 84K$	29,28-31	30-34
Barabási- Albert (5)	70,000	839928	879-	880-
			1273	1279

Set 2 experiments. In the second set of experiments, we use unlabeled networks; these networks are in rows 4 through 8 of Table 1. Because there are no explicit labels, each edge is treated as possessing a unique label. Hence, we operate on the original networks; consequently, we omit the operations in the first row of Figure 2 and execute the procedures in the second and third rows. While these networks are much smaller than the networks of Set 1, they are bigger than the subgraphs that we operate on in Set 1, and hence represent a different kind of scalability assessment of Algorithm 2.

Types of experimental analyses. In both sets of experiments, we evaluate Algorithm 2 using two performance measures: (i) the quality of the inferred model and (ii) the total number of queries used to obtain the model. Further, the quality in (i) is assessed in two ways: (a) how close the estimated probabilities are to the actual model and (b) how IC (contagion) dynamics on networks compare when using true versus estimated edge probabilities. Due to the stochastic nature of Algorithm 2, for each combination of IC model, ϵ and δ , we obtained 10 estimates of the IC model probabilities. All networks in this study are taken as directed.

5.2 Large Graph Experimental Results

The starting point for this work is step 1 of Figure 2, and we progress through step 5 in generating and evaluating different networks.

Labeled network G generation. The networks in the first three rows of Table 1 are analyzed in this section. We explain the network generation process, which is the first row of Figure 2. These

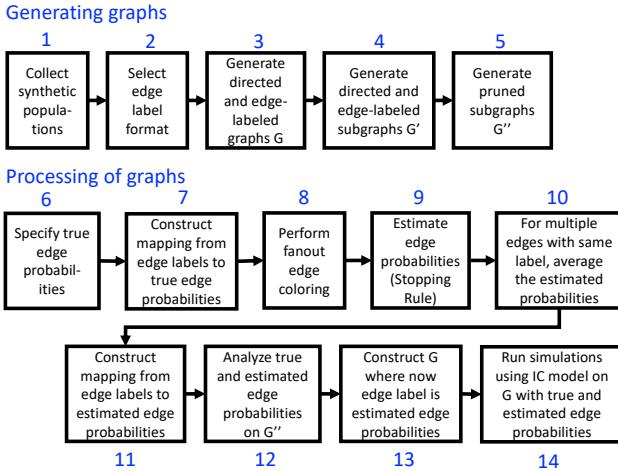


Figure 2: Workflow for experiments per graph of Table 1. The first five steps produce the directed graphs G , G' and G'' from synthetic populations, and are applicable to the first set of experiments. The last nine steps operate on graphs to infer model properties, compare true and estimated IC models (i.e., edge probabilities), and compare dynamics on networks through simulation and are applied to the networks in both sets of experiments. This workflow uses the parameter values in Table 2.

networks G , which have been used to make epidemic and economic assessments to inform public policy [4, 21], are constructed from synthetic human populations that have per-person attributes such as age, gender, family composition, and home location, and a set of daily activities (e.g., go to work, go to school). See [3] for population construction details.

To construct edge labels, we are guided by the influence literature, where age and context affect how individuals influence each other. For example, the degree of influence on a person by others changes between ages 10 to 30 [24, 28], although there are clearly additional complicating factors [8]. Furthermore, a person's interactions with others vary with context [30], which we represent through activity types of individuals' activities. We first bin the ages of people in the population in 10-year increments per row 2 of Table 2. Then, people in a population can have any number of activities of the six types, as specified in the third row of Table 2. A network G is induced on the population in the following way: there is an undirected edge $\{n_1, n_2\}$ between humans n_1 and n_2 if they are co-located (in the same room of the same building during an activity) and the times of their visits to the location overlap. Each undirected edge produces two directed edges. For our purposes, we specify a 4-tuple as an edge label for the directed edge (n_1, n_2) ; this is shown in the fourth row of Table 2. The set of all possible edge labels is denoted by L , and we let $n_L = |L|$. Our methods are agnostic to the particular form of an edge label.

Unpruned subgraph G' generation. For each network, we construct a subgraph according to Algorithm 3. We call this the *unpruned* subgraph, designated G' . The intuition behind this subgraph is as follows. Each edge label in L in the original graph represents a corresponding edge probability in the IC model. Our goal is to determine these probabilities at a minimal cost. To do this, we identify a

Table 2: Summary of the parameters and their values used in the experiments. These are used in conjunction with Figure 2.

Parameter	Description
Networks	Networks (graphs and subgraphs) in Table 1.
G .	
Age bins.	Nodes in labeled graphs are humans. Ages are binned [0,9], [10,19], [20,29], . . . , [80,89], 90+. This gives 10 age bins.
Activity types.	Humans can have six different activity types: home, work, school, college, shopping, other [3].
Edge label.	For directed edge (n_1, n_2) , from node n_1 to node n_2 , the edge label is the 4-tuple (activity type of n_1 , activity type of n_2 , age bin of n_1 , age bin of n_2).
True edge probabilities.	These are edge probabilities for the IC model. Set 1 experiments: p_e values assigned from {0.1, 0.2, 0.3, . . . , 0.8, 0.9}. Set 2 experiments: {0.1, 0.2, . . . , 0.5}.
Estimated edge probabilities.	These are edge probabilities for the IC model produced by Algorithm 2.
True instance, t_i .	For each network G' and G'' , there are five independent probability assignments made to edges, called true probabilities, to account for stochasticity. For each edge label, a true probability p_e is assigned uniformly at random, for each instance. Values for instances are labeled 0 through 4.
Epsilon, ϵ .	Used in Algorithm 2 that estimates edge probabilities; three values are used: $\epsilon = 0.1, 0.3, 0.5$.
Delta, δ .	Used in Algorithm 2 that estimates edge probabilities; three values are used: $\delta = 0.1, 0.3, 0.5$.
Estimated solutions, t_e .	For each assignment of true edge probabilities, ten estimated solutions are computed according to Algorithm 2 to account for stochasticity. We average results over all ten instances.

small subset of nodes of G such that among *all* of the edges between pairs of these nodes, there is at least one edge with each label from L . This enables us to determine the probability corresponding to each edge type. For simplicity, we assume the cost of each node to be 1 so that the goal is to produce a subgraph with a small number of nodes. Also, for the specified nodes, we cannot intervene into the system; e.g., we cannot remove edges (interactions) between pairs of nodes, even when these edges have redundant labels, because we must not disrupt the interactions among people. This is called the **non-intervention** condition. The fanout coloring is used precisely to deal with this constraint, providing a method for specifying system configurations to query.

There are several additions to the implementation of Algorithm 3 in producing G' from G . First, both G and G' are directed. Second, when selecting node v , after the first criterion of selecting a node incident on the greatest number of labels yet to be covered, we prefer out-edges from v rather than in-edges because this can enable more efficient queries: all out-edges from v can be evaluated simultaneously by setting all of the corresponding in-nodes on these edges to state 0. We break ties by selecting a node with the minimum number of edges. The last criterion is to minimize the number of additional edges introduced into G' (that may have redundant labels). After selecting a node v and its neighbors that contribute new edge labels to G' , all edges among these nodes, and between these nodes to the existing nodes of G' are formed. This is to satisfy the criterion to not intervene in the system G' by removing interactions.

Pruned subgraph G'' generation. There is one more step. We may be able to prune edges from G' . A node is **critical** if it is incident on an edge that is the only representative of a label in G' . If a node is only critical as an in-node (i.e., as v in directed edge (u, v)), then any edge for which v is an out-node (i.e., v in (v, u)) can be deleted, because it is redundant. Similarly, for nodes that are only critical as out-nodes, all incoming edges can be removed from G' . The insight for critical out-nodes is as follows. These nodes will be set to state 1 in some queries to determine their effect on their corresponding in-neighbors, to infer those edge probabilities. They never need to be set to state 0 and used to determine the probability on an incoming edge because there is another edge in G' that can be used to infer this probability. The same intuition applies for in-nodes. In this way, we prune G' , producing the pruned graph G'' . This process only removes edges; no nodes are removed. Also, this process does not violate our non-intervention condition above: these edges are removed because we will never set the out-nodes of these edges to state 1 in any query, so that they will not be used to infer edge probabilities. A similar argument holds for in-nodes.

Results on generating labeled subgraphs. Table 3 shows our results on generating subgraphs G' and G'' . The number n_ℓ of labels in G for each network is shown in the rightmost column of that table. The numbers of nodes in these networks are far less (about $5\times$ less) than $2n_\ell \approx 900$, which is the number of nodes required if n_ℓ edges with unique labels formed a perfect matching. Assuming unit cost per node in Algorithm 3, this is a roughly $5\times$ cost savings. Further, the pruning process reduced the number of edges in G'' compared that of G' , by about 17%. In turn, this will reduce the number of queries for inferring edge probabilities.

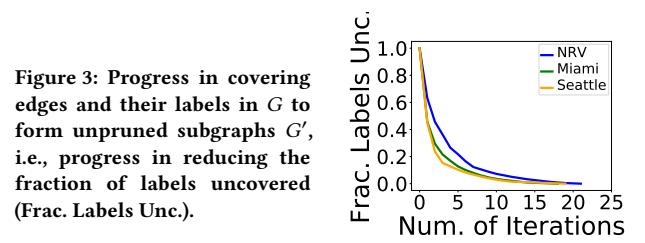
Table 3: Results on subgraphs G' and G'' generated with Algorithm 3 based on the labeling of the directed networks in Set 1 of Table 1. The subgraphs are 10^4 to 10^6 times smaller than the original graphs, in terms of numbers of edges. The number of edges in G'' is further reduced by pruning edges in G' . The numbers of edge labels n_ℓ are the same for the original graphs.

Network	Properties					
	n	$ E' $, $ E'' $	$\frac{ E' }{n_\ell}$, $\frac{ E'' }{n_\ell}$	Δ', Δ''	Fanout', n_ℓ	Fanout''
NRV	193	792, 659	1.74, 1.45	70, 66	36, 32	456
Miami	172	872, 731	1.82, 1.53	82, 78	42, 38	479
Seattle	165	850, 698	1.78, 1.46	86, 79	44, 37	477

The performance of Algorithm 3 is shown in Figure 3. This figure shows that for all three networks, all n_ℓ edge labels can be covered by selecting nodes in about 20 iterations of the while loop in Algorithm 3 to produce G' . The first couple of iterations reduce the number of uncovered edge labels by about 50%, with asymptotic progress thereafter.

Note that the results in Table 3 and Figure 3 are for evaluating the top row in Figure 2. We now turn to evaluating the IC model inference, the second and third rows of Figure 2.

Results on comparing true p_e and estimated \hat{p}_e edge probabilities in IC model. This experiment covers steps 6 through 12 of Figure 2. The parameters evaluated begin with the *True edge probabilities* in Table 2. A true edge probability in $\{0.1, 0.2, \dots, 0.9\}$



is assigned to each edge label in G' and G'' . Fanout edge coloring is performed on G' and G'' , and Algorithm 2 is used to estimate edge probabilities for the five different true edge probability instances t_i , and for all nine combinations of (ϵ, δ) . Because Algorithm 2 is stochastic, we generate 10 estimated edge probability \hat{p}_e solutions for each combination of $(G, t_i, \epsilon, \delta)$ of Table 2. Note that this algorithm is edge-label unaware, so that each edge's probability is estimated independently. Thus, while there is a unique mapping from edge label to p_e , there is no unique mapping from edge label to \hat{p}_e . To create this latter mapping, we average the computed \hat{p}_e values on edges that have the same label. We then use these mappings to produce edge probabilities $(p_e \text{ and } \hat{p}_e)$ for all edges in the first three original networks G of Table 1. We will discuss the larger (original) networks below momentarily. Now, we provide results for the comparisons of p_e and \hat{p}_e , and for numbers of queries.

Figure 4 provides average absolute errors, computed by averaging values of $|p_e - \hat{p}_e|$ across all edges, as a function of ϵ . These data points, represented as squares, have values around 0.02. That is, the average error between true and estimated probabilities is quite small, considering that $p_e \in [0.1, 0.9]$. However, the maximum error for an edge can be large—roughly 0.8 in this figure. These results indicate that while the maximum error in estimated probability can be large, most probability estimates are in very good agreement with the true probabilities, because the average absolute error is quite small. We consider probabilities on G'' (and G') because these are the graphs on which the edge probabilities are estimated.

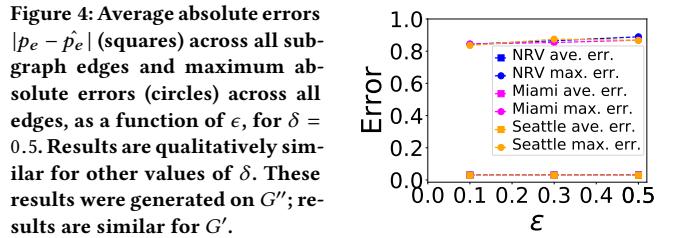


Figure 5 shows for G'' of Miami and Seattle networks the number of queries used to achieve the level of accuracy in \hat{p}_e conveyed by ϵ and δ . We note that as ϵ increases the numbers of queries decrease, and for a fixed ϵ , the numbers of queries decrease as δ increases. Results for G'' of NRV are similar. These results make for interesting comparisons with the graphs in Set 2 as discussed in Section 5.3.

Results on comparing simulations on networks using true p_e and estimated \hat{p}_e edge probabilities: a usage scenario. To evaluate the effects of p_e and \hat{p}_e on *population dynamics*, we run simulations on the large NRV, Miami, and Seattle networks of Table 1. We used the following combinations of (ϵ, δ) values: $(0.1, 0.1)$,

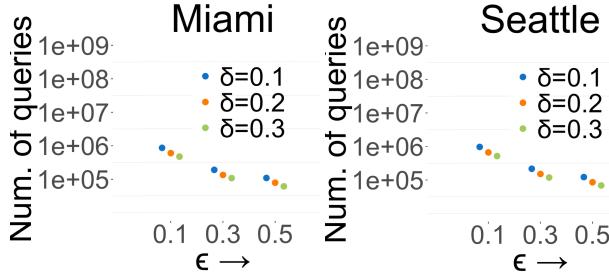


Figure 5: The number of queries required to compute edge probability estimates. The numbers of queries decrease as ϵ and δ increase. The results for NRV are similar to those for Miami and Seattle.

(0.3, 0.3), and (0.5, 0.5). For each (G, p, ϵ, δ) 4-tuple (where p indicates whether true p_e or estimated \hat{p}_e probabilities are used), we run 100 diffusion instances; the initial condition for each run is that one randomly-chosen node is in state 1 and all others are in state 0. The seed node for run j ($1 \leq j \leq 100$) is the same for all (G, p, ϵ, δ) .

Results are shown in Figure 6 for the Miami network. Figure 6(a) shows the cumulative fraction of nodes in the network that reach state 1 as a function of time. The three curves that each use estimated edge probabilities are in excellent agreement with the curve generated with true probabilities; the curves overlap. In Figure 6(b), we assess the fraction of nodes changing to state 1 at each time step, and compare these values from the simulations using estimated probabilities to those generated in simulations with the true probabilities. The errors are small—about 1% or less in the predicted fraction of newly infected nodes as a function of time—with the largest errors corresponding to the greatest spreading rate of the contagion. The plots for the other two networks are similar. This use case illustrates how inferred threshold systems can be used to produce real-world predictions of contagion dynamics.

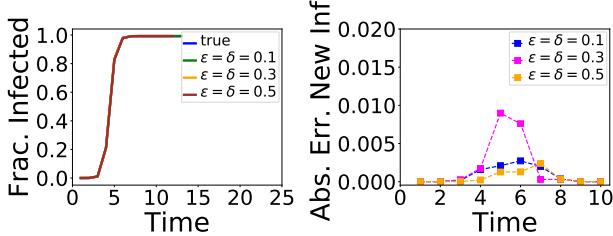


Figure 6: (a) Comparison of agent-based simulation results, using the IC model per agent, for true edge probabilities, versus estimated edge probabilities for the (ϵ, δ) combinations in the legend. It shows the cumulative fraction of nodes (agents) in the Miami network that are infected/activated as a function of time. It can be seen that \hat{p}_e values produce contagion dynamics results similar to those produced by p_e . Each curve is the average of 100 simulation instances. Variances at each time, for the 100 simulations, are very small and are not shown for clarity. The largest variance over all data is 0.028 (at $t = 6$); the great majority of variance values are $< 10^{-4}$. (b) Plot of absolute error in number of new infections/activations per time unit; hence, these data are *temporal* errors in simulation results. These errors are small: less than 1% in the fraction of infected/activated nodes.

5.3 Small Graph Experimental Results

The experiments were performed on two sets of synthetic networks and three real-world networks [17], all of comparable size. They

are listed in the last five rows of Table 1. Among the synthetic networks, one set consists of five replicates of directed Erdős-Rényi graphs. The other set has five replicates of the Barabási-Albert graph; each replicate was obtained by first constructing an undirected graph with 70K nodes and average degree 6, and then replacing each edge with a pair of bidirectional edges. We used the same approach to obtain directed versions of the real-world networks. For space reasons, we present representative results for selected networks. Other networks exhibit similar behavior. For our experiments, each IC model corresponding to a network was obtained by drawing the edge probabilities uniformly at random from the discrete set $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. For each network, we considered five replicates of the IC model. It should be noted that the probability inference algorithm (Algorithm 2 in Section 3) was run on these graphs directly; we didn't use the algorithm to find a subgraph (Algorithm 3 of Section 4).

Fan-out edge coloring. The number of colors used τ is listed in Table 1. In all cases, the quantities τ and $\Delta_{in} + 1$ were very close; the maximum difference between τ and $\Delta_{in} + 1$ was 9. Recalling the discussion on the optimal number of queries required (Section 3), we note that in practice the approach of constructing the line graph \mathcal{E}_G and using the greedy vertex coloring strategy seems to yield near-optimal results.

Accuracy of the estimates. In Figure 7(a), we compare the estimated influence probabilities with the reference IC model using two measures, namely mean absolute error (boxes) and maximum error (vertical lines). Here, the error corresponding to an edge probability p and its estimate \hat{p} is $|\hat{p} - p|$. For each (ϵ, δ) pair, the mean absolute error is much lower than ϵ ; it is roughly 0.1ϵ . Also, we observe that even the maximum error is comparable to ϵ , indicating that the estimates are much more accurate than the performance guarantees given by our theoretical results even for high values of δ . Also, we did not see much variation across networks since network structure has no role to play in this analysis. (Network structure only affects the number of queries required.)

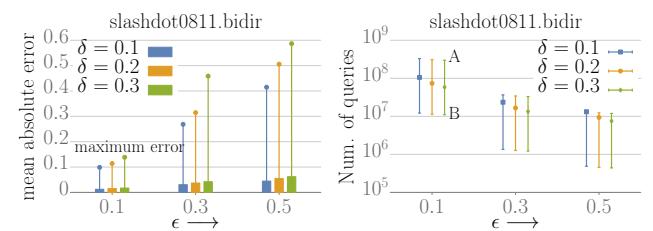


Figure 7: (a) The accuracy of the estimated models and (b) the number of queries required to obtain these estimates. The results shown here for Slashdot0811 are representative of the results for all networks. In (b) $A = 3\tau T(\epsilon, \delta/|E|)\rho/p_{min}$ and $B = \Delta_{in}\tau T(\epsilon, \delta/|E|)\rho/p_{max}$, the upper and lower bounds for the expected number of queries respectively.

Number of queries required. In Figure 7(b), the average number of queries required to infer the given absolute model for different ϵ, δ values is shown for a representative network. This is compared with two values based on the discussion in Section 3. Consistently, we note that the number of queries required is close

to $3\tau T(\epsilon, \delta/|E|)\rho/p_{\min}$. Even though the number of colors τ is close to the lower bound, since the probabilities were drawn uniformly from $\{0.1, \dots, 0.5\}$, under the assumption that most color classes have many edges in them, the probability that every color class has an edge with probability $p_{\min} = 0.1$ is high. Under this observation, the lower bound significantly improves (p_{\min} replaced by p_{\max}). We note that the number of queries used here (from about 10^7 to 5×10^8) is much larger compared to those in Section 5.2 since the subgraph generation algorithm was not used.

Dynamical analysis. For every IC model (both true and inferred), we simulated the spread on all networks by seeding one vertex at random in each instance. The results are averaged over 100 iterations. Figure 8 (left) summarizes the error in the fraction of influenced nodes in the inferred models when compared with the reference model. We recall that there are five IC models for every network with probabilities drawn from the same distribution. The large variance in mean absolute error suggests that the probability assignment plays an important role in inference. Also, an increase in ϵ does not show a corresponding increase in the error. Figure 8 (right) shows average curves of cumulative fraction of nodes influenced as a function of time for the 5 networks. The variance in inferred models rises at large outbreak sizes.

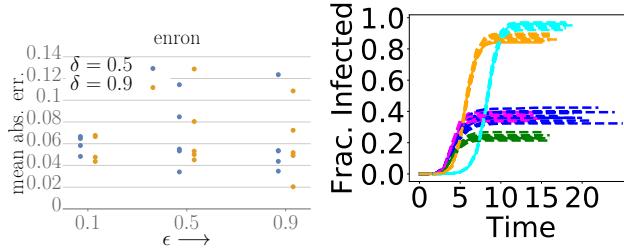


Figure 8: Comparison of dynamics between inferred models and reference model: We show the error in the average fraction of infections. Representative plots for Enron network are shown (left). Also shown are the cumulative infection plots (right) for various networks as a function of time for $\epsilon = \delta = 0.5$ (Enron-blue, Epinion-green, Slashdot-magenta, ER-cyan, and BA-orange). Variances within 100 iterations of one simulation are about 0.08 maximum; we show differences across runs with multiple curves.

6 FUTURE WORK

Our work suggests several future research directions under the active query framework. For example, one can investigate the edge probability inference problem for other diffusion models such as the SIR model [10] and its variants (e.g., SEIR model). Another direction is to obtain probability estimates for a maximum number of edges under a budget on the number of queries. Finally, it is of interest to develop a more sophisticated stopping criterion to further reduce the number of queries needed to obtain provably good performance guarantees.

Acknowledgments: We thank the referees of CIKM 2018 for providing valuable suggestions. We thank Dominik Borkowski, William Miles Gentry, Jeremy Johnson, William Marmagis, Douglas McMaster, Kevin Shinpaugh, and Robert Wills. This work has been partially supported by DTRA CNIMS (Contract HDTRA1-11-D-0016-0001), NSF DIBBS Grant ACI-1443054, NSF BIG DATA Grant IIS-1633028 and NSF EAGER Grant CMMI-1745207. The U.S. Government is

authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

REFERENCES

- [1] A. Adiga, V. Cedeno-Mieles, et al. 2018. Active Query Based Inference of Probabilistic Contagion Models Over Networks. NIDSSL Tech. Rep. 2018-48, Virginia Tech, Blacksburg, VA.
- [2] A. Adiga, C. J. Kuhlman, M. V. Marathe, S. S. Ravi, D. J. Rosenkrantz, and R. E. Stearns. 2017. Inferring local transition functions of discrete dynamical systems from observations of system behavior. *Theor. CS* 679 (2017), 126–144.
- [3] C. Barrett, R. Beckman, M. Khan, V. S. A. Kumar, M. Marathe, P. Stretz, T. Dutta, and B. Lewis. 2009. Generation and Analysis of Large Synthetic Social Contact Networks. In *Proceedings of the Winter Simulation Conference*. 1003–1014.
- [4] C. Barrett, K. Bisset, J. Leidig, A. Marathe, and M. Marathe. 2011. Economic and social impact of influenza mitigation strategies by demographic class. *Epidemics* 3 (2011), 19–31.
- [5] C. Barrett, H. B. Hunt, M. V. Marathe, S. S. Ravi, D. J. Rosenkrantz, and R. E. Stearns. 2011. Modeling and analyzing social network dynamics using stochastic discrete graphical dynamical systems. *Theor. CS* 412, 30 (2011), 3932–3946.
- [6] D. Centola. 2010. The Spread of Behavior in an Online Social Network Experiment. *Science* 329 (2010), 1194–1197.
- [7] P. Dagum, R. Karp, M. Luby, and S. Ross. 2000. An optimal algorithm for Monte Carlo estimation. *SIAM Journal on computing* 29, 5 (2000), 1484–1496.
- [8] L. M. S. Dekkers, A. Beekens, A. D. Hofman, et al. 2017. Formal Modeling of the Resistance to Peer Influence Questionnaire: A Comparison of Adolescent Boys and Girls With and Without Mild-to-Borderline Intellectual Disability. *Assessment* 43 (2017), 1–14.
- [9] R. Durrett. 1988. *Lecture Notes on Particle Systems and Percolation*. Wadsworth.
- [10] D. Easley and J. Kleinberg. 2010. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press.
- [11] M. R. Garey and D. S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-completeness*. W. H. Freeman & Co., San Francisco.
- [12] J. Goldenberg, B. Libai, and E. Muller. 2001. Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. *Marketing Letters* 12, 3 (2001), 211–223.
- [13] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, and Y. Moreno. 2011. The dynamics of protest recruitment through an online network. *Scientific Reports* (2011).
- [14] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan. 2010. Learning Influence Probabilities in Social Networks. In *WSDM*. 241–250.
- [15] D. Kempe, J. Kleinberg, and E. Tardos. 2003. Maximizing the Spread of Influence Through a Social Network. In *Proc. 9th ACM SIGKDD*. 137–146.
- [16] J. Kleinberg, S. Mullanathan, and J. Ugander. 2017. Comparison-Based Choices. arXiv:1705.05735v1 [cs.DS]. 20 pages.
- [17] J. Leskovec and A. Krevl. 2014. SNAP Datasets: Stanford Large Network Dataset Collection.
- [18] S. Li, X. Gao, W. Bao, and G. Chen. 2017. FM-Hawkes: A Hawkes Process Based Approach for Modeling Online Activity Patterns. In *CIKM*. 1119–1128.
- [19] T. M. Liggett. 1985. *Interacting Particle Systems*. Springer.
- [20] L. Liu, J. Tang, J. Han, and S. Yang. 2012. Learning Influence from Heterogeneous social networks. *Data Mining and Knowledge Discovery* 25 (2012), 511–544.
- [21] A. Marathe, B. Lewis, J. Chen, and S. Eubank. 2011. Sensitivity of Household Transmission to Household Contact Structure and Sizes. *PLoS One* 6 (2011), e22461–1–e22461–7.
- [22] W. Mason and D. J. Watts. 2012. Collaborative Learning in Networks. *PNAS* 109, 3 (2012), 764–769.
- [23] M. Mitzenmacher and E. Upfal. 2005. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press.
- [24] K. C. Monahan, L. Steinberg, and E. Cauffman. 2009. Affiliation With Antisocial Peers, Susceptibility to Peer Influence, and Antisocial Behavior During the Transition to Adulthood. *Dev. Psych.* 45 (2009), 1520–1530.
- [25] H. Mortveit and C. Reidys. 2007. *An Introduction to Sequential Dynamical Systems*. Springer, New York, NY.
- [26] D. Romero, B. Meeder, and J. Kleinberg. 2011. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter. In *Proc. 20th WWW*. ACM, 695–704.
- [27] K. Saito, R. Nakano, and M. Kimura. 2008. Prediction of Information Diffusion Probabilities for Independent Cascade Model. In *Proc. KES*. 67–75.
- [28] L. Steinberg and K. C. Monahan. 2007. Age Differences in Resistance to Peer Influence. *Dev. Psych.* 43 (2007), 1531–1543.
- [29] H. Tong, B. A. Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos. 2012. Gelling and melting, large graphs by edge manipulation. In *CIKM*. 245–254.
- [30] J. C. Turner, P. J. Oakes, S. A. Haslam, and C. McGarty. 1994. Self and Collective: Cognition and Social Context. *Personality and Social Psychology Bulletin* 20, 5 (1994), 454–463.
- [31] D. B. West. 2001. *Intro. to Graph Theory*. Prentice-Hall.

Generative Modeling of Human Behavior and Social Interactions Using Abductive Analysis

Yihui Ren*, Vanessa Cedeno-Mieles*[†][§], Zhihao Hu[‡], Xinwei Deng[‡],
 Abhijin Adiga*, Christopher Barrett*, Saliya Ekanayake*, Brian J. Goode*, Gizem Korkmaz*, Chris J. Kuhlman*,
 Dustin Machi*, Madhav V. Marathe*[†], Naren Ramakrishnan*[¶][†], S. S. Ravi*, Parang Saraf*[¶][†], Nathan Self*[¶][†]
 Biocomplexity Institute*, Discovery Analytics Center¶, Departments of Computer Science[†] and Statistics[‡]
 Virginia Tech - Blacksburg, VA*
 Escuela Superior Politécnica del Litoral, ESPOL - Guayaquil, Ecuador[§]

Noshir Contractor
 Department of Industrial Engineering
 and Management Sciences
 Northwestern University
 Evanston, IL

Joshua Epstein
 College of Global Public Health
 New York University
 New York, NY

Michael W. Macy
 Department of Sociology
 Cornell University
 Ithaca, NY

Abstract—Abduction is an inference approach that uses data and observations to identify plausible (and preferably, best) explanations for phenomena. Applications of abduction (e.g., robotics, genetics, image understanding) have largely been devoid of human behavior. Here, we devise and execute an iterative abductive analysis process that is driven by the social sciences: behaviors and interactions among groups of human subjects. One goal is to understand intra-group cooperation and its effect on fostering collective identity. We build an online game platform; perform and analyze controlled laboratory experiments; form hypotheses; build, exercise, and evaluate network-based agent-based models; and evaluate the hypotheses in multiple abductive iterations, improving our understanding as the process unfolds. While the experimental results are of interest, the paper’s thrust is methodological, and indeed establishes the potential of iterative abductive looping for the (computational) social sciences.

I. INTRODUCTION

A. Background and Motivation

Abduction is an inference approach that uses data and observations to identify plausible (preferably, best) explanations for phenomena [1]. Abduction has broad application in robotics, genetics, automated systems, and image understanding [2]–[5].

However, in contrast to this notion of abduction, our focus is the specification and implementation of an abductive *looping* process, wherein abduction is executed in successive iterations. Every iteration builds off of all previous ones, so that explanations may evolve from accumulated data. As a differentiator from previous work, our interests are behaviors and human interactions within networked groups in the social sciences.

In particular, our exemplar is to understand whether a cooperative game can produce collective identity (CI) within a group. CI is an individual’s cognitive, moral, and emotional IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain
 978-1-5386-6051-5/18/\$31.00 © 2018 IEEE

connection with a broader community, category, practice, or institution [6]. There are many applications and contexts in which CI is important, including team formation, maintenance, and behavior in organizations, communities, and marginalized groups (e.g., [7]–[9]).

Inspired by [10], we use a group anagrams (word construction) game to engender CI, where players work cooperatively to form words by sharing letters, and use the Dynamic Identity Fusion Index (DIFI) score [11] as our measurement of CI. Identity fusion is a feeling of oneness with the group that induces people to tether their feelings of personal agency to the group [11]. Our first novelty is that this is the first work on performing and analyzing *online* experiments, and developing and evaluating agent-based models (ABMs), for the group anagrams game and CI. (The first *face-to-face* experiments were conducted recently [10], with no modeling work done; their setup is somewhat different than ours.) Our experimental findings below constitute our second novelty.

The abductive loop (AL) process is described in Section II, but among its components are experiments and modeling, and we make note of experiment-modeling interactions here. There have been several controlled experimental studies of comparable size to our experiments (e.g., [12]–[14]). Also, empirically grounded, data-driven modeling of human behavior is done [15]–[18]. We combine these two ideas, in a particular way that is guided by abduction, and perform them iteratively. The proposed abductive analysis is to form hypotheses to evaluate theories as part of the looping process, and develop new insights about CI. Thus, our third novelty is abductive iterations where data drive models, and model predictions drive new experiments in a principled approach. Looping over abductive analyses is relatively rare (see the robotics work [2] as an exception), and the use of abduction and abductive iterations

in the social sciences is very rare. Our approach provides an exemplary case of coupling theory development/evaluation with real problems [19]: real data guide our theory evaluations.

This work was motivated in part by the DARPA Next Generation Social Science (NGS2) program. Goals of the program include devising methods to identify theories that are and are not applicable for explaining societal events.

B. Contributions

1. Specification and demonstration of iterative abductive analysis process. Using [20], [21] as a starting point, we explicitly incorporate modeling and iterations into the abductive process; the latter necessitates specifying what is to be done in the next iteration. The iterative process is successfully demonstrated through the anagrams experiments, agent-based modeling, and hypothesis generation and testing. The proposed abductive process can be considered as a general methodology for other social science researches. For example, our method of model construction from data (see Contribution 2 below) can be used to capture other temporal human action sets among interacting agents.

2. Data-driven networked agent-based models (ABMs) of experiments: design, construction, and evaluation. We design, construct, and evaluate three data-driven ABMs as part of the iterative abductive analysis. We adapt a conditional random fields (CRF) [22] modeling approach with four parameters to flexibly incorporate history effects on agent actions that evolve in time. It can alleviate the overfitting problem that would arise with, e.g., a static Markov model that would require capturing many more state transitions. ABM is used as our simulation modeling approach because of its fine granularity and for its generative properties [23]; that is, local interactions produce population-level dynamics. We use inductive and deductive inference in three ways, use KL-divergence to compare model predictions with experimental data, compare results from multiple ABMs, and evaluate the transition matrices of our ABMs using a statistical approach.

3. New experimental understanding of the formation of collective identity (CI). We discover three novel insights on the formation of CI by coupling the team anagrams game and DIFI score. First, players' DIFI scores increase with increasing numbers of neighbors in the anagrams game. Second, the number of interactions increases as number of neighbors (i.e., a player's network degree) increases from 2 to 4. However, the numbers of interactions, relatively speaking, saturates with further increases in degree. Third, despite this saturation, the DIFI score continues to increase with degree, suggesting complicated interactions among game parameters. Our analysis is a first work on quantifying the formation of CI since little work has been conducted on this subject in the literature. It is important to note that this experimental work (like the modeling work) takes place within the abductive loop framework.

Organization. An overview of the abductive loop process is presented in Section II, providing a framework for the rest of the paper. Related work is in Section III. In Sections IV and V,

the experiment is described and modeling is presented, which address the respective components of the abductive loop. This enables a more streamlined description of the abductive loop for CI in Section VI. Section VII summarizes.

II. OVERVIEW OF ABDUCTIVE LOOP

Figure 1 illustrates our iterative abductive process, which includes inductive and deductive steps and hypothesis testing. This structure follows that of [20], [21], which are based on Piercian abduction [1], but augments it in key areas. Note that in contrast to confirmatory (deductive) analyses, where theories, hypotheses, and models are developed *first*, and used to predict subsequent experimental results, one-step abduction first generates data through experiments or observations. (Abduction uses data to drive the scientific discovery process.) Then, data analysis consists of searching for *patterns* and generalizing these into *phenomena*, which is an inductive step. These results are used to formulate hypotheses based on theories whose purpose is to explain the data. Hypotheses may exist (from a previous loop) or may be proposed in this step, and can be removed (e.g., via falsification). Multiple candidate theories may be posed for a given phenomena. Models are developed from the data, with the objective of generating outputs that help evaluate hypotheses and theories, and/or help guide experiments for the next loop. The best explanation, or hypothesis/theory appraisal, is the process of identifying the best explanation for the phenomena [24]; this includes hypothesis falsification. Finally, the last step in an iteration is to determine what to do next, in terms of designing new experiments. The iterative process may terminate for any number of reasons; e.g., a best explanation has been found.

Several variations on Figure 1 are possible. For example, the modeling, hypothesis, and theory steps may be interchanged, e.g., modeling results may be used to formulate hypotheses and devise theories. Also, as a model matures, a deductive analysis may be executed: (quantitative) model predictions in one iteration are evaluated with new experiments in the next iteration.

Furthermore, our abductive loop process makes modeling a much more prominent feature of the process than in [20], [21]: in hypothesis evaluation and in “What is Next?,” which specifies experiments for the next loop. Moreover, we build generative ABMs, while models are based predominantly on similarity in [20]. Consequently, inference plays a large role in our work. We execute this loop in Section VI in evaluating CI.

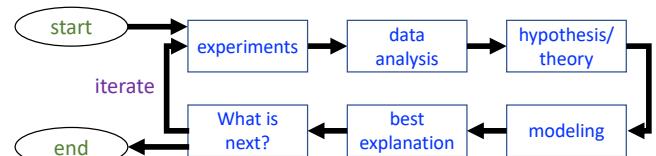


Fig. 1: Steps in our iterative abductive analysis/loop.

III. RELATED WORK

Anagrams and CI Experiments. Over 20 experimental works (e.g., [25]) use anagrams games—with *individual* players. The only cooperative team-play of an anagrams game is reported

in [10]; their goal, like ours, is to *produce* CI. While this motivated our experiment, there are several differences in procedures and context, e.g., theirs is a face-to-face experiment; ours is online. They measure CI with the proxy of public goods game contributions, while we use DIFI score. There are several in-person experiments studying the *implications* of CI where team members interact (e.g., [26]).

Agent Based Models of Anagrams and CI. The closest work to ours is [27], [28], where ABMs of identity diffusion are presented in which an agent adopts (changes) her type of identity to that of a neighbor with a stronger (higher valued) type of identity. Hence, this is a contagion process much like a voter model. We, in contrast, model the process of producing CI. There are no ABMs (or models of any kind) of group anagrams games, to our knowledge.

Data-Driven Modeling. Data-driven modeling works include [15]–[18]. These works cover explore-exploit networked experiments with limited modeling [15]; individual models of single-choice (i.e., one-shot) evacuation decisions [17]; ABM of emotion and information contagions spreading on a network and comparisons with a single event [16]; and ABM of solar panel adoption and comparisons with data in San Diego county [18]. None of these works use ABMs to model networked experiments where individuals take a series of actions (that may be repeated) over time, to study CI.

Abduction and Abductive Loop. Constructive procedures for implementing abductive analyses include [20], [21]. We extend those works by making modeling a first-class process, and by adding the process of what to do in the next iteration. In addition to the applications cited in the Introduction, abduction was used to understand emergency room personnels' efforts to save injured people in terms of "social viability" [29]. Perhaps the work closest to ours is [30] in that they develop models and make predictions based on data. However, their data are either artificially generated or address isolated individuals, and they use abduction rather than abductive iterations.

IV. GAME PLATFORM AND EXPERIMENT (GAME)

Web App Platform for Experiments. Owing to space limitations, we provide a very brief overview of the web app game platform that we built. The platform consists of the oTree infrastructure [31] for recruiting players from Amazon Mechanical Turk (AMT) and interactions during the game; Django Channels for player interactivity; and JavaScript and HTML for generating the screens for a consent form, instructions, information, a survey, and game interactions in a 2-phase game. Experiments and data analyses are part of the abductive loop of Sections II and VI and Figure 1.

2-Phase Game Description. Phase-1 is the group anagrams (word construction) game, where n players cooperate in sharing letters to form and submit words of length ≥ 3 . Communication channels between pairs of agents mean that they can share letters with each other, and this induces a graph on the players. We use random regular graphs of degree k on the n players so that everyone has the same number of neighbors. Over all

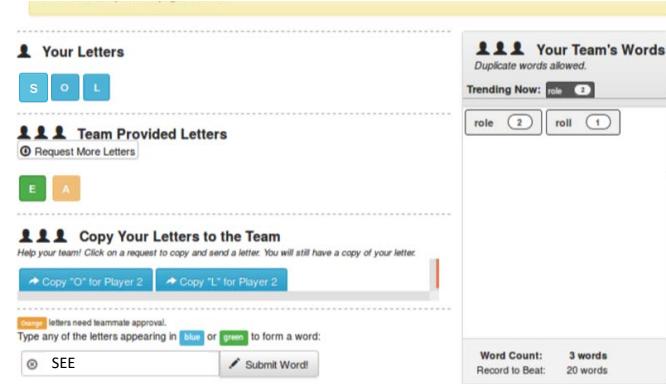


Fig. 2: The anagrams game screen, phase-1, for one player. This player has own letters "S," "O," and "L" and has requested an "E" and "A" from neighbors. The "E" is green, so this player's request has been fulfilled and so "E" can be used in forming words. But the request for "A" is still outstanding so cannot be used in words. Below these letters, it shows that Player 2 has requested "O" and "L" from this player; this player can reply to these requests, if she so chooses. Below that is a box where the player types and submits new words, like "SEE."

abductive loops, experiments are run in groups with nominal values of $10 \leq n \leq 20$ and with regular degree $2 \leq k \leq 8$.

A screen shot of one player's screen at one point in time, is shown in Figure 2. Each player is given $n_l = 3$ letters that she can use to form words and that she can share with others. She has an infinite supply of letters so that sharing letters does not inhibit her own use of letters. A player can also request letters from her neighbors and if the neighbors provide those letters, then she can use those letters in words, but she cannot pass on the received letters. Also, letters can be used any number of times in a word, meaning that if a player forms the word *tat* using *a* and one *t* (used twice), the player still has the *a* and *t* to form more words. Hence, a player needs to receive a requested letter only once.

Initially, a player sees her n_l own letters and those of all of her neighbors, but has access only to her own letters. Over the 5-minute anagrams game duration, players can idle, form words, request letters from their neighbors, and reply to requests.

Team members earn money by forming as many words as possible. Players are told that the total team earnings e_t are split evenly; each player receives e_t/n , so that it is in their interest to assist their neighbors. After the phase-1 anagrams game, each player is told the total number of words formed by the team, and each player's individual earnings.

The phase-2 DIFI procedure follows immediately after phase-1. Each player executes individually the DIFI procedure [11], to measure the degree to which the player feels part of the team (i.e., associates their identity with that of the team). Each player does this by moving a circle in a browser, representing herself, relative to a fixed team circle. The DIFI score is in the range [-100,125], with a score < 0 representing no overlap of circles, and therefore indicating no CI; = 0 representing the

circles just touching; and > 0 indicating overlap of the two circles and hence formation of some level of CI. Plots of data are in Sections V and VI.

V. AGENT-BASED MODELS OF ANAGRAMS GAME AND MODELING RESULTS

We present three ABMs of the anagrams game that are used in the abductive loop analyses to follow in Section VI. All models were developed *as part of* the abductive loop process, but are presented here to emphasize their construction and evaluation, and to obviate the need for a large digression for the models in the description of the AL process in Section VI. All models, wherein each player is an agent, are data-driven, and hence *inductive inference* is used with data in three ways: to inform model structure, to characterize model parameters, and to estimate parameter values.

In all models, we represent the set V of players and their communication channels E as an undirected graph $G(V, E)$. The game is modeled as a discrete-time stochastic process, where at each time step, a player performs one of the actions from the action set $A = \{a_1, a_2, a_3, a_4\}$, consisting of: (i) a_1 : idling (i.e., thinking); (ii) a_2 : replying to a neighbor with a requested letter, (iii) a_3 : requesting a letter from a neighbor, and (iv) a_4 : forming and submitting a word.

ABM M0 is a baseline model that is presented after M1 for ease of exposition. ABMs M1 and M2 model the actions of A , but are generic in that request a_3 and reply a_2 letters and submit word a_4 are not associated with particular letters. For example, if the player action is a_4 , then the model assumes that the player will form a word. In all ABMs, actions are taken at integer numbers of seconds; that is, simulations of interacting agents take place as time advances in discrete 1-second increments from 0 to 300. This time increment is based on the data.

A. Agent-Based Model M1, and then Baseline Model M0

ABM M1 Development. The goal is to accurately quantify the transition probability from one action $a(t) = a_i$ at time t to the next action $a(t+1) = a_j$ for each agent v ; $i, j \in [1..4]$; and $a(t) \in A$. For clarity, we use i and j to represent the actions a_i and a_j . Agent v executes a stochastic process driven by transition probability matrix $\Pi = (\pi_{ij})_{m \times m}$, where $m \equiv |A|$ (here, $= 4$) and

$$\pi_{ij} = \Pr(a(t+1) = j | a(t) = i) \text{ with } \sum_{j=1}^m \pi_{ij} = 1. \quad (1)$$

To make Π dynamic in time and account for history effects, four variables are introduced that evolve in time: number $z_L(t)$ of letters that v has available to use (i.e., in hand) at t ; number $z_W(t)$ of valid words that v has formed; size $z_B(t)$ of the buffer of letter requests that v has yet to reply to; and number $z_C(t)$ of consecutive time increments that v has taken the same action. Thus, letting $z = (1, z_L, z_W, z_B, z_C)_{(m+1) \times 1}$, we can model π_{ij} as a function of these covariates, i.e., $\pi_{ij} = f_{ij}(z)$. We use multinomial logistic regression to model π_{ij} as

$$\pi_{ij} = \frac{\exp(z' \beta_j^{(i)})}{1 + \sum_{h \neq i} \exp(z' \beta_h^{(i)}),} \quad (2)$$

where $\beta_j^{(i)} = (\beta_{j1}^{(i)}, \dots, \beta_{j,m+1}^{(i)})'$, $\beta_i^{(i)} = \mathbf{0}$, and prime indicates transpose. For a given i , the parameter set can be expressed as

$$\mathbf{B}^{(i)} = \begin{pmatrix} \beta_{11}^{(i)} & \beta_{12}^{(i)} & \dots & \beta_{1,m+1}^{(i)} \\ \beta_{21}^{(i)} & \beta_{22}^{(i)} & \dots & \beta_{2,m+1}^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{41}^{(i)} & \beta_{42}^{(i)} & \dots & \beta_{4,m+1}^{(i)} \end{pmatrix}. \quad (3)$$

Baseline ABM M0. ABM M0 is a simplification of M1. The transition matrix Π is formed from the data so that the π_{ij} in Equation (1) are *constant*; time-invariant, independent of z .

Inductive Inference. We address the three dimensions of inference stated above. First, the model structure is informed by the $k = 2$ data. Second, and briefly, the parameters used in the feature vector z are justified as follows: $z_L(t)$ captures the idea that the more letters v has in-hand, the more likely the agent is to form words; $z_W(t)$ captures the notion that the more words that have been formed, the larger the vocabulary of the player. $z_B(t)$ captures the notion that the more letter requests that have not been replied to, the more likely v is to reply; and $z_C(t)$ captures the notion that the more time v is idle (thinking), the more likely v will take some other action at the next timestep. Third, parameters in Equation (3) are inferred from the $k = 2$ experimental data using the framework of maximum likelihood estimation for the multinomial distribution.

Results. Throughout, we use k to denote the number of neighbors (degree) of an agent v . Also, we evaluate five variables and their distributions, across all players in a set of games, in comparing models and experiments: $x = (x_1, x_2, x_3, x_4, x_5)$, where x_1 is the number of letter replies received ($RplR$); x_2 is the number of replies sent ($Rpls$); x_3 is the number of letter requests received ($RqsR$); x_4 is the number of requests sent ($RqsS$); and x_5 is the number of words formed ($Wrds$). To measure the performance of our models, we use *KL-divergence* between our model prediction on x and the experimental observation of x , *throughout this manuscript*.

Figure 3 provides model predictions and comparisons with experimental results. The first two plots show distributions of experimental data (in gray) and ABM M1 predictions in red. The green curves are from the baseline model M0. Clearly, ABM M1 is in better agreement with the experimental data compared to M0 in Figures 3a and 3b. From KL-divergence in Figure 3c, it is clear that the predictions of M1 represent the experimental data better than those of the baseline model. In addition, we use M1 to make predictions for graphs with larger $k > 2$, resulting in more interactions. Counterintuitively, as shown in Figure 3d, the number of replies does not change as k increases. These results call for more experiments at larger k . Note that we exercise M1 learned from experiments with $k = 2$ to predict the case of $k = 2$ (self-consistency checks), and to predict results of other cases with $k > 2$, as in Figure 3d.

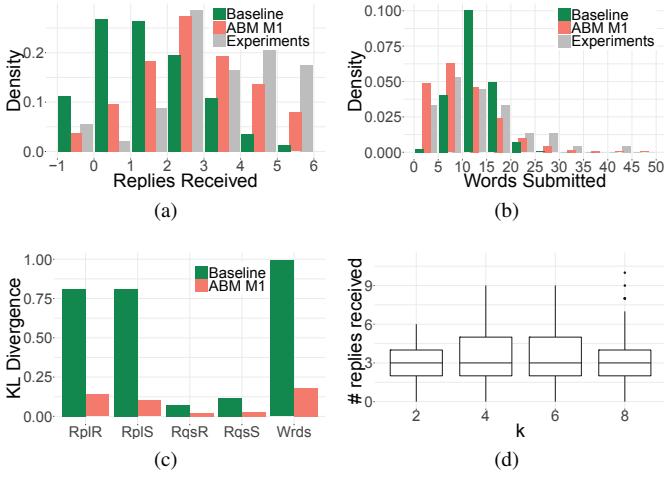


Fig. 3: ABM M0 and M1 predictions of the $k = 2$ experiments, and other simulation results. (a) Distribution of replies received, and (b) distribution of words formed, each at the end of the 5-minute anagrams game (gray bars) for all $k = 2$ experiments, compared to M1 predictions (red) for 100 simulations of an $n = 10$ player game. A baseline model M0 is shown in green for comparison. (c) KL-divergence values for the baseline (M0) and M1 models across the five parameters of x : lower values are better. These figures show that M1 generates predictions much closer to the experimental data than does M0. (d) M1 model distributions predicted for the number of replies received at the end of game ($n = 10$, 100 simulations), for different regular degrees k of the game network G .

We remark that we also fitted M1 using experimental data with $k = 4$, and consequently made predictions for the case of $k = 2$. We compared the distributions of x between prediction and experimental results using KL-divergence, whose values range from 0.11 to 0.46, indicating good predictions.

B. Agent-Based Model M2

Enhancement from ABM M1: arbitrary network topology.

Model M1 was developed with data where all game players have the same degree $k = 2$. To generalize M1 to incorporate various k , we conducted additional experiments with $2 < k \leq 8$ as a part of the second AL (Section VI-B below).

Development for Arbitrary Degree. We build a hierarchical model to incorporate the effect of agent degree k . For different values of k , the parameter coefficients in $\mathbf{B}^{(i)}$ in Equation (3) will be a function of k , denoted as $\mathbf{B}^{(i)}(k)$. We use an orthogonal polynomial basis to construct a continuous and smoothing function for $\beta_{jh}^{(i)}(k)$ for any given i, j, h , as

$$\beta_{jh}^{(i)}(k) = \alpha_0^{(i,j,h)} + \alpha_1^{(i,j,h)}\xi_l(k) + \alpha_2^{(i,j,h)}\xi_q(k), \quad (4)$$

where ξ_l and ξ_q are the linear and quadratic functions of the orthogonal basis in terms of k . This formulation provides a means to capture nonlinear effects. We have

$$\mathbf{B}^{(i)}(k) = \mathbf{C}_0^{(i)} + \mathbf{C}_1^{(i)}\xi_l(k) + \mathbf{C}_2^{(i)}\xi_q(k), \quad (5)$$

where

$$\mathbf{C}_r^{(i)} = \begin{pmatrix} \alpha_{11}^{(i,r)} & \alpha_{12}^{(i,r)} & \dots & \alpha_{1,m+1}^{(i,r)} \\ \alpha_{21}^{(i,r)} & \alpha_{22}^{(i,r)} & \dots & \alpha_{2,m+1}^{(i,r)} \\ \vdots & \ddots & \ddots & \vdots \\ \alpha_{41}^{(i,r)} & \alpha_{42}^{(i,r)} & \dots & \alpha_{4,m+1}^{(i,r)} \end{pmatrix}, r = 0, 1, 2, \dots, (6)$$

with $\alpha_{ih}^{(i,r)} = 0$ for any r and h .

Inductive Inference. To estimate the parameters sets $\mathbf{C}_0^{(i)}, \mathbf{C}_1^{(i)}, \mathbf{C}_2^{(i)}$, we use maximum likelihood estimation across the observations for $k = 2, 4, 6$, and 8.

Results. Figure 4 provides comparison results between ABMs M1 (red) and M2 (blue). KL-divergence values for distributions of replies received are shown in Figure 4a.

Although M1 performs well for $k = 2$, M2 betters M1 for $k > 2$, as M2 incorporates experimental data with $2 \leq k \leq 8$. This improvement is consistent among other x variables as shown in Figure 4b.

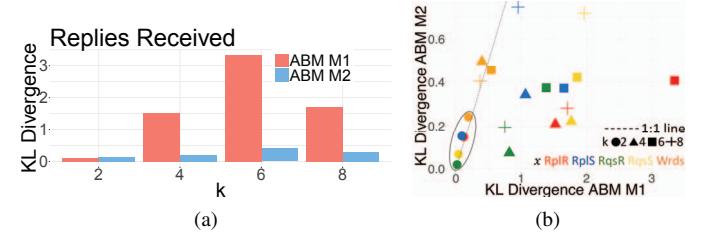


Fig. 4: Comparisons of KL-divergence values generated with models ABM M1 and ABM M2. (a) KL-divergence values for ABM M1 (red) and M2 (blue) for distributions of replies received, for experiments with $k = 2, 4, 6$, and 8. M2 gives much better performance, as expected, as it explicitly accounts for agent degree. (b) A scatter plot of KL-divergence for M1 (x-axis) and M2 (y-axis) for 4 k values and 5 x variables (different scales on x,y axes). For $k > 2$, M2 performs better. As expected, M1 and M2 perform equally well (highlighted) for $k = 2$ as M1 is learned from $k = 2$ experimental data.

C. Model Evaluation

To evaluate the goodness of fitting for the proposed hierarchical model in Equation (4), we compare the estimated (model) transition probability matrix $\hat{\Pi} = (\hat{\pi}_{ij})$ for M2 with the empirical (data) transition probability matrix $\tilde{\Pi} = (\tilde{\pi}_{ij})$ under different settings of covariates. Here the empirical transition probability matrices $\tilde{\Pi}$ are obtained under the settings by grouping the value of each covariate with three levels, as described in Table I, to provide comparable numbers of samples across bins. Under each setting, a level combination of the four covariates, we compute a counting matrix $\mathcal{N} = (n_{ij})$, where n_{ij} is the number of data instances with transition from state i to state j . Consequently, we calculate the empirical probability $\hat{\pi}_{ij} = \frac{n_{ij}}{\sum_j n_{ij}}$. There are 324 settings in total, and 279 of them have valid empirical transition probability matrices. For the estimated transition probability matrix $\hat{\Pi} = (\hat{\pi}_{ij})$, the value

TABLE I: Three bins and ranges of values for the z variables from Section V-A.

Level	Buffer (z_B)	Letter (z_L)	Word (z_W)	Consec. (z_C)
1	0	0-3	0-1	0-3
2	1	4-6	2-8	4-11
3	≥ 2	≥ 7	≥ 9	≥ 12

of $\hat{\pi}_{ij}$ is estimated by the proposed model under each setting of covariates, where the averaged value at each level of the covariate is used in the estimated model.

The Root of Mean Squared Errors (RMSE) is used to quantify the difference between $\hat{\Pi} = (\hat{\pi}_{ij})$ and $\tilde{\Pi} = (\tilde{\pi}_{ij})$. The RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{1}{4|\mathcal{I}|} \sum_{i \in \mathcal{I}} \sum_{j=1}^4 (\hat{\pi}_{ij} - \tilde{\pi}_{ij})^2} \quad (7)$$

where $\mathcal{I} = \{i : \min_j n_{ij} > 0\}$ is the index set of the rows where the empirical probability can be obtained.

Under each setting of covariates, we define the *Min.Count* as $n_{min} = \min_{i,j} \{n_{ij} : n_{ij} > 0\}$ as the smallest nonzero counts among transitions from state i to state j . It is known that when n_{min} is small, the empirical probability $\tilde{\pi}_{ij}$ is not accurate. Figure 5 shows the scatter plot between the RMSE and n_{min} for the 279 settings. From the figure, the proposed method generally provides an accurate estimation of probability transition matrix in most settings. Clearly, the value of RMSE decreases as the Min.Count n_{min} increases. When $n_{min} \geq 100$, the value of RMSE is smaller than 0.069, showing very good model fitting. When n_{min} is small, the RMSE is relatively high. One explanation is that the empirical probabilities cannot be calculated accurately when n_{min} is small.

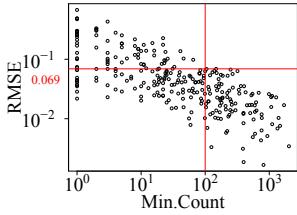


Fig. 5: Scatter plot of RMSE against Min.Count for different settings of covariates in Table I. See Equation (7) for RMSE and text for Min.Count.

VI. ABDUCTIVE LOOP (AL) ANALYSES AND RESULTS

In this section, we present the results of iterative abductive analyses, using the steps in Figure 1. We present two ALs. We also present an abductive tree with more hypotheses, that puts these two loops in context. We note that the experiments (Section IV) and modeling (Section V) are major components of the abductive looping process, and were separated out to make this section more streamlined.

A. Abductive Loop 1 (AL-1)

Experiments. A set of 18 experiments with a total of 87 players was completed where $k = 2$.

Data Analysis. Data were explored for patterns [20]. Time-series of actions in A , per player, show that players tend to

request particular letters with the goal of forming specific words: when a requested letter is received, there is often a burst of words formed with that letter. The other extreme, of requesting all letters initially and then figuring out words, is done far less frequently. The time-histories of actions, per player, also led to the model structures of the ABMs. Data for all players, for the five components of x in Section V-A, were combined to produce distributions of numbers of actions, against which models were compared. Measurements of these variable values are selected for correlation with DIFI scores.

Hypothesis/Theory. Hypothesis H₁₁: *In the team-based anagrams game, the sense of CI formed is driven more by the number of words a player forms than the number of interactions of a player (requests and replies).* Social Exchange Theory [32] focuses on the individual and suggests that the number of words resonates more in forming CI because they are directly related to reward in the game. Theory of social interactions [33] indicates that interactions are important for forming an interdependent organization. Reciprocity Theory suggests that v_i will respond to v_j 's requests because v_i wants v_j to respond to hers, so that interactions are important.

Model. The model of Section V-A was constructed from the time histories of actions of players for experiments with $k = 2$. The results relevant to this iteration are provided in Figure 3. ABM M1 is much better at capturing the dynamics in the experiments than is baseline model M0.

Best Explanation. Results of a linear regression in Table II indicate that hypothesis H₁₁ is falsified because Wrds, the number of words formed, is not significant, while RplR, RplS and RqsS are significant. Thus, Social Exchange Theory can be eliminated as theory of CI formation in this experiment. It is somewhat surprising that Wrds is not significant because it is the variable that is most closely associated with the reward (earnings). In the social sciences, and in many domains, eliminating candidate theories is a valuable result (that is, an analysis does not always have to identify the best theory).

TABLE II: Results of linear regression of variables in x against dependent variable DIFI score, indicating that interactions are more significant than number of words formed in producing CI. These data are generated in AL-1.

Var.	Interc.	RplR	RplS	RqsR	RqsS	Wrds
est.	103.	15.0	-13.0	6.41	-16.4	-0.213
p-val.	0.001	0.019	0.011	0.332	0.011	0.735

What is Next? Figure 3d indicates that the model predicts behavior that is invariant with the degrees of players [and hence the number of letters that neighbors possess] (plots of other variables of x are similar). We want to determine whether there is an effect of k , and hence the next experiments are specified as using increasing k (i.e., $k > 2$). Thus, the ABM M1 (driven by the data) is guiding what to do next. However, we are not using M1 to predict specific quantitative experimental results, as in a deductive analysis. Rather, M1 is used in a qualitative manner.

B. Abductive Loop 2 (AL-2)

In reality, we executed multiple abductive iterations, studying, in turn, $k = 4$ and then $k = 6, 8$. However, in the interest of space, we combine them into one iteration.

Experiments. A set of 16 experiments with a total of 137 players was completed where $k = 4, 6$, and 8 , respectively.

Data Analysis. We continued the same types of analyses described in AL-1, but with the added dimension of k .

Hypothesis/Theory. Hypothesis H₂₂: (a) As the number of neighbors increases in the anagrams game, the level of CI increases because there are more interactions. (b) However, beyond four neighbors (equivalently, for more than 12 neighbor letters) there is no benefit of additional neighbors. The theory of social interactions states that interactions with more neighbors creates more interdependence. Theory of cognitive load [34] suggests that cognitive load might be too great at some point, resulting in a player being unable to take advantage of more input.

Model. The model of Section V-B was constructed from the time histories of actions of players, from the combined data from both iterations. Model results relevant to this iteration are provided in Figure 4. ABM M2 captures trends in degree k much more effectively than ABM M1, for all parameters of x . Although not shown, ABM M2 shows effects of increasing k : the predicted numbers of actions increases as k increases, which is far different from the predictions in Figure 3d for ABM M1, where k has no effect.

Best Explanation. Figure 6 provides results that address hypotheses H₂₂(a) and H₂₂(b). Note that H₂₂(b) has two interpretations because “benefit” may be in terms of numbers of interactions (from AL-1) or in terms of DIFI score. We first examine H₂₂(b) in terms of numbers of interactions. Figure 6a shows the frequency distributions for replies received, for the four values of k . Focusing on interactions, note the large change in distributions in going from $k = 2$ to $k = 4$, but relatively minor changes for further increases in k . Thus, the saturation in the distributions (and others are similar), supports hypothesis H₂₂(b): the number of neighbors increases, but the number of interactions does not, for $k > 4$. This is consistent with cognitive load theory.

Now we evaluate H₂₂(a) and (b) in terms of DIFI score. Figure 6b shows that as k increases from 2 to 8, the probability density of DIFI scores shifts demonstrably to increasing DIFI. That is, greater numbers of neighbors produce more CI, as measured by the DIFI score. This does not wholly support H₂₂(a); while increasing k does correlate with increasing DIFI score, it is not because of the number of interactions, which does not increase appreciably for $k > 4$. These data falsify H₂₂(b): there is additional benefit, in terms of increased DIFI score, with increasing number of neighbors. The applicability of the theory of social interactions is not clear, but the data suggest that it is the number of different people with whom one interacts that is important, rather than the total number of interactions. More experiments are needed.

What is Next? At this point we halt the iterative abduction process for this paper, although it could continue. In a next

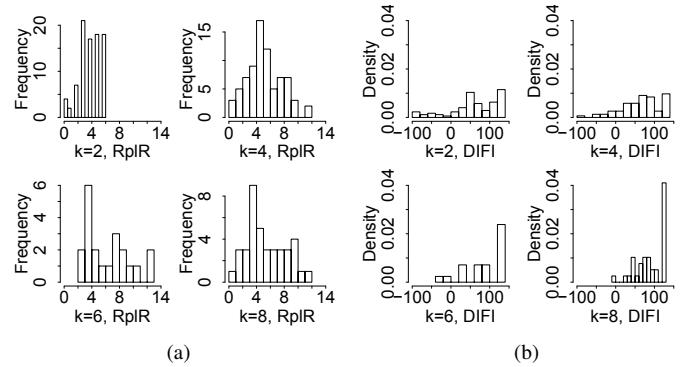


Fig. 6: Statistical analysis correlation results of the anagrams game parameters and DIFI score. (a) Frequency distributions of replies received change markedly from $k = 2$ to $k = 4$, but relatively little for further increasing k . (b) Probability density of DIFI score moves dramatically to larger DIFI score with increasing k . Each of these results is novel—these results are new. All the more novel is the combination of the two: while game measurable saturate (other data besides replies received), the DIFI score does not. These data are generated in AL-2.

iteration, we could (i) try to isolate the effects of number of interactions versus the number of neighbors in different experiments, or (ii) study the effects of different degrees of players and different numbers and qualities of letters initially assigned to players within the same experiment. We could also perform a deductive (confirmatory) analysis by making specific quantitative predictions for experiments using ABM M2 as part of AL-2, and running corresponding experiments in AL-3.

C. Abductive Loops: Role of Analyst and Bigger Picture

Two ALs have been demonstrated. Many additional loops are possible, as illustrated in Figure 7, which depicts several hypotheses, including the two addressed above (in orange). Figure 7 and Table III make clear the important role of an analyst in this process, as she guides the direction of the looping. So, while a plan such as that in Figure 7 may be useful, the actual tree structure will evolve with analyst decisions as the looping progresses and as data are generated, because hypotheses are based on newly-generated data in abduction.

VII. SUMMARY

We formalize an abductive loop, implement it computationally, and exercise it in an experimental setting (the anagram game) designed to induce CI, as operationalized by Swann’s DIFI score. However, our abductive looping process is not tied to CI. As part of the abductive iterations, we provide novel experimental insights into CI and build and evaluate three ABMs. This work establishes the potential of iterative abductive looping for the (computational) social sciences.

ACKNOWLEDGMENT

We thank the reviewers of ASONAM 2018 for providing valuable suggestions. We also thank the computer systems

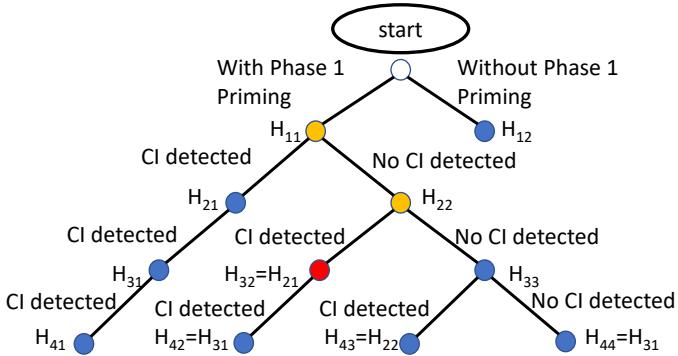


Fig. 7: Abductive tree representing candidate abductive loops with dependencies. Hypotheses are nodes, and are provided in Table III; edges are outcomes of ALs. The orange colored nodes correspond to abductive iterations presented herein. The red node is a candidate next loop. This tree is not unique; different analysts can devise different trees. Note that the actual hypotheses are not specified a priori; they are based on newly-generated data from the abductive iterations per Figure 1.

TABLE III: Candidate hypotheses to be evaluated in abductive iterations of Figure 7. Hypotheses H_{11} and H_{22} are given in the text. H_{43} is the same as H_{22} .

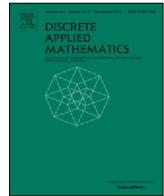
Hypothesis Number	Description
H_{12}	Playing the Phase 1 Anagrams game will produce greater individual DIFI scores than not playing Phase 1.
H_{21}	As the number and quality of letters assigned to a person decreases (i.e., as the letters assigned to a player occur less frequently in common words), collective identity of the player will increase.
H_{31}	Playing the game with players face to face will produce greater individual DIFI scores in Phase 2 (by enabling players to communicate and pick up on visual and verbal cues).
H_{42}	
H_{44}	
H_{33}	Lesser payouts in the Phase 1 anagrams game means that players do not have enough incentive to engage their neighbors.
H_{41}	Having the Phase 1 game score of another team displayed during Phase 1 will increase CI because it will create a stronger in-group/out-group paradigm.

administrators and managers at the Biocomplexity Institute for their help in this and many other works: Dominik Borkowski, William Miles Gentry, Jeremy Johnson, William Marmagias, Douglas McMaster, Kevin Shinpaugh and Robert Wills. This work is supported in part by DARPA Cooperative Agreement D17AC00003 (NGS2), DTRA CNIMS (Contract HDTRA1-11-D-0016-0001), National Science Foundation via grants DGE-1545362, IIS-1633363, and DIBBS ACI-1443054, and Army Research Laboratory under grant W911NF-17-1-0021. The US Government is authorized to reproduce and distribute reprints of this work for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA, DTRA, NSF, Army Research Laboratory, or the U.S.

Government.

REFERENCES

- [1] C. S. Pierce, "Elements of logic," in *Collected Papers of Charles Sanders Pierce*, C. Hartshorn *et al.*, Eds., 1931.
- [2] M. Shanahan, "Perception as abduction: Turning sensor data into meaningful representation," *Cognitive Science*, pp. 103–134, 2005.
- [3] E. A. M. Andrews and A. J. Bonner, "Explaining genetic knock-out effects using cost-based abduction," in *IJCAI*, 2011, pp. 1635–1640.
- [4] F. Vanderhaegen and P. Caulier, "A multi-viewpoint system to support abductive reasoning," *Information Sciences*, pp. 5349–5363, 2011.
- [5] B. Juba, "Learning abductive reasoning using random examples," in *AAAI*, 2016.
- [6] F. Polletta and J. M. Jasper, "Collective identity and social movements," *Annual Review of Sociology*, vol. 27, pp. 283–305, 2001.
- [7] S. Atran, R. Axelrod, and R. Davis, "Sacred barriers to conflict resolution," *Science*, vol. 317, pp. 1039–1040, 2007.
- [8] A. Silke, "Holy warriors: Exploring the psychological processes of jihadi radicalization," *European Journal of Criminology*, pp. 99–123, 2008.
- [9] L. A. DeChurch and J. R. Mesmer-Magnus, "The cognitive underpinnings of effective teamwork: a meta-analysis," *J. App. Psych.*, pp. 32–53, 2010.
- [10] G. Charness, R. Cobo-Reyes, and N. Jimenez, "Identities, selection, and contributions in a public-goods game," *Games and Econ. Beh.*, 2014.
- [11] J. William B. Swann, A. Gomez *et al.*, "Identity fusion and self-sacrifice: Arousal as a catalyst of pro-group fighting, dying, and helping behavior," *J. Pers. and Soc. Psych.*, 2010.
- [12] M. Kearns, S. Judd, J. Tan, and J. Wortman, "Behavioral experiments on biased voting in networks," *PNAS*, vol. 106, pp. 1347–1352, 2009.
- [13] S. Judd, M. Kearns *et al.*, "Behavioral dynamics and influence in networked coloring and consensus," *PNAS*, pp. 14978–14982, 2010.
- [14] M. Kearns, S. Judd, and Y. Vorobeychik, "Behavioral experiments on a network formation game," in *EC*, 2012, pp. 690–704.
- [15] W. Mason and D. J. Watts, "Collaborative learning in networks," *PNAS*, vol. 109, no. 3, pp. 764–769, 2012.
- [16] B. Li, D. Sun, Z. Lin, and C. Ou, "Agent-based simulation research on group emotion evolution of public emergency," in *ASONAM*, 2014.
- [17] C. Nguyen, K. J. Schlesinger, and J. M. Carlson, "Data-driven models for individual and group decision making," in *ASONAM*, 2017, pp. 852–859.
- [18] H. Zhang, Y. Vorobeychik *et al.*, "Data-driven agent-based modeling, with application to rooftop solar adoption," *JAAMAS*, 2016.
- [19] D. J. Watts, "Should social science be more solution-oriented?" *Nat. Hum. Behav.*, pp. 1–5, 2017.
- [20] B. D. Haig, "An abductive theory of scientific method," *Psychological Methods*, vol. 10, pp. 371–388, 2005.
- [21] S. Timmermans and I. Tavory, "Theory construction in qualitative research: From grounded theory to abductive analysis," *Sociological Theory*, vol. 30, pp. 167–186, 2012.
- [22] C. Sutton and A. McCallum, "An introduction to conditional random fields," *Foundations and Trends in Machine Learning*, pp. 267–373, 2011.
- [23] J. M. Epstein, *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton University Press, 2007.
- [24] P. Thagard, "Explanatory coherence," *Behavioral and Brain Sciences*, vol. 12, pp. 435–502, 1989.
- [25] I. G. Sarason, "Test anxiety and social influence," *Journal of Personality*, vol. 41, pp. 261–271, 1973.
- [26] M. Goldman, J. W. Stockbauer, and T. G. McAuliffe, "Intergroup and intragroup competition and cooperation," *J. Exp. Soc. Psych.*, 1977.
- [27] I. Lustick, "Agent-based modelling of collective identity: Testing constructivist theory," *J. of Artif. Soc. and Soc. Sim.*, vol. 3, 2000.
- [28] D. Rousseau and A. M. van der Veen, "The emergence of a shared identity: An agent-based computer simulation of idea diffusion," *Journal of Conflict Resolution*, vol. 49, no. 5, pp. 686–712, 2005.
- [29] S. Timmermans, "Social death as a self-fulfilling prophecy: David Sudnow's passing on revisited," *Sociological Quarterly*, 1999.
- [30] P. Singla and R. J. Mooney, "Abductive markov logic for plan recognition," in *AAAI*, 2011.
- [31] D. L. Chen, M. Schonger *et al.*, "otree—an open-source platform for laboratory, online and field experiments," *J. Beh. and Exper. Fin.*, 2016.
- [32] G. Homans, *Social Behavior: Its Elementary Forms*. Harcourt, 1961.
- [33] G. S. Becker, "A theory of social interaction," *Journal of Political Economy*, vol. 82, pp. 1063–1093, 1974.
- [34] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cognitive Science*, vol. 12, pp. 257–285, 1988.



Sublinear approximation algorithms for boxicity and related problems[☆]



Abhijin Adiga^a, Jasine Babu^{b,*}, L. Sunil Chandran^c

^a Virginia Tech., USA

^b Indian Institute of Technology, Palakkad, India

^c Indian Institute of Science, Bangalore, India

ARTICLE INFO

Article history:

Received 9 August 2015

Received in revised form 20 October 2017

Accepted 30 October 2017

Available online 1 December 2017

Keywords:

Boxicity

Approximation algorithm

Partial order dimension

Threshold dimension

Parameterization

ABSTRACT

The boxicity of a graph $G(V, E)$ is the minimum integer k such that G can be represented as the intersection graph of axis parallel boxes in \mathbb{R}^k . Cubicity is a variant of boxicity, where the axis parallel boxes in the intersection representation are restricted to be of unit length sides. Deciding whether the boxicity (resp. cubicity) of a graph is at most k is NP-hard, even for $k = 2$ or 3 . Computing these parameters is inapproximable within $O(n^{1-\epsilon})$ -factor, for any $\epsilon > 0$ in polynomial time unless $NP = ZPP$, even for many simple graph classes.

In this paper, we give a polynomial time $\kappa(n)$ factor approximation algorithm for computing boxicity and a $\kappa(n)\lceil\log\log n\rceil$ factor approximation algorithm for computing the cubicity, where $\kappa(n) = 2\lceil n\sqrt{\log\log n}/\sqrt{\log n}\rceil$. These $o(n)$ factor approximation algorithms also produce the corresponding box (resp. cube) representations. As a special case, this resolves the question posed by Spinrad (2003) about polynomial time construction of $o(n)$ dimensional box representations for boxicity 2 graphs. Other consequences of our approximation algorithm include $O(\kappa(n))$ factor approximation algorithms for computing the following parameters: the partial order dimension (poset dimension) of finite posets, the interval dimension of finite posets, minimum chain cover of bipartite graphs, Ferrers dimension of digraphs and threshold dimension of split graphs and co-bipartite graphs. Each of these parameters is inapproximable within an $O(n^{1-\epsilon})$ -factor, for any $\epsilon > 0$ in polynomial time unless $NP = ZPP$ and the algorithms we derive seem to be the first $o(n)$ factor approximation algorithms known for all these problems. We note that obtaining a $o(n)$ factor approximation for poset dimension was also mentioned as an open problem by Felsner et al. (2017).

In the second part of this paper, parameterized approximation algorithms for boxicity using various edit distance parameters are derived. We also present a parameterized approximation scheme for cubicity, using minimum vertex cover number as the parameter.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Let $G(V, E)$ be a graph. If I_1, I_2, \dots, I_k are (unit) interval graphs on the vertex set V such that $E(G) = E(I_1) \cap E(I_2) \cap \dots \cap E(I_k)$, then $\{I_1, I_2, \dots, I_k\}$ is called a box (cube) representation of G of dimension k . Boxicity (cubicity) of a non-complete graph G , denoted by $\text{box}(G)$ (respectively, $\text{cub}(G)$), is defined as the minimum integer k such that G has a box (cube) representation

[☆] A part of this work was done at IISc Bangalore. A preliminary version of this work appeared in IPEC 2012.

* Corresponding author.

E-mail addresses: abhijin@vbi.vt.edu (A. Adiga), jasine@iitpkd.ac.in (J. Babu), sunil@csa.iisc.ernet.in (L.S. Chandran).

Table 1

Parameterized approximations for computing boxicity and corresponding box representations using [Theorems 3](#) and [4](#).

Parameter k	Approximation guarantee	Running time
Interval completion No.	Additive 2	$2^{O(k^2 \log k)} n^{O(1)}$
Feedback vertex set size	$2 + \frac{2}{\text{box}(G)}$ factor	$2^{O(k^2 \log k)} n^{O(1)}$
Proper interval vertex deletion No.	$2 + \frac{1}{\text{box}(G)}$ factor	$2^{O(k^2 \log k)} n^{O(1)}$
Proper interval edge deletion No.	Additive 2	$2^{O(k^2 \log k)} n^{O(1)}$
Planar vertex deletion No.	$2 + \frac{3}{\text{box}(G)}$ factor	$f(k)n^{O(1)}$
Crossing number	Additive 6	$f(k)n^{O(1)}$
Planar edge deletion No.	Additive 6	$f(k)n^{O(1)}$
Max leaf No.	Additive 2	$2^{O(k^2 \log k)} n^{O(1)}$

of dimension k . For a complete graph, it is defined to be zero. Equivalently, boxicity (cubicity) is the minimum integer k such that G can be represented as the intersection graph of axis parallel boxes (cubes) in \mathbb{R}^k . Boxicity was introduced by Roberts [29] in 1969 for modeling problems in social sciences and ecology. Some well known NP-hard problems like the max-clique problem are polynomial time solvable, if low dimensional box representations are known [30].

For any graph G on n vertices, $\text{box}(G) \leq \lfloor \frac{n}{2} \rfloor$ and $\text{cub}(G) \leq \lfloor \frac{2n}{3} \rfloor$. Upper bounds of boxicity in terms of parameters like maximum degree [2] and tree-width [12] are known. It was shown by Scheinerman [31] in 1984 that the boxicity of outer planar graphs is at most two. In 1986, Thomassen [34] proved that the boxicity of planar graphs is at most 3.

Computation of boxicity is a notoriously hard problem. Even for $k = 2$ or 3, deciding whether boxicity (resp. cubicity) of a graph is at most k is NP-complete [37,24,6]. Recently, Chalermsook et al. [10] proved that no polynomial time algorithm for approximating boxicity of bipartite graphs with approximation factor within $O(n^{1-\epsilon})$ for any $\epsilon > 0$ is possible unless $\text{NP} = \text{ZPP}$. Same non-approximability holds in the case of split graphs and co-bipartite graphs too. Since cubicity and boxicity are equal for co-bipartite graphs, these hardness results extend to cubicity as well.

Boxicity is also closely related to other dimensional parameters like poset dimension, interval dimension, threshold dimension, minimum chain cover number of bipartite graphs, and Ferrers dimension of digraphs [13,25,37]. These parameters also have $O(n^{1-\epsilon})$ approximation hardness results for $\epsilon > 0$, assuming $\text{NP} \neq \text{ZPP}$. Further, unless $\text{NP} \subseteq \text{ZPTIME}(n^{\text{poly log } n})$, for any $\gamma > 0$ there is no $n/2^{(\log n)^{3/4+\gamma}}$ factor approximation algorithm for any of these problems including boxicity and cubicity [10] (for more details, see Section 5.1).

Main results

1. If G is a graph on n vertices, containing a clique of size $n - k$ or more, then $\text{box}(G)$ and an optimal box representation of G can be computed in time $n^2 2^{O(k^2 \log k)}$.
2. Using the above result, we derive a polynomial time $2 \lceil n \sqrt{\log \log n} / \sqrt{\log n} \rceil$ factor approximation algorithm for computing boxicity and a $2 \lceil n (\log \log n)^{\frac{3}{2}} / \sqrt{\log n} \rceil$ factor approximation algorithm for computing the cubicity. To our knowledge, no approximation algorithms for approximating boxicity and cubicity of general graphs within $o(n)$ factor were known previously.
3. The above algorithms also give us the corresponding box (resp. cube) representations. As a special case, this answers the question posed by Spinrad [33] about polynomial time construction of $o(n)$ dimensional box representations for boxicity 2 graphs in the affirmative.
4. As a consequence of our $o(n)$ factor approximation algorithm for boxicity, we derive polynomial time $o(n)$ factor approximation algorithms for computing several related parameters: poset dimension, interval dimension of finite posets, minimum chain cover of bipartite graphs, Ferrers dimension of digraphs, and threshold dimension of split graphs and co-bipartite graphs. These algorithms seem to be the first $o(n)$ factor approximation algorithms known for each of these problems. We note that obtaining an $o(n)$ factor approximation algorithm for poset dimension was described as an open problem in Felsner et al. [21].
5. In the second half of this paper, we discuss a general method for obtaining some parameterized approximation algorithms for boxicity, using various edit distance parameters. A summary of the results derived using this method are given in [Table 1](#). Notice that, if the parameter value is below $\sqrt{\log n} / \sqrt{\log \log n}$, the corresponding algorithms run in time polynomial in n , in the case of the first four parameters mentioned in the table.
6. A $(1 + \epsilon)$ -factor FPT approximation scheme for computing the cubicity of graphs using vertex cover number as the parameter is derived, for any $\epsilon > 0$, by allowing the running time of the algorithm to vary depending on ϵ and k .

2. Outline

In Section 3, we discuss some basic properties of boxicity which will be required in later sections. In Section 4, an algorithm for computing the boxicity and optimal box representations of graphs with large cliques is developed. This algorithm is one

of the main tools that is used in deriving many of our other algorithms in this paper. In Section 5, $o(n)$ factor approximation algorithms for boxicity and cubicity are developed and some corollaries of these results are derived. In Section 6, we develop two algorithms which provide general frameworks for obtaining parameterized approximation algorithms for boxicity with vertex (edge) edit distance parameters. Various parameterized approximation algorithms for boxicity which can be derived as corollaries of these two general algorithms are also discussed in this section. In Section 7, we discuss parameterized approximations for cubicity, using minimum vertex cover number as the parameter.

3. Prerequisites

In this section, we give some basic facts necessary for the later part of this paper. For a vertex $v \in V$ of a graph G , we use $N_G(v)$ to denote the set of neighbors of v in G . We use $G[S]$ to denote the induced subgraph of $G(V, E)$ on the vertex set $S \subseteq V$. Unless specified otherwise, graphs discussed in this paper are assumed to be undirected and without self-loops or parallel edges. The *intersection* of two graphs $G_1(V, E_1), G_2(V, E_2)$ is the graph with vertex set V and edge set $E_1 \cap E_2$ and is denoted as $G_1 \cap G_2$. Similarly, the *union* of two graphs $G_1(V, E_1), G_2(V, E_2)$ is the graph with vertex set V and edge set $E_1 \cup E_2$ and is denoted as $G_1 \cup G_2$. If $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$ are two graphs with $V_1 \subseteq V_2$ and $E_1 \subseteq E_2$, then G_1 is a *subgraph* of G_2 and G_2 is a *supergraph* of G_1 . We use the notation $G_1 \subseteq G_2$ to denote G_1 is a subgraph of G_2 and use $G_1 \supseteq G_2$ to denote G_1 is a supergraph of G_2 .

If I is an interval representation of an interval graph $G(V, E)$, we use $l_v(I)$ and $r_v(I)$, respectively, to denote the left and right end points of the interval corresponding to $v \in V$ in I . The interval corresponding to v is denoted as $[l_v(I), r_v(I)]$. We interchangeably use the same symbol for representing an interval graph and its interval representation, when the meaning of the usage is clear from the context. We use $\tau(n)$ to denote $n\sqrt{\log \log n} / \sqrt{\log n}$.

From an interval representation I of an interval graph G , it is easy to locally modify the interval end points and derive another interval representation of G in which the $2|V|$ interval end points are all distinct. Hence, every interval graph has an interval representation in which all $2|V|$ interval end points are distinct. Moreover, it is easy to observe that any interval representation of an interval graph with distinct end points induces an ordering of the $2|V|$ end points. From these observations, it can be concluded that there are at most $(2n)! = 2^{O(n \log n)}$ interval graphs on n vertices and hence, there are at most $\binom{2n!}{b} = 2^{O(nb \log n)}$ distinct b -dimensional box representations of a graph G on n vertices. Further, all these box representations can be enumerated in time $2^{O(nb \log n)}$ ensuring that in every box representation B of G enumerated, in the interval representation of each interval graph in B , the $2n$ interval end points are distinct [4, Proposition 1]. This enumeration gives a brute force method to obtain an optimal box representation of G . In linear time, it is also possible to check whether a given graph is a unit interval graph and if so, generate a unit interval representation of it [5]. Hence, we can generalize [4, Proposition 1] as stated below.

Proposition 1. *If $G(V, E)$ is a graph on n vertices and boxicity b , an optimal box representation of G can be computed in $2^{O(nb \log n)}$ time, ensuring that in the interval representation of each interval graph in the box representation, the $2n$ interval end points are all distinct. If $G(V, E)$ is a graph on n vertices and cubicity b , an optimal cube representation of G can be computed in $2^{O(nb \log n)}$ time.*

Lemma 1 (Roberts [29]). *Let $G(V, E)$ be any graph. For any $x \in V$, $\text{box}(G) \leq 1 + \text{box}(G \setminus \{x\})$.*

The following lemma is based on a well-known technique of producing box representations.

Lemma 2. *Let $G(V, E)$ be a graph on n vertices. Let $S \subseteq V$ be such that $\forall v \in V \setminus S$ and $u \in V$ such that $u \neq v$, $(u, v) \in E$. If a k -dimensional box representation B_S of $G[S]$ is known, then, in $O(kn)$ time we can construct a box representation B of G of dimension $|B_S|$. Moreover, $\text{box}(G) = \text{box}(G[S])$.*

We include a proof of this lemma, for the sake of completeness.

Proof. Let $B_S = \{I_1, I_2, \dots, I_p\}$ be a box representation of $G[S]$. For $1 \leq i \leq p$, let $l_i = \min_{u \in S} l_u(I_i)$ and $r_i = \max_{u \in S} r_u(I_i)$. For $1 \leq i \leq p$ let I'_i be the interval graph on vertex set V obtained by assigning to each vertex $v \in V$ the interval

$$[l_v(I'_i), r_v(I'_i)] = \begin{cases} [l_v(I_i), r_v(I_i)] & \text{if } v \in S, \\ [l_i, r_i] & \text{if } v \in V \setminus S. \end{cases}$$

It is easy to see that $B_2 = \{I'_1, I'_2, \dots, I'_p\}$ is a box representation of G and $\text{box}(G) \leq \text{box}(G[S])$. Since $G[S]$ is an induced subgraph of G , we also have $\text{box}(G) \geq \text{box}(G[S])$. The whole construction can be done in $O(kn)$ time. \square

Lemma 3. *Let $G(V, E)$ be a graph on n vertices and let $A \subseteq V$. Let $G_1(V, E_1)$ be a supergraph of G with $E_1 = E \cup \{(x, y) \mid x, y \in A, x \neq y\}$. If a box representation B of G is known, then in $O(n|B|)$ time we can construct a box representation B_1 of G_1 of dimension $2|B|$. In particular, $\text{box}(G_1) \leq 2 \text{box}(G)$.*

Proof. Let $\mathcal{B} = \{I_1, I_2, \dots, I_b\}$ be a box representation of G . For each $1 \leq i \leq b$, let $l_i = \min_{u \in V} l_u(I_i)$ and $r_i = \max_{u \in V} r_u(I_i)$. For $1 \leq i \leq b$, let I_{i_1} be the interval graph on vertex set V obtained by assigning to each vertex $v \in V$ the interval

$$[l_v(I_{i_1}), r_v(I_{i_1})] = \begin{cases} [l_i, r_v(I_i)] & \text{if } v \in A, \\ [l_v(I_i), r_v(I_i)] & \text{if } v \in V \setminus A, \end{cases}$$

and let I_{i_2} be the interval graph on vertex set V obtained by assigning to each vertex $v \in V$ the interval

$$[l_v(I_{i_2}), r_v(I_{i_2})] = \begin{cases} [l_v(I_i), r_i] & \text{if } v \in A, \\ [l_v(I_i), r_v(I_i)] & \text{if } v \in V \setminus A. \end{cases}$$

It is easy to see that this construction can be done in $O(nb)$ time. For an illustration of the above construction, refer to Fig. 1.

Note that, in constructing I_{i_1} and I_{i_2} we have only extended some of the intervals of I_i and therefore, I_{i_1} and I_{i_2} are supergraphs of I_i and in turn of G . By construction, A induces cliques in both I_{i_1} and I_{i_2} , and thus they are supergraphs of G_1 too.

Now, consider $(u, v) \notin E$ with $u \in V \setminus A, v \in A$. Then $\exists i \in \{1, 2, \dots, b\}$ such that either $r_v(I_i) < l_u(I_i)$ or $r_u(I_i) < l_v(I_i)$. If $r_v(I_i) < l_u(I_i)$, then clearly the intervals $[l_i, r_v(I_i)]$ and $[l_u(I_i), r_u(I_i)]$ do not intersect and thus $(u, v) \notin E(I_{i_1})$. Similarly, if $r_u(I_i) < l_v(I_i)$, then $(u, v) \notin E(I_{i_2})$. If both $u, v \in V \setminus A$ and $(u, v) \notin E$, then $\exists i$ such that $(u, v) \notin E(I_i)$ for some $1 \leq i \leq b$ and clearly by construction, $(u, v) \notin E(I_{i_1})$ and $(u, v) \notin E(I_{i_2})$.

It follows that $G_1 = \bigcap_{1 \leq i \leq b} I_{i_1} \cap I_{i_2}$ and $\mathcal{B}_1 = \{I_{i_1}, I_{i_2}, I_{2_1}, I_{2_2}, \dots, I_{b_1}, I_{b_2}\}$ is a box representation of G_1 of dimension $2b$. If $|\mathcal{B}| = \text{box}(G)$ to start with, then we get $|\mathcal{B}'| \leq 2 \text{box}(G)$. Therefore, $\text{box}(G_1) \leq 2 \text{box}(G)$. \square

If $S \subseteq V$ induces a clique in G , then it is easy to see that the intersection of all the intervals in I corresponding to vertices of S is nonempty. This property is referred to as the *Helly property of intervals* and we refer to this common region of intervals as the *Helly region* of the clique S .

Definition 1. Let $G(V, E)$ be a graph in which $S \subseteq V$ induces a clique in G . Let $H(V, E')$ be an interval supergraph of G . Let p be a point on the real line. If H has an interval representation I satisfying the following conditions:

- (1) p belongs to the Helly region of S in I .
- (2) The end points of intervals corresponding to vertices of $V \setminus S$ are all distinct in I .
- (3) For each $v \in S$,

$$\begin{aligned} l_v(I) &= \min \left(p, \min_{u \in N_G(v) \cap (V \setminus S)} r_u(I) \right) \text{ and} \\ r_v(I) &= \max \left(p, \max_{u \in N_G(v) \cap (V \setminus S)} l_u(I) \right) \end{aligned}$$

then we call I a *nice interval representation of H with respect to S and p* . If H has a nice interval representation with respect to clique S and some point p , then H is called a *nice interval supergraph of G with respect to clique S* .

Fig. 2 shows some nice interval supergraphs of a graph along with their corresponding nice interval representations.

Lemma 4. Let $G(V, E)$ be a graph in which $S \subseteq V$ induces a clique in G . For every interval supergraph H of G , we can derive a graph H' such that $H \supseteq H' \supseteq G$ and H' a nice interval supergraph of G with respect to S .

Proof. Without loss of generality, we can assume that all $2|V|$ interval end points are distinct in (the interval representation of) H . (Otherwise, we can always alter the end points locally and make them distinct.) Let $p \in \mathbb{R}$ be a point belonging to the Helly region corresponding to S in H . Let H' be the interval graph on vertex set V obtained by assigning to each vertex $v \in V$ the interval

$$[l_v(H'), r_v(H')] = \begin{cases} [l_v(H), r_v(H)] & \text{if } v \in V \setminus S, \\ [l'_v, r'_v] & \text{if } v \in S, \end{cases}$$

where $l'_v = \min \left(p, \min_{u \in N_G(v) \cap (V \setminus S)} r_u(H) \right)$ and $r'_v = \max \left(p, \max_{u \in N_G(v) \cap (V \setminus S)} l_u(H) \right)$. Note that since $l'_v \leq p \leq r'_v$, the interval $[l'_v, r'_v]$ is well-defined.

We claim that $H \supseteq H' \supseteq G$. Since for any vertex $v \in V$, the interval of v in H contains the interval of v in H' , we have $H \supseteq H'$. It directly follows from the definition of H' that $H'[V \setminus S] = H[V \setminus S]$. For any $(u, v) \in E(G)$, with $u \in V \setminus S$ and $v \in S$, the interval of v intersects the interval of u in H' , by the definition of $[l'_v, r'_v]$. Intervals of vertices of S share the common point p in H' . Thus, $H \supseteq H' \supseteq G$. Now, from the definition of H' it follows that it is a nice interval supergraph of G with respect to the clique S and point p . \square

Fig. 2 gives an illustration of the above construction.

Corollary 1. If $G(V, E)$ has a box representation \mathcal{B} of dimension b and $S \subseteq V(G)$ induces a clique in G , then G also has a box representation \mathcal{B}' of the same dimension, in which $\forall H' \in \mathcal{B}', H'$ is a nice interval supergraph of G with respect to S .

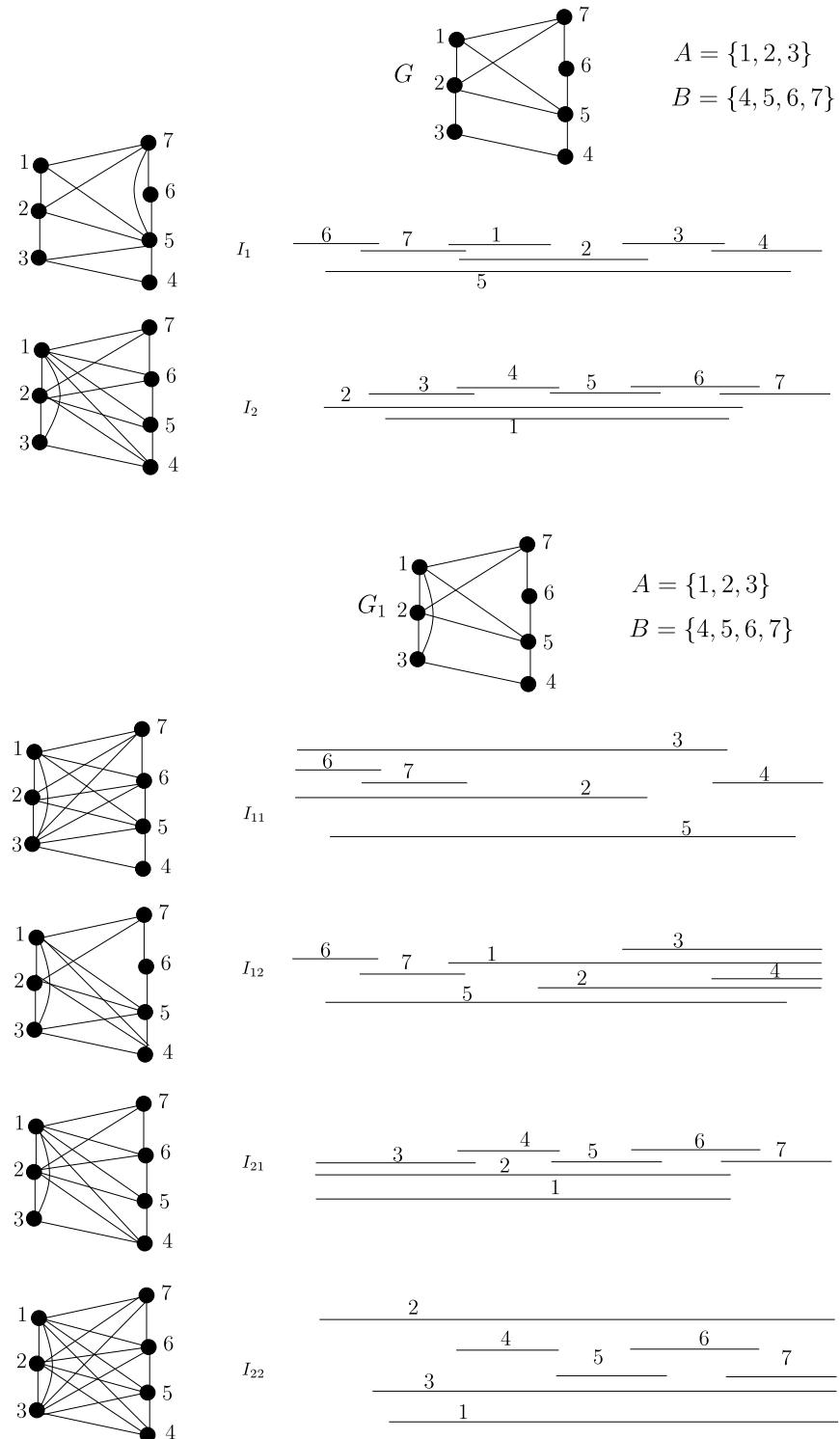


Fig. 1. A graph G , and a supergraph G_1 of G are shown. A two-dimensional box representation $\mathcal{B} = \{I_1, I_2\}$ of G and a four-dimensional box representation $\mathcal{B}' = \{I_{11}, I_{12}, I_{21}, I_{22}\}$ of G_1 derived from \mathcal{B} , using the construction given in the proof of Lemma 3 are also shown.

Proof. Let $\mathcal{B} = \{H_1, H_2, \dots, H_b\}$ be a box representation of G . For each $1 \leq i \leq b$, let H'_i be the nice interval supergraph of G with respect to S , derived from H_i , as stated in Lemma 4. Since, by Lemma 4 we have $H_i \supseteq H'_i \supseteq G$, for each $1 \leq i \leq b$, it

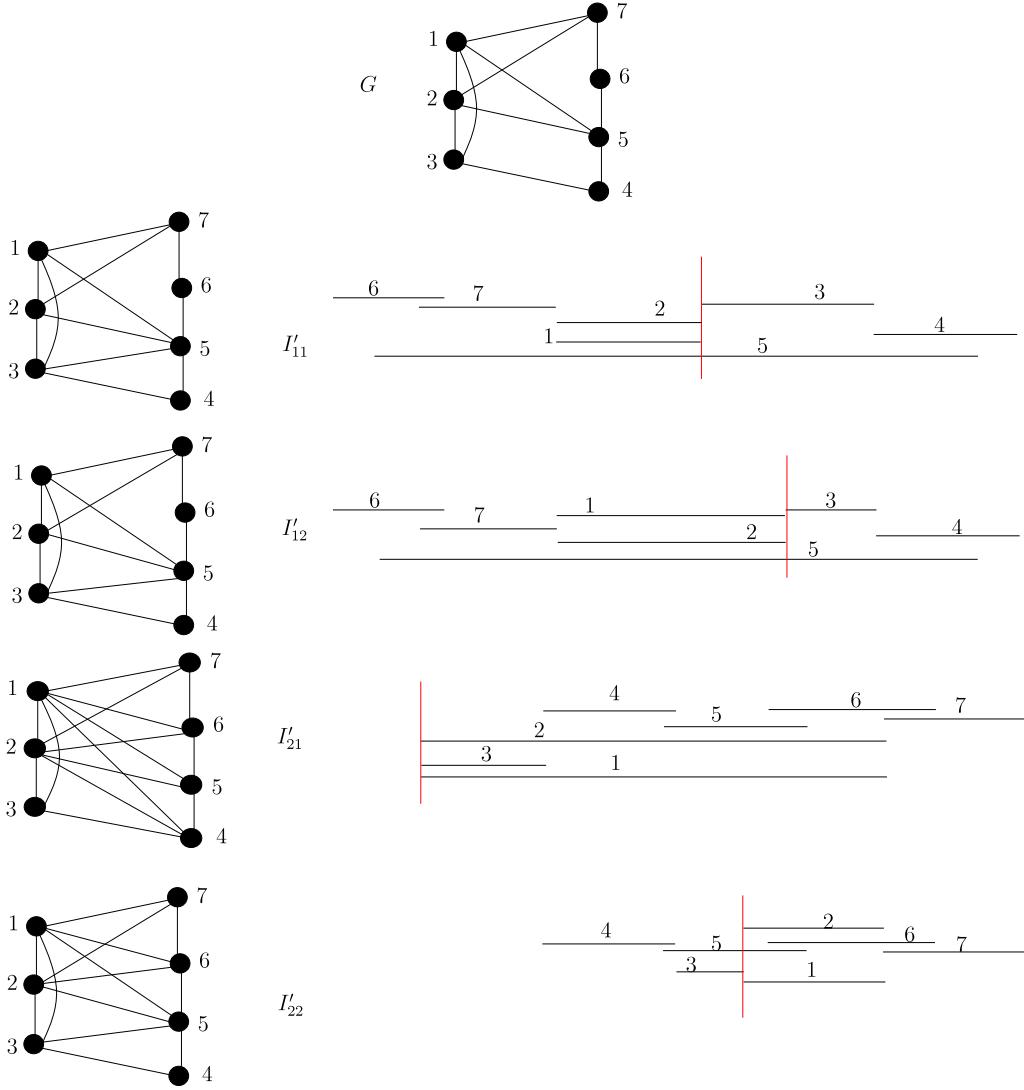


Fig. 2. $I'_{11}, I'_{12}, I'_{21}, I'_{22}$ are nice interval supergraphs of G with respect to the clique $S = \{1, 2, 3\}$. These nice interval supergraphs are, respectively, derived from the interval supergraphs $I_{11}, I_{12}, I_{21}, I_{22}$ of G given in Fig. 1 using the construction given in the proof of Lemma 4. In each interval representation, a point belonging to the Helly region of clique S is indicated by a vertical line. $\mathcal{B}' = \{I'_{11}, I'_{12}, I'_{21}, I'_{22}\}$ is a box representation of G .

follows that $\mathcal{B}' = \{H'_1, H'_2, \dots, H'_b\}$ is also a box representation of G . Notice that \mathcal{B}' satisfies our requirement. An illustration of this corollary is given in Fig. 2. \square

Lemma 5. Let G be a graph on n vertices, with its vertices arbitrarily labeled as $1, 2, \dots, n$. If G contains a clique of size $n - k$ or more, then:

- (a) A subset $A \subseteq V$ such that $|A| \leq k$ and $G[V \setminus A]$ is a clique, can be computed in $O(n2^k + n^2)$ time.
- (b) There are at most $2^{O(k \log k)}$ nice interval supergraphs of G with respect to the clique $V \setminus A$. These can be enumerated in $n^2 2^{O(k \log k)}$ time.

Proof.

- (a) We know that, if G contains a clique of size $n - k$ or more, then the complement graph \bar{G} has a vertex cover of size at most k . We can compute \bar{G} in $O(n^2)$ time and a minimum vertex cover A of \bar{G} in $O(n2^k)$ time [27]. We have $|A| \leq k$ and $G[V \setminus A]$ is a clique because $V \setminus A$ is an independent set in \bar{G} .
- (b) Let H be any nice interval supergraph of G with respect to $V \setminus A$. Let I be a nice interval representation of H with respect to $V \setminus A$ and a point p . Let P be the set of end points (both left and right) of the intervals corresponding to vertices of A in

H . Clearly $|P| = 2|A| \leq 2k$. The order of end points of vertices of A in I from left to right corresponds to a permutation of elements of P and therefore, there are at most $(2k)!$ possibilities for this ordering. Moreover, note that the points of P divide the real line into $|P| + 1$ regions and that p can belong to any of these regions. From the definition of nice interval representation, it is clear that, once the point p and the end points of vertices of A are fixed, the end points of vertices in $V \setminus A$ get automatically decided.

Thus, to enumerate every nice interval supergraph H of G with respect to clique $V \setminus A$, it is enough to enumerate all the $(2k)! = 2^{O(k \log k)}$ permutations of elements of P and consider $|P| + 1 \leq 2k + 1$ possible placements of p in each of them. Some of these orderings may not produce an interval supergraph of G though. In $O(n^2)$ time, we can check whether the resultant graph is an interval supergraph of G and output the interval representation. The number of supergraphs enumerated is only $(2k + 1)2^{O(k \log k)} = 2^{O(k \log k)}$. \square

It may be noted that since the vertices of G are labeled in [Lemma 5](#), we can retain the same labeling of vertices in the definition and construction of nice interval supergraphs of G , while using the method described in the proof of [Lemma 5](#). Therefore, we have the following:

Remark 1. By construction, vertices of the nice interval supergraphs obtained by the proof of [Lemma 5\(b\)](#) retain their original labels as in G .

4. Boxicity of graphs with large cliques

One of the central ideas in this paper is the following theorem about computing the boxicity of graphs which contain very large cliques. Using this theorem, in Section 5 we derive $o(n)$ factor approximation algorithms for computing the boxicity and cubicity of graphs.

Theorem 1. Let G be a graph on n vertices, containing a clique of size $n - k$ or more. Then, $\text{box}(G) \leq k$ and an optimal box representation of G can be found in time $n^2 2^{O(k^2 \log k)}$.

Proof. Let $G(V, E)$ be a graph on n vertices containing a clique of size $n - k$ or more. We can assume that G is not a complete graph; otherwise, the problem becomes trivial. Arbitrarily label the vertices of G as $1, 2, \dots, n$. Using part (a) of [Lemma 5](#), we can compute in $O(n2^k + n^2)$ time, $A \subseteq V$ such that $|A| \leq k$ and $G[V \setminus A]$ is a clique. It is easy to infer from [Lemma 1](#) that $\text{box}(G) \leq \text{box}(G \setminus A) + |A| = k$, since $\text{box}(G \setminus A) = 0$ by definition.

Let \mathcal{F} be the family of all nice interval supergraphs of G with respect to the clique $V \setminus A$. By [Corollary 1](#), if $\text{box}(G) = b$, then there exists a b -dimensional nice box representation of G , i.e., a box representation $\mathcal{B}' = \{I'_1, I'_2, \dots, I'_b\}$ of G in which $I'_i \in \mathcal{F}$, for each $1 \leq i \leq b$. By part (b) of [Lemma 5](#), $|\mathcal{F}| = 2^{O(k \log k)}$ and all graphs in \mathcal{F} can be enumerated in $n^2 2^{O(k \log k)}$ time. Given an integer d , $1 \leq d \leq b$, verifying whether there exists a d -dimensional nice box representation of G , and producing if one exists, can be done in $n^2 2^{O(k \cdot d \log k)}$ time, as follows: consider every subfamily $\mathcal{F}' \subseteq \mathcal{F}$ with $|\mathcal{F}'| = d$ and check if \mathcal{F}' gives a valid box representation of G (this validation is straightforward because vertices of supergraphs in \mathcal{F}' retain their original labels as in G , as explained in [Remark 1](#)). We might have to repeat this process for $1 \leq d \leq b$ in that order, to identify the optimum dimension b . Hence the total time required to compute an optimal box representation of G is $bn^2 2^{O(k \cdot b \log k)}$, which is $n^2 2^{O(k^2 \log k)}$, because $b \leq k$ by the first part of this theorem. \square

5. Approximation algorithms for computing boxicity and cubicity

In this section, we use [Theorem 1](#) and derive an $o(n)$ factor approximation algorithms for boxicity and cubicity. Let $G(V, E)$ be the given graph with $|V| = n$. Without loss of generality, we can assume that G is connected. Recall the notation, $\tau(n) = n\sqrt{\log \log n} / \sqrt{\log n}$. Let $k = \lceil n/\tau(n) \rceil$ and $t = \lceil \tau(n) \rceil \geq \lceil \frac{n}{k} \rceil$. The algorithm proceeds by defining t supergraphs of G and computing their optimal box representations. Let the vertex set V be partitioned arbitrarily into t sets V_1, V_2, \dots, V_t where $|V_i| \leq k$, for each $1 \leq i \leq t$. We define supergraphs G_1, G_2, \dots, G_t of G with $G_i(V, E_i)$ defined by setting $E_i = E \cup \{(x, y) \mid x, y \in V \setminus V_i \text{ and } x \neq y\}$, for $1 \leq i \leq t$.

Lemma 6. Let G_i be as defined above, for $1 \leq i \leq t$. An optimal box representation \mathcal{B}_i of G_i can be computed in $n^{O(1)}$ time, where $n = |V|$.

Proof. Noting that $G[V \setminus V_i]$ is a clique and $|V_i| \leq k$, by [Theorem 1](#), we can compute an optimal box representation \mathcal{B}_i of G_i in $n^2 2^{O(k^2 \log k)}$ time, where $n = |V|$. By the definition of k , we have $n^2 2^{O(k^2 \log k)} = n^{O(1)}$. \square

Lemma 7. Let \mathcal{B}_i be as computed above, for $1 \leq i \leq t$. Then, $\mathcal{B} = \bigcup_{1 \leq i \leq t} \mathcal{B}_i$ is a valid box representation of G such that $|\mathcal{B}| \leq t' \text{box}(G)$, where t' is $2 \lceil \tau(n) \rceil$. The box representation \mathcal{B} is computable in $n^{O(1)}$ time.

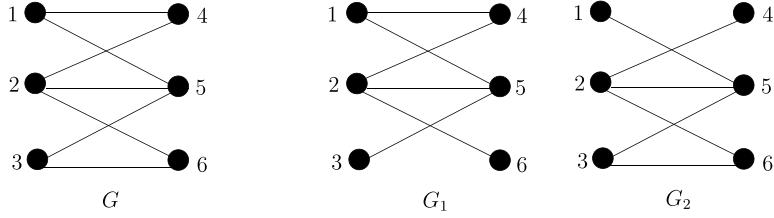


Fig. 3. A bipartite graph G and its chain cover.

Proof. We can compute optimal box representations \mathcal{B}_i of G_i , for $1 \leq i \leq t = \lceil \tau(n) \rceil$ as explained in [Lemma 6](#) in total $n^{O(1)}$ time. Observe that $E(G) = E(G_1) \cap E(G_2) \cap \dots \cap E(G_t)$. Therefore, it is a trivial observation that the union $\mathcal{B} = \bigcup_{1 \leq i \leq t} \mathcal{B}_i$ gives us a valid box representation of G .

We will prove that this representation gives the approximation ratio as required. By [Lemma 3](#) we have, $|\mathcal{B}_i| = \text{box}(G_i) \leq 2 \text{box}(G)$. Hence, $|\mathcal{B}| = \sum_{i=1}^t |\mathcal{B}_i| \leq 2t \text{box}(G)$. \square

The box representation \mathcal{B} obtained from [Lemma 7](#) can be extended to a cube representation \mathcal{C} of G as stated in the following lemma.

Lemma 8. *A cube representation \mathcal{C} of G , such that $|\mathcal{C}| \leq t' \text{cub}(G)$, where t' is $2 \lceil \tau(n) \log \log n \rceil$, can be computed in $n^{O(1)}$ time.*

Proof. We can compute optimal box representations \mathcal{B}_i of G_i , for $1 \leq i \leq t = \lceil \tau(n) \rceil$ as explained in [Lemma 6](#) in $O(n^4)$ time. By [3, Corollary 1] we know that, from the optimal box representation \mathcal{B}_i of G_i we can construct a cube representation \mathcal{C}_i of G_i of dimension $\text{box}(G_i) \lceil \log \alpha(G_i) \rceil$, where $\alpha(G_i)$ is the independence number of G_i which is at most $|V_i|$. The construction of cube representation given in [3] can be done in $n^{O(1)}$ time.

(Recall the assumption that G is connected.)

It is easy to see that $\mathcal{C} = \bigcup_{1 \leq i \leq t} \mathcal{C}_i$ gives us a valid cube representation of G . We will prove that this cube representation gives the approximation ratio as required. We have,

$$|\mathcal{C}| = \sum_{i=1}^t |\mathcal{C}_i| \leq \sum_{i=1}^t |\mathcal{B}_i| \lceil \log \alpha(G_i) \rceil \leq \sum_{i=1}^t |\mathcal{B}_i| \lceil \log k \rceil \leq 2t \text{box}(G) \log \log n \leq 2t \log \log n \text{cub}(G). \quad \square$$

Combining [Lemmas 7](#) and [8](#), we get the following theorem which gives $o(n)$ factor approximation algorithms for computing boxicity and cubicity.

Theorem 2. *Let $G(V, E)$ be a graph on n vertices. Then a box representation \mathcal{B} of G , such that $|\mathcal{B}| \leq t \text{box}(G)$, where t is $2 \lceil \tau(n) \rceil$, can be computed in polynomial time. Further, a cube representation \mathcal{C} of G , such that $|\mathcal{C}| \leq t' \text{cub}(G)$, where t' is $2 \lceil \tau(n) \log \log n \rceil$, can also be computed in polynomial time.*

5.1. Consequences of [Theorem 2](#)

Now, we describe how [Theorem 2](#) can be used to derive sublinear approximation algorithms for some well-known problems whose computational complexity is closely related to that of boxicity.

Chain cover of bipartite graphs. A bipartite graph is a *chain graph*, if it does not contain an induced matching of size 2. Given a bipartite graph $G(V, E)$, the *minimum chain cover number* of G , denoted by $\text{ch}(G)$ is the smallest number of chain graphs on the vertex set V such that the union of their edge sets is $E(G)$. [Fig. 3](#) shows a bipartite graph G and two chain graphs G_1 and G_2 such that $E(G_1) \cup E(G_2) = E(G)$. Therefore, $\text{ch}(G) \leq 2$. Since G contains an induced matching of size 2, $\text{ch}(G) > 1$ and from this, it follows that $\text{ch}(G) = 2$.

It is well-known that G is a chain graph if and only if its complement is a co-bipartite interval graph [37, Lemma 4]. From this, it immediately follows that $\text{ch}(G) = \text{box}(\bar{G})$ [37, Lemma 4]. For an illustration of this fact, see [Figs. 3 and 4](#).

Corollary 2. *There is a polynomial time $2 \lceil \tau(n) \rceil$ factor approximation algorithm to compute the minimum chain cover number of an n -vertex bipartite graph.*

Threshold dimension of split graphs. The concept of threshold graphs and threshold dimension was introduced by Chvátal and Hammer [14] while studying some set-packing problems. A graph $G(V, E)$ is called a *threshold graph* if there exists $s \in \mathbb{R}$ and a labeling of vertices $w : V \mapsto \mathbb{R}$ such that $\forall u, v \in V, (u, v) \in E \Leftrightarrow w(u) + w(v) \geq s$. Threshold graphs are also characterized

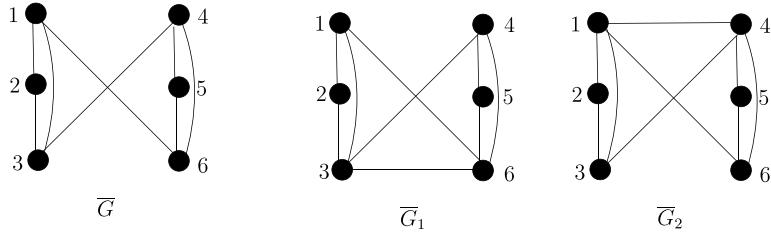


Fig. 4. Co-bipartite graphs \bar{G} , \bar{G}_1 , \bar{G}_2 are the respective complements of graphs G , G_1 , G_2 of Fig. 3. It is easy to verify that \bar{G}_1 , \bar{G}_2 are interval graphs and $\{\bar{G}_1, \bar{G}_2\}$ is a box representation of \bar{G} .

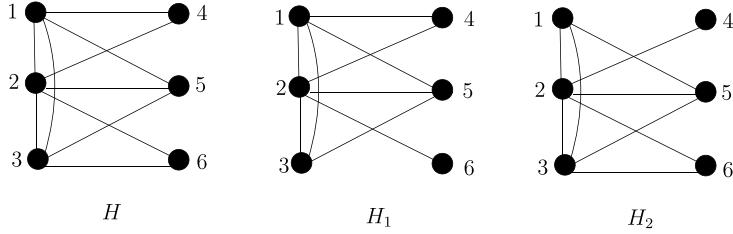


Fig. 5. A split graph H and its threshold cover. Note that bipartite graphs G , G_1 , and G_2 of Fig. 3 are obtained, respectively, from split graphs H , H_1 and H_2 , by converting the clique side of the split graphs into independent sets.

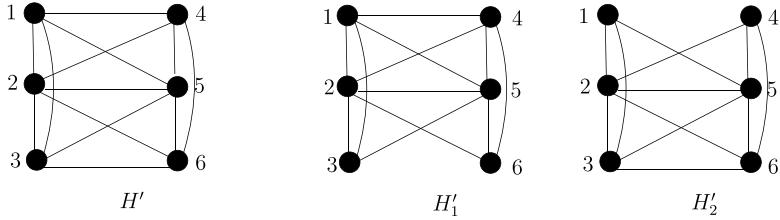


Fig. 6. A co-bipartite graph H' and its threshold cover. Note that split graphs H , H_1 , and H_2 of Fig. 5 are obtained, respectively, from co-bipartite graphs H' , H'_1 and H'_2 , by converting one of the clique sides of the split graphs into an independent set.

as graphs which have neither an induced matching of size 2, nor an induced path on 4 vertices nor an induced cycle on 4 vertices [14].

The *threshold dimension* of G , denoted by $t(G)$ is the minimum number of threshold subgraphs required to cover $E(G)$. Fig. 5 shows a graph H and two threshold graphs H_1 and H_2 that cover $E(H)$. Since H contains an induced path on 4 vertices, H is not a threshold graph and therefore $t(H) = 2$. Even for split graphs, threshold dimension is hard to approximate within an $O(n^{1-\epsilon})$ factor for any $\epsilon > 0$, unless $\text{NP} = \text{ZPP}$ [10,25].

Corollary 3. *There is a polynomial time $2 \lceil \tau(n) \rceil$ factor approximation algorithm to compute the threshold dimension of any split graph on n vertices.*

Proof. Given any split graph G , there is a polynomial time method to construct a bipartite graph H on the same vertex set such that $t(G) = \text{ch}(H)$ [25, Page 149, Lemma 7.3.4]. From the approximation algorithm for computing $\text{ch}(H)$, the result follows. \square

Threshold dimension of co-bipartite graphs. Cozzens et al. [16] showed that if G is a co-bipartite graph, an associated split graph G' on the same vertex set can be constructed in polynomial time, such that for any $k \geq 2$, $t(G) \leq k$ if and only if $t(G') \leq k$ (see Fig. 6).

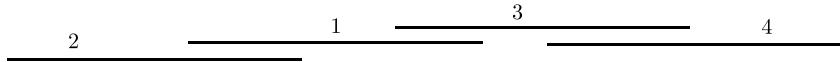
This reduction shows that the hardness result of threshold dimension of split graphs is also applicable for the threshold dimension of co-bipartite graphs. Moreover, we get the following.

Corollary 4. *There is a polynomial time $2 \lceil \tau(n) \rceil$ factor approximation algorithm to compute the threshold dimension of any co-bipartite graph on n vertices.*

Partial order dimension. This concept was introduced by Dushnik and Miller in 1941 [17]. A partially ordered set (poset) $\mathcal{P} = (X, P)$ consists of a nonempty set X and a binary relation P on X that is reflexive, antisymmetric and transitive. If every pair of

$$X = \{1, 2, 3, 4\}$$

$$I_1 = \{(1, 1), (2, 2), (3, 3), (4, 4), (1, 4), (2, 3), (2, 4)\}$$



$$X = \{1, 2, 3, 4\}$$

$$I_2 = \{(1, 1), (2, 2), (3, 3), (4, 4), (1, 3), (1, 4), (2, 3)\}$$

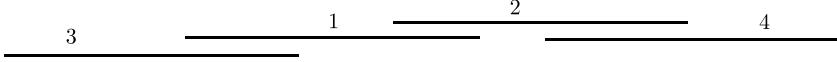


Fig. 7. Interval realization of two interval extensions (X, I_1) and (X, I_2) of the poset (X, P) , where $X = \{1, 2, 3, 4\}$ and $P = \{(1, 1), (2, 2), (3, 3), (4, 4), (1, 4), (2, 3)\}$.

distinct elements of X are comparable under the relation P , then (X, P) is called a *total order* or a *linear order*. A *linear extension* of a partial order (X, P) is a linear order (X, P') such that $\forall x, y \in X, (x, y) \in P \Rightarrow (x, y) \in P'$. For example, if $X = \{1, 2, 3, 4\}$, $P = \{(1, 1), (2, 2), (3, 3), (4, 4), (1, 2), (3, 4)\}$, $P_1 = \{(1, 1), (2, 2), (3, 3), (4, 4), (1, 2), (1, 3), (1, 4), (2, 3), (2, 4), (3, 4)\}$ and $P_2 = \{(1, 1), (2, 2), (3, 3), (4, 4), (1, 2), (3, 1), (4, 1), (3, 2), (4, 2), (3, 4)\}$, then (X, P) is a partial order and $(X, P_1), (X, P_2)$ are its linear extensions.

The *dimension* of a poset $\mathcal{P} = (X, P)$, denoted by $\dim(\mathcal{P})$ is defined as the smallest integer k such that \mathcal{P} can be expressed as the intersection of k linear extensions $(X, P_1), (X, P_2), \dots, (X, P_k)$ of \mathcal{P} : i.e., if $\forall x, y \in X, (x, y) \in P \Leftrightarrow (x, y) \in P_i$, for each $1 \leq i \leq k$. In the example given in previous paragraph, the partial order (X, P) is not a total order; however, it is the intersection of total orders (X, P_1) and (X, P_2) . Therefore, $\dim((X, P)) = 2$.

A *height-two poset* is a poset (X, P) in which all elements of X are either minimal elements or maximal elements under the relation P . Even in the case of height-two posets, partial order dimension is hard to approximate within an $O(n^{1-\epsilon})$ factor for any $\epsilon > 0$, unless $NP = ZPP$ [10]. A height-two poset $\mathcal{P} = (X, P)$ in which X_1 is the set of minimal elements and X_2 is the set of maximal elements can be associated with a bipartite graph $B(\mathcal{P})$ with vertex set X and edge set given by $\{(x, y) : x \in X_1, y \in X_2, (x, y) \notin P\}$ [25, Page 147].

Corollary 5. There is a polynomial time $O(\tau(n))$ factor approximation algorithm to compute the partial order dimension of a poset $\mathcal{P} = (X, P)$ defined on an n -element set X .

Proof. Let $\mathcal{P} = (X, P)$ be a poset with $|X| = n$. By a construction given by R. Kimble [35, Theorem 5], given a poset $\mathcal{P} = (X, P)$ of arbitrary height, we can construct a height-two poset $\mathcal{P}' = (Y, P')$ from $\mathcal{P} = (X, P)$ in polynomial time so that $\dim(\mathcal{P}) \leq \dim(\mathcal{P}') \leq 1 + \dim(\mathcal{P})$ and $|Y| = 2|X|$. It is also known that $\dim(\mathcal{P}) = \text{ch}(B(\mathcal{P}'))$ [37, Footnote, Page 354]. Therefore, by computing $\text{ch}(B(\mathcal{P}'))$ using the algorithm given by Corollary 2, we can compute a $O(\tau(n))$ approximation of $\dim(\mathcal{P})$. \square

Interval dimension of posets. A poset (X, P) is an *interval order*, if each $x \in X$ can be assigned an open interval (l_x, r_x) of the real line such that $(x, y) \in P$ for $x \neq y$ if and only if $r_x \leq l_y$. Fig. 7 shows assignment of intervals to vertices of two interval orders (X, I_1) and (X, I_2) .

An *interval order extension* of a partial order (X, P) is an interval order (X, P') such that $\forall x, y \in X, (x, y) \in P \Rightarrow (x, y) \in P'$. The *interval dimension* of a poset $\mathcal{P} = (X, P)$, denoted by $\text{idim}(\mathcal{P})$, is defined as the smallest integer k such that \mathcal{P} can be expressed as the intersection of k interval order extensions of \mathcal{P} . For example, consider $X = \{1, 2, 3, 4\}$ and the partial order (X, P) , where $P = \{(1, 1), (2, 2), (3, 3), (4, 4), (1, 4), (2, 3)\}$. It is not difficult to see that (X, P) is not an interval order. Fig. 7 shows two interval order extensions (X, I_1) and (X, I_2) of (X, P) such that their intersection is (X, P) . Therefore, $\text{idim}((X, P)) = 2$. Since linear orders are interval orders, it follows that $\text{idim}(\mathcal{P}) \leq \dim(\mathcal{P})$. On the other hand, the poset dimension of an interval order can be large.

Since the height-two¹ poset \mathcal{P}' given by Kimble's construction [35, 25] from an arbitrary finite poset \mathcal{P} satisfies $\dim(\mathcal{P}) = \text{ch}(B(\mathcal{P}'))$ [37, Footnote, Page 354] and $\text{ch}(B(\mathcal{P}')) = \text{idim}(\mathcal{P}')$ [37, Lemma 4], from the approximation hardness of poset-dimension [10], we can see that interval dimension is hard to approximate within an $O(n^{1-\epsilon})$ factor for any $\epsilon > 0$, unless $NP = ZPP$. Felsner et al. [20] showed that given a poset (X, P) , it is possible to construct another poset (Y, P') in polynomial time, such that $|Y| = 2|X|$ and $\text{idim}((X, P)) = \dim((Y, P'))$.

Corollary 6. There is a polynomial time $O(\tau(n))$ factor approximation algorithm for computing the interval dimension of any poset $\mathcal{P} = (X, P)$ defined on a set X of n elements.

¹ The notion of height-two posets as per our definition is the same as the notion of height-one posets in [37].

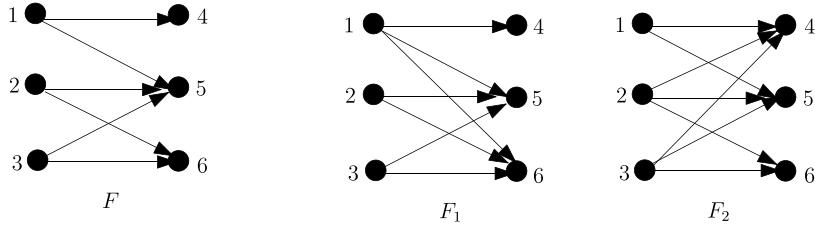


Fig. 8. A digraph F and two Ferrers digraphs F_1 and F_2 whose intersection is F .

Ferrers dimension of digraphs Ferrers relations were introduced by Riguet in 1950s [28]. A digraph $G(V, E)$ is called a *Ferrers digraph* if it does not contain vertices x, y, z, w (not necessarily distinct), satisfying $(x, y) \in E, (z, w) \in E$, but $(x, w) \notin E, (z, y) \notin E$ [25, Page 33]. The *Ferrers dimension* [15] of a digraph G is the smallest number of Ferrers digraphs whose intersection is G . Fig. 8 shows a digraph F and two Ferrers digraphs F_1 and F_2 whose intersection is F . Since vertices $\{1, 4, 2, 6\}$ in F violate the condition for a Ferrers digraph, the Ferrers dimension of F is 2.

Since a partial order \mathcal{P} has $\dim(\mathcal{P})$ equal to the Ferrers dimension of its underlying digraph [15], Ferrers dimension is also hard to approximate within an $O(n^{1-\epsilon})$ factor for any $\epsilon > 0$, unless $\text{NP} = \text{ZPP}$. Cogis [15] showed that given a digraph $G(V, E)$, a poset $\mathcal{P} = (X, P)$ can be constructed in polynomial time, such that $|X| \leq 2|V|$ and the poset dimension of \mathcal{P} is equal to the Ferrers dimension of G .

Corollary 7. *There is a polynomial time $O(\tau(n))$ factor approximation algorithm for computing the Ferrers dimension of a digraph on n vertices.*

6. Parameterized approximations for toxicity with edit distance parameters

The algorithm for finding the toxicity of graphs with large cliques obtained from Theorem 1 was used in the previous section to derive sublinear approximation algorithms for toxicity and cubicity problems and also for some related dimensional parameters. In this section, we will derive some parameterized approximation algorithms for toxicity using Theorem 1 and some techniques that are similar in nature to those used in the proof of Theorem 1.

A study of parameterized algorithms for toxicity was initiated by Adiga et al. [4]. Later, Bruhn et al. [7] also considered various structural parameterizations of toxicity (which appeared after the initial version of this article [1]).

A framework for parameterizing problems with edit distance as the parameter was introduced by Cai [8]. For a family \mathcal{F} of graphs, and $k \geq 0$ an integer, Cai used $\mathcal{F} + ke$ (respectively, $\mathcal{F} - ke$) to denote the family of graphs that can be converted to a graph in \mathcal{F} by deleting (respectively, adding) at most k edges, and $\mathcal{F} + kv$ to denote the family of graphs that can be converted to a graph in \mathcal{F} by deleting at most k vertices. This framework was used by Cai [8], for studying the parameterized complexity of the vertex coloring problem on $\mathcal{F} - ke$, $\mathcal{F} + ke$ and $\mathcal{F} + kv$ graphs, with k as the parameter, for various families of graphs \mathcal{F} .

In a similar way, we consider the parameterized complexity of computing toxicity of $\mathcal{F} + k_1e - k_2e$ and $\mathcal{F} + kv$ graphs for families \mathcal{F} of bounded toxicity graphs, using $k_1 + k_2$ (the edge edit distance) and k (the vertex edit distance) as parameters. Note that, parameters like interval completion number, minimum vertex cover size and minimum feedback vertex set size are examples of vertex edit distance parameters from some bounded toxicity graph families (respectively, consider \mathcal{F} to be the family of interval graphs, the family of graphs with no edges, and the family of trees).

A subset $S \subseteq V$ such that $|S| \leq k$ is called a **modulator** for an $\mathcal{F} + kv$ graph $G(V, E)$ if $G \setminus S \in \mathcal{F}$. Similarly, a set E_k of pairs of vertices such that $|E_k| \leq k$ is called a modulator for an $\mathcal{F} - ke$ graph $G(V, E)$ if $G'(V, E \cup E_k) \in \mathcal{F}$. Modulators for graphs in $\mathcal{F} + ke$ and $\mathcal{F} + k_1e - k_2e$ are defined in a similar manner.

The following theorem gives us a parameterized algorithm for computing the toxicity of $\mathcal{F} + kv$ graphs. The proof of this theorem uses Theorem 1.

Theorem 3. *Let \mathcal{F} be a family of graphs such that $\forall G' \in \mathcal{F}$, $\text{box}(G') \leq b \leq n$. Let $T(n)$ denote the time required to compute a b -dimensional box representation of a graph belonging to \mathcal{F} on n vertices. Let G be an $\mathcal{F} + kv$ graph on n vertices. Given a modulator of G , a box representation \mathcal{B} of G , such that $|\mathcal{B}| \leq 2 \text{box}(G) + b$ can be computed in time $T(n - k) + n^2 2^{O(k^2 \log k)}$.*

Proof. Let \mathcal{F} be the family of graphs of toxicity at most b . Let $G(V, E)$ be an $\mathcal{F} + kv$ graph on n vertices, with a modulator S_k on k vertices such that $G' = G \setminus S_k \in \mathcal{F}$. We define two supergraphs of G , namely $H_1(V, E_1)$ and $H_2(V, E_2)$ such that $E = E_1 \cap E_2$ with $\text{box}(H_1) \leq 2 \text{box}(G)$, $\text{box}(H_2) \leq b$ and their required valid box representations are computable within the time specified in the theorem. It is easy to see that the union of valid box representations of H_1 and H_2 will be a valid box representation \mathcal{B} of G and hence $|\mathcal{B}| \leq \text{box}(H_1) + \text{box}(H_2) \leq 2 \text{box}(G) + b$. This will complete our proof of Theorem 3.

We define H_1 to be the graph obtained from G by making $V \setminus S_k$ a clique on $n - k$ vertices, without altering other adjacencies in G . Formally, $E_1 = E \cup \{(x, y) \mid x, y \in V \setminus S_k, x \neq y\}$. Using [Theorem 1](#), we can get an optimal box representation \mathcal{B}_1 of H_1 in $n^2 2^{O(k^2 \log k)}$ time. By [Lemma 3](#), $|\mathcal{B}_1| \leq 2 \text{box}(G)$.

We define H_2 to be the graph obtained from G by making each vertex in S_k adjacent to every other vertex in the graph and leaving other adjacencies in G unaltered. Formally, $E_2 = E \cup \{(x, y) \mid x \in S_k, y \in V, x \neq y\}$. Let \mathcal{B}' be a box representation of G' of dimension at most b (computed in time $T(n - k)$). Then, \mathcal{B}' is a box representation of $H_2[V \setminus S_k]$ as well, because $H_2[V \setminus S_k] = G'$. By [Lemma 2](#), $\text{box}(H_2) = \text{box}(H_2[V \setminus S_k])$ and a box representation \mathcal{B}_2 of H_2 of dimension at most $|\mathcal{B}'| \leq b$ can be produced in $O(n^2)$ time.

Since $G = H_1 \cap H_2$, $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ is a valid box representation of G , of dimension at most $2 \text{box}(G) + b$. All computations were done in $T(n - k) + n^2 2^{O(k^2 \log k)}$ time. \square

Using a similar method, we also get a parameterized approximation algorithm for computing the boxicity of $\mathcal{F} + k_1e - k_2e$ graphs.

Theorem 4. *Let \mathcal{F} be a family of graphs such that $\forall G' \in \mathcal{F}$, $\text{box}(G') \leq b \leq n$. Let $T(n)$ denote the time required to compute a b -dimensional box representation of a graph belonging to \mathcal{F} on n vertices. Let G be an $\mathcal{F} + k_1e - k_2e$ graph on n vertices and let $k = k_1 + k_2$. Given a modulator of G , a box representation \mathcal{B} of G , such that $|\mathcal{B}| \leq \text{box}(G) + 2b$, can be computed in time $T(n) + O(n^2) + 2^{O(k^2 \log k)}$.*

Proof. Let \mathcal{F} be the family of graphs of boxicity at most b . Let $G(V, E)$ be an $\mathcal{F} + k_1e - k_2e$ graph on n vertices, where $k_1 + k_2 = k$. Let $E_{k_1} \cup E_{k_2}$ be a modulator of G such that $|E_{k_1}| = k_1$, $|E_{k_2}| = k_2$ and $G'(V, (E \cup E_{k_2}) \setminus E_{k_1}) \in \mathcal{F}$. Let $T \subseteq V(G)$ be the set of vertices incident with edges in $E_{k_1} \cup E_{k_2}$.

As in the proof of [Theorem 3](#), we define two supergraphs of G , namely $H_1(V, E_1)$ and $H_2(V, E_2)$ such that $E = E_1 \cap E_2$ with $\text{box}(H_1) \leq 2b$, $\text{box}(H_2) \leq \text{box}(G)$ and their required valid box representations are computable within the time specified in the theorem. As earlier, the union of valid box representations of H_1 and H_2 will be a valid box representation of \mathcal{B} of G and hence $|\mathcal{B}| \leq \text{box}(H_1) + \text{box}(H_2) \leq 2b + \text{box}(G)$. This will complete our proof of [Theorem 4](#).

Let $H_1(V, E_1)$ be the graph obtained from G' by making T a clique, without altering other adjacencies in G' . Formally, $E_1 = E' \cup \{(x, y) \mid x, y \in T, x \neq y\}$. Let \mathcal{B}' be a box representation of G' of dimension at most b computed in time $T(n)$. From the box representation \mathcal{B}' of G' , in $O(b \cdot n) = O(n^2)$ time we can construct (by [Lemma 3](#)) a box representation \mathcal{B}_1 of H_1 with dimension $2b$.

Let $H_2(V, E_2)$ be the graph obtained from G by making each vertex in $V \setminus T$ adjacent to every other vertex in the graph and leaving other adjacencies in G unaltered. Formally, $E_2 = E \cup \{(x, y) \mid x \in V \setminus T, y \in V, x \neq y\}$. Clearly, $|T| \leq 2k$ and therefore, using the construction in [Proposition 1](#), an optimal box representation \mathcal{B}_T of $H_2[T]$ can be computed in $2^{O(k^2 \log k)}$ time. By [Lemma 2](#), $\text{box}(H_2) = \text{box}(H_2[T])$ and a box representation \mathcal{B}_2 of H_2 of dimension $\text{box}(H_2[T])$ can be computed from the box representation \mathcal{B}_T of $H_2[T]$ in $O(n^2)$ time. Observe that $H_2[T] = G[T]$. Therefore, $|\mathcal{B}_2| = \text{box}(G[T]) \leq \text{box}(G)$, because $G[T]$ is an induced subgraph of G .

Since $G = H_1 \cap H_2$, $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ is a valid box representation of G , of dimension at most $\text{box}(G) + 2b$. All computations were done in $T(n) + O(n^2) + 2^{O(k^2 \log k)}$ time. \square

Remark 2. Though in [Theorems 3](#) and [4](#) it is assumed that a modulator of G for \mathcal{F} is given, in several important special cases (as in the case of corollaries discussed below), the modulator for \mathcal{F} can be computed from G in FPT time. Moreover, in those cases, $T(n)$ is a polynomial in n . Thus, the algorithms given by [Theorems 3](#) and [4](#) turns out to be FPT approximation algorithms for boxicity.

6.1. Corollaries of [Theorems 3](#) and [4](#)

A general approach for obtaining parameterized approximation algorithms for the boxicity of $\mathcal{F} + k_1e - k_2e$ and $\mathcal{F} + kv$ graphs for families \mathcal{F} of bounded boxicity graphs are given by [Theorems 3](#) and [4](#). Parameterized approximation algorithms for boxicity with various vertex and edge edit distance parameters can be derived using the above approach.

Given an input graph G and a parameter of interest, the general procedure is as follows:

- Use known FPT algorithms to compute the parameter of interest for the input graph G and obtain from G a modulator S_k for the corresponding family \mathcal{F} .
- Compute a low dimensional box representation for the graph $G' = (G \setminus S_k) \in \mathcal{F}$, in polynomial time.
- Use our algorithm of [Theorem 3](#) ([Theorem 4](#)) to derive an FPT approximation algorithm for computing boxicity of G using the parameter of interest.

Some corollaries of [Theorem 3](#) are discussed below.

Corollary 8. *FVS as the parameter: The minimum number of vertices to be deleted from a graph G so that the resultant graph is acyclic is called the feedback vertex set size (FVS) of G . If $\text{FVS}(G) \leq k$, we get a $\left(2 + \frac{2}{\text{box}(G)}\right)$ factor approximation for boxicity with FVS as the parameter k , which runs in time $2^{O(k^2 \log k)} n^{O(1)}$.*

Proof. If $\text{FVS}(G) \leq k$, using existing FPT algorithms [9], in $O(3.83^k kn^2)$ time we can compute a minimum feedback vertex set S of $G(V, E)$ such that $G' = G(V \setminus S)$ is a forest. Thus, with modulator S , $G \in \mathcal{F} + kv$, where \mathcal{F} is the family of graphs which are forests. Since a box representation of dimension two can be computed in polynomial time for any forest [32], using our algorithm of [Theorem 3](#), we get a $2 + \frac{2}{\text{box}(G)}$ factor approximation for boxicity with FVS as the parameter k , which runs in time $2^{O(k^2 \log k)} n^{O(1)}$. \square

Remark 3. For the boxicity problem parameterized by FVS, Adiga et al. [4] gave an algorithm with the same approximation factor as the algorithm discussed above, but with running time $2^{O(2^k k^2)} n^{O(1)}$. The running time of our algorithm is better.

Corollary 9. Proper Interval Vertex Deletion number (PIVD) as the parameter: The minimum number of vertices to be deleted from the graph G , so that the resultant graph is a proper interval graph, is called PIVD(G). If $\text{PIVD}(G) \leq k$, we get a $2 + \frac{1}{\text{box}(G)}$ factor approximation for boxicity with PIVD as the parameter k , which runs in time $2^{O(k^2 \log k)} n^{O(1)}$.

Proof. If $\text{PIVD}(G)$ is at most k , we can use the FPT algorithm running in $O(6^k kn^6)$ time for proper interval vertex deletion [23] to compute a $S \subseteq V$ with $|S| \leq k$ such that $G \setminus S$ is a proper interval graph. Thus, with modulator S , $G \in \mathcal{F} + kv$, where \mathcal{F} is the family of all proper interval graphs. Since a box representation of dimension one can be computed in polynomial time for any proper interval graph [5], using our algorithm of [Theorem 3](#), we get a $2 + \frac{1}{\text{box}(G)}$ factor approximation for boxicity with PIVD as the parameter k , which runs in time $2^{O(k^2 \log k)} n^{O(1)}$. \square

Remark 4. It is easy to see that $\text{PIVD}(G) \leq \text{MVC}(G)$, where $\text{MVC}(G)$ is the size of the minimum vertex cover number of G . Hence, PIVD(G) is a better parameter than the parameter MVC(G). This algorithm has the same running time as the additive one approximation algorithm for boxicity with MVC(G) as the parameter, discussed in [3].

Corollary 10. Planar Vertex Deletion number (PVD) as the parameter: The minimum number of vertices to be deleted from G to make it a planar graph, is called the planar vertex deletion number of G . If $\text{PVD}(G) \leq k$, we get an FPT algorithm for boxicity, giving a $\left(2 + \frac{3}{\text{box}(G)}\right)$ factor approximation for boxicity using planar vertex deletion number as the parameter.

Proof. If $G \in \text{Planar} + kv$, we can use the FPT algorithm running in $O(f(k)n^2)$ time for planar deletion [26] to compute a $S \subseteq V$ with $|S| \leq k$ such that $G \setminus S$ is planar. Thus, with modulator S , $G \in \mathcal{F} + kv$, where \mathcal{F} is the family of planar graphs. Since planar graphs have 3 dimensional box representations computable in polynomial time [34], using our algorithm of [Theorem 3](#), we get an FPT algorithm for boxicity, giving a $2 + \frac{3}{\text{box}(G)}$ factor approximation for boxicity of graphs that can be made planar by deleting at most k vertices, using planar vertex deletion number as the parameter. \square

[Theorem 4](#) also gives us FPT approximation algorithms for computing boxicity with various parameters of interest.

Corollary 11. Interval Completion number as the parameter: The minimum number of edges to be added to a graph G , so that the resultant graph is an interval graph, is called the interval completion number of G . If the interval completion number G is at most k , we get an FPT algorithm that achieves an additive 2 approximation for box(G) which runs in time $2^{O(k^2 \log k)} n^{O(1)}$.

Proof. If the interval completion number of a graph $G(V, E)$ is at most k , we can use the FPT algorithm for interval completion [36] with running time $O(k^{2k} n^{O(1)}) = 2^{O(k \log k)} n^{O(1)}$ to compute E_k such that $|E_k| \leq k$ and $G'(V, E \cup E_k)$ is an interval graph. Thus, with modulator E_k , $G \in \mathcal{F} - ke$, where \mathcal{F} is the class of interval graphs. Since a box representation of dimension one can be computed in polynomial time for any interval graph [5], combining with our algorithm of [Theorem 4](#), we get an FPT algorithm that achieves an additive 2 factor approximation for box(G), with interval completion number k as the parameter which runs in time $2^{O(k^2 \log k)} n^{O(1)}$. \square

Corollary 12. Proper Interval Edge Deletion number (PIED) as the parameter: The minimum number of edges to be deleted from the graph G , so that the resultant graph is a proper interval graph, is called PIED(G). If $\text{PIED}(G)$ is at most k , we get an FPT algorithm that achieves an additive 2 approximation for box(G), with PIED(G) as the parameter k , which runs in time $2^{O(k^2 \log k)} n^{O(1)}$.

Proof. If $\text{PIED}(G)$ is at most k , we can use the FPT algorithm running in $O(9^k n^{O(1)})$ time for proper interval edge deletion [23] to compute a $E_k \subseteq E$ with $|E_k| \leq k$ such that $G'(V, E \setminus E_k)$ is a proper interval graph. Thus, with modulator S , $G \in \mathcal{F} + ke$, where \mathcal{F} is the family of all proper interval graphs. Since a box representation of dimension one can be computed in polynomial time for any interval graph, combining with our algorithm of [Theorem 4](#), we get an FPT algorithm that achieves an additive 2 factor approximation for box(G), with PIED as the parameter k , which runs in time $2^{O(k^2 \log k)} n^{O(1)}$. \square

Corollary 13. Planar Edge Deletion number (PED) as the parameter: The minimum number of edges to be deleted from G so that the resultant graph is planar is called PED(G). If $\text{PED}(G) \leq k$, we get an FPT algorithm that gives an additive 6 approximation for box(G) with PED(G) as the parameter.

Proof. If $\text{PED}(G) \leq k$, we can use the FPT algorithm for planar edge deletion [22] to compute $E_k \subseteq E$ such that $|E_k| \leq k$ and $G'(V, E \setminus E_k)$ is a planar graph. Thus, with modulator E_k , $G \in \mathcal{F} + ke$, where \mathcal{F} is the class of planar graphs. Since planar graphs have 3 dimensional box representations computable in polynomial time [34], using our algorithm of [Theorem 4](#), we get an FPT algorithm that gives an additive 6-factor approximation for $\text{box}(G)$ with $\text{PED}(G)$ as the parameter. \square

Corollary 14. *Max Leaf number (ML) as the parameter: The number of the maximum possible leaves in any spanning tree of a graph G is called $\text{ML}(G)$. If $\text{ML}(G) \leq k$, we get an FPT algorithm that achieves an additive 2 approximation for $\text{box}(G)$ which runs in time $2^{O(k^3 \log k)} n^{O(1)}$.*

Proof. The underlying algorithm here same as that in the proof of [Theorem 4](#).

We assume that G is connected and the max leaf number of G is at most k . If the graph G is just a cycle on n vertices ($n \geq 3$), we know that $\text{box}(G) = 1$ if $n = 3$ and $\text{box}(G) = 2$ if $n > 3$. Thus, we can also assume that G is not a cycle. Moreover, the maximum degree of any vertex in G is at most k ; otherwise we can start with a vertex of degree at least $k+1$ and grow it to a spanning tree with more than k leaves, which is a contradiction.

In the proof of [Theorem 4](#), supergraphs H_1 and H_2 were obtained by modifying a certain graph G' whose edge edit distance to G is small. Here, we will define G' in a slightly different way and then define H_1 and H_2 in a similar way as we did in the proof of [Theorem 4](#).

We start by defining a graph G_1 , such that G is a subdivision of G_1 . For this, we use the following result.

Property 1 (Fellows et al. [19]). *If the max leaf number of a graph G is at most k , then G is a subdivision of a graph $G_1(V', F)$ with $|V'| \leq 4k - 2$ and $V' \subseteq V$. (G_1 may contain multi edges and self loops.)*

Let $G_1(V', F)$ the graph given by [Property 1](#). Since G is not a cycle, we can eliminate all degree two vertices from G_1 one by one, by edge contractions. Therefore, without loss of generality, we can assume that there are no degree two vertices in G_1 and V' is precisely the set of vertices of G whose degree is not equal to 2.

Claim 1. *There are at most $4k - 2$ vertices in G , whose degree is not equal to 2.*

Proof. As explained above, we assume that there are no degree two vertices in G_1 . Since G is a subdivision of G_1 and a subdivision only introduces degree 2 vertices, we can conclude that there are at most $4k - 2$ vertices in G , whose degree is not equal to 2. \square

Let $E_k \subseteq E$ be the set of edges of G which have at least one of its incident vertices belonging to V' . Now, we will define G' as the graph with vertex set V and edge set $E \setminus E_k$.

Claim 2. *The graph $G'(V, E \setminus E_k)$ is an interval graph and it can be computed in polynomial time from G .*

Proof. Since G is a subdivision of $G_1(V', F)$, it is easy to see that, the graph $G'(V, E \setminus E_k)$ is just a collection of vertex disjoint paths and isolated vertices. It is straightforward to verify that G' is an interval graph. To compute G' , we just need to compute E_k . Since E_k is defined from V' and G , we only need to compute the set V' , which is precisely the set of vertices of G whose degree is not equal to 2. This can be done in polynomial time. \square

Since G' is an interval graph, we have $\text{box}(G') \leq 1$ and an interval representation of G' can be constructed in linear time [5]. Let T be the set of vertices of G , which are incident to at least one edge in E_k . In other words, $T = V' \bigcup_{v \in V'} N_G(v)$. Since maximum degree of G is at most k (as explained at the beginning of this proof), we get $|T| \leq |V'| + k \cdot |V'| \leq (4k - 2) + k \cdot (4k - 2) = O(k^2)$. From the proof of [Theorem 4](#), we can notice that the proof goes through with this definition of T and the complexity of the algorithm depends only on $|T|$ and not on the number of edges being modified in G .

For clarity, we just repeat some important points of the algorithm of [Theorem 4](#) here, with modifications occurring mainly in the running time analysis. Let $H_1(V, E_1)$ be the graph obtained from G' by making T a clique, without altering other adjacencies in G' . From the box representation of G' of dimension one, in $O(n)$ time we can construct (by [Lemma 3](#)) a box representation B_1 of H_1 with dimension 2.

Let $H_2(V, E_2)$ be the graph obtained from G by making vertices in $V \setminus T$ adjacent to every other vertex in the graph and maintaining other adjacencies in G unaltered. As in the proof of [Theorem 4](#), we have $H_2[T] = G[T]$. Hence, $\text{box}(H_2[T]) = \text{box}(G[T]) \leq \text{treewidth}(G[T]) + 2 \leq \text{treewidth}(G) + 2 \leq 2 \cdot \text{ML}(G) + 2 \leq 2k + 2$ [12,19]. We know that $|T| = O(k^2)$ and therefore, using the construction in [Proposition 1](#), an optimal box representation B_T of $H_2[T]$ can be computed in $2^{O(k^3 \log k)} n^{O(1)}$ time and from B_T , an optimal box representation of H_2 of dimension at most $\text{box}(G)$ is computed in polynomial time.

Union of box representations of H_1 and H_2 gives a $2 + \text{box}(G)$ dimensional box representation for G , obtained in $2^{O(k^3 \log k)} n^{O(1)}$ time. \square

Remark 5. If $\text{ML}(G) \leq k$, we get an FPT algorithm that achieves an additive 2 approximation for $\text{box}(G)$ which runs in time $2^{O(k^3 \log k)} n^{O(1)}$, the running time and approximation ratio being the same as in Adiga et al. [4]. Bruhn et al. [7] gave an additive 1 approximation algorithm for the same problem. However, the running time of their algorithm is very high, compared to that of the algorithm presented here.

7. An FPT approximation scheme for computing cubicity with parameter MVC

Fellows et al. [18, Corollary 5] proved an existential result that for every fixed pair of integers k and b , there is an $f(k) \cdot n$ time algorithm which determines whether a given graph G on n vertices and $\text{MVC}(G) \leq k$ has cubicity at most b . In the theorem below, we derive a FPT approximation algorithm, for computing the cubicity of graphs, using their vertex cover number as the parameter. Our algorithm is constructive.

Theorem 5. Let G be a graph on n vertices. A cube representation of G which is of dimension at most $2 \text{cub}(G)$ can be computed in time $2^{O(2^k k^2)} n^{O(1)}$, where $k = \text{MVC}(G)$. By allowing a larger running time of $2^{O(g(k, \epsilon))} n^{O(1)}$, we can achieve a $(1 + \epsilon)$ approximation factor, for any $\epsilon > 0$, where $g(k, \epsilon) = \frac{1}{\epsilon} k^3 2^{\frac{4k}{\epsilon}}$.

Proof. Let $G(V, E)$ be a graph on n vertices. Without loss of generality, we can assume that G is connected. We can compute a minimum vertex cover of G in time $2^{O(k)} n^{O(1)}$ [27]. Let $S \subseteq V$ be a vertex cover of G of cardinality k . We define two supergraphs of G , namely $H_1(V, E_1)$ and $H_2(V, E_2)$ such that $E = E_1 \cap E_2$ with $\text{cub}(H_1) \leq \text{cub}(G)$ and $\text{cub}(H_2) \leq \text{cub}(G)$.

Let $S \subseteq V$ be a vertex cover of G of cardinality k . First we define an equivalence relation on the vertices of the independent set $V \setminus S$ such that vertices u and v are in the same equivalence class if and only if $N_G(u) = N_G(v)$. Let A_1, A_2, \dots, A_t be the equivalence classes. We define H_1 to be the graph obtained from G by making each A_i into a clique and maintaining other adjacencies as it is in G . Formally, $E_1 = E \cup \{(u, v) \mid u \neq v \text{ and } u, v \text{ belong to the same } A_i, \text{ for some } 1 \leq i \leq t\}$.

For each A_i , let us consider the mapping $n_{A_i} : A_i \mapsto \{1, 2, \dots, |A_i|\}$, where $n_{A_i}(v)$ is the unique number representing $v \in A_i$. (Note that if $u \in A_i$ and $v \in A_j$, where $i \neq j$, then, $n_{A_i}(u)$ and $n_{A_j}(v)$ could potentially be the same.) Let $s = \max_{1 \leq i \leq t} |A_i|$.

We define one more partitioning of the independent set $V \setminus S$ into equivalence classes B_1, B_2, \dots, B_s such that for $1 \leq i \leq s$, $B_i = \{v \mid n_{A_i}(v) = i, \text{ for some } 1 \leq j \leq t\}$. We define H_2 to be the graph obtained from G by making each B_i into a clique, and making each vertex in S adjacent to every other vertex in V . Formally, $E_2 = \{(u, v) \mid u \neq v \text{ and } u \in S, v \in V \} \cup \{(u, v) \mid u \neq v \text{ and } u, v \text{ belong to the same } B_i, \text{ for some } 1 \leq i \leq s\}$.

If u, v are two adjacent vertices of a graph G such that $N_G(u) \cup \{u\} = N_G(v) \cup \{v\}$, we call them as twin vertices. G' is called a reduced graph of G if G' is obtained from G by repeatedly contracting the edges among pairs of twin vertices.

Claim 3. If G' is a reduced graph of G , then, $\text{cub}(G) = \text{cub}(G')$ and from an optimal cube representation C' of G' , in polynomial time, we can obtain an optimal cube representation C of G .

Proof. Let $C' = \{I'_1, I'_2, \dots, I'_p\}$ be an optimal cube representation of G' . For each $1 \leq i \leq p$, define the interval graph I_i as follows: If $u \in V(G')$, then the interval corresponding to u in I_i is same as it is in I'_i . If $u \in V(G) \setminus V(G')$, then $\exists v \in V(G')$ such that u, v are twins in G . In this case, define the interval corresponding to u in I_i is same as the interval of its twin v in I'_i . It can be verified that $C = \{I_1, I_2, \dots, I_p\}$ is a valid cube representation of G . Thus, $\text{cub}(G) \leq p$. Since G' is an induced subgraph of G , we also have $\text{cub}(G) \geq \text{cub}(G') = p$. \square

Observe that in graph H_1 , vertices in each A_i , $1 \leq i \leq t$ are twins of each other. We can construct a reduced graph H'_1 of H_1 by contracting all vertices in A_i to a single vertex, for each $1 \leq i \leq t$. Now, H'_1 has only $t + |S|$ vertices, which is at most $2^k - 1 + k$. It is known that $\text{cub}(H'_1) \leq \text{MVC}(H'_1) + \lceil \log(|V(H'_1)|) - \text{MVC}(H'_1) \rceil - 1$ [11]. Since $\text{MVC}(H'_1) = k$, we get $\text{cub}(H'_1) \leq 2k - 1$. Using the construction in Proposition 1, we can compute an optimal cube representation C'_1 of H'_1 in time $2^{O(2^k k^2)}$. By the claim above, from C'_1 we can get an optimal cube representation C_1 of H_1 in polynomial time, with $|C_1| = \text{cub}(H'_1)$. Observe that H'_1 is an induced subgraph of G , which implies $|C_1| \leq \text{cub}(G)$.

Similarly, observe that, in graph H_2 , vertices in each B_i , $1 \leq i \leq s$ are twins of each other. We can construct a reduced graph H'_2 of H_2 by contracting all vertices in B_i to a single vertex, for each $1 \leq i \leq s$ and contracting S to a single vertex. Now, H'_2 is a graph on $s + 1$ vertices. We can also observe that H'_2 is in fact a star graph with s leaves. In polynomial time, we can construct an optimal cube representation C'_2 of H'_2 which is of dimension $\lceil \log s \rceil$ [32]. As earlier, from C'_2 we can get an optimal cube representation C_2 of H_2 in polynomial time, with $|C_2| = \text{cub}(H'_2) = \lceil \log s \rceil$. Observe that H'_2 is an induced subgraph of G , which implies $|C_2| \leq \text{cub}(G)$.

It can be easily verified that $E = E_1 \cap E_2$ and hence $C_1 \cup C_2$ is a valid cube representation of G of dimension $|C_1| + |C_2| \leq 2 \text{cub}(G)$, constructible in $2^{O(2^k k^2)} n^{O(1)}$ time.

We can also achieve a $(1 + \epsilon)$ approximation factor, for any $\epsilon > 0$ by allowing a larger running time as explained below. Define $f(k_\epsilon) = k(1 + 2^{(2k-1)/\epsilon})$, where $k = \text{MVC}(G)$. If $|V(G)| = n \leq f(k_\epsilon)$, then, by Proposition 1, we can get an optimal cube representation of G in time $2^{O(f^2 k_\epsilon \log f(k_\epsilon))}$. Otherwise, we have $(2k - 1)/\lceil \log \lceil (n - k)/k \rceil \rceil \leq \epsilon$. In this case, we use the construction described above, to get a cube representation of G of dimension $|C_1| + |C_2|$. We prove that in this case, $|C_1| + |C_2| \leq \text{cub}(G)(1 + \epsilon)$.

It is known that $\text{cub}(G) \geq \lceil \log \psi(G) \rceil$, where $\psi(G)$ is the number of leaf nodes in the largest induced star in G [3]. By the pigeon hole principle, $\max_{v \in S} |N_G(v) \cap (V \setminus S)| \geq \lceil (n - k)/k \rceil$. Therefore, $\text{cub}(G) \geq \lceil \log \psi(G) \rceil \geq \lceil \log \lceil (n - k)/k \rceil \rceil$. Recall that

$|C_1| \leq 2k - 1$. Therefore, $|C_1| + |C_2| \leq 2k - 1 + \text{cub}(G) \leq \text{cub}(G) \left(\frac{2k-1}{\text{cub}(G)} + 1 \right) \leq \text{cub}(G) \left(\frac{2k-1}{\lceil \log \lceil (n - k)/k \rceil \rceil} + 1 \right) \leq \text{cub}(G)(1 + \epsilon)$.

The total running time of this algorithm is $2^{O(\frac{1}{\epsilon} k^3 2^{\frac{4k}{\epsilon}})} n^{O(1)}$. \square

8. Conclusion

We have presented $o(n)$ factor approximation algorithms for computing the boxicity and cubicity of graphs. Using these algorithms, we also derived $o(n)$ factor approximation algorithms for some related well-known problems, including poset dimension and Ferrers dimension. To the best of our knowledge, for none of these problems polynomial time sublinear factor approximation algorithms were known previously. Since polynomial time approximations within an $O(n^{1-\epsilon})$ factor for any $\epsilon > 0$ are considered unlikely for any of these problems, no significant improvement in the approximation factor can be expected. We have also presented a general method of obtaining parameterized approximation algorithms for boxicity using vertex and edge edit distance parameters.

References

- [1] A. Adiga, J. Babu, L.S. Chandran, Polynomial time and parameterized approximation algorithms for boxicity, in: IPEC, 2012, pp. 135–146.
- [2] A. Adiga, D. Bhowmick, L.S. Chandran, Boxicity and poset dimension, *SIAM J. Discrete Math.* 25 (4) (2011) 1687–1698.
- [3] A. Adiga, L.S. Chandran, Cubicity of interval graphs and the claw number, *J. Graph Theory* 65 (2010) 323–333.
- [4] A. Adiga, R. Chitnis, S. Saurabh, Parameterized algorithms for boxicity, in: ISAAC, 2010, pp. 366–377.
- [5] K.S. Booth, G.S. Lueker, Testing for the consecutive ones property, interval graphs, and graph planarity using pq-tree algorithms, *J. Comput. System Sci.* 13 (3) (1976) 335–379.
- [6] H. Breu, D.G. Kirkpatrick, Unit disk graph recognition is NP-hard, *Comput. Geom.* 9 (1998) 2–34.
- [7] H. Bruhn, M. Chopin, F. Joos, O. Schaudt, Structural parameterizations for boxicity, *Algorithmica* 74 (4) (2016) 1453–1472.
- [8] L. Cai, Parameterized complexity of vertex colouring, *Discrete Appl. Math.* 127 (3) (2003) 415–429.
- [9] Y. Cao, J. Chen, Y. Liu, On feedback vertex set: New measure and new structures, *Algorithmica* 73 (1) (2015) 63–86.
- [10] P. Chalermsook, B. Laekhanukit, D. Nanongkai, Graph products revisited: Tight approximation hardness of induced matching, poset dimension and more, *SODA* 2013, 2013, pp. 1557–1576.
- [11] L.S. Chandran, A. Das, C.D. Shah, Cubicity, boxicity, and vertex cover, *Discrete Math.* 309 (8) (2009) 2488–2496.
- [12] L.S. Chandran, N. Sivadasan, Boxicity and treewidth, *J. Combin. Theory Ser. B* 97 (2007) 733–744.
- [13] S. Chaplick, P. Hell, Y. Otachi, T. Saitoh, R. Uehara, Ferrers dimension; boxicity; segment-ray graph; unit grid intersection graph, *Discrete Appl. Math.* 216 (Part 1) (2017) 130–135.
- [14] V. Chvátal, P.L. Hammer, Aggregation of inequalities in integer programming, *Ann. Discrete Math.: Stud. Integer Program.* 1 (1977) 145–162.
- [15] O. Cogis, On the Ferrers dimension of a digraph, *Discrete Math.* 38 (1) (1982) 47–52.
- [16] M.B. Cozzens, M.D. Halsey, The relationship between the threshold dimension of split graphs and various dimensional parameters, *Discrete Appl. Math.* 30 (2–3) (1991) 125–135.
- [17] B. Dushnik, E.W. Miller, Partially ordered sets, *Amer. J. Math.* 63 (3) (1941) 600–610.
- [18] M.R. Fellows, D. Hermelin, F.A. Rosamond, Well-quasi-orders in subclasses of bounded treewidth graphs, in: IWPEC, 2009, pp. 149–160.
- [19] M. Fellows, D. Lokshtanov, N. Misra, M. Mnich, F. Rosamond, S. Saurabh, The complexity ecology of parameters: An illustration using bounded max leaf number, *Theoret. Comput. Sci.* 45 (4) (2009) 822–848.
- [20] S. Felsner, M. Habib, R.H. Möhring, On the interplay between interval dimension and dimension, *SIAM J. Discrete Math.* 7 (1) (1994) 32–40.
- [21] S. Felsner, I. Mustata, M. Pergel, The complexity of the partial order dimension problem - closing the gap, *SIAM J. Discrete Math.* 31 (2017) 172–189.
- [22] M. Grohe, Computing crossing numbers in quadratic time, *J. Comput. System Sci.* 68 (2004) 285–302.
- [23] P. Hof, Y. Villanger, Proper interval vertex deletion, *Algorithmica* 65 (4) (2013) 845–867.
- [24] J. Kratochvíl, A special planar satisfiability problem and a consequence of its NP-completeness, *Discrete Appl. Math.* 52 (3) (1994) 233–252.
- [25] N.V.R. Mahadev, U.N. Peled, Threshold Graphs and Related Topics, in: Annals of Discrete Mathematics, vol. 56, Elsevier, 1995.
- [26] D. Marx, I. Schlotter, Obtaining a planar graph by vertex deletion, *Algorithmica* 62 (3–4) (2012) 807–822.
- [27] R. Niedermeier, Invitation to Fixed-Parameter Algorithms, in: Oxford Lecture Series in Mathematics and Its Applications, vol. 31, Oxford University Press, 2006.
- [28] J. Riguet, Les relations de Ferrers, *C. R. Séances Hebd. Acad. Sci. (Paris)* 232 (1951) 1729–1730.
- [29] F.S. Roberts, On the boxicity and cubicity of a graph, in: Recent Progresses in Combinatorics, Academic Press, New York, 1969, pp. 301–310.
- [30] B. Rosgen, L. Stewart, Complexity results on graphs with few cliques, *Discrete Math. Theoret. Comput. Sci.* 9 (2007) 127–136.
- [31] E.R. Scheinerman, Intersection Classes and Multiple Intersection Parameters of Graphs, Princeton University, 1984 (Ph.D. Thesis).
- [32] C.D. Shah, Boxicity, Cubicity, and Vertex Cover, (M.Sc Thesis), IISc, Bangalore, 2008 http://clweb.csa.iisc.ernet.in/chintan/chintan_thesis.pdf.
- [33] J.P. Spinrad, Efficient Graph Representations, in: Fields Institute Monographs, vol. 19, American Mathematical Society, 2003.
- [34] C. Thomassen, Interval representations of planar graphs, *J. Combin. Theory Ser. B* 40 (1986) 9–20.
- [35] W.T. Trotter Jr., Combinatorial problems in dimension theory for partially ordered sets, in: Problèmes Combinatoires et Théorie des Graphes, Colloque Internationaleaux C.N.R.S. (260), 1978, pp. 403–406.
- [36] Y. Villanger, P. Heggernes, C. Paul, J.A. Telle, Interval completion is fixed parameter tractable, *SIAM J. Comput.* 38 (5) (2008) 2007–2020.
- [37] M. Yannakakis, The complexity of the partial order dimension problem, *SIAM J. Algebr. Discrete Methods* 3 (3) (1982) 351–358.

Abhijin Adiga, Shuyu Chu, Stephen Eubank, Christopher J Kuhlman, Bryan Lewis, Achla Marathe, Madhav Marathe, Eric K Nordberg, Samarth Swarup, Anil Vullikanti, Mandy L Wilson

To cite: Adiga A, Chu S, Eubank S, et al. Disparities in spread and control of influenza in slums of Delhi: findings from an agent-based modelling study. *BMJ Open* 2018;8:e017353. doi:10.1136/bmjopen-2017-017353

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-017353>).

Received 18 April 2017

Revised 31 October 2017

Accepted 3 November 2017

ABSTRACT

Objectives This research studies the role of slums in the spread and control of infectious diseases in the National Capital Territory of India, Delhi, using detailed social contact networks of its residents.

Methods We use an agent-based model to study the spread of influenza in Delhi through person-to-person contact. Two different networks are used: one in which slum and non-slum regions are treated the same, and the other in which 298 slum zones are identified. In the second network, slum-specific demographics and activities are assigned to the individuals whose homes reside inside these zones. The main effects of integrating slums are that the network has more home-related contacts due to larger family sizes and more outside contacts due to more daily activities outside home. Various vaccination and social distancing interventions are applied to control the spread of influenza.

Results Simulation-based results show that when slum attributes are ignored, the effectiveness of vaccination can be overestimated by 30%–55%, in terms of reducing the peak number of infections and the size of the epidemic, and in delaying the time to peak infection. The slum population sustains greater infection rates under all intervention scenarios in the network that treats slums differently. Vaccination strategy performs better than social distancing strategies in slums.

Conclusions Unique characteristics of slums play a significant role in the spread of infectious diseases. Modelling slums and estimating their impact on epidemics will help policy makers and regulators more accurately prioritise allocation of scarce medical resources and implement public health policies.

INTRODUCTION

Infectious disease is one of the leading causes of human morbidity and mortality worldwide. Reports from the Centers for Disease Control (CDC) show that over 200 000 people in the USA are hospitalised with influenza-like illness (ILI) symptoms each year, and the mortality on average is over 36 000 annually.^{1,2} In Delhi, India, a joint study by CDC, All India Institute of Medical Sciences and the National Institute of Virology has shown

Strengths and limitations of this study

- We show that the unique attributes of slums must be accounted for in understanding the spread and control of infectious diseases.
- We demonstrate that the granularity afforded by the agent-based model enables extraction of subpopulations, and subsets of interactions, to help interpret results.
- This study does not consider age-specific susceptibility or immunity from past infections; all individual persons are assumed to be equally susceptible.
- The disease transmission risk does not change across activity types; for example, an hour with an infected person at home or at work carries the same risk.
- Colocation-based contact time is used as a proxy for physical proximity and short-distance, environmentally mediated transmission.

that ILI cases are present throughout the year, although they peak in rainy and winter seasons.³ It carries a significant economic burden through reduced productivity and high costs of healthcare.^{4–7} A CDC study finds that for outpatient and non-medically attended individuals, acute respiratory infections cost 1%–5% of monthly per capita income in India. In contrast, cost of inpatient care can be as high as 6%–34% of monthly per capita income.⁸ For developed countries, the annual cost of influenza is estimated to be between \$1 million and \$6 million per 100 000 people, according to the WHO.⁹

In 2007, India established an Integrated Disease Surveillance Programme, which included a network of 12 regional laboratories, to minimise the threat of avian influenza and other highly infectious zoonotic diseases.¹⁰ India faces some unique challenges in surveillance, prevention and control because of the seasonality of influenza at subregional levels. This seasonal variation depends on latitude,



Network Dynamics and Simulation Sciences Laboratory, Biocomplexity Institute, Virginia Tech, Blacksburg, Virginia, USA

Correspondence to
Dr Achla Marathe;
amarathe@vt.edu

monsoon season, humidity and climatic factors of the regions. Acute respiratory infections are estimated to be 43 million per year, of which 4%–12% are due to influenza.^{11 12} Chadha *et al*¹³ estimated hospitalisations due to respiratory illnesses to be 160 per 10 000 persons in year 2011, and children under the age of 5 had the highest incidence of them.

Given that influenza is environmentally mediated and spreads through close proximity, population density is an important factor in its spread. In India, the average population density is about 1000 people per square mile; in the slums, it can be 10–100 times higher.¹⁴ Larger household size and crowding make it easier to transmit infections.^{15–18} For example, Baker *et al*¹⁶ find that meningococcal disease risk among children doubles with the addition of two adolescents or adults (10 years or older) to a six-room house. Other than overcrowding, slums are characterised by their lack of medical services,^{19 20} which makes slum residents highly vulnerable to infectious diseases. Diseases like cholera, malaria, dengue and HIV are common in slums across the world.^{21–23}

This research uses Delhi, the National Capital Territory of India, where 13% of its 13.8 million people live in slum areas, as an example city to study the spread and control of influenza. Delhi is an interesting case study. It ranks fourth in the world in urban population, and among the top 25 largest urban areas, it ranks tenth in population density. Although Delhi is our target population, the results are likely to be useful in studying other slum areas within and outside of India because of the wide range of intervention types and parameter values examined.

This paper is an extension of the work done in Chen *et al*,⁴ which shows that slum populations have a significant effect on influenza transmission in urban areas. Ignoring the influence of slum characteristics underestimates the speed of an outbreak and its extent. However, Chen *et al*⁴ do not consider any interventions on the epidemic spread. *The focus of this research is to study the effect of different intervention strategies on several subpopulations (slum, age and gender) in two different Delhi networks, that is, original (referred to as network 1) and refined (network 2).*

The original network used in Xia *et al*²⁴ studied the spread and control of influenza in Delhi using network 1, which did not take into account the special attributes of the slum population, such as larger family sizes and different types of daily activity schedules. Chen *et al*⁴ used network 2, the refined social network of Delhi, which accounted for slum demographics and slum activities, but did not study intervention strategies. In network 2, there are 298 slum regions in Delhi, containing about 1.8 million people.

The goals of this work focus on understanding the effects of pharmaceutical interventions (PI) and non-pharmaceutical interventions (NPI) on epidemic outcomes. PIs include vaccinations, and NPIs are social distancing measures such as school closure, quarantine and staying home. These effects are studied comparatively: (1) in network 1 versus network 2, overall and for

subpopulations in each; and (2) in the slum and non-slum regions of network 2. Additionally, in a scenario where interventions can be applied to a limited number of individuals, we explore how resources should be split between slum and non-slum subpopulations in order to achieve the best outcomes with respect to total infection rate (ie, the cumulative fraction of a population infected).

METHODS

We use an agent-based modelling (ABM) approach to simulate the spread and containment of influenza in social contact networks of Delhi, India. We compare two networks: one considers slum-specific attributes, and the other does not. In this section, we describe the networks, the disease model for each agent, the interventions and the heterogeneities of the problem that make ABM uniquely suited to study epidemics. Throughout this paper, each agent in the ABM is an individual human.

Social contact networks

This study uses two synthetic social networks of Delhi, created in Xia *et al*²⁴ and in Chen *et al*.⁴ Details on their construction can be found in Xia *et al*,²⁴ Chen *et al*,⁴ Barrett *et al*,²⁵ Bissett *et al*²⁶ and references therein. The synthetic social network by Xia *et al*²⁴ is called *network 1*, and the more refined network developed in Chen *et al*,⁴ *network 2*.

It is important to note that while the social contact networks are inputs in epidemiological simulations, these networks are not specified directly. Rather, these networks are the outputs of population generation methods that are overviewed below and cited immediately above, and include activity surveys and demographic data, both inside and outside of slums. Thus, the topologies of the networks arise from the population generation process and its inputs.

Network 1 was developed in part from LandScan and Census data for Delhi, a daily set of activities of individuals, and the locations of those activities including geolocations of residential areas, shopping centres and schools, collected through surveys by MapmyIndia.com. By assigning activity locations to individuals' activities, people are located at particular times at particular geographical coordinates (including office buildings, schools and others) and within particular rooms of buildings. Next, contacts between individuals are estimated when each person is deemed to have made contact with a subset of other people simultaneously present at the same location. This gives rise to a synthetic social contact network where network edges represent these contacts.

Network 2 models the slum regions in Delhi and assigns slum-specific attributes to the individuals whose homes reside in the slum polygons. Slum residents' attributes and their daily sets of activities are collected through a ground survey in Delhi slums, by a vendor, IndiaMART (www.Indiamart.com/trips). The slum polygons are obtained from MapMechanic.com. Individuals living in the

slum regions are a part of the slum population. All other individuals are part of the non-slum population. Network 2 is a geolocated and contextualised social contact network of Delhi with slums integrated in it.

Following are the main differences between the original network (network 1) and the refined network (network 2). The original social contact network 1 treats the slum regions like any other region in Delhi in terms of assignment of demographics and individual activities, that is, no special consideration is given to slum residents. The refined network 2 identifies 298 slum polygons (zones) in Delhi and assigns slum-specific demographics and activities to the individuals whose homes reside inside these polygons. Thus, the number of individuals is the same in both populations. The slum population constitutes about 13% (1.8 million) of the entire Delhi population of 13.8 million people. The main effect of integrating slums is that network 2 has more home-related contacts due to larger family sizes and more outside contacts due to more daily activities outside home. Also, those individuals who reside outside of slum zones have the same activities in both networks (but their contacts may change if their interactions include slum residents). Overall, there are over 231 million daily interactions between pairs of individuals. Online supplementary table 1 compares those two networks as well as data sources for slum and non-slum Delhi, India. For example, the average degree increases from 30.4 to 33.4 from network 1 to network 2, and the maximum degree increases from 170 to 180. We refer to Chen *et al*⁴ for more detailed information about the two networks. Several plots of properties and structural characteristics of networks 1 and 2 are given in Chen *et al*.²⁷

Disease model

An SEIR (Susceptible (S), Exposed (E), Infectious (I) and Removed or Recovered (R)) model is considered within each individual. Each node in the network represents an individual, and each edge represents a contact on which the disease can spread. A contact represents possible transmission between two people who are colocated for some duration (based on their activity schedules). This is an approximation to model direct contact and short-distance, environmentally mediated transmission that might include direct physical contact, fomite-mediated and airborne transmission.²⁸

We start each epidemic simulation with 20 index cases, randomly chosen. (We find that results are not sensitive to the number of initial infections.) The detailed description of the SEIR model as well as the choices of transmissibility value, R_0 , the explicit incubation and exposed periods can be found in the online supplementary information. This disease model has been used in other works such as Liao *et al*⁶ and Marathe *et al*.²⁹

The transmissibility value for disease transmission is that for the strong influenza model in Chen *et al*.⁴ That work used mild, strong and catastrophic influenza models, so we chose the intermediate transmissibility.

This corresponds to base attack rates (ie, cumulative infection fractions) of 0.42 and 0.48, respectively, in networks 1 and 2. These rates are generally higher than those in some other studies that either compute experimental attack rates from cases or compute them in modelling studies such as this one. Attack rates used by past researchers for different strains of influenza include Asia (0.22–0.50),³⁰ South-East Asia (0.11–0.31 in children³¹; 0.05–0.65³²) and India (0.111–0.235³³; 0.074–0.424³⁴; 0.045–0.294³⁵; 0.008–0.100³⁶; 0.209 for various strains¹³). The results in Chen *et al*⁴ indicate that the results here, for this particular transmissibility, will be qualitatively the same for other transmissibilities, but will scale down or up as transmissibility changes in the same direction.

Interventions

This work considers three vaccination scenarios, that is, vaccinate when cumulative infection rate is 0% (VAX0, ie, vaccinate on day 1), 1% (VAX1) and 5% (VAX5). Three classes of social distancing strategies are considered: (1) stay-home (SHO) if infected, that is, eliminate all non-home-related contacts but continue to maintain contacts within the household; (2) close-schools when cumulative infection rate has reached 1% (CS1) and when it has reached 5% (CS5), that is, eliminate school-related contacts; and (3) isolation (ISO), in which all contacts, including home contacts, of a person are eliminated when a person becomes infectious. For vaccination, five different compliance rates (10%, 30%, 50%, 70% and 90%) and two different vaccine efficacies (30% and 70%) are considered.

VAX0, SHO and ISO are all fairly aggressive interventions because they are implemented either before a person gets infected or immediately on becoming infectious. These are actions taken at the individual or family level. For example, vaccination before the influenza season or isolating a sick child at home is a family decision. Even CS1 is an aggressive intervention in the sense that this action is taken by government officials based on aggregate school sickness levels—closing schools before any outbreaks is typically not done. From these starting points, vaccinations when 1% or 5% of the population is infected (VAX1, VAX5) and closing schools when 5% of the population is infected (CS5) are less aggressive treatments. The five levels of compliance are also variations on aggressiveness in treatments.

These conditions and parameters are consistent with results from other studies and guidelines put out by international organisations. A meta-study of immunisation and slums³⁷ identifies several vaccination-related studies of slums in India. Unfortunately, these studies are for other diseases such as hepatitis B, measles, mumps, malaria and typhoid fever. Nonetheless, slum vaccination rates for children over these ailments range from 25% to 69% for full immunity and from 15% to 55% for partial immunity. Vaccination effectiveness for ILI in India was determined to be about 33%–36%.³⁸ In 2012–2013, of 1000 pregnant women in Srinagar, India, none were vaccinated against

influenza.³⁹ With regard to school closures, the WHO states that school closures may be undertaken proactively (before an outbreak) or reactively (after influenza starts to spread).⁴⁰ WHO recommends that school closure occurs before 1% of the population becomes infected. It also recommends that people (students and staff) stay home when they feel ill. In another meta-study,⁴¹ it was found that school closure, effected when 0.1% of the population was infected, was twice as effective in reducing the total attack rate as school closure occurring after 1% of the population was infected. Moreover, the percentage of people infected before school closure was triggered varied between 0.02% and 10% across several studies.

When a susceptible node is vaccinated, its probability of getting infected by an infectious node is scaled down by the efficacy. If it becomes infectious, its probability of infecting susceptible nodes is also scaled down by the efficacy. In other words, both incoming and outgoing infection probabilities of vaccinated individuals are reduced by the vaccine efficacy. Interventions are applied to slum residents, non-slum residents and the entire region of Delhi.

For each experiment, 25 replicates are simulated for 400 days, and their mean results are reported. The averages are timepoint-wise averages; for example, the mean infection rate at day 100 is calculated by taking the average of the 25 infection rates that occur on day 100 of each replicate. Online supplementary table 2 summarises all the interventions considered, and online supplementary table 3 contains all variables in simulations, including intervention parameters.

Heterogeneities captured

There are several heterogeneous aspects to this problem that motivate the use of an ABM approach: (1) the 298 slum zones have populations that vary by more than four orders of magnitude in size; (2) the geographical extents of slum zones differ; (3) the slum zones are located at irregular spatial intervals throughout Delhi; (4) the activity patterns of people living in slums are different from those in the non-slum region; and (5) each individual interacts with specific others based on colocation.

The implications of these heterogeneities include the following: First, the particular synthetic households that live within slums are predicated on the number of slum zones, their locations and their spatial geometries. These homes have larger family size and hence more home contacts. Second, slum individuals have different activity patterns which change the colocated contacts of each slum person: that is, with whom they interact and for how long. For example, see the supplementary information of Chen *et al.*²⁷ The average total contact durations by activity type and by slum/non-slum residents are provided, which show that non-slum people have greater contact durations for work, school and college activities, but less for home and other types. Overall, a slum person has about 50% greater total contact duration per day compared with a non-slum person. The same supplementary information

shows that in the age range 20–60 years (by year), women who live in slums have more contacts per day than their male counterparts. However, women whose homes are outside of slum regions have average number of daily contacts that are below their male counterparts.

RESULTS AND ANALYSIS

Our results are grouped as follows: (1) comparison of network 1 and network 2 for base case and intervention cases; (2) results for both networks based on demographic classes, such as slum/non-slum, gender and age groups, for a wider range of intervention strategies; (3) comparison of network 1 with the non-slum population of network 2; (4) effects of PIs and NPIs for a wide range of parameter values; and (5) effects of different resource allocation strategies.

All differences are tested with the two-sample t-test and they are all statistically significant with P values smaller than 2.2e-16. The 95% CIs are given for each comparison. Here is a brief summary of selected results with examples of mechanisms, to provide a high-level overview. Details of results follow this summary, and these details matter because there are many factors (inputs) in a simulation whose interactions change results.

1. Ignoring the unique attributes of slums in a population overestimates the benefits of the interventions. For example, in the case of vaccination intervention (efficacy 30% and compliance 30%), the values for the epidemic size (ie, cumulative percentage of infected), peak infection rate (ie, maximum percentage of a population infected on any day) and time to peak are 33.1%, 3.0% and 184 days, respectively, in network 2, whereas they are 23.3%, 1.34% and 286 days in network 1. In relative terms, the epidemic size and peak infection rate are underestimated by 42.2% and 123.2%, respectively, while the time to peak is overestimated by 35.7% in network 1 (see figures 1 and 2 and online supplementary table 4). The larger family sizes for slum families in network 2 and the increased number of edges result in larger outbreaks and faster time to peak infections.
2. Interventions are more effective in network 1 than in network 2 for all types of interventions: vaccination, closing schools, staying home and isolation. These trends also hold over wide ranges of efficacy and compliance (see figures 3, online supplemental figures 1 and 2). Hence, not accounting for slums gives overly optimistic results for the effectiveness of the interventions. The reduced average family size in network 1 means fewer within-home edges, which slows infection and reduces spreading. Closing schools and staying home interventions do not affect home edges. However, the magnitude of this effect varies with intervention conditions (eg, compliance rate, time at which intervention is applied).
3. Cumulative infection rates by subpopulation in network 2 show that slums sustain greater infection rates

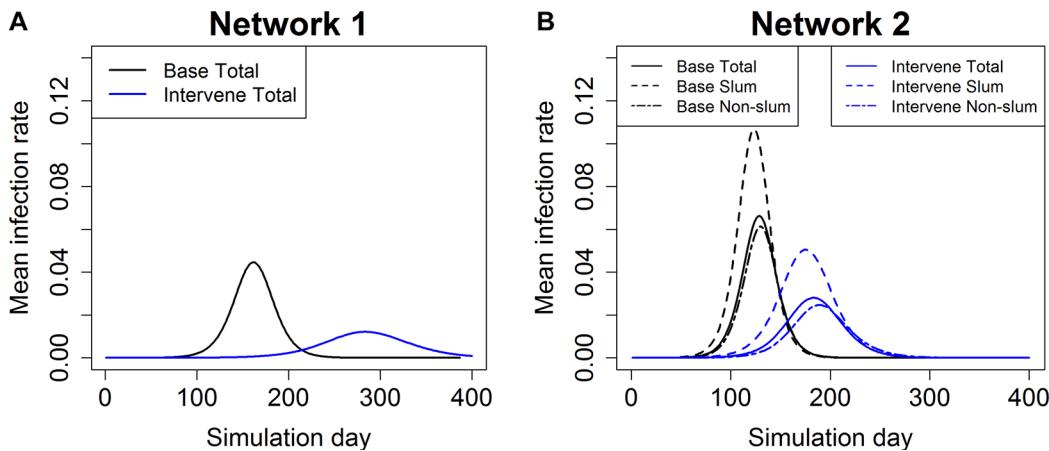


Figure 1 Epidemic curves for base case and vaccination case in (A) network 1 and (B) network 2. Each timepoint in the curve is an average over 25 replicates. The vaccines are given randomly to 30% of the entire population and the vaccine efficacy is 30%. For network 2, epidemic curves are shown for total population and slum and non-slum subpopulations. ‘Intervene Total’ refers to the epidemic curve of the entire Delhi population when the vaccine intervention is applied. ‘Intervene Slum’ refers to the epidemic curve for just the slum population, and ‘Intervene Non-slum’ refers to the epidemic curve for just the non-slum population for the intervention case. Epidemic curves for a variety of compliances and efficacies are reported in online supplementary figures 1 and 2. (A) Total Delhi network 1 and (B) Total Delhi network 2.

- than non-slums under all intervention scenarios, sometimes by as much as 44.0% (see figure 4 and online supplementary table 5 for more details). This is due to the greater household sizes in slums.
- For network 2, under a wide range of intervention compliance rates (10%–90%), the ISO strategy is up to 32% more effective in containing an outbreak than vaccination (for 30% efficacy). Staying home is up to 18% more effective than vaccination at 50% compliance (see figure 3 and online supplementary table 6 for more details). Isolation, although hard to implement from practical considerations, is most effective because edges to susceptible individuals are removed (isolation also provides a good comparative case). Dif-

ferences between staying home and vaccination depend on compliance rates.

- For network 2, delay in triggering interventions has 7.3%–44.0% more adverse effect in slums than in non-slum regions across compliance rates from 10% to 90% (see figure 4 and online supplementary table 7 for more details). Early interventions mean actions are taken when outbreaks are smaller and are therefore more readily contained.
- Comparison of network 1 (figure 3A) with the non-slum population (figure 4B) of network 2 shows that just the presence of slum-specific activities and interactions with non-slum population makes social dis-

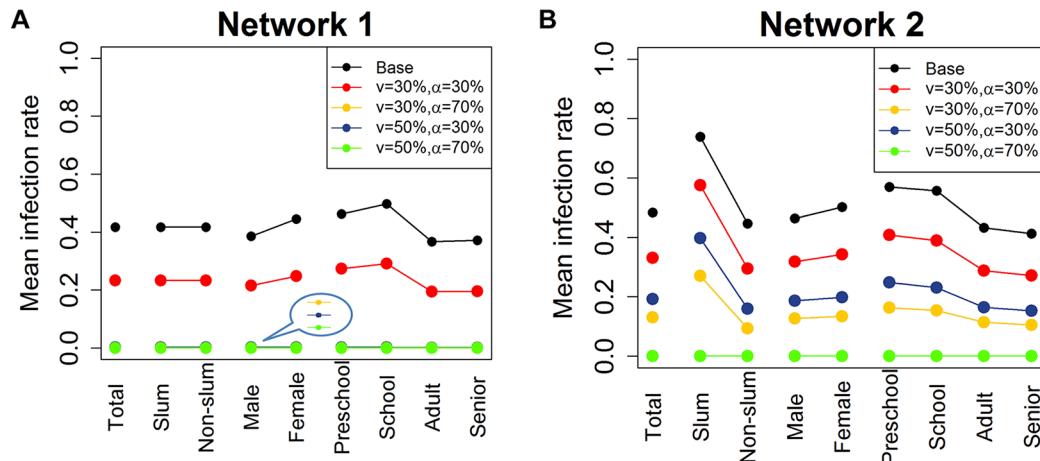


Figure 2 Mean cumulative infection rates for different subgroups in the two networks. Two vaccination rates ($v=30\%$, 50%) and two vaccine efficacy rates ($\alpha=30\%$, 70%) are considered. Individuals are chosen at random in the entire network for vaccination on day 0. Mean infection rates are calculated within each group. The last several lines in the plot for network 1 are overlapping at the bottom because the mean infection rates are almost 0 under those scenarios. ‘Total’ refers to the entire population of Delhi. ‘Slum’ and ‘Non-slum’ refer to slum and non-slum regions, respectively. ‘Male’ and ‘Female’ denote the total number of men and women in Delhi, respectively. Age groups are denoted by ‘Preschool’, ‘School’, ‘Adult’ and ‘Senior’. (A) Total Delhi network 1 and (B) Total Delhi network 2.

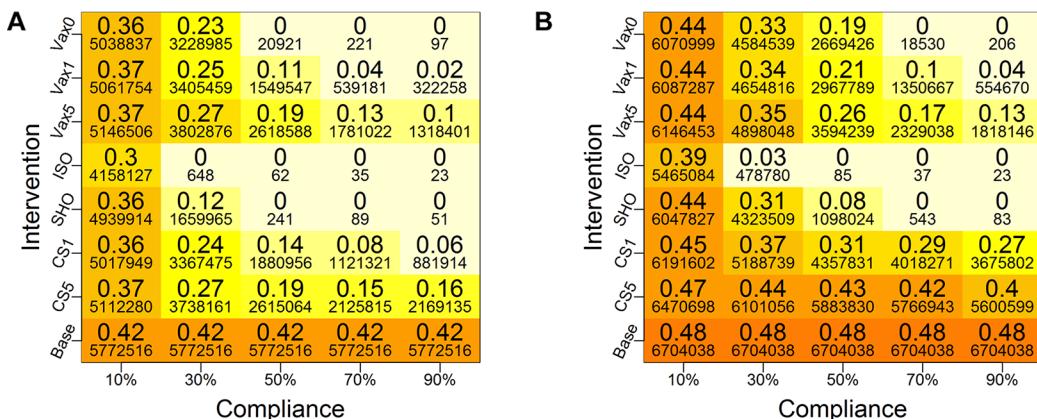


Figure 3 Mean cumulative infection rates under different interventions for (A) network 1 and (B) network 2. The larger font numbers are fractions of populations that are infected and the smaller font numbers are counts of infected individuals. Colours of the boxes correspond to the values of the large numbers (ie, fractions of infected), and the same scheme is used for both plots for comparisons—and for all plots in this paper. Five different compliance rates are examined (10%, 30%, 50%, 70% and 90%), and four types of intervention strategies (vaccination (VAX), close-schools (CS), stay-home (SHO) and isolation (ISO)) are considered. For vaccines, three different trigger points are considered: when the cumulative infection rate reaches 0% (VAX0), 1% (VAX1) and 5% (VAX5) of the total population. The vaccine efficacy is set at 30%. For CS, two trigger points are used: when cumulative infection rate reaches 1% (CS1) and 5% (CS5). Compliant individuals are selected at random from the entire Delhi population, and the cumulative infection rates are calculated for each network. Base is the baseline case with no interventions. (A) Total Delhi network 1 and (B) Total Delhi network 2.

- tancing-based interventions less effective in the non-slum regions of network 2.
7. A full-factorial design that splits resources between slum and non-slum regions indicates that the most effective intervention is to give vaccines to slums and apply social distancing to non-slums. Applying vaccine and social distancing to slum regions is the next most effective intervention (see figure 5). By applying social distancing to non-slums, these individuals are

kept isolated from slum individuals who are infected. The greatest benefits accrue to the slum populations.

Comparison between networks 1 and 2: base case versus interventions

We start with a comparative analysis of the influenza epidemic, with and without interventions, on network 1 and network 2, to measure the impact of integrating slums in the population on epidemic measures. Figure 1

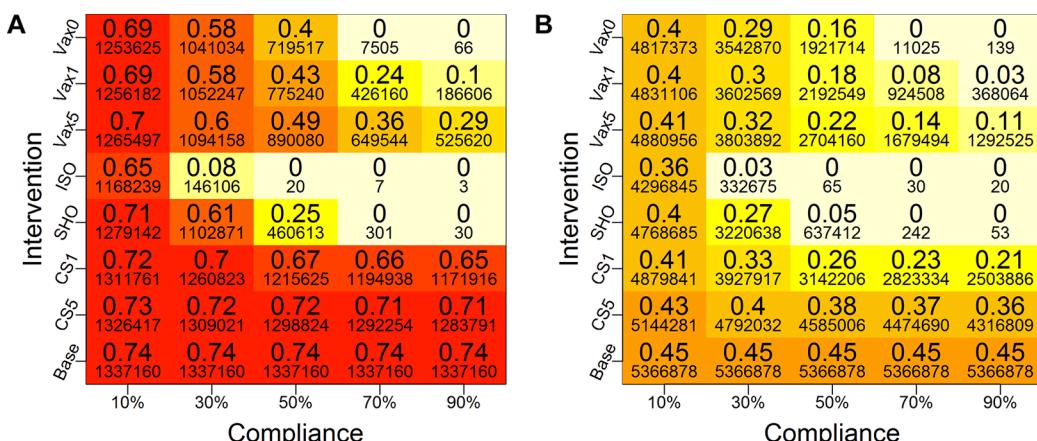


Figure 4 Heat map of cumulative infection rates in (A) slum and (B) non-slum regions of network 2 under different intervention conditions. The colours of boxes correspond to the larger numbers in the boxes—the cumulative infection rates—and the two plots use the same scheme for comparisons. Darker colours correspond to higher infection rates. The smaller font numbers are counts of infected individuals. The vaccination efficacy is fixed at 30%. Five different compliance rates (10%, 30%, 50%, 70% and 90%) and four types of intervention strategies (vaccination (VAX), close-schools (CS), stay-home (SHO) and isolation (ISO)) are considered. For vaccines, three different trigger points are considered: when cumulative infection rate reaches 0% (VAX0), 1% (VAX1) and 5% (VAX5). For CS, two trigger points are used: when the cumulative infection rate reaches 1% (CS1) and 5% (CS5). Compliant individuals are selected randomly from the entire Delhi population, and the mean infection rates are calculated separately for the slum and non-slum subpopulations. Although not reported here, qualitatively similar results are found for other transmission rates, as well as for higher vaccine efficacy (70%). Base is the baseline case with no interventions. (A) Slum region of network 2 and (B) Non-slum region of network 2.

	Total Network 2	Slum Network 2	Non-slum Network 2	Total Network 1
VsSs	0.44 6043049	0.55 989079	0.42 5053971	0.37 5176345
VnSs	0.43 5938919	0.58 1042116	0.41 4896803	0.36 4995753
SsVn	0.44 6023678	0.67 1217415	0.40 4806263	0.36 4986302
SvSn	0.44 6104571	0.72 1309577	0.40 4794993	0.36 5016324
Base	0.48 6704038	0.74 1337160	0.45 5366878	0.42 5772516

Figure 5 Mean cumulative infection rates for each category listed on the x-axis, for network 2 and network 1, under four different intervention scenarios. The colour scheme of the boxes is based on the large values in the boxes—the cumulative infection rates. Darker colours correspond to higher infection rates. Smaller font values are the number of infected individuals. The vaccine efficacy is set at 30%. VsSs refers to the case when vaccines and social distancing are both applied to slum residents; VnSs refers to the case when vaccines and social distancing are applied to non-slum residents. Similarly, VsSn means vaccines are given to slums and stay-home is applied to non-slums; and VnSs means vaccines are given to non-slums and stay-home is applied to slums. Base refers to the case where no intervention is applied.

shows the average simulation time histories for the base case, and when vaccination is applied randomly to 30% of the population in each network with vaccine efficacy set at 30%. Simulations for other vaccine efficacies and compliance rates give qualitatively similar results. Two sets of those results are shown in online supplementary figures 1 and 2. Note that network 1 does not distinguish between slum and non-slum individuals, so the epidemic curve is not split by subpopulation.

Results in network 2 differ significantly from results in network 1 for both the base case and intervention case. In network 2, the epidemic starts earlier, peaks earlier, has a larger epidemic size and has higher peaks compared with the corresponding epidemic quantities in network 1. Thus, if policy planners ignore slums and use network 1 to plan, there will be a false sense of security and lack of urgency to implement interventions. For both the base case and the intervention case, ignoring unique characteristics of the slums will result in an underestimation of the infections and the speed of spread.

For the intervention cases, the time to peak infection decreases by 35.7%, that is, from 286 days for network 1 to 184 days for network 2, meaning an influenza epidemic would peak roughly 100 days earlier than one would expect based on the results from network 1. For the base case, time to peak infection drops by 20.8%, that

is, 34 days' reduction for network 2 as compared with network 1.

Percentage changes and differences must be viewed cautiously, and to illustrate this point we present data for the key parameters in online supplementary tables 4 and 8. The difference in the peak infection rate (ie, the maximum fraction of daily infected individuals during the simulation) between networks 1 and 2 for the base case is 2.2%, or 47.6% in percentage change (see online supplementary table 8). For the intervention case shown in online supplementary table 4, the difference between the two networks is less (1.7%), but the percentage change is more (123.2%) because the magnitudes of the peak infection rates are reduced when effective interventions are used. We make note of this here and mainly use the percentage change values in discussing results. For more detailed comparison between vaccination intervention and the base case in network 1 and network 2, we refer to supplementary tables 7 and 9 and supplementary figures 4 and 5.

Comparison between networks 1 and 2 based on individual demographic information

We divide the Delhi population into strata by age, gender and geographical home location (ie, slum and non-slum), and analyse the mean cumulative infection rates by subpopulation for the two networks. In simulations, individuals are chosen at random in the entire network for vaccination. Various vaccination scenarios are investigated.

Figure 2 displays the cumulative infection rate results. On the x-axis, 'Total' refers to the entire population of Delhi. There are three breakdowns of the entire population. 'Slum' and 'Non-slum' refer to slum and non-slum regions, respectively. 'Male' and 'Female' denote the total number of men and women in Delhi, respectively. Four age groups are considered: 'Preschool' (0–4), 'School' (5–18), 'Adult' (19–64) and 'Senior' (65+). The black lines correspond to the mean cumulative infection rates for the base case. Other curves indicate vaccination strategies under different levels of vaccination rate (v) and vaccine efficacy (α). Two vaccination rates (30% and 50%) and two vaccine efficacy rates (30% and 70%) are shown in the figure.

For network 1, vaccination rate of 50% or higher stops the epidemic for all categories of individuals, regardless of vaccine efficacy. An efficacy of 70% also contains the epidemic, given a vaccination rate of at least 30%. In comparison, for network 2, either a vaccination rate of 70% is required (not shown in plot for clarity), or a vaccination rate of 50% combined with a vaccine efficacy of 70% is required to stop the epidemic for all categories of individuals.

In network 1, slum and non-slums are treated the same so the infection rates are identical in figure 2. However, all scenarios in network 2 show a higher burden of disease on the slum population. This is due to the fact that slum households have larger family size and more contacts on

average than households in non-slum areas (see Chen *et al*²⁷). As shown later, we find similar patterns of infection in slum and non-slum subpopulations for other interventions such as 'close-schools' and 'stay-home'.

The results in both [figures 1 and 2](#) indicate that ignoring the effect of slums results in overestimation of the benefits of interventions in terms of reduction in the mean cumulative infection rate and peak infection rate, as well as the time to peak. This optimism holds for slum, non-slum and total population under various levels of vaccination rates and efficacy rates in network 2 (see online supplementary table 10 for more detailed comparison of results between slum and non-slum in network 2).

Comparison between networks 1 and 2 across a wide range of intervention strategies

Next, we consider a variety of intervention strategies for comparative analysis. We consider vaccination (VAX), school closure (CS), stay home (SHO) and isolation (ISO) strategies. For vaccines, three different trigger points are considered: when cumulative infection rate reaches 0% (VAX0), 1% (VAX1) and 5% (VAX5). For CS, two trigger points are used: when the cumulative infection rate reaches 1% (CS1) and 5% (CS5). Under the SHO strategy, all non-home activities and interactions are eliminated, but all contacts within the household are maintained. Under ISO an individual has no contact with other individuals (even home interactions are eliminated). The SHO and ISO interventions are implemented for compliant infectious individuals, after they become infectious, for the entire infectious duration.

[Figure 3](#) displays average cumulative infection rates in network 1 and network 2 for a wide range of intervention strategies. For each strategy, five different compliance rates are considered, that is, 10%, 30%, 50%, 70% and 90%. The cumulative infection rates (ie, fractions) are displayed as larger numbers in boxes, while smaller font numbers are the actual number of infected individuals. Darker colours correspond to higher infection rates. Note that compliance rate is simply the vaccination rate for strategies VAX0, VAX1 and VAX5. Compliant individuals are selected at random from the entire population. The 'Base' values do not vary with compliance because the base case has no intervention. Note that all heat maps in this paper use the same colour scheme so that colours can be compared across figures.

Since network 1 does not distinguish between slum and non-slum populations, we only compare the two networks for the whole of Delhi. The general pattern is similar for both networks. However, all interventions have a larger effect on network 1 under the same compliance rate (ie, corresponding numbers are uniformly lower for network 1 than for network 2). The infection rates drop to 0 at a smaller compliance rate for VAX0, SHO and ISO strategies in network 1 as compared with those for network 2.

At a high level, among all intervention strategies, early vaccination (VAX0 and VAX1), ISO and SHO are

more effective than the other strategies, and this is more readily observed at higher compliance rates. For these more effective strategies, the interventions per person are implemented right after (or very shortly after) the person is infected. For example, SHO is implemented immediately after a person becomes infectious. Thus, a person who becomes infectious can infect their family members, but if these other members become infectious, then they, too, will be confined to home. Thus, home-bound people can infect their family members, but no one beyond their family (for 100% compliance). As compliance rate increases, this effect approaches, roughly, a 'family-based' isolation intervention (similar to ISO), consistent with the results in [figure 3](#) and in subsequent results.

Effect of vaccination versus social distancing on slum and non-slum subpopulations

We now compare the impact of vaccination and social distancing on slum and non-slum subpopulations from network 2. Social distancing interventions are CS, SHO and ISO.

The mean cumulative infection rates (and actual numbers of infections underneath) for each compliance level are shown in the heat maps in [figure 4](#) for slum and non-slum populations in network 2. The axis labels are identical to those in [figure 3](#), as is the colour scheme of the cells. The base case values are constant since there is no intervention and hence no compliance. Darker colours correspond to higher infection rates.

Compared with the base case, all interventions reduce infection rates to some extent. As the compliance rate increases, infection rates drop for all interventions. Infection rates drop to 0 in slum and non-slum regions at a compliance level of 70% or higher, under SHO, ISO and VAX0 strategies. Early interventions or lower trigger levels reduce the infection rates significantly, and this effect increases with compliance rate.

The following observations can be made from [figure 4](#). Social distancing, that is, SHO, at low and intermediate compliance and CS at all compliance levels, is less effective in slum regions as compared with non-slum regions. This is because CS only eliminates school interactions for those attending school, and there are fewer school edges in slums compared with non-slum areas, as shown in online supplementary figure 6. The effectiveness of CS in slums is mitigated by the greater average number and duration of interactions at home in slums as compared with non-slums (see online supplementary figure 6 and Chen *et al*²⁷). Thus, if a person is sick, there is a greater chance of transmitting contagion to family members, who then may have activities outside of school, thus circumventing the CS intervention. At high compliance, SHO is effective because all interactions outside home (including school) are eliminated.²⁷

These observations are also supported by supplementary figure 7, which contains numbers of edges used to transmit contagion for a base case run of [figure 1](#). There are several effects that bear on the above observations. First, in the

cases of activities ‘work’, ‘other’ and ‘school’, the number of edges transmitting contagion from slums to non-slums is greater than the reverse: from non-slum to slum. Second, in two of these three activity categories, there are more slum to non-slum transmissions than slum to slum transmissions. Edges of transmission for slum dwellers are dominated by home interactions. The infected homes in slums serve as launching points to drive disease to non-slums through slum to non-slum interactions. (There are no ‘mixed’ edges at homes, and shopping and college activities have low levels of slum activity because of socioeconomic factors.) We will see the effects of these mechanisms in figure 5, but we now return to figure 4.

Isolation works well at 30% or higher compliance rates, but it is a much harder strategy to implement, especially in slums. However, it is considered here for comparative analysis. Vaccination also produces marked decreases in cumulative outbreak sizes as compliance increases. However, CS is generally less effective because this intervention removes only a fraction of interactions for a fraction of the population, that is, school-aged children. Simulations were also run for 70% vaccine efficacy. Since results are qualitatively similar for those parameters, these plots are provided in online supplementary figure 3.

Comparison between network 1 and non-slum areas of network 2

Note that network 1 treats all parts of the region as non-slum, that is, all individuals follow non-slum activities and demographics. In order to capture the additional disease risk to the non-slum population that arises from the interactions with the slum population, we compare network 1 in figure 3A with the non-slum population of network 2 in figure 4B. In base case, the additional disease risk to the non-slum population goes up from 42% to 45%. However, the beneficial effects of social distancing strategies drop by a large amount; for example, CS strategies are 5 to 20 percentage points less effective in the non-slum areas of network 2. This effect changes non-linearly with the compliance rate. As compliance rate goes up, the difference between performance of network 1 and non-slum parts of network 2 goes up in CS1 and CS5. This implies that in network 2, non-slum population requires much higher levels of compliance to achieve the same results as in network 1. This difference is less stark for vaccination-based interventions, that is, VAX0, VAX1 and VAX5. This is expected since the effect of vaccination is less dependent on interactions; it is only through herd immunity that interactions come into play.

Constrained resource allocation among slum and non-slum areas

We consider a specific scenario under network 2. If only a limited number of vaccines are available, and only a certain fraction of individuals can be kept home during an epidemic, how should these interventions be applied to the slum and non-slum regions so that

the epidemic can be controlled effectively? Given that slum residents’ attributes differ from those of non-slum residents, is there a strategy that works better in slums than in non-slum areas? The total population in Delhi is about 13.8 million, which includes about 1.8 million slum residents. We assume that only 10% of the total population can be covered by interventions, half through vaccination and the other half through stay home. Enough vaccines are available to cover 5% of the total population (ie, 692 183 vaccinated, corresponding to about 38.25% of slum or 5.75% of non-slum population), and 5% of the individuals can stay home (692 183 individuals; this is applied to only the infected individuals). Note that an individual may receive a vaccine and also stay at home if this individual, in spite of being vaccinated, gets infected.

We consider four different ways of applying interventions to 10% of the total population: (1) apply both interventions to slums, that is, give all vaccines to slums and apply SHO only in the slums (VsSs); (2) apply all interventions to non-slum areas (VnSn); (3) give vaccines to slums and SHO to non-slums (VsSn); and (4) give vaccines to non-slums and apply SHO to slums (VnSs).

For both types of intervention, the same number of individuals is chosen randomly from slum or non-slum areas. Ten per cent of the total Delhi population amounts to 76.5% of slum population, 11.5% of the non-slum population, or a combination of 38.25% of the slum and 5.75% of the non-slum population (ie, half from slums and half from non-slums). Figure 5 shows the mean cumulative infection rates, as well as the number of infected from the entire population of Delhi, and the slum and non-slum areas under each of the four scenarios. The first three columns refer to network 2, and the last column shows results for network 1. Since network 1 does not distinguish between slum and non-slum areas, the infection rates in each subpopulation remain the same as for the total population.

Comparison of the last two columns in figure 5 indicates that the non-slum population in network 2 faces 3%–5% additional disease risk compared with network 1 in all cases. This is primarily driven by the increased interactions within slum populations and between slum and non-slum populations in network 2.

In figure 5, all four intervention strategies produce essentially the same total attack rates (around 43%–44%), a drop of 4%–5% over the base case. The dominant effect on network 2 is the benefits that primarily accrue to the slum population for the VsSs and VsSn strategies because they drive down the fraction of infected slum residents from 0.74 to 0.55 or 0.58. Also, as described in the context of figure 4 and online supplementary figure 6 above, social distancing of the non-slum residents helps to isolate them from the infected slum residents. Results such as these may be helpful to policy makers in breaking the poverty trap in economically poor regions.⁴²

Also, the strategy of vaccinating non-slums and social distancing slums (VnSs) is not as effective as the interventions in rows 1 and 2 of figure 5. This is a counterintuitive

result, since the density of population is much higher in the slums, which may lead to the belief that social distancing in slums will break up the dense clusters. However, a careful examination shows that keeping slum residents home is not an effective social distancing strategy because their family size is, on average, almost three times the family size of non-slum households.²⁷ The high level of mixing at home makes social distancing ineffective in slums unless the infected individual is completely isolated. However, complete isolation is not viable in slum areas where the entire household may live in a single room.

DISCUSSION

With slum populations expected to grow to two billion by 2030,⁴³ it is becoming increasingly urgent to understand how to control the spread of infectious diseases in slum areas and measure its effect on urban populations. To our knowledge, a detailed study of interventions to control influenza epidemics in slums, using an agent-based simulation model, has never been done before. Slum conditions are important for a city beyond the direct effects of disease transmission. For example, civil wars may be precipitated or exacerbated by disease outbreaks because they decrease social health and welfare.⁴⁴

Even though slum regions contain only 13% of the total population of Delhi, Chen *et al*⁴ show that omitting their attributes leads to underestimation of the overall infection rate and the peak infection rate of the epidemic. This paper extends that work by evaluating the differential impact of interventions on slum and non-slum regions. Various vaccination and social distancing strategies are analysed under different scenarios that show that the slum population is more prone to infections under the same control measures. Furthermore, taking account of slum populations significantly alters the disease dynamics in the *entire* population. Differences in key measures are demonstrated between the cases of accounting for slum populations and not: for example, a 100% increase in the peak attack rate in some cases when slum regions' characteristics are taken into account, compared with the case when they are ignored.

Figure 4, which compares infections in slum with non-slum areas, shows that at very high compliance rates, some interventions can be equally effective in both slums and non-slums. However, such high compliance rates are typically not feasible due to practical realities on the ground, and also because they require timely diagnosis of infected cases. For SHO to be effective, the coverage rate needs to be 70% or more in both slums and non-slums, and the diagnosis of the infected individuals needs to be correct and immediate. In other words, effective control of a contagious epidemic in a high-density place like Delhi would require either early and drastic action (eg, ISO) or a highly compliant set of individuals, or a combination of these features.

This work overall demonstrates the power of agent-based and population modelling to evaluate complicated interaction-based epidemiological phenomena. Clearly, there are limitations to this work (several are itemised below). But these agent and population approaches provide a platform for adding additional complexity. All of the figures demonstrate that quantitative results depend on complicated interplay among inputs. These results are important because they inform policy decisions. An equally important benefit of this type of work, but not often stated, is developing intuition about epidemic dynamics (in this case, with the effects of slums) to enable decision makers to reason about nuanced interactions among effects to a degree that is hard to obtain with other approaches that lack this level of detail. However, we believe that other modelling approaches may also be valuable in understanding epidemic dynamics in slum populations.

Despite the detailed modelling effort, there are limitations to this work and areas for improvement in the future. For example, this model assumes that both slum and non-slum individuals have the same level of immunity. This may not be true for seasonal infections. Previous researchers have argued that individuals who live in smaller family sizes, who have access to household amenities and maintain a high level of personal cleanliness face declining microbial exposure, which can modify their immune response and reduce their level of tolerance to respiratory infections.⁴⁵ Slum households characterised by larger family size and overcrowding are likely to encounter much higher microbial exposure and therefore may be protected by their greater immunity.^{16 17}

Areas for future work include (1) examination of different population level base attack rates derived from different transmission probabilities; (2) different susceptibilities and infectivity for individual agents, for example, based on age; (3) effects of asymptomatic infections (although we have addressed this to some extent with compliance and efficacy of interventions); (4) seasonal effects^{46 47}; (5) effects of immunity for an individual from previous infections (in previous seasons); (6) evaluation of interaction of different strains from season to season; (7) comparison of tropical versus subtropical factors; (8) evaluation of a specific outbreak scenario; (9) impact of sickness on absenteeism from work and its economic ramifications; (10) effects on rural versus urban populations; (11) using combinations of interventions rather than one at a time (this was only done here in figure 5)—however, to disambiguate results, it is prudent to first examine individual interventions; (12) effect of changing disease transmission rate for different activity types; (13) effect of changing contact times at different locations; and (14) to capture close proximity transmission, one could use actual physical proximity (here, we use colocation). Finally, just as changes in modelling details can change model results, so too changes in the conditions in actual outbreaks can change results; some of

these factors are listed above. It is essentially impossible to capture all of these effects—many of which are unknown—down to the level of individual humans.

Public health implications

This research demonstrates that modelling slum populations is important for understanding disease dynamics and for designing effective control measures. Ignoring the influence of slum characteristics on their urban environment will significantly underestimate the speed of an outbreak and its extent, and hence will lead to misguided interventions by public health officials and policy planners. Lessons from this research can be applied in the field, and observations collected from the field can provide valuable data to improve the models and validate the results. For example, our results show that a slum resident has about 50% greater total contact duration per day compared with a non-slum resident. This makes social distancing-based interventions more taxing in the slum population. Public health policy makers may want to subsidise pharmaceutical resources for the slum population to make them more affordable. Similarly, we find women in slums have a higher number of contacts per day than their male counterparts, whereas in non-slum regions women have a fewer number of daily contacts than their male counterparts. This kind of information can be used to prioritise the distribution of limited resources; for example, women could be given preference over men for vaccination in slum areas. This research provides simulation-based evidence that in general social distancing strategies are ineffective in slums because of a large number of contacts at home. Unless one applies complete isolation, which is not feasible in slums, just staying at home still keeps a large number of contacts and pathways of spread intact.

Contributors AA, SE, CJK, AM, MM, SS, AV designed and conceived the study. SC carried out the experiments and simulations. SC, CJK, AM performed data analysis. CJK, BL, AM, MM, EKN, MLW helped with reviewing the results and writing the paper.

Funding This work has been partially supported by the Defense Threat Reduction Agency (DTRA) (grant no. HDTRA1-11-1-0016 and HDTRA1-11-D-0016-0001), National Institutes of Health (NIH) (grant no. 1R01GM109718), National Science Foundation (NSF) (grant no. CCF-1216000, CNS-1011769 and NRT-DESE-154362), and NIH Models of Infectious Disease Agent Study (MIDAS) (grant no. 2U01GM070694-11 and 3U01FM070694-09S1).

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Data pertaining to figures and statistical analysis are partially provided in the supplementary file, and also can be obtained by contacting the corresponding author through email.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

- Thompson WW, Shay DK, Weintraub E, et al. Mortality associated with influenza and respiratory syncytial virus in the United States. *JAMA* 2003;289:179–86.
- Thompson WW, Shay DK, Weintraub E, et al. Influenza-associated hospitalizations in the United States. *JAMA* 2004;292:1333–40.
- Chandra N. Influenza viruses circulate in Delhi throughout the year, 2012. <http://indiatoday.intoday.in/story/influenza-viruses-circulate-in-delhi-throughout-the-year/1/179252.html>.
- Chen J, Chu S, Chungbaek Y, et al. Effect of modelling slum populations on influenza spread in Delhi. *BMJ Open* 2016;6:e011699.
- Barrett C, Bisset K, Leidig J, et al. Economic and social impact of influenza mitigation strategies by demographic class. *Epidemics* 2011;3:19–31.
- Liao S, Ma Y, Chen J, et al. Paid Sick-Leave: Is It a Good Way to Control Epidemics? In: Glass K, Colbaugh R, Ormerod P, et al. (eds). *Complex Sciences. Complex 2012. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol 126. Cham: Springer, 2013.
- Dorratolaj N, Marathe A, Lewis B, et al. 2015. *Cost-benefit analysis of vaccine-based interventions to control pandemic influenza* 3rd Conference on Modeling of Infectious Diseases. Chennai, India: Institute of Mathematical Sciences.
- Peasah SK, Purakayastha DR, Koul PA, et al. The cost of acute respiratory infections in Northern India: a multi-site study. *BMC Public Health* 2015;15:330.
- WHO Influenza vaccines. WHO position paper. *Weekly Epidemiol Rec* 2005;80:279–87.
- Integrated Disease Surveillance Program. Influenza Surveillance Lab Network. <http://idsp.nic.in/index1.php?lang=1&level=1&sublinkid=5789&lid=3722>.
- Broor S, Krishnan A, Roy DS, et al. Dynamic patterns of circulating seasonal and pandemic a(H1N1)pdm09 influenza viruses from 2007–2010 in and around Delhi, India. *PLoS One* 2012;7:e29129.
- Chadha MS, Broor S, Gunasekaran P, et al. Multisite virological influenza surveillance in India: 2004–2008. *Influenza Other Respir Viruses* 2012;6:196–203.
- Chadha MS, Hirve S, Dawood FS, et al. Burden of seasonal and pandemic influenza-associated hospitalization during and after 2009 A(H1N1)pdm09 pandemic in a rural community in India. *PLoS One* 2013;8:e55918.
- Nijman J. A study of space in mumbai's slums. *Tijdschr Econ Soc Geogr* 2010;101:4–17.
- Lofgren E, Fefferman NH, Naumov YN, et al. Influenza seasonality: underlying causes and modeling theories. *J Virol* 2007;81:5429–36.
- Baker M, McNicholas A, Garrett N, et al. Household crowding a major risk factor for epidemic meningococcal disease in Auckland children. *Pediatr Infect Dis J* 2000;19:983–90.
- Burström B, Diderichsen F, Smedman L. Child mortality in Stockholm during 1885–1910: the impact of household size and number of children in the family on the risk of death from measles. *Am J Epidemiol* 1999;149:1134–41.
- Souza LS, Ramos EA, Carvalho FM, et al. Viral respiratory infections in young children attending day care in urban Northeast Brazil. *Pediatr Pulmonol* 2003;35:184–91.
- Gulyani S, Bassett EM, Talukdar D. A tale of two cities: A multidimensional portrait of poverty and living conditions in the slums of Dakar and Nairobi. *Habitat Int* 2014;43:98–107.
- Acolin A, Chattaraj S, Wachter SM. Urban Governance and Development of Informality in China and India. In: Birch EL, Chattaraj S, Wachter SM, eds. *Slums: How Informal Real Estate Markets Work*: University of Pennsylvania Press, 2014.
- Madise NJ, Ziraba AK, Inungu J, et al. Are slum dwellers at heightened risk of HIV infection than other urban residents? Evidence from population-based HIV prevalence surveys in Kenya. *Health Place* 2012;18:1144–52.
- Nossiter A. Cholera epidemic envelops coastal slums in West Africa. *New York Times* 2012.
- Desai VK, Kapadia SJ, Kumar P, et al. Study of measles incidence and vaccination coverage in slums of Surat city. *Indian J Community Med* 2003;28.
- Xia H, Nagaraj K, Chen J, et al. Synthesis of a high resolution social contact network for Delhi with application to pandemic planning. *Artif Intell Med* 2015;65:113–30.
- Barrett CL, Beckman RJ, Khan M, et al. Generation and analysis of large synthetic social contact networks. In *Winter Simulation Conference*: IEEE, 2009:1003–14.
- Bissett K, Cadena J, Khan M, et al. An integrated agent-based approach for modeling disease spread in large populations to support health informatics. In *2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*: IEEE, 2016:629–32.



27. Chen J, Chu S, Chungbaek Y, et al. Supplemental information for effect of modelling slum populations on influenza spread in Delhi. *BMJ Open* 2016;6:e011699.
28. Koopman J. Modeling infection transmission. *Annu Rev Public Health* 2004;25:303–26.
29. Marathe A, Lewis B, Chen J, et al. Sensitivity of household transmission to household contact structure and size. *PLoS One* 2011;6:e22461.
30. Kaji M, Watanabe A, Aizawa H. Differences in clinical features between influenza A H1N1, A H3N2, and B in adult patients. *Respirology* 2003;8:231–3.
31. Meeyai A, Praditsitthikorn N, Kotirum S, et al. Seasonal influenza vaccination for children in Thailand: a cost-effectiveness analysis. *PLoS Med* 2015;12:e1001829.
32. Longini IM, Nizam A, Xu S, et al. Containing pandemic influenza at the source. *Science* 2005;309:1083–7.
33. Choudhry A, Singh S, Khare S, et al. Emergence of pandemic 2009 influenza A H1N1, India. *Indian J Med Res* 2012;135:534–7.
34. Gurav YK, Pawar SD, Chadha MS, et al. Pandemic influenza A(H1N1) 2009 outbreak in a residential school at Panchgani, Maharashtra, India. *Indian J Med Res* 2010;132:67–71.
35. Broor S, Gupta S, Mohapatra S, et al. Emergence of 2009A/H1N1 cases in a tertiary care hospital in New Delhi, India. *Influenza Other Respir Viruses* 2011;5:e552–557.
36. Ramamurti N, Pillai LC, Gunasekaran P, et al. Influenza activity among the paediatric age group in Chennai. *Indian J Med Res* 2005;121:776–9.
37. Crocker-Buque T, Mindra G, Duncan R, et al. Immunization, urbanization and Slums—a Systematic Review of factors and Interventions. *BMC Public Health* 2017;17:1–16.
38. Vashisht VM, Kalra A, Choudhury P. Influenza vaccination in India: position paper of Indian Academy of Pediatrics, 2013. *Indian Pediatr* 2013;50:867–74.
39. Koul PA, Bali NK, Ali S, et al. Poor uptake of influenza vaccination in pregnancy in northern India. *Int J Gynaecol Obstet* 2014;127:234–7.
40. World Health Organisation. Measures in School Settings. Pandemic (H1N1) 2009 Briefing note, 2009;10. http://www.who.int/csr/disease/swineflu/notes/h1n1_school_measures_20090911/en/
41. Jackson C, Mangtani P, Hawker J, et al. The effects of school closures on influenza outbreaks and pandemics: systematic review of simulation studies. *PLoS One* 2014;9:e97297.
42. Bonds MH, Keenan DC, Rohani P, et al. Poverty trap formed by the ecology of infectious diseases. *Proc Biol Sci* 2010;277:1185–92.
43. Un-Habitat. The challenge of slums: global report on human settlements 2003. *Manag Environ Qual Int J* 2004;15:337–8.
44. Letendre K, Fincher CL, Thornhill R. Does infectious disease cause global variation in the frequency of intrastate armed conflict and civil war? *Biol Rev Camb Philos Soc* 2010;85:669–83.
45. Björkstén B. Effects of intestinal microflora and the environment on the development of asthma and allergy. *Springer Semin Immunopathol* 2004;25:257–70.
46. Koul PA, Broor S, Saha S, et al. Differences in influenza seasonality by latitude, northern India. *Emerg Infect Dis* 2014;20:1746–9.
47. Hirve S, Newman LP, Paget J, et al. Influenza seasonality in the tropics and subtropics - when to vaccinate? *PLoS One* 2016;11:e0153003.



FEED THE FUTURE

The U.S. Government's Global Hunger & Food Security Initiative

Innovation Lab for Integrated Pest Management

Pest Risk Assessment of the Fall Armyworm, *Spodoptera frugiperda* in Egypt



AID-OAA-L-15-00001



Contributors

- E.A. Heinrichs, IPM IL Asia Program Manager and Entomologist
- Jaspreet Sidhu, IPM IL Entomologist
- R. Muniappan, IPM IL Director and Entomologist
- Amer Fayad, IPM IL Assoc. Director and Plant Pathologist
- Abhijin Adiga, Research Assistant Professor and Modeler, Biocomplexity Institute
- Achla Marathe, Professor, Biocomplexity Institute of Virginia Tech and the Department of Agriculture and Applied Economics, Virginia Tech
- Joseph Mcnitt, Graduate Research Assistant, Biocomplexity Institute of Virginia Tech
- Srinivasan Venkatramanan, Postdoctoral Associate, Biocomplexity Institute of Virginia Tech

Acknowledgements

- John Bowman, AOR, IPM IL, USAID BFS, Washington
- Annie Steed, USAID Mission Egypt
- Walid Sallam, ACDI VOCA representative, Chief of Party, AMAL Project, Egypt
- Ahmed Heneidy, Plant Protection Research Institute, Agricultural Research Center, Cairo, Egypt
- Asraf Arnaouty, Head of the Department of Economic Entomology, Cairo University, Egypt

Table of Contents

I.	INTRODUCTION	5
II.	BACKGROUND	6
III.	HOST PLANTS.....	8
IV.	TAXONOMY.....	8
V.	IDENTIFICATION	9
VI.	BIOLOGY	10
	Natural Enemies	14
VII.	DAMAGE.....	20
VIII.	HOW TO IDENTIFY AND DIFFERENTIATE THE FAW FROM OTHER SIMILAR SPECIES IN EGYPT	22
IX.	MOBILITY AND DISPERSAL.....	20
X.	SPREAD AND ESTABLISHMENT	20
	Establishment potential	21
	Pathways of introduction and spread	22
	Natural spread (unaided dispersal) based on flight biology	23
	Modeling the natural spread based on ecological suitability, hosts etc.	26
	Wind patterns and its effect on the spread of FAW	28
	Travel	30
	Trade.....	31
XI.	RISK TO OTHER COUNTRIES	32
	Natural spread	32
	Travel	32
	Trade.....	32
XII.	ECONOMIC IMPACT OF FAW	32
XIII.	DEVELOPMENT OF A MANAGEMENT PLAN FOR THE FAW IN EGYPT	35
	Outline for an awareness and management workshops in Egypt (Before and after FAW arrival)	36
XIV.	PREPAREDNESS STEPS.....	36
XV.	REFERENCES	38
	Appendix 1	41

Data	41
Spatiotemporal spread model using cellular automata.....	41
Role of temperature and humidity.....	42
Model description.....	42
Parameter choices.....	43
CA transition rules	43
Metric for evaluating models	43
Appendix 2.....	45
FAW monitoring	45
Appendix 3.....	46
Control measures.....	46

I. INTRODUCTION

The purpose of this document is to conduct a pest risk assessment of the invasive fall armyworm (FAW), *Spodoptera frugiperda* (J. E. Smith) (Lepidoptera: Noctuidae) for Egypt. Invasive species are a major cause of crop loss and can adversely affect food security (Cook et al., 2011). Morimoto and Kiritani (1995) defined exotic (invasive) insect species as "those species reproducing naturally in a designated area where they were brought by unusual means, such as air, ocean current, and accidental or intentional introduction". One of the important problems in pest control is that many invasive insects already have resistance to some pesticides and arrive without their natural enemies, which keep them under control in their native countries.

Quantifying the threat presented by the FAW to Egypt and developing effective biosecurity policy requires: 1) an understanding of the potential sources of the FAW; 2) its likelihood of arriving/entering at a particular location in Egypt; 3) the likelihood of its establishment at specific locations within Egypt; and 4) an estimate of the possible impact.

Pest risk analysis and assessment are performed to determine whether a pest should be regulated and the strength of any phytosanitary measures to be taken against it. The assessment is the evaluation of the likelihood of entry, the likelihood of establishment upon arrival, spread of a pest or disease within the country, an estimate of the possible impact, the resources available to tackle the pest, and a road map for its management. It is the evaluation of the potential yield loss and also includes recommendations to reduce the adverse effects on human or animal health and biodiversity arising from the use of toxic pesticides.

The assessment of the economic impact resulting from FAW invasion requires an integration of information on: 1) the biology, ecology and damage caused by the FAW; 2) its entry; 3) establishment; 4) spread; 5) valuation of assets at risk; and 6) market consequences.

II. BACKGROUND

The Fall Armyworm is a polyphagous pest that is native to the tropics in North and South America. In North America, the FAW will move north in the late summer and early fall, which is when it does most of its damage. It then dies off in the cold weather. It affects all stages of plant development and is difficult to control. The pest can survive year-round in the southeastern United States due to the warm and humid climate. In Africa, the FAW was first detected in Nigeria in January 2016. After that, it has spread to other West, Central and East African countries.

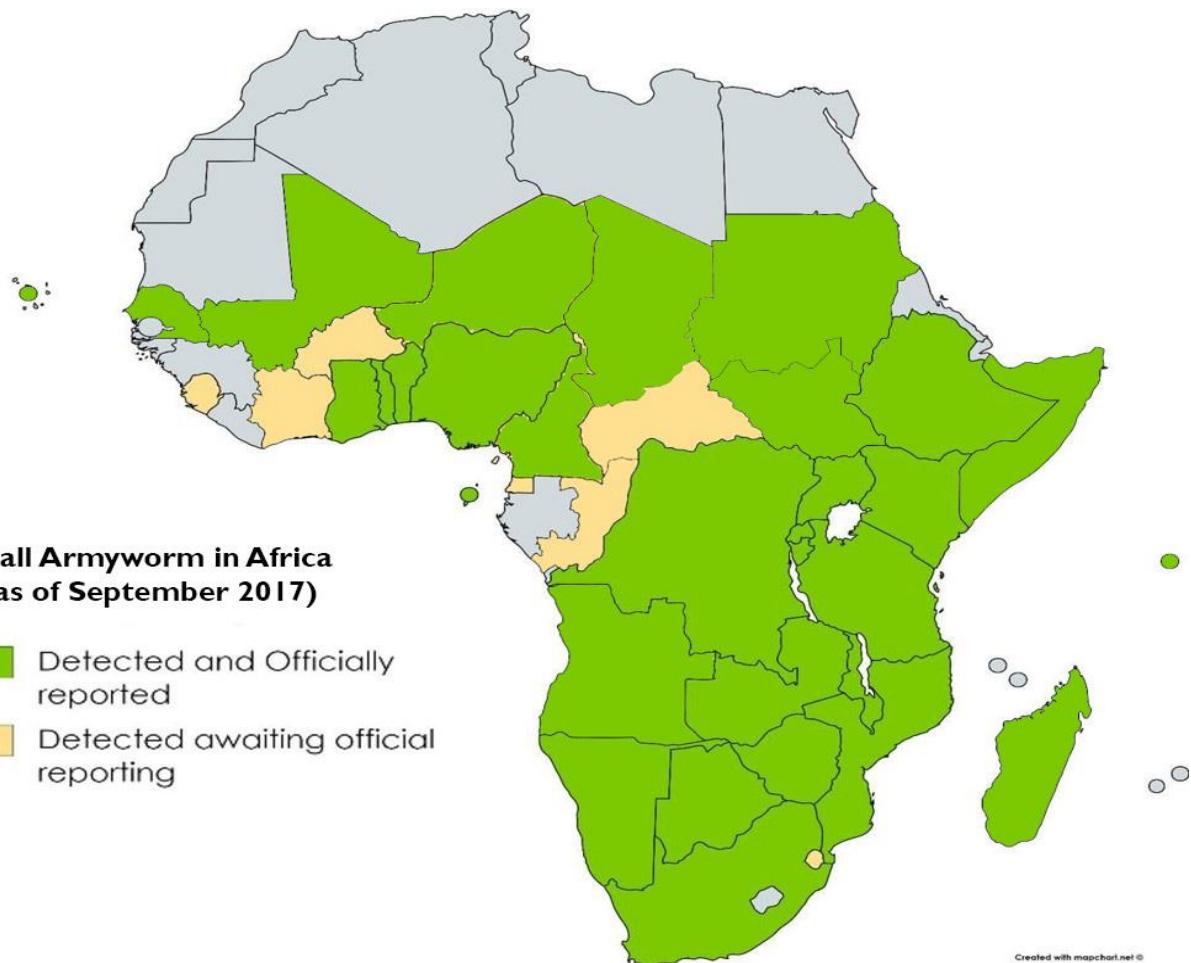


Fig. 1 FAW spread in Africa (CABI)

As of December 2017, 38 countries including Angola, Benin, Botswana, Cameroon, the Democratic Republic of Congo, Ethiopia, Ghana, Guinea, Ivory Coast, Kenya, Madagascar, Mali, Malawi, Mozambique, Namibia, Nigeria, Rwanda, Sao Tome and Principe, Senegal, South Africa, South Sudan, Sudan, Swaziland, Tanzania, Togo, Uganda, Zambia, and Zimbabwe have confirmed the presence of the FAW (Fig. 1). However, information on FAW presence or absence from the remaining countries is yet to be reported.

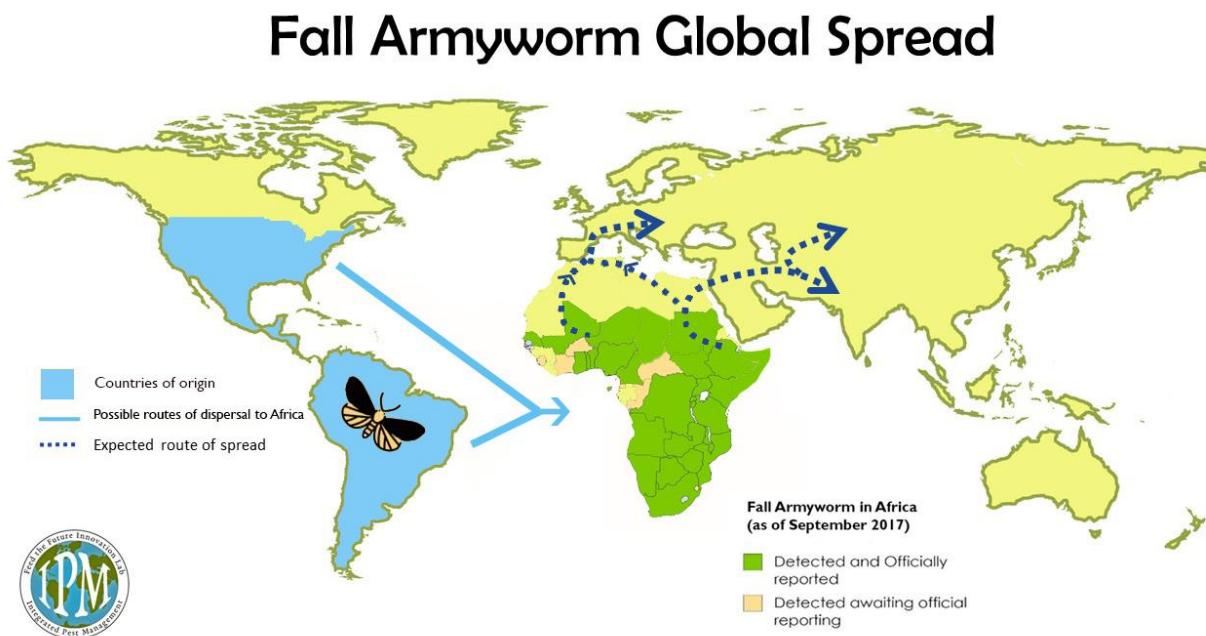


Fig. 2 Potential routes of FAW spread from Africa

S. frugiperda is a transboundary pest, therefore its appearance in some countries in Africa raises the level of threat to other regions of the continent not yet infested and other tropical and subtropical regions of the old world. The environmental suitability on the Mediterranean coast in Morocco, Algeria, Tunisia, and Libya, increases the possible spread of this insect to Southern

Europe, while climatic suitability in East Africa makes the Middle East and Asia more vulnerable to the spread of the FAW (Fig. 2).

III. HOST PLANTS

Although the FAW has a wide host range, with over 80 plant species in 27 families, it shows a preference for the Poaceae. Its preferred host plants are Bermuda grass, crabgrass, maize, millet, rice, sorghum, sugar cane, and wheat. This pest also attacks other non-graminaceous crops such as apple, cowpea, cotton, grape, groundnut, orange, papaya, peach, potato, soybean, strawberry and a number of ornamental plants. Weeds known to serve as hosts include bentgrass, Johnson grass, morning glory, nutsedge, pigweed and sand spur. It is challenging to manage the FAW due to its polyphagous behavior and ability to survive on diverse alternate hosts. Its impact on cotton, maize, sorghum, sugarcane, tomatoes, wheat, and some ornamental plants in Egypt is likely to be significant (refer to section on Economic Impact).

IV. TAXONOMY

For more than 30 years, it has been known that in the Americas, *S. frugiperda* occurs in two races: a ‘rice strain’ (R strain) and a ‘maize strain’ (C strain) (Pashley *et al.* 1985); the former is thought to preferentially feed on rice and various pasture grasses and the latter on maize (maize), cotton and sorghum. However, this may be geographically variable – for example, this is not consistent in Argentina (Juárez *et al.*, 2012). It should be noted that both strains will feed on maize. The strains are morphologically identical, but can be distinguished using DNA barcodes. The FAW strain in Togo appears to be the haplotype found in southern Florida and the

Caribbean (Nagoshi *et al.*, 2017). However, both the maize strain and the rice strain are now confirmed in Africa (Cock *et al.*, 2017). The knowledge about FAW strains is important for two reasons: 1) different haplotypes have different host ranges, 2) different biotypes carry different pesticide resistance genes.

V. IDENTIFICATION

Identification of larvae in the field requires expertise and skills as the FAW is easily confused with similar species such as the African armyworm (*Spodoptera exempta*), and the cotton leafworm (*Spodoptera littoralis*), as well as species of other noctuid genera, such as the African maize stalk borer (*Busseola fusca*). However, there are certain identification guidelines developed by taxonomists in the United States that are useful for identifying the FAW (Fig. 3).

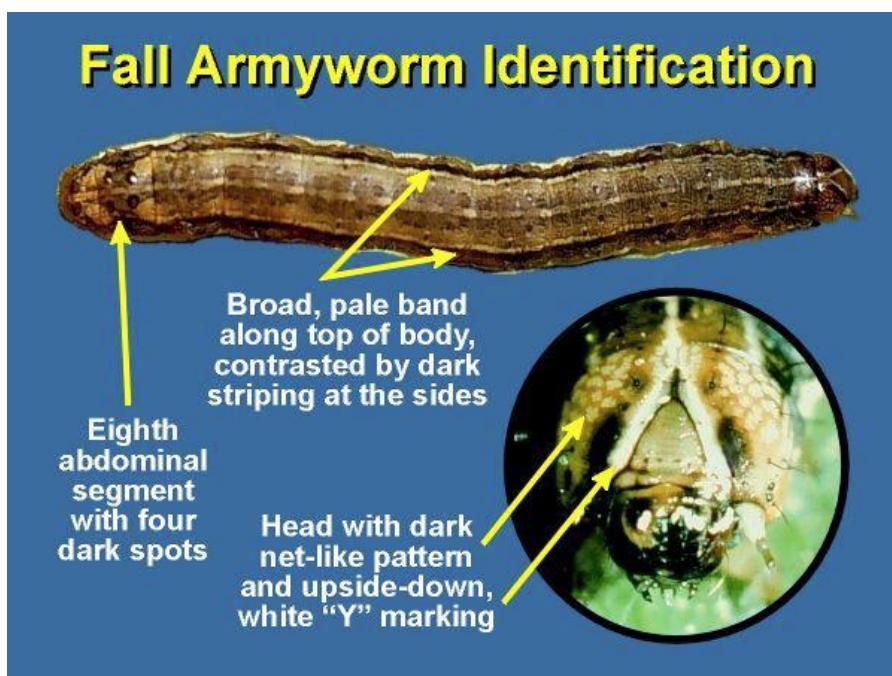


Fig. 3 FAW ID info (Farmbiz Africa / University of Nebraska)

VI. BIOLOGY

Eggs are dome shaped and are dirty-white to gray in color (Fig. 4). Eggs are laid in groups or clusters of about 10-200 eggs per egg mass, usually on the underside of leaves. Sometimes the eggs are deposited in layers but usually eggs are laid in a single layer attached to foliage. After oviposition, the female deposits a layer of grayish scales or hairs over the eggs and covers the egg mass giving it a hairy or moldy appearance (Fig. 5). Depending on environmental conditions, eggs hatch in two to five days in optimum temperatures.



Fig. 4 Newly laid FAW eggs (Bugguide.net)



Fig. 5 FAW eggs with hairs (Bugguide.net)

In its native regions, the FAW goes through six larval instars with the final instar being most devastating and consuming up to 80% of the plant material. The newly hatched larvae are greenish with a black head, which turns orange-brown in the 2nd instar. Newly hatched larvae first feed near where the egg mass was laid, then move upwards on the maize plants, and then disperse by wind using silk threads.

The larvae exhibit cannibalistic behavior, and under heavy infestations, larval densities can be reduced to one or two per plant. Cannibalism was found to account for approximately 40% mortality when maize plants were infested with more than one fourth-instar larvae over a three-day period (Chapman *et al.*, 2000).

Fully-grown larvae are 3.1 – 3.8 cm long and vary in color from pale green to almost black, with three yellowish stripes running down the back. There is a wider dark stripe and a wavy yellow-red blotched stripe on each side (Fig. 6). The FAW's head has a predominant white, inverted Y-shaped suture between the eyes (Fig. 3). In its native range, developmental times of immature stages vary with temperature (an acceptable range of between 11°C and 30°C).



Fig. 6 Fully grown FAW larva (Holly Schwarting)



Fig. 7 FAW pupa (Bugguide.net)

Pupation normally takes place in the soil, but may also occur on plant parts under high population densities. The pupa is reddish brown (Fig. 7). The pupal stage is 8-9 days in the summer and longer than 2 weeks under winter conditions. The adult moths have a wingspan of 32 to 40 mm. The male moth has dark gray and brown shaded mottled forewings with conspicuous triangular white spots at the tip and near the center of the wing (Fig. 8). These

markings are less distinct in female moths (Fig. 9). The hind wing is iridescent, silver-white, with a narrow dark border in both sexes.



Fig. 8 FAW adult male
(Lyle J. Buss, University of Florida)



Fig. 9 FAW adult female
(Lyle J. Buss, University of Florida)

Adults are nocturnal, and are most active at dusk when they mate. Like other noctuids, FAW adult females have a high fecundity rate. Females deposit most of their eggs during the first four to five days of life, but some eggs may be laid for up to three weeks. Females can mate multiple times during this period and lay multiple egg masses, with a potential fecundity of up to 1,000 eggs per female. At low population densities, females normally lay eggs on the underside of leaves. However, at high densities, oviposition is indiscriminate over the entire plant or on non-host plant objects. Adults can live up to an average of 10 days but sometimes the duration extends up to three weeks in the temperate region of the U.S. The larvae are nocturnal feeders. Unlike other armyworm species, FAW larvae are typically found damaging maize in patches throughout a field. In the northern parts of the U.S. they appear in maize fields late in the season, from mid-July through the fall harvest, but in Africa/Egypt, due to the tropical climate, it can multiply year-round.

In the summer months, the FAW completes its life cycle in about 30 days (Fig. 10); however, during winter months in the southern U.S. it takes 80 to 90 days to complete its life cycle. Development is slow in cooler climates and the number of generations in an area varies. The FAW does not have the ability to diapause and frost kills the insect. Since frost is not an issue in the African continent, this fact needs to be considered when conducting surveys and seeking better understanding of FAW biology in Africa.

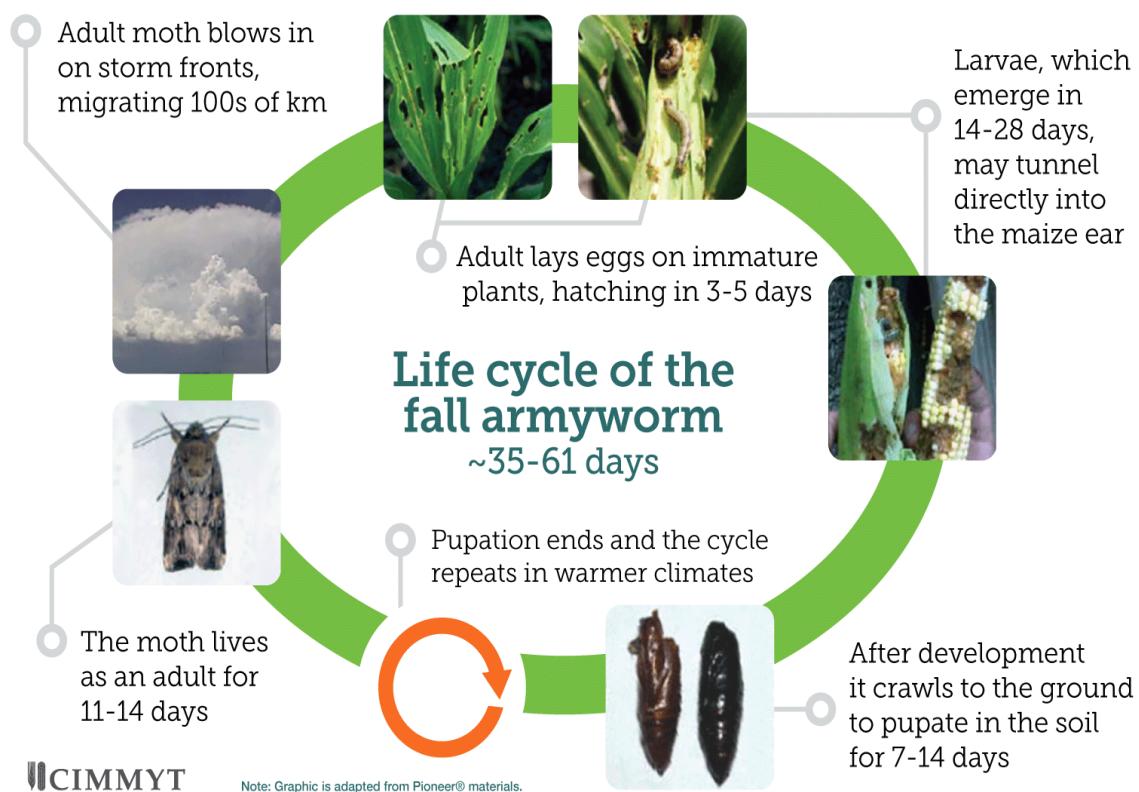


Fig. 10 Typical life cycle of the FAW

Natural Enemies

In its native region, FAW eggs, larvae, and pupae are attacked by several species of parasitoids (Table 1). Among several groups of parasitoids, the egg parasitoids *Telenomus* spp. (Hymenoptera: Platygastridae) and *Trichogramma* spp. (Hymenoptera: Trichogrammatidae), are deemed important in several countries. Egg parasitoid *Telenomus remus*, an introduced parasitoid from the Pacific, has proven very effective in South America and Florida. These egg parasitoids *Trichogramma* spp. and *Telenomus* spp. are easy to rear under laboratory conditions. However, the presence of scales/hairs over the egg masses acts as a barrier against parasitism by *Trichogramma* spp. but not to *Telenomus* spp. This physical barrier could be overcome by selecting a more aggressive species of *Trichogramma*, capable of breaking the physical barrier imposed by scales on the eggs. It is therefore essential to know the species/strains present in the agro-ecosystem when choosing the *Trichogramma* species to be used for applied biological control of the FAW. The wasp parasitoids most frequently reared from larvae in the U.S. are *Cotesia marginiventris* and *Chelonus texanus* (Hymenoptera: Braconidae). In Argentina, larval parasitoids collected were *Campoletis grioti* (Hymenoptera: Ichneumonidae), *Chelonus insularis* (Hymenoptera: Braconidae), *Archytas marmoratus* and *A. incertus*, (Diptera: Tachinidae) *Ophion* spp. (Hymenoptera: Ichneumonidae), *Euplectrus platyhypenae* (Hymenoptera: Eulophidae), and *Incamyia chilensis* (Diptera: Tachinidae). In Mexico, 13 genera of hymenopteran larval parasitoids, belonging to three families-Braconidae, Ichneumonidae and Eulophidae were recovered.

Table 1. Common parasitoids of FAW found worldwide

Parasitoid	Type	Country
<i>Telenomus</i> spp. Hymenoptera: Platygastridae	Egg	Antigua, Barbados, Brazil, Colombia, Dominican Republic, Guadeloupe, Ecuador, Guyana, Honduras, Nicaragua, Puerto Rico, Suriname, Trinidad, U.S., Venezuela, Israel, Cuba, Mexico
<i>Trichogramma</i> spp. Hymenoptera: Trichogrammatidae	Egg	Barbados, Nicaragua, Brazil, Chile, Colombia, Argentina, Cuba, U.S., Guadeloupe, Mexico
<i>Chelonus</i> spp. Hymenoptera: Braconidae	Egg, Larval	Barbados, Nicaragua, Honduras, Mexico, Trinidad, Argentina, Brazil, Chile, Colombia, Cuba, Haiti, Puerto Rico, U.S., Uruguay, Venezuela
<i>Agathis stigmatera</i> Hymenoptera: Braconidae	Larval	Argentina, Peru, U.S.
<i>Archytas</i> spp. Diptera: Tachinidae	Larval	Argentina, Barbados, Honduras, Mexico, Nicaragua, U.S., Venezuela, Brazil, Chile, Puerto Rico, Suriname, Trinidad, Uruguay, Cuba, Ecuador, Guadeloupe, Lesser Antilles, Peru
<i>Campoletis grioti</i> (Hymenoptera: Ichneumonidae)	Larval	Argentina
<i>Cotesia marginiventris</i> Hymenoptera: Braconidae	Larval	Honduras, Barbados, Nicaragua, Argentina, Brazil, Chile, Lesser Antilles, Mexico, Puerto Rico, Suriname, U.S., Uruguay, Venezuela, Trinidad & Tobago, Colombia, Guyana
<i>Euplectrus</i> spp. Hymenoptera: Eulophidae	Larval	Nicaragua, U.S., Argentina, Puerto Rico, Panama, Honduras, Barbados, Brazil, Chile, Colombia, Cuba, Guyana, Lesser Antilles, Mexico, Trinidad, Venezuela, Honduras
<i>Lespesia</i> spp. Diptera: Tachinidae	Larval	Brazil, Honduras, U.S., Argentina, Brazil, Chile, Cuba, Guadeloupe, Guatemala, Honduras, Lesser Antilles,

		Mexico, Nicaragua, Puerto Rico, Uruguay, Venezuela, Colombia
<i>Ophion</i> spp. Hymenoptera: Ichneumonidae	Larval	Argentina, Uruguay, Chile, U.S., Brazil, Honduras, Mexico, Nicargua, Peru,
<i>Brachymeria</i> spp. Hymenoptera: Chalcididae	Pupal	Argentina, U.S.
<i>Cryptus albitarsis</i> Hymenoptera: Ichneumonidae	Pupal	U.S.
<i>Diapetimorpha introit</i> Hymenoptera: Ichneumonidae	Pupal	Honduras, U.S.
<i>Ichneumon promissorius</i> , <i>I. ambulatorius</i> . Hymenoptera: Ichneumonidae	Pupal	U.S.
<i>Trichospilus pupivora</i> Hymenoptera: Eulophidae	Pupal	Barbados

Five species of Ichneumonidae: *Diapetimorpha introit*, *Cryptus albitarsis*, *Ichneumon promissorius*, *Ichneumon ambulatorius* and *Vulgichneumon brevicinctor*, two species of Chalcididae: *Brachymeria ovata* and *B. robusta* and one Eulophid species, *Trichospilus pupivora* have been reported on FAW pupae from the U.S., Argentina, and Barbados but they are of limited effectiveness. Another biological control agent, *Doru luteipes* (Dermaptera: Forficulidae) has been used as an agent for the biological control of FAW eggs in Brazil.

Although several pathogens have been shown to reduce the abundance of FAW larvae in maize, only *Bacillus thuringiensis* (Bt) is currently used, and success depends on having the product on the foliage when the larvae first appear. Bt sprays tend to be short-lived as they are very susceptible to UV degradation and require multiple sprays. Another option for biological control of FAW is *S. frugiperda* nuclear polyhedrosis virus (SFNPV). A large number of isolates

of NPV have been obtained from the field and some have been detected as promising isolates. A commercial formulation for *S. frugiperda* NPV, SPOBIOL, prepared by CORPOICA, the Colombian public-private ag research partnership, is available and has been licensed with Certis LLC, a U.S company. Some studies have also shown that *Metarhizium anisopliae* and *Beauveria bassiana* have potential as microbial control agents against FAW.

There are a number of egg and larval parasitoids found in Africa that could attack FAW eggs and larvae. There are 11 species of *Telenomus* and 26 species of *Trichogramma/Trichogrammatoidea* found in Africa (Tables 2 and 3). There are two larval parasitoids, *Habrobracon hebetor* in Niger, and *Cotesia* spp. in Kenya that can attack FAW larvae. Another parasitoid, *Bracon mellitor* was introduced into Egypt to control *Spodoptera littoralis* may also attack FAW.

Table 2: List of *Telenomus* spp. recorded in Africa

<i>Telenomus</i> spp.		Host	Distribution
<i>T. applanatus</i>		<i>Eldana saccharina</i>	Gabon, Ghana, Ivory coast
<i>T. bini</i>		<i>Maliarpha separatella, Chilo</i> spp., <i>Scirpophaga</i> spp.	Ghana, Ivory Coast, Madagascar, Malawi, Senegal, Tanzania.
<i>T. busseolae</i>		<i>Busseola fusca, Sesamia</i> spp., <i>Coneista ignefusalis</i>	Cameroon, Egypt, Ghana, Kenya, Nigeria, Reunion, Senegal, South Africa, Sudan, Uganda
<i>T. creusa</i>		<i>Chilo diffusilineus</i>	Malawi
<i>T. etielliphaga</i>		<i>Etiella zinckenella,</i>	Senegal
<i>T. nemesis</i>		<i>Chilo orichalcociliellus</i>	Ghana, Kenya, Mozambique, Senegal

<i>T. nephele</i>	<i>Scirpophaga melanoclista, S. occidentella, S. subumbrosa</i>	Cameroon, Ghana, Ivory Coast, Malawi, Mali, Senegal.
<i>T. procas</i>	<i>Antigastra catalaunalis</i>	Senegal, Sudan
<i>T. soudanensis</i>	<i>Chilo zacconius</i>	Niger
<i>T. thestor</i>	<i>Chilo orichalcociliellus</i>	Ivory Coast, Kenya, Senegal, Uganda, Zaire
<i>T. versicolor</i>	<i>Scirpophaga melanoclista</i>	Ghana, Ivory Coast, Malawi, Senegal

Table 3. Trichogrammatidae egg parasitoids recorded in Africa

Trichogrammatidae	Host	Distribution
<i>Trichogramma bourarachae</i>	<i>Helicoverpa armigera</i>	Morocco
<i>Trichogramma bournieri</i>	<i>Chilo partellus</i>	Comoros, Kenya
<i>Trichogramma cacoeciae</i>	-	Morocco
<i>Trichogramma chilonis</i>	<i>Eldana saccharina, Busseola fusca, Chilo partellus</i>	South Africa
<i>Trichogramma ethiopicum</i>	-	Cameroon
<i>Trichogramma evanescens</i>	<i>Chilo Agamemnon, Helicoverpa armigera, Pectinophora gossypiella, Spodoptera littoralis</i>	Egypt, Madagascar
<i>Trichogramma japonicum</i>	<i>Chilo partellus</i>	Malawi
<i>Trichogramma kalkae</i>	<i>Diopsis macrophthalma</i>	Malawi
<i>Trichogramma</i> sp. nr <i>kalkae</i>	-	Zimbabwe
<i>Trichogramma kayo</i>	-	Sudan
<i>Trichogramma mandelai</i>	<i>Diparopsis watersi</i>	Chad
<i>Trichogramma</i> sp. nr <i>mwanzai</i>	<i>Chilo diffusilineus, Chilo partellus, Busseola fusca, Eldana saccharina, Sitotroga cerealella</i>	Malawi, Kenya

<i>Trichogramma ostriniae</i>	<i>Busseola fusca, Chilo partellus</i>	South Africa
<i>Trichogramma papilionidis</i>		Angola
<i>Trichogramma pretiosum</i>	<i>Apple leaf roller</i>	South Africa
<i>Trichogramma pinneyi</i>	<i>Diopsis macroura</i>	Malawi
<i>Trichogramma</i> spp. nr <i>exiguum</i>	<i>Chilo partellus</i>	Kenya
<i>Trichogramma voegeli</i>	-	Morocco
<i>Trichogrammatoides armigera</i>	<i>H. armigera,</i> <i>Heliocheilus albipunctella</i>	Kenya, Niger
<i>Trichogrammatoides bactrae</i>	<i>P. gossypiella</i>	Egypt
<i>Trichogrammatoides citri</i>	-	Madagascar
<i>Trichogrammatoides combreti</i>	-	Senegal
<i>Trichogrammatoides cryptophlebia</i>	<i>Cryptophlebia batrochopa,</i> <i>C. leucotreta</i>	Malawi, South Africa
<i>Trichogrammatoides eldanae</i>	<i>E. saccharina, Sesamia calamistis</i>	South Africa, Nigeria, Kenya
<i>Trichogrammatoides lutea</i>	<i>C. partellus, B. fusca,</i> <i>H. armigera</i>	South Africa, Kenya, Ivory Coast, Ethiopia, Mali, Mozambique, Senegal
<i>Trichogrammatoides simmondsi</i>	<i>Diopsis macroura,</i> <i>C. partellus,</i> <i>Thaumatotibia leucotreta,</i> <i>H. armigera, Atherigona soccata</i>	Malawi, South Africa, Kenya, Burkina Faso

VII. DAMAGE

The developing larvae feed on different parts of the host plant, depending on the crop, the stage of crop development, and the age of the larvae. The FAW generally feeds on foliage, but during heavy infestations, larvae also feed on maize ears. Young larvae initially feed near where the egg mass was laid and superficially feed on one side of the



Fig. 11a Young FAW larvae feeding on a maize leaf. (D Visser ARC-VOP)



Fig. 11b Young FAW larvae dispersing by using silk threads (D Visser ARC-VOP)

leaves leaving the epidermis intact on other side (Fig. 11a). Then the larvae disperse using silk threads blown by wind, a phenomenon known as ballooning (Fig. 11b).



Fig. 12 Ragged appearance of leaves due to FAW larval feeding (CABI)

Foliar damage to maize is usually characterized by ragged feeding (Fig. 12), and moist sawdust-like frass near the whorl and upper leaves of the plant (Fig. 13).



Fig. 13 Sawdust-like frass near the whorl due to FAW feeding (CABI)

Later instars feed by making holes in leaves and eat from the edge of the leaves inward. Feeding in the whorl of maize often produces a characteristic row of perforations in the leaves. Due to young larval dispersal and the cannibalistic behavior by late instars, larval numbers are reduced to few larvae per plant. Fully-grown larvae cause

extensive defoliation, often leaving only the ribs and stalks of maize plants. Larvae can also burrow into the growing point and affect the growth of plants. In maize, larvae sometimes also bore into the ear through the husk and feed on the tip of ears (Fig. 14) and on kernels (Fig. 15). When boring through the husk they produce holes (Fig. 16).



**Fig. 14 FAW larva feeding at tip of a maize ear
(P. Chinwada)**



**Fig. 15 FAW larvae feeding on a maize ear.
(CABI)**



**Fig. 16 FAW feeding hole in maize ear
(P. Chinwada)**

VIII. HOW TO IDENTIFY AND DIFFERENTIATE THE FAW FROM OTHER SIMILAR SPECIES IN EGYPT

Spodoptera littoralis - Eggs are translucent with few hairs or scales (Fig. 17).



Fig. 17 *Spodoptera littoralis* eggs. (EPPO.int)

The larvae are usually brown colored with distinct black spots. Sometimes the larvae may be yellowish or blackish with light spots. Caterpillars have dark and light longitudinal bands and two dark, semi-



Fig. 18 *Spodoptera littoralis* larva (Pyrgus.de)

circular spots laterally on each segment, except for the prothorax (Fig. 18). Moths are

grey-brown and have a characteristic “scratch like” pattern on forewings. The tip of the forewing is light brown, with a distinct white marking shaped like an “A” and a white, three-branched, fork-like pattern. Hind wings are whitish with grayish-brown margins and veins as well as fringe hairs (Fig. 19).

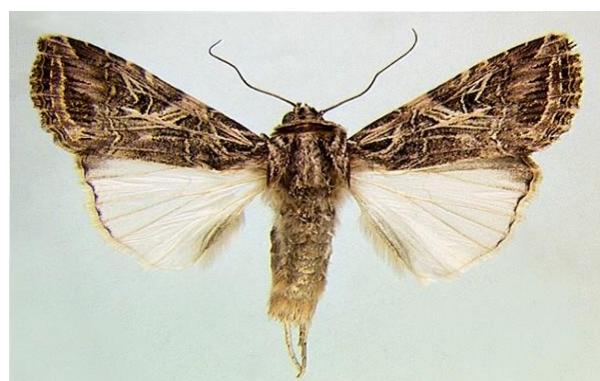


Fig. 19 *Spodoptera littoralis* adult (CABI)

Spodoptera exempta- Young larvae are light colored, while the older ones are usually blackish in color. They are velvety-black on the upper body surface with green, black, yellow, and white lateral stripes. The underside of the body is green or yellow and the larvae do not have hairs on the body (Fig. 20).



Fig. 20 *Spodoptera exempta* larvae (CABI)

The adult moths are similar in appearance to the FAW and lay eggs in groups or layers covered with hairs.

Spodoptera exigua: Larvae are pale green or yellow in color when young (Fig. 21). Older larvae are darker in color and develop lateral stripes and sometimes dots (Fig. 22). Sometimes a characteristic pink line or spots are seen on the sides of larvae. The larvae

are smooth without any hairs. The adults have a mottled grey and brown forewings



Fig. 21 Young *Spodoptera exigua* larva (John Capinera)



Fig. 22 Mature *Spodoptera exigua* larva (Pyrgus.de)

with an irregular banding pattern and a characteristic light colored bean shaped spot (Fig. 23). Eggs are laid in groups covered with hairs or scales.



Fig. 23 *Spodoptera exigua* moth (John Capinera)

IX. MOBILITY AND DISPERSAL

Noctuids are generally considered strong fliers and are assumed to migrate at night and downwind. Fall armyworm adults are nocturnal and their early evening movement near fields is generally with the wind. There are records of 16-30 hour tethered flight by FAW males (van Handel, 1974).

In Central America, FAW moths generally disperse about 500 km before oviposition, from seasonally dry habitats to wet habitats (Johnson ,1987). Moths fly downwind above the boundary layer (the lowest part of the atmosphere, above which the wind direction and strength may be different), so the direction of movement depends largely on prevailing winds. The data indicates that FAW follow this pattern and can move variable distances on weather fronts (Rose *et al.*, 1975, Young, 1979). There is one documented incidence of long-distance migration by FAW on a weather front where FAW travelled 1,600 km from Mississippi to southern Canada in 30 hours (Rose *et al.*, 1975). Therefore, FAW has the potential to spread rapidly and has already spread to the western, eastern and southern regions of Africa in a span of around 18 months since its discovery in the western region.

X. SPREAD AND ESTABLISHMENT

Using multiple modeling methods, and data sources, we identified different routes and pathways of possible FAW introduction to Egypt, its spread within the country, and the threat this scenario poses to other countries. To this end, we have accounted for ecological factors, spatiotemporal

variation in vegetation, and production of major hosts of FAW, its flying capacity, wind patterns, and international trade and travel. To study its spatiotemporal spread, we developed a cellular automata model (CA) based on a recent work by Guimapi and others (Guimapi *et al.*, 2016). To study the role of wind, we used the TAPPAS (Tool for Accessing Pest and Pathogen Ariel Spread) (Durr *et al.*, 2015) interactive framework for modeling pest and pathogen spread through wind. This work derives from a recent study of migration patterns of FAW by Westbrook and others (Westbrook *et al.*, 2016). To analyze international trade and travel, we used datasets from Food and Agricultural Organization (FAO) and WorldPop (<http://www.worldpop.org.uk/>).

Establishment potential

To assess the suitability of FAW to establish in different parts of Egypt, work by Abrahams *et al.* (2017) was adopted. They used seven species distribution models to assess the environmental suitability of Africa for the establishment of FAW (Fig. 24a). The models account for climatic factors and the biological properties of the pest. Overall, their results indicate that most southern parts of Egypt are less suitable for FAW, while the central and northern parts are moderately suitable.

However, there seems to be high variation in the predictions across the models for this African region (see Fig. 24(b) inset) (Abrahams *et al.*, 2017). Also, to the best of our knowledge, their approach does not seem to account for crop production or man-made diversity. For example, several hosts of FAW are grown along the Nile River (suitable for establishment), even though, in general, the areas some distance from the Nile may not be suitable for establishment.

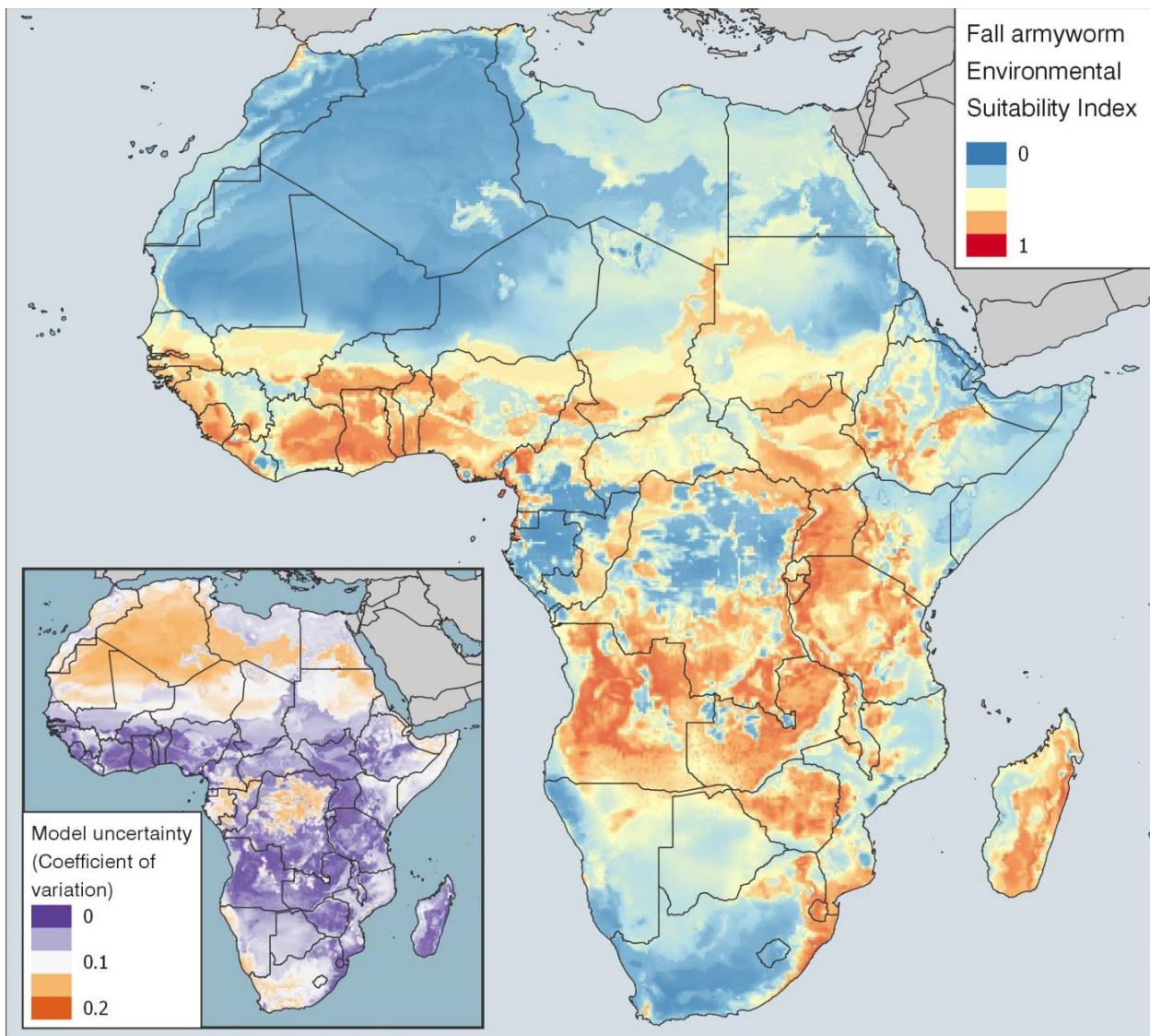


Fig. 24 Environmental suitability for the establishment of FAW in Africa including Egypt and surrounding countries (a) and 24 (b) (inset) Model uncertainty Abrahams *et al.* (2017). In fig. 24a dark blue is least suitable and dark red, the most suitable to the establishment of the FAW. Notice that Abraham's model does not take into account the crop areas suitable for FAW movement north along the Nile River running from Sudan to the Nile Delta.

Pathways of introduction and spread

There are three potential pathways of introduction into Egypt that we accounted for: 1) natural spread; 2) travel; and 3) trade. We first predicted the natural spread (unaided dispersal) from

Sudan and Ethiopia to Egypt based on the flight biology of the insect in North America and wind directions from the sources in Sudan and Ethiopia. This exercise was followed by modeling the flight of the FAW from Ethiopia and Sudan to Egypt.

Natural spread (unaided dispersal) based on flight biology

The FAW has spread onward to southern and eastern parts of Africa since its introduction in West Africa in 2016. There is no documented evidence on the possible methods or pathways of its spread within Africa. It seems likely that of the three pathways; 1) unaided dispersal through flight, 2) as a stowaway in aircraft or other transportation or 3) through trade. Unaided dispersal through insect flight may be the most probable means of introduction into Egypt. In order to assess the risks of the FAW spread into Egypt, it is appropriate to assess the potential pathways of entry, especially when the pest is already in Sudan and other countries in East Africa.

The following calculations were employed to determine the rate of migration (flight) movement from a source point (nearest source locations to Egypt) to a specific location in Egypt and the rate of FAW movement within Egypt once it arrives.

A= Distance from a source to a potential location in Egypt (miles)

B= Migration potential [X miles/generation (30 days)]

C= Number of months to reach a specific destination in Egypt

C= A/B

Note (**A**) In the US, movement from S. Texas to Canadian border (1,740 mi.) occurs in 105 days or 3.5 months (**C**)

where $B = 1740 / 105 = 16.5$ mi./day = 497 mi./mo. or one insect generation $(1740/497) = 3.5$ months. or 3.5 insect generations

A conservative estimate for the US would be 250 miles /generation /month. (Johnson 1987). In North America, moths fly downwind above the boundary layer, so the direction of movement depends largely on prevailing winds. When the wind pattern is right, moths can move much larger distances: for example, 1,600 km from Mississippi to southern Canada in 30 hours has been recorded (Rose *et al.*, 1975). The FAW clearly, has the potential to spread rapidly from Ethiopia and Sudan to Egypt, if the prevailing winds are in a northerly direction and of sufficient speed. There is no published research however, as to the factors affecting FAW movement in Africa. Indeed, the FAW may already be near Egypt or approaching Egypt soon.

Based on wind movements, the migration potential [(X miles/generation (30 days)] from Sudan is expected to be lower than that in the US. Because the FAW is expected to move from Lake Tana, Ethiopia to Khartoum via the Blue Nile, and is already in Khartoum, we only calculated the migration potential from Khartoum to Lake Nasser and from Lake Nasser to the Nile Delta.

Calculations:

1. Khartoum, Sudan to Lake Nasser in Egypt

$A = \text{Distance from source (Khartoum, Sudan) to Lake Nasser in Egypt} = 700 \text{ miles}$

$B = \text{Migration potential: Miles/generation-Conservative} = 150/\text{Liberal} = 300$

$\text{Conservative estimate} = C = A/B = 700/150 = \underline{\textbf{4.7 months}}$ or 4.7 generations

$\text{Liberal estimate} = C = A/B = 700/300 = \underline{\textbf{2.3 months}}$ or 2.3 generations

2. Lake Nasser (Aswan) in Egypt to the Nile Delta near Cairo

A= Distance from source (Lake Nasser in Egypt) to Nile River Delta (near Cairo)= 500 miles

B= Migration potential: Miles/generation-Conservative= 150/ Liberal=300

Conservative estimate= C= A/B= 500/150= **3.3 months** or 3.3 generations

Liberal estimate= C= A/B=500/300= **1.7 months** or 1.7 generations

Sudan is one potential source of FAW to invade Egypt from the south. It would follow the watershed of the Nile River as suitable crops for feeding and reproduction of the FAW are only available along the watershed (Fig. 25). Beyond the irrigated areas along the Nile there is no availability of crops for the FAW to feed on.



Fig. 25 Possible routes of introduction and spread of FAW in Egypt based on model output and suitability.

The movement from Sudan is expected to follow the watershed of the White Nile River as suitable crops are only grown along the watershed. Other areas do not have crops on which the FAW can feed as it moves north. The FAW feeds on 80 different host plants and the most suitable host plants along the Nile are cotton, cowpeas, groundnut, maize, millet, rice, sorghum, sugarcane, tomatoes, and wheat. At the upper portion of the White Nile, near Khartoum, where the White Nile and the Blue Nile merge, the Gezieira Scheme has extensive areas of cotton, maize, sorghum and sugarcane, the preferred hosts. From Khartoum, the FAW would be expected to follow the Nile moving north, as the watershed has extensive irrigated areas providing the required host plants for the FAW. The FAW would be expected to enter Egypt just south of Lake Nasser. The wind direction in Sudan is north, northwest or northeast 89 % of the time on an annual basis (N-32%, NW- 44% and NE- 13%). The distance from Sudan to Lake Nasser is approximately 700 miles.

The distance from Lake Nasser to the Nile River Delta at Cairo is approximately 500 miles. A conservative estimate is that it will take 3.3 months (500 miles ÷ 150 miles per 30 days) and a liberal estimate is 1.7 months (500 mi. ÷ 300 mi. per 30 days). There is also a slight possibility that the FAW from Ethiopia could enter Egypt via Yemen and Saudi Arabia to the east side of the delta in Egypt (Fig. 26).

Modeling the natural spread based on ecological suitability, hosts etc.

To assess the threat of FAW through natural spread, we developed a cellular automata model accounting for vegetation and availability of host crops. The output of the model is the spatiotemporal spread of the FAW. The model parameters include vegetation and production thresholds that determine establishment potential, the pest's flying capacity and time to complete

a life cycle. One of the instances of the model output (which was a close fit to pest reports) is shown in Fig. 26.

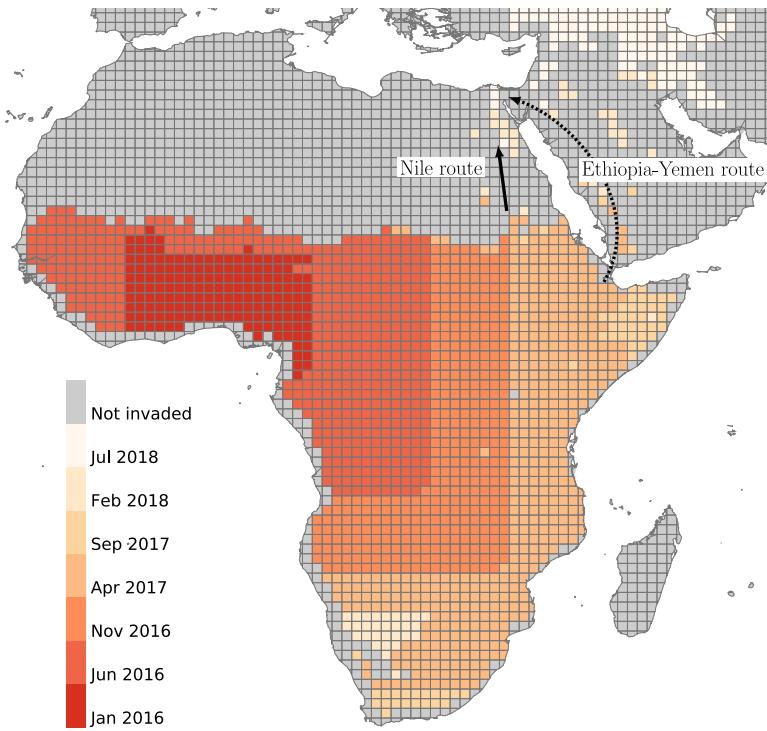


Fig. 26 Spatiotemporal spread of the FAW as predicted by the cellular automata model for 36 months starting January 2016.

The country closest to Egypt with reports of the FAW is Sudan (Muniappan, personal communication). The movement from Sudan would be expected to follow the watershed of the White Nile River. This is because suitable crops are grown only along the watershed. Also, areas further from it have very low vegetation in general. At the upper portion of the White Nile, near Khartoum, where the White Nile and the Blue Nile, coming from Ethiopia merge, the Gezieira Scheme has extensive areas of cotton, maize, sugarcane and sorghum as preferred hosts. Fig. 26 illustrates this route. According to our models, it will take between 4-8 months for FAW to reach Lake Nasser through this route.

The movement from Ethiopia can happen in two ways (Fig. 25). The most probable one would start near Lake Tana and follow the Blue Nile watershed to Kartoum where it merges with the White Nile and then move north up the Nile watershed to Lake Nasser, similar to the Sudan population (Fig. 26). There are suitable food crops for the FAW along the Blue Nile. The other route is along the coastal regions of Yemen and Saudi Arabia. The latter would take longer (greater than 9 months) and directly affect the Nile Delta.

Our models also indicate the possibility of the FAW spreading to Southeastern Europe and Western Asia within 3-4 months after establishing in the Nile Delta region. This is under the assumption that these regions are not already invaded from other sources or routes.

Wind patterns and its effect on the spread of FAW

As mentioned earlier, the FAW has been known to cover thousands of miles aided by wind. Recently, Westbrook *et al.* (2016) studied the wind aided migratory flight of FAW in the continental USA using the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Draxler and Hess, 1997). Our study is based on this work. For simulations, we used the TAPPAS online software tool (Durr *et al.*, 2015) which is developed to simulate the long-distance spread of pests and pathogens. It uses the HYSPLIT framework in the backend.

We studied monthly wind patterns with particular emphasis on wind directions from pest reported areas of Sudan and Ethiopia. In the simulations, the insect population was initiated at two places: one in the Khartoum area of Sudan and the other near Addis Ababa in Ethiopia. Each insect is modeled as a particle. The properties of the particles were set based on the work of Westbrook *et al.* (2016). These include the release altitude which is between 500m-AGL (Above

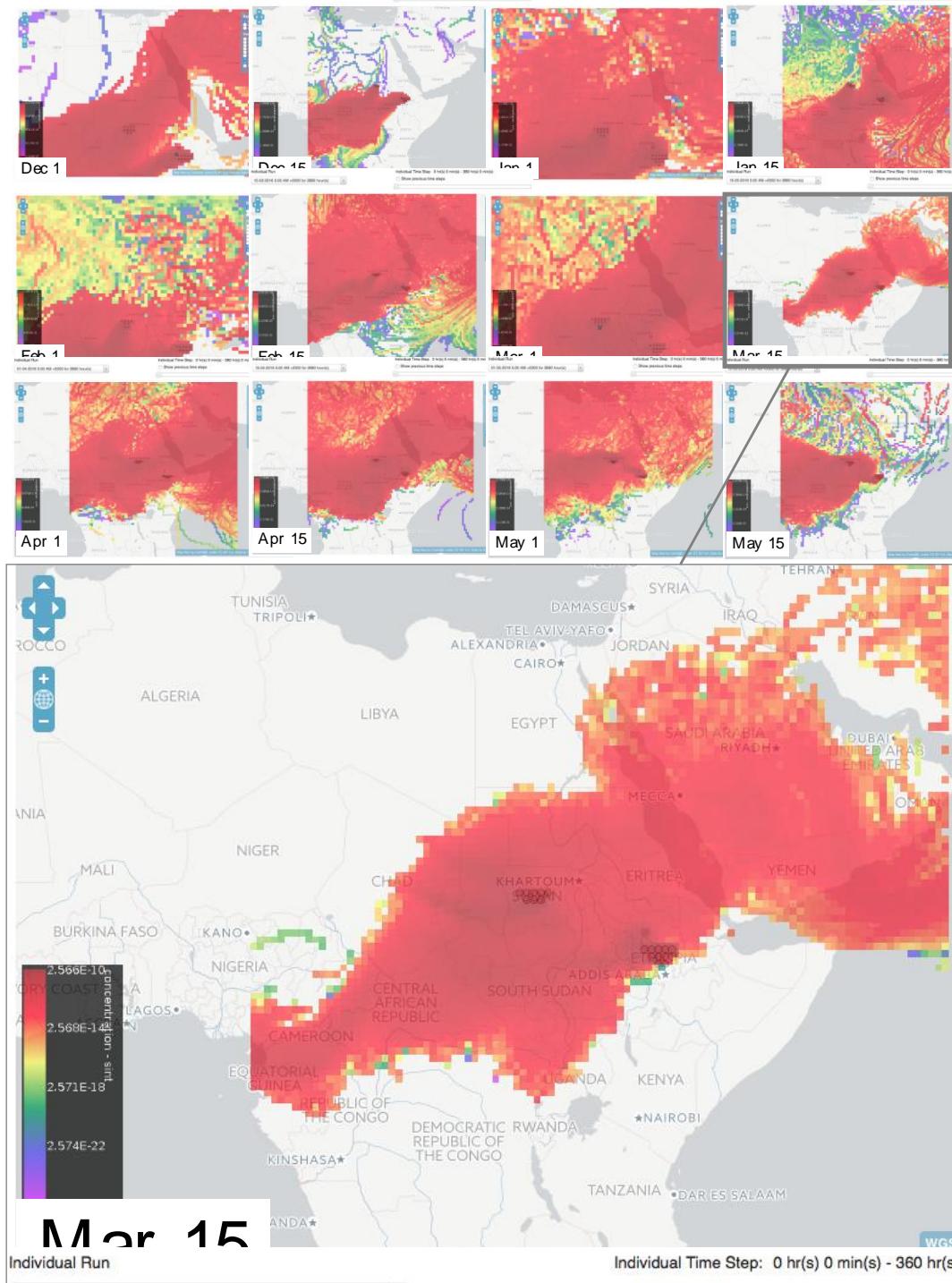


Figure 27. Wind patterns studied using the TAPPAS simulation tool: The results of 12 simulations are presented. Each simulation is identical except for the start date which varies from December 1 to May 15 in 15-day interval. For clarity, one of the results is expanded. The concentration of insect population is color coded (from red, yellow, blue to violet in descending order of concentration). This is the concentration 120 days after the start date.

Ground Level) to 3000m-AGL, diameter, release quantity and release time. Each simulation was run for a duration of 120 days with different start dates accounting for multi-generational migration.

Here, we would like to note that in addition to biology and physical properties of the pest, Westbrook *et al.* (2016) incorporate many other details including but not limited to production cycles of corn in the continental USA, time to complete life cycle in terms of degree days, presence of water bodies, etc. Also, in their framework, after every 12-hour flight the moths were run through a biological model tied to corn growth. However, to mimic such a detailed spread model, we do not have required data for Africa. Hence, in this work, we focus only on the effect of wind on particles which are endowed with the properties of the insects.

The results are shown in Fig. 27. We note a general trend of Northeasterly wind flow from Sudan and Ethiopia towards Southern and Central Egypt, Yemen and Saudi Arabia. This observation strongly suggests that wind can aid in the introduction and spread of FAW in Egypt. In particular, it can help the pest in two ways: 1) enable it to cross the low vegetation areas between Northern Sudan and Southern parts of Egypt and, 2) speed up the spread process.

Travel

From a travel perspective, Egypt's air travel passenger volume, when restricted to Africa, is dominated by domestic flow, followed by other countries from Northern Africa (Fig. 28a). Among the countries with highest passenger inflow, Nigeria (13%), Sudan (10%), and South Africa (8%) report the presence of FAW. It accounts for 20% of the inflow among countries in this region, excluding Egypt (approximately 900,000 passengers per year). Cairo accounts for the

highest inflow of passengers from the rest of Africa (around 2,000,000 or 75%) followed by Alexandria (12%) and Hurghada (6%). However, for a more accurate assessment, it is important to consider the type of travelers. If most travelers are tourists visiting Egypt, it is possible that the risk of introduction through baggage is much less. However, if the introduction is through flights, then these passenger flow volumes are a good indication of relative threat. Most of the international air travel from Egypt is to Asia (40%), Europe (31%) and Africa (28%) (Fig. 28 b).

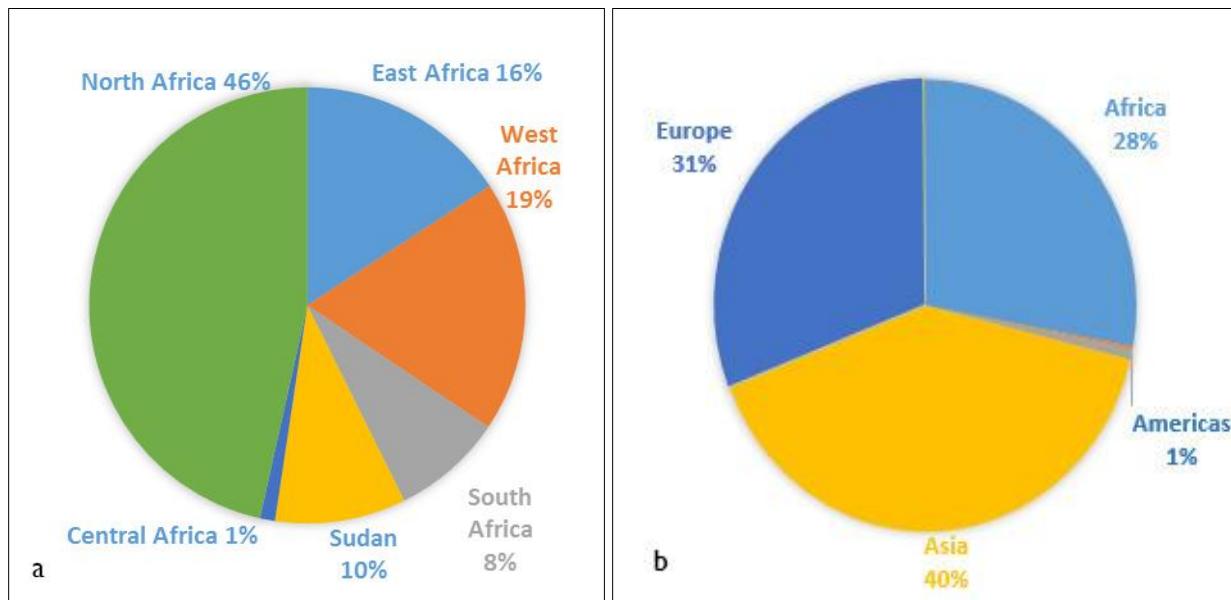


Fig. 28. International air travel: Volume of passengers travelling annually (a) to Egypt from different parts of Africa and (b) from Egypt to different regions of the world. (Most travelers from the Americas travel through Europe or the Middle East, therefore the America account for low percentage)

Trade

Among the identified host crops of FAW, Egypt imports less than 1% from Africa. Though more than 30% of its imports are from the Americas, historically, it does not seem to be a threat. Hence, trade does not appear to be an important pathway.

XI. RISK TO OTHER COUNTRIES

We assessed the possible risk of FAW spread from Egypt to other parts of the world that are currently free from FAW. Here again, we considered the three possible pathways: natural spread, trade, and travel.

Natural spread

Our models also indicate the possibility of the FAW spreading from Egypt to Western Asia within 3-4 months after establishing in the Nile Delta region. This prediction is based on the assumption that these regions have not already been invaded as of this writing.

Travel

Based on the volume of passenger flow from Egypt, several countries in Western Asia, Western and Europe and Northern Africa are at risk from air travel (Fig. 27 b). The top countries are Saudi Arabia (11%), United Arab Emirates (9%), Germany (7%), United Kingdom (6%), and Italy (5%). Again, as discussed previously, the risk would depend on the type of traveler and whether the mode of invasion is baggage or as a stowaway on a flight.

Trade

The region under major threat of FAW due to imports from Egypt is Western Asia, in particular the Middle-Eastern countries and Russia. Several countries import from Egypt. Among these countries, Saudi Arabia (21%) and Russia (18%) are the top importers. It is of interest to note here that FAW eggs have been detected in quarantine, in the Netherlands, on cut roses and vegetables shipped by plane from Kenya, and Zambia (EPPO, 2017).

XII. ECONOMIC IMPACT OF FAW

To compute the impact, we will explore two different measures: 1) the direct impact and, 2) the total impact, following the methodology in Soliman *et al.*, 2012. The direct impact measures the

direct revenue loss from the invasion of FAW on each of the crops. This depends on the loss encountered by each crop, the total cultivated area, yield per unit of land, and the price of the crop. The cultivated area multiplied by the yield equals total production. The direct impact, however, does not account for the change in the market price of the crops due to crop loss (hence drop in market supply) or the impact of price change on consumers' and producers' welfare.

To calculate a more comprehensive economic impact, we use the partial equilibrium approach. This approach assumes that the price for substitute and complementary goods remain unchanged. The partial equilibrium method accounts for the shift in the supply curve and the resulting change in market clearing price. To calculate the new equilibrium price, data is needed on demand and supply elasticities for each crop. Once the new price is determined, changes in consumer and producer surplus can be calculated.

Table 4 provides the input data used in the calculation of the economic impact for each crop. Table 5 shows the direct and total economic impact of FAW on each of the eight crops. The economic impact is measured in terms of change in social welfare from before to after the pest invasion. The change in social welfare is measured by the sum of change in consumers' surplus and change in producers' surplus. In order to calculate the change in surplus, we first find the new equilibrium price. Table 5 shows the new equilibrium price for all crops, which is higher than the original price. The consumers' surplus drops for all the crops. The producers' surplus (profits) increases for all crops except for sorghum and soybean.

The reason producers' surplus drops for sorghum and soybean is that the increase in revenue due to higher price is more than compensated by the decrease in revenue due to reduced demand, resulting in net drop in revenue and hence profits. The change in profits depends upon

the interplay between supply and demand elasticities, the change in price, loss due to invasion and the amount sold. Note that even though the profits are higher for all other crops, the total social welfare still drops because it is the sum of consumers' and producers' surplus, and the drop in consumers' surplus is higher in magnitude than the increase in producers' surplus. This model does not consider the impact of exports and imports. Only the domestic demand and supply of each crop is considered.

Direct economic loss from a crop = production*proportion lost due to FAW

Total direct economic loss across all 8 crops = \$2.68 billion/year (first year of infestation)

Total (direct plus indirect) economic loss= \$37.5 billion/year (first year of infestation)

Table 4. Input data used in the calculation of economic impact.

Crop	Proportion lost due to FAW	Production (in tons)	Original Price (USD/ton)	Demand Elasticity	Supply Elasticity
Wheat	0.2	9,279,804	387	0.47	0.38
Maize	0.2	8,059,906	321	0.24	0.57
Rice	0.3	5,467,392	301	0.66	0.21
Sorghum	0.075	804,051	308	0.44	0.5
Sugarcane	0.4	16,055,013	55	0.57	0.09
Cotton	0.3	252,504	1028	0.59	0.67
Soybean	0.1	39,872	565	0.44	0.5
Tomato	0.3	8,288,043	200	0.12	0.5

Data sources:

Production, original price: "Market Information" file.

Proportional loss: These figures are estimates. Same as given in "FAW risk assessment" file.

Demand and supply elasticity: <https://www.ers.usda.gov/data-products/commodity-and-food-elasticities/> ; http://ageconsearch.umn.edu/bitstream/59510/2/10-WP_506.pdf

Table 5. Direct and total economic impact of FAW on each of the eight crops.

Crop	New Equilibrium Price (USD)	Direct Loss (millions of USD)	Total Loss (millions of USD)
Wheat	503	718.26	6,769.60
Maize	423	517.45	14,411.04
Rice	454	493.71	1,298.31
Sorghum	335	18.57	670.54
Sugarcane	119	353.21	502.59
Cotton	1,364	77.87	5,592.15
Soybean	632	2.25	85.35
Tomato	356	497.28	8,177.43
Total Loss		2,678.60	37,507.02

XIII. DEVELOPMENT OF A MANAGEMENT PLAN FOR THE FAW IN EGYPT

The invasion of Egypt by the FAW is inevitable. It may only be a matter of a few months as it has already invaded Sudan, a country south of Egypt. It is expected that the FAW will reach Aswan in the south and then proceed to spread northwards to the Nile Delta. There is also a slight possibility that it could reach the eastern part of the Delta through Yemen and Saudi Arabia east of the Red Sea.

In the development of a management plan, we recommend organization of a network of administrators, scientists, NGOs, extension personnel, and farmers to develop and communicate FAW management strategies. This should be coordinated by the value chain project (ACDI VOCA). To prepare government officials and farmers to combat this pest invasion, it is suggested that awareness workshops be conducted at Aswan, Luxor, and Alexandria and in the eastern part of Delta (Port Said or Mansour) as soon as possible. Workshop participants should

include officials from the Ministry of Agriculture, scientists from the Agricultural Research Center, and universities and members of farmers associations.

Outline for an awareness and management workshops in Egypt (Before and after FAW arrival)

1. Taxonomy of *Spodoptera* species
2. FAW biology, distribution in the new world and Africa
3. FAW strains and identification
4. FAW host plants
5. FAW monitoring (see Appendix 2)
6. FAW field observations
7. Phytosanitary and sanitary measures to mitigate the risk
8. Control measures (See appendix 3)
 - Cultural control
 - Mechanical control:
 - Host plant resistance
 - Biological control
 - Botanical pesticides
 - Microbial pesticides
 - Chemical insecticides
9. PERSUAP (Pesticide evaluation report and safer use action plan)

XIV. PREPAREDNESS STEPS

- a) Organize a network of administrators, scientists, NGOs, extension personnel and farmers to develop and communicate FAW management strategies.
- b) Organize awareness and management training workshops at Aswan, Luxor, Alexandria and Port Said or Mansoura.

- c) Farmer training.
- d) Strengthen existing biocontrol laboratories in the production and distribution of natural enemies.
- e) Provide FAW alerts as it spreads through Egypt.
- f) Prepare bulletins and other print media and distribute.
- g) Utilize mass media for dissemination of information on management of FAW to the public.
- h) Identify safe pesticides to be used in the IPM program (PERSUAP).

XV. REFERENCES

1. Abrahams, P., Bateman, M., Beale, T., Clottey, V., Cock, M., Colmenarez, Y., Corniani, N., Day, R., Early, R., Godwin, J., Gomez, J., Moreno, P. G., Murphy, S. T., Oppong-Mensah, B., Phiri, N., Pratt, C., Richards, G., Silvestri, S. and Witt, A. (2017). Fall Armyworm: Impacts and Implications for Africa. *Outlooks for Pest Management*, doi: 10.1564/v28_oct_02
2. Carrasco, L.R., Mumford, J.D., Harwood, T., Macleod, A., Grabenweger, G., Leach, A.W., Knight, J.D., Baker, R.H.A. (2010). Unveiling human-assisted dispersal mechanisms in invasive insects: integration of spatial stochastic simulations and phenology models. *Ecological Modeling* 221: 2068-2075
3. Chapman, J. W., Williams, T., Martínez, A. M., Cisneros, R., Caballero, P., Cave, R. D., and Goulson, D. (2000). Does cannibalism in *Spodoptera frugiperda* reduce the risk of predation? *Behavioral Ecology and Sociobiology*. 48, 321–327.
4. Cock, M.J.W., Beseh, P.K., Buddie, A.G., Cafá, G. and Crozier, J. (2017) Molecular methods to detect *Spodoptera frugiperda* in Ghana, and implications for monitoring the spread of invasive species in developing countries. *Scientific Reports* 7(4103), 10 pp. doi:10.1038/s41598-017-04238-y
5. Cook D. C., Fraser, R. W., Paini, D. R., Warden, A. C., Lonsdale, W. M., Barro, P. J. D. (2011). Biosecurity and yield improvement technologies are strategic complements in the fight against food insecurity. *PLoS One* 6(10):e26084
6. Draxler, Roland R., and G. D. Hess. "Description of the HYSPLIT4 modeling system." (1997).
7. Durr. P., Graham, K., Freeman, J., Backett, D., and van Klinkan, R.D. (2015). TAPPAS: Tools for Accessing Pest and Pathogen Aerial Spread. Version 1.0
8. Early, R., Bradley, B.A., Dukes, J.S., Lawler, J.J., Olden, J.D., Blumenthal, D.M., Gonzalez, P., Grosholz, E.D., Ibanez, I., Miller, L.P., Sorte, C.J.B., Tatem, A.J. (2016). Global threats from invasive alien species in the twenty-first century and national response capacities. *Nature Communications* 7: 12485. doi:10.1038/ncomms12485.
9. FAOSTAT, “Trade matrix”, Accessed in October 2017,

<http://www.fao.org/faostat/en/#data>.

10. Guimapi, R.Y.A., Mohamed, S.A., Okeyo, G.O., Ndjomatchoua, F.T., Ekesi, S. Tonnang, H.E.Z. (2016). Modeling the risk of invasion and spread of *Tuta absoluta* in Africa. *Ecological Complexity* 28: 77–93.
11. HYSPLIT: <https://ready.arl.noaa.gov/HYSPLIT.php>
12. Johnson S.J. (1987). Migration and the life history strategy of the fall armyworm, *Spodoptera frugiperda* in the Western Hemisphere. *Insect Science and its Application* 8: 543–549
13. Juárez, M. L., Murúa, M. G., García, M. G., Ontivero, M., Vera, M. T., Vilardi, J. C., Groot, A. T., Castagnaro, A. P. and Gastaminza, G., Willink, E. (2012). Host Association of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) Corn and Rice Strains in Argentina, Brazil, and Paraguay, *Journal of Economic Entomology* 105 (2): 573–582, <https://doi.org/10.1603/EC11184>
14. Mao, L., Wu, X., Huang, Z., Tatem, A.J. (2015.). Modelling monthly flows of global air travel passengers: An open access data source. *Journal of Transport Geography* 56: 60
15. Morimoto, N. and Kiritani, K. (1995). Fauna of exotic insects in Japan. *Bulletin of National Institute for Agro-Environmental Sciences* 12: 87-120
16. Nagoshi, R.N., Koffi, D., Agboka, K., Tounou, K.A., Banerjee, R., Jurat-Fuentes, J. L., Meagher, R.L. (2017). Comparative molecular analyses of invasive fall armyworm in Togo reveal strong similarities to populations from the eastern United States and the Greater Antilles. *PLoS ONE* : <https://doi.org/10.1371/journal.pone.0181982>
17. Pashley, D. P., Johnson, S. J. and Sparks, A. N. (1985). Genetic population structure of migratory moths: the fall armyworm (Lepidoptera: Noctuidae). *Annals of the Entomological Society of America* 78: 756–762
18. Rose A. H., Silversides R. H. and Lindquist O. H. (1975). Migration flight by an aphid, *Rhopalosiphum maidis* (Hemiptera: Aphididae), and a noctuid, *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *Canadian Entomologist* 107: 567-576
19. Soliman, T., Mourits, M.C.M, van der Werf, W., Hengeveld, G.M., Robinet, C. and Oude Lansink, A.G.J.M. (2012). Framework for modelling economic impacts of

- invasive species, applied to pine wood nematode in Europe. PLoS ONE7:e45505
20. Van Handel, E. (1974). Lipid utilization during sustained flights of moths. Journal of Insect Physiology 20: 2329-2332
21. Westbrook, J. K., Nagoshi, R. N., Meagher, R. L., Fleischer, S. J., Jairam, S. (2016). Modeling seasonal migration of fall armyworm moths. International Journal of Biometeorology 60: 255-267.
22. Worldpop: <http://www.worldpop.org.uk/>
23. Young, J. R. (1979) Assessing the movement of fall armyworm (*Spodoptera frugiperda*) using insecticide resistance and wind patterns. In: Movement of Highly Mobile Insects: Concepts and Methodology in Research. (Edited by Rabb R. L. and Kennedy G. G.), pp. 344-351. North Carolina State Univ. Graphics, Raleigh

Appendix 1

Data

We used diverse datasets in this analysis: vegetation, climate, production, international trade, and travel. For natural vegetation, we used the normalized difference vegetation index (NDVI) (NEO 2017). NDVI is a numerical indicator that uses near-infrared radiation (NIR) and visible radiation (VIS) of the electromagnetic spectrum. The time resolution for this dataset is one month at a resolution of 0.1 arc degree x 0.1 arc degree. For production, we used spatial distribution data from MAPSPAM. The data is obtained by using machine-learning techniques to estimate global distribution of more than 40 crops using partially available data on production and climatic factors. We identified five major hosts among the data available: wheat, rice, maize, sorghum, and sugarcane. Harvest area was used as an indicator of presence of host. The resolution is 5 minutes x 5 minutes.

We used FAOSTAT's database (FAOSTAT 2017) to analyze Egypt's trade with other countries. For exports, we considered a number of fresh fruits and vegetables: cauliflowers, peppers, cucumbers, eggplants, tomatoes, apples, grapes, oranges, papaya, strawberries and peaches. For international travel, we used the Worldpop and open flights datasets (Mao *et al.*, 2015), which provide information on passenger flows between airports and information about the airports (location, country, etc.) respectively. Finally, to group figures according to regions, we used the UN country grouping scheme.

Spatiotemporal spread model using cellular automata

We adapted a cellular automata (CA)-based diffusion model developed by Guimapi and others (Guimapi *et al.*, 2016). It is intended to capture the natural spread of the FAW accounting for

environmental factors, presence of natural vegetation, production of major host plants, the pest's flying capacity, and time to complete a lifecycle.

Role of temperature and humidity

Our model assumes that host presence is an indicator of the suitability for the pest to establish. As discussed by (Abrahams *et al.*, 2017), it is not adequate to consider only temperature data to reasonably estimate the environmental suitability of the pest. Therefore, we used the presence of vegetation and host plants as the indicator of suitability.

Model description

The Cellular Automata (CA) model consists of four components: 1) a grid of cells overlaid on the focus region, 2) cell states, 3) cell state transition rules, and 4) time steps. The focus area is the bounding box -20° to 60° latitude and -40° to 40° longitude encompassing the continent of Africa and parts of Europe and West Asia adjacent to Africa (Fig. 26). A grid of cell size 1°×1° (approximately 110km×110km at the equator) was overlaid on the focus region. Each cell can be in one of the following two states: susceptible (S) or invaded (I). Susceptible means the region covered by the cell has not been invaded by the FAW and invaded state corresponds to the situation that the region is infested.

The simulation proceeds in discrete time steps. Each time step corresponds to t months, where t can range from 0.5 to 2 months. At any time step, a cell's state is influenced by its closed Moore neighborhood of range r . Here, "closed" implies that the current cell's state is also taken into account. When range $r=1$, it corresponds to the 3°×3° cell that includes the current cell and its eight neighbors, when $r=2$, it is the 5°×5° cell including the current cell, its neighbors and all their neighbors, and so on. The current cell's state can change only if its closed Moore

neighborhood has an invaded cell, and if this is true, the state transition is governed by a set of simple transition rules. To evaluate the cell's state, the corresponding month's data is applied.

Parameter choices

The model is determined by four inputs: The threshold for vegetation and production, range r , the parameter t that determines the length in months a simulation step corresponds to.

CA transition rules

A cell is evaluated only if it has a neighbor (depending on range r) in state I. If the cell has a NDVI greater than the threshold and if at least the harvested area of one crop (maize, rice, sorghum, sugarcane, and wheat,) is greater than that of the production threshold, the cell is assigned an invaded status. Also, if the NDVI is greater than a second threshold (higher than the first threshold), its state is set to I following Guimapi *et al.* (2016).

Metric for evaluating models

To calibrate our models and compare them we adapt the maximum-likelihood approach of Carrasco *et al.* (2010). Uncertainties and delays in identifying and reporting of the pest's presence (or absence) depends on several factors. This can be mainly attributed to lack of knowledge and infrastructure to monitor and report, which varies from one country to another. Typically, pest reports become more accurate as awareness of invasion spreads. We use the following general framework to compare the model output to pest reports.

Let $t(x, S)$ denote the probability that the pest invades location x at time t in the simulation output. In our case, x corresponds to an administrative region (state, governorate, province, etc.) and t corresponds to a month. Let $t(x, G)$ denote the probability that the pest actually invaded x at time t (G denoting ground truth). While it is impossible to ascertain this value exactly, we can model this based on expert judgement. There are three types of locations;

1) those which are invaded in both simulation output as well as ground-truth, 2) those which are invaded in simulation output, but do not yet report pest presence (false positives), and 3) those which report pest presence, but are not invaded in the simulation output (false negatives).

We evaluated each simulation output based on false positives, false negatives and a score that is computed as follows for locations that correspond to case 1) Let w_u denote the uncertainty interval. Greater the w_u , the lesser the penalty for mismatch between $t(x, G)$ and $t(x, S)$. Let c_x denote the confidence we place in the report from the location. For all locations x such that both $t(x, S)$ and $t(x, G)$ exist, the total score is $score(S, w_u) = \sum c_x fl(|t(x, S) - t(x, G)|/w_u)$ where $fl()$ corresponds to the floor function. If the model matches the ground truth, then the score is 0. Therefore, lower the score, the better the fit.

For ground truth, we chose administrative regions of two or more countries each from West, East and Southern Africa. The criteria were confidence in reports based on several factors: reports from EPPO and FAO, monitoring quality (Early *et al.*, 2016), and general awareness in the region. In addition, we also included some administrative regions of Egypt, Morocco, and even Oman as regions that FAW has not invaded. We used the same confidence for all locations that report FAW ($c_x = 1$).

Appendix 2

FAW monitoring

Populations can be sampled using blacklight traps and pheromone traps. Pheromone traps are very efficient and should be suspended at canopy height during the whorl stage of maize growth. Insect catches indicate the presence of moths in the area but are not necessarily good indicators of density. Once the moths are detected, it is recommended to search for eggs and larvae. A random sampling of 20 plants in five locations, or 10 plants in 10 locations, is generally considered to be adequate to assess the proportion of plants infested. Sampling to determine larval density often requires large sample sizes, especially when larval densities are low or larvae are young, so it is not often used. Traps containing the FAW pheromone should be set up along the Nile riverbanks starting at Aswan (20 traps), Luxor (15 traps), Alexandria (10 traps) and the eastern part of the Delta (10 traps). The traps should be examined weekly. Collected moths should be sent to a specialist for identification and confirmation of FAW. When the FAW is found in a trap(s), neighboring maize fields should be surveyed for FAW infestation. Suspected FAW larvae should be collected, placed in alcohol vials and sent to a specialist for identification and confirmation. Other crop fields such as cotton, rice, sorghum, and sugarcane should also be examined for FAW infestation, as these are also potential host crops

Appendix 3

Control measures

a) Cultural control:

1. Planting border rows with maize, a month after planting the main field.

Trap Cropping with Young Corn Plants

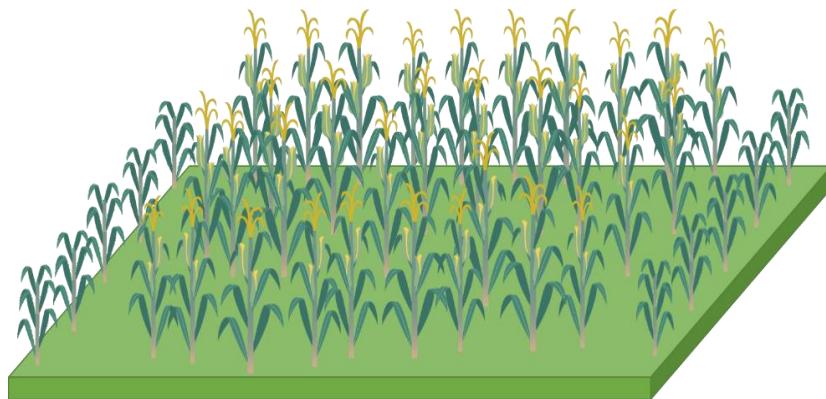


Fig. 29 Trap cropping with young plants

2. Planting a couple of plants taller than maize (such as castor) in the middle of the field to attract moths to lay eggs on them.

Trap Cropping with Castor Plant



Fig. 30 Trap cropping with Castor plants

3. Intercropping with beans has shown to reduce the FAW infestations by 20-30 percent.
 4. Push and pull technique (ICIPE).

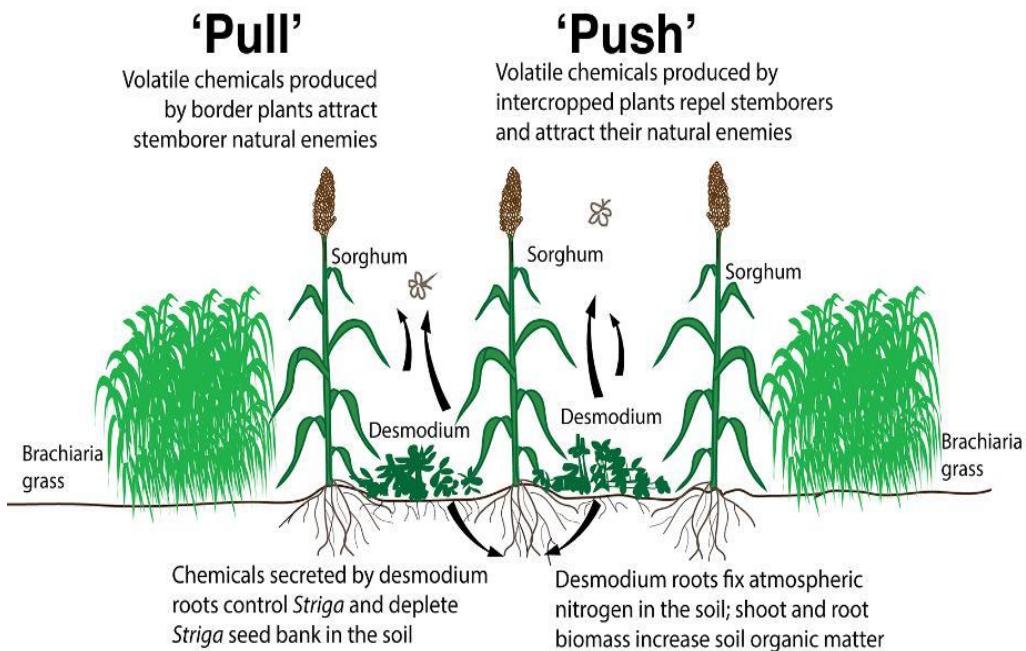


Fig. 31 Push-pull strategy for managing FAW in Africa

b) Mechanical control

Hand picking and squashing eggs and caterpillars of FAW.

c) Host Plant Resistance

(CIMMYT is developing resistant varieties but it will take a couple of years at the minimum)

d) Biological Control

Classical biological control: Classical biological control is not to be considered at this stage. Augmentative and conservation biological control options are to be tested before resorting to classical biological control. Egypt is more advanced in augmentative biological control compared to other countries in Africa.

Augmentative biological control: There are 24 government and private laboratories in Egypt that produce and supply *Trichogramma* spp. and other natural enemies to farmers for control of various pest species. The biological control laboratory at the Department of Economic Entomology and Pesticides in Cairo University, headed by Dr. Ashraf Arnaouty, is producing *Trichogramma aceae*, *T. euproctides*, *Chrysoperla* sp., *Orius* sp., coccinellids, and other natural enemies. Additionally, parasitoids *Trichogrammatoidea bactrae*, *Trichogramma evanescens* and *Telenomus busceola* have been recorded from Egypt. These could be field collected, multiplied and tested for parasitism efficacy on FAW. This laboratory is capable of providing training to technicians from government and private agencies on production of the natural enemies. It could even serve as a regional center for providing training for technicians from East and North African countries. In the past, the IPM Innovation Lab has arranged technicians from Niger and Mali to undergo training on *Trichogramma* spp. production at this facility.

Trichogramma spp. and *Chrysoperla* sp. should be released immediately after the discovery of FAW in an area to minimize the damage caused by FAW. Surveys for local natural enemies recruited by FAW should be started and continued. Natural enemies collected should be identified and evaluated for their efficacy against FAW. Effective ones can be mass multiplied and released in the field.

Conservation biological control: Broad spectrum chemical pesticides should be avoided. Pesticides that are compatible with natural enemies should be selected and used.

e) Botanical Pesticides

One of the private pesticide companies in Egypt producing neem extract and using it for control of pests in the fields. This company should be supported to enhance its production of neem

products and their quality control and distribution. Experiments need to be conducted to integrate neem products in the management of FAW in the maize IPM program.

f) Microbial Pesticides

Use of *Beauveria bassiana*, *Metarhizium anisopliae*, *Bacillus thuringiensis*, and NPVs should be explored.

g) Chemical insecticides

Insecticides are considered a main control option in response to FAW outbreaks. However, there are major limitations to the use of chemicals. The FAW larvae are often inaccessible to insecticides because of their tendency to hide in the whorls and reproductive parts of the host plant, limiting the efficacy of spraying.

Under African conditions, insecticides can be expensive and many subsistence farmers cannot afford chemical control methods. Spraying large areas of food crops and pastures with insecticides can be problematic in low income countries, as appropriate safety procedures may not be implemented on a regular basis. Personal protective equipment may not be widely available or affordable to subsistence farmers, which increases the risk of pesticide exposure and pesticide poisoning. Management using insecticides should be considered when substantial damage occurs on at least 25 percent of the plants. If high levels of damage are noted in isolated areas of a field, spot treatments may be warranted. For an effective control and an adequate penetration by insecticides, spraying should be done in the late afternoon or early evening, before the larvae burrow into the whorls or ears. Various insecticides recommended for FAW include pyrethroids, carbamates, and organophosphates. Granular insecticides can also be applied over the young plants because the particles fall deep into the whorl. However, a reliance on chemical

control to manage pest populations has become increasingly ineffective as regional populations develop resistance to several toxicological groups of insecticides.

PERSUAP (Pesticide evaluation report and safer use action plan). A PERSUAP for the FAW in Africa has been developed by IPM Innovation Lab and has been approved by the Bureau of Food Security, USAID. If needed, it could be modified for use in Egypt.

Learning the Behavior of a Dynamical System via a “20 Questions” Approach

Abstract

Developing techniques to infer the behavior of networked social systems has attracted a lot of attention in the literature. Using a discrete dynamical system to model a networked social system, the problem of inferring the behavior of the system can be formulated as the problem of learning the local functions of the dynamical system. We investigate the problem assuming an active form of interaction with the system through queries. We consider two classes of local functions (namely, symmetric and threshold functions) and two interaction modes, namely batch mode (where all the queries must be submitted together) and adaptive mode (where the set of queries submitted at a stage may rely on the answers received to previous queries). We develop complexity results which suggest that, in general, the problem of generating query sets of minimum size is computationally intractable. We present efficient heuristics that produce query sets under both batch and adaptive query modes. Our results show that a small number of appropriately chosen queries are provably sufficient to learn all the node functions. We also present experimental results to demonstrate the performance of our heuristics on over 20 well known networks.

Introduction

Background and Motivation. Discrete dynamical systems are used in a variety of settings to understand population-level contagion dynamics in terms of individual (human) agent behavior. Examples include the spread of health behaviors (Valente 2010) such as overdose prevention (Sherman et al. 2009); viruses like Ebola (Siettos et al. 2015); obesity (Christakis and Fowler 2007); segregation (Schelling 1971); becoming a user of an online communications tool (Karsai et al. 2014); coordination (Rosenthal et al. 2015); and financial contagions (Gai and Kapadia 2010). The frameworks in these works and in ours here are network representations of populations, where nodes and edges represent entities such as humans and pairwise interactions, respectively. Each of the cited works can be viewed as capturing influence through **threshold models** (Granovetter 1978; Schelling 1978), where a node v_i contracts a contagion if at least a particular number of its neighbors has already contracted it. This number for v_i is called its **threshold** t_i . We

are interested in **complex contagions** (Centola and Macy 2007) that are characteristic of social contagions, where agents need multiple reinforcing interactions to adopt a contagion; i.e., for cases where $t_i \geq 1$. (Watts 2002) argues that threshold models are used in a host of settings where incomplete information exists or when there is insufficient time to make more deliberate decisions.

In particular, we note that small changes in the thresholds of nodes can make large differences in population dynamics. An example is provided in (Granovetter 1978), where a change in one node’s threshold by a value of 1, in an arbitrarily large graph, changes population-level collective action from non-existent to full collective action. Several works have used mined data to infer thresholds for applications ranging from protests, to Twitter messaging, to joining social media (González-Bailón et al. 2011; Romero, Meeder, and Kleinberg 2011; Ugander et al. 2012); see also (Easley and Kleinberg 2010). Importantly, in all of these cases, *heterogeneous* (i.e., non-uniform) thresholds among agents have been inferred. *Thus, node thresholds must be determined based on a node’s individual behavior, its (local) neighborhood structure, and behaviors (and state) of nodes in this neighborhood.* Symmetric functions, which generalize threshold functions, also serve as natural models in game theoretic settings (Papadimitriou and Roughgarden 2003).

Some works have studied threshold inference in a **passive** setting (e.g., (Adiga et al. 2017)) where observations are *given* and the problem is to infer thresholds from these observations. In this work, we study the case where an algorithm has *control* over what information it extracts from the system via **querying** the system for desired information. In particular, the algorithm gives a set of configurations (or queries) to the system and infers the system properties based on the observed outputs. We study two query modes, namely **batch** and **adaptive** modes, that differ in their degrees of control. Under the batch mode, all the queries must be submitted together. In the adaptive mode, queries can be submitted in several stages, and queries at a stage can depend on the answers to previous queries, a strategy similar to that used in games such as “Twenty question”¹.

Our work is similar in spirit—but quite different in prob-

¹See Wikipedia entry on this game.

lem domain and results—to some of the recent works on inference (e.g. (Kleinberg, Mullainathan, and Ugander 2017)). To the best of our knowledge, this is the first work which approaches the problem of inference of dynamical systems from a combinatorial and algorithmic perspective. In doing so, we relate it to well-studied graph theoretic problems such as coloring. The formulation also enables us to quantify rigorously the complexity of inferring such systems.

Summary of Results. Our focus is on the following problem: given the underlying graph of a dynamical system, construct queries to identify all the local functions. The optimization goal is to minimize the number of queries. We present both theoretical and experimental results as summarized below.

1. We develop algorithms for generating query sets under both batch and adaptive modes to identify local functions of dynamical systems. As can be expected, adaptive query mode can produce significantly smaller query sets compared to the batch mode. We also show that if the goal is to find a query set which can identify symmetric functions with high probability, the size of the query set can be further reduced.
2. We prove lower bounds on the number of queries needed under both batch and query modes. We also present complexity results that point out the difficulty of efficiently generating small query sets.
3. We present an approximation algorithm that reduces the size of a query set (or makes it compact) by eliminating redundant queries and establish its performance guarantee.
4. We evaluate the proposed algorithms on a large number of real-world and synthetic networks. For the batch mode, one of our approaches based on greedy graph coloring generated query sets of minimum size for most of the real-world networks. We also demonstrate the effectiveness of a simple approach based on sampling queries from a particular distribution followed by a compaction algorithm.
5. We develop a greedy adaptive heuristic based on binary search and evaluate it by generating query sets for various settings of networks and threshold assignments. Our results show that for most cases, it significantly outperforms the batch mode algorithms.

All proofs for propositions, lemmas, and theorems appear in the supplement.

Related Work. There are several works on the passive mode of inference. Many researchers have studied the problem of learning automata; e.g., (Murphy 1996). (Kearns and Vazirani 1994) study the problem of learning normal forms and Boolean functions. Works such as (González-Bailón et al. 2011; Romero, Meeder, and Kleinberg 2011) infer thresholds from social media data. Learning the source nodes of infection for contagion spreading is addressed in (Zhu, Chen, and Ying 2017). Many of these problems are formally hard even for simple local functions. The work of (Adiga et al. 2017) provides several problems aimed at inferring thresholds in threshold-based discrete dynamical systems.

Active querying is studied in (Kleinberg, Mullainathan, and Ugander 2017) in the context of determining user choices from a finite set of ranked options—the choice set problem. The goal is to minimize the number of queries of

arbitrary subsets S of size k , of a universal set U , to learn a user’s choice from among the elements of each set S . With these results, the algorithm can then predict the user’s choice for any subset $S \subseteq U$ of size k . They show that this can be accomplished with $O(n \log n)$ queries where $n = |U|$.

Although error-tolerant approaches for querying systems are beyond the scope of this work, there are several works that include allowance for errors in inferring system properties. These include (Valiant 1984; Juba 2016; He et al. 2016; Zhang, Mathew, and Juba 2017; Kleinberg, Mullainathan, and Ugander 2017).

There are several challenges in determining individual node thresholds in realistic settings: (i) data are collected at discrete time intervals (not continuously), (ii) there may be time delay effects in agents observing their neighborhoods, and (iii) inherent stochasticity (Valente 1996; Berry and Cameron 2017). Practical guidelines and issues for threshold measurement are discussed in (Berry and Cameron 2017). Here, we investigate problems of inferring local functions using rigorous formulations, supplementing them with experimental results from heuristics.

Synchronous Dynamical Systems (SyDSs)

Formal Definitions

Let \mathbb{B} denote the Boolean domain $\{0,1\}$. A **Synchronous Dynamical System** (SyDS) \mathcal{S} over \mathbb{B} is specified as a pair $\mathcal{S} = (G, \mathcal{F})$, where (a) $G(V, E)$, an undirected graph with $|V| = n$, represents the underlying graph of the SyDS, with node set V and edge set E , and (b) $\mathcal{F} = \{f_1, f_2, \dots, f_n\}$ is a collection of functions in the system, with f_i denoting the **local function** associated with node v_i , $1 \leq i \leq n$.

Each node of G has a state value from \mathbb{B} . Each function f_i specifies the local interaction between node v_i and its neighbors in G . The inputs to function f_i are the state of v_i and those of the neighbors of v_i in G ; function f_i maps each combination of inputs to a value in \mathbb{B} . This value becomes the next state of node v_i .

At any time t , the **configuration** \mathcal{C} of a SyDS is the n -vector $(s_1^t, s_2^t, \dots, s_n^t)$, where $s_i^t \in \mathbb{B}$ is the state of node v_i at time t ($1 \leq i \leq n$). In a SyDS, all nodes compute and update their next state *synchronously*.

Classes of Local Functions

We consider two classes of local functions, namely **threshold** and **symmetric** functions. They are defined below.

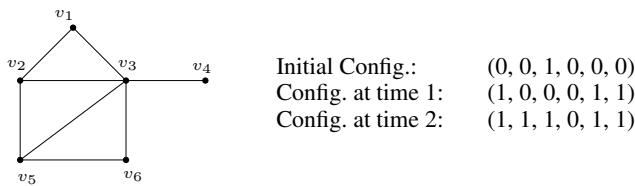
(i) **Threshold functions:** The local function f_v associated with node v of a SyDS \mathcal{S} is a t_v -**threshold** function for some integer $t_v \geq 0$ if the following condition holds: the value of f_v is 1 if the number of 1’s in the input to f_v is *at least* t_v ; otherwise, the value of the function is 0. Let d_v denote the degree of node v , and let t_v denote the threshold of node v . The number of inputs to the function f_v is $d_v + 1$. Thus, we assume that $0 \leq t_v \leq d_v + 2$. (The threshold values 0 and $d_v + 2$ allow us to realize local functions that always output 1 and 0 respectively.)

(ii) **Symmetric functions:** A local function f_v at node v is **symmetric** if the value of the function depends only on the number of 1’s in the input. Thus, a symmetric function f_v

with k inputs can be specified using a table with $k + 1$ rows, with row i specifying the value of the function when the number of 1's in the input to the function is exactly i , $0 \leq i \leq k$. Note that each threshold function is also a symmetric function.

We will use the term “symmetric SyDS” (“threshold SyDS”) to refer to a SyDS whose local functions are all symmetric (threshold).

Example: Consider the graph of a threshold SyDS shown in Figure 1. Suppose the local transition functions at each of the nodes v_1, v_5 and v_6 is the 1-threshold function and the functions at v_2, v_3 and v_4 are 2-threshold functions. Assume that initially, v_3 is in state 1 and all other nodes are in state 0. During the first time step, the states of nodes v_1, v_5 and v_6 change to 1 since each of them has a neighbor (namely, v_3) in state 1. Also, the state of v_3 changes to 0 since its threshold is 2 and none of its neighbors is in state 1. The states of v_2 and v_4 don't change; they continue to be 0. During time step 2, v_2 and v_3 change to 1 but v_4 remains at 0. Once the system reaches the configuration $\mathcal{C} = (1, 1, 1, 0, 1, 1)$ at time step 2, it remains in that configuration forever; that is, \mathcal{C} is a **fixed point** for this system. \square



Note: Each configuration has the form $(s_1^t, s_2^t, s_3^t, s_4^t, s_5^t, s_6^t)$, where s_i^t is the state of node v_i at time t , $1 \leq i \leq 6$. The configuration at time 2 is a fixed point.

Figure 1: An Example of a SyDS.

Additional Terminology: If a given SyDS can transition in one step from a configuration \mathcal{C}' to a configuration \mathcal{C} , then \mathcal{C} is a **successor** of \mathcal{C}' and \mathcal{C}' is a **predecessor** of \mathcal{C} . Since our local functions are deterministic, each configuration has a unique successor; however, a configuration may have zero or more predecessors. A **fixed point** is a configuration \mathcal{C} for which the successor is \mathcal{C} itself.

Given a graph $G(V, E)$ and a node $v_i \in V$, the **closed neighborhood** of v_i , denoted by $N[v_i]$, is defined by $N[v_i] = \{v_i\} \cup \{v_j : \{v_i, v_j\} \in E\}$. Thus, the inputs to the local function f_i at v_i are the states of the nodes in $N[v_i]$.

Query Model

The general problem addressed in this paper is that of correctly identifying the local functions of a SyDS by querying the system. We assume that the underlying network is known. Each query specifies a configuration \mathcal{C} and the response from the system is the successor \mathcal{C}' of \mathcal{C} . Since the state of each node is either 0 or 1, each query q and the response to q are bit vectors. We consider two query modes. In the **batch** query mode, a user must submit all the queries at the same time as a single batch. In the **adaptive** query mode, a user may submit the queries in several batches; the queries chosen in a batch may rely on the responses received from

the system for the previous batches of queries. As will be seen, for threshold SyDSs, the adaptive query mode can significantly decrease the number of queries. The following additional definitions regarding queries will be used throughout this paper.

Given a query q and a node v_i , the **score** of q with respect to v_i , denoted by $\text{score}(q, v_i)$, is the number of nodes in the closed neighborhood $N[v_i]$ of v_i that are set to 1 by q . Thus, $\text{score}(q, v_i)$ gives the number of 1's in the input provided by q to the local function f_i at v_i .

Definition 1 Let \mathcal{S} be a symmetric SyDS. For any node v_i , let d_i denote the degree of v_i .

- (a) A query set Q **covers a node** v_i if for each j , $0 \leq j \leq d_i + 1$, there is a query $q \in Q$ such that $\text{score}(q, v_i) = j$.
- (b) A query set Q **covers a set** B of nodes if Q covers every node $v_i \in B$.
- (c) A query set Q is **complete** if it covers the node set V .

When a query set Q covers a node v , the local symmetric function f_v can be correctly inferred from the responses to the queries in Q . Thus, complete query sets have the following property.

Observation 1 Let \mathcal{S} be a symmetric SyDS. If Q is a complete query set for \mathcal{S} , then each local function of \mathcal{S} can be determined given the successor of each query in Q . \blacksquare

Theoretical Results

In this section, we first present an algorithm for generating query sets under the batch mode for symmetric SyDSs. We then show that for threshold SyDSs, the number of queries can be substantially reduced under the adaptive query mode. We also establish lower bounds on the number of queries needed under both modes. We present complexity results that suggest that in general, generating complete query sets of minimum size is computationally intractable. We also develop an efficient heuristic to reduce the size of query sets by eliminating redundant queries and prove its performance guarantee.

Generating Query Sets Under the Batch Mode

We begin by defining the notion of a **monotone query sequence**. The sequence of queries constructed can be submitted as a batch to learn all the local functions of a symmetric SyDS. Using the notion of “sequence” allows us to point out an interesting connection between the problem of identifying local symmetric functions and a variant of the node coloring problem for the underlying graph.

Definition 2 (a) Given two queries q_1 and q_2 , we use the notation $q_1 \leq q_2$ to mean that every bit which is 1 in q_1 is also 1 in q_2 .

(b) A query sequence $\langle q_1, q_2, \dots, q_r \rangle$ is **monotone** if for each i , $1 \leq i \leq r - 1$, $q_i \leq q_{i+1}$.

(c) Let \mathcal{S} be a SyDS in which each local function is symmetric and let M be a monotone query sequence. If M is also a complete query set for \mathcal{S} (i.e., each node v of \mathcal{S} is covered by M), then M is a **complete monotone query sequence**.

Figure 2: Steps of the Algorithm ALG-MONOTONE-SEQ

Input: Graph $G(V, E)$ of a symmetric SyDS \mathcal{S} .

Output: A monotone complete query sequence M for \mathcal{S} .

Steps:

1. Construct the graph $G^2(V, E')$.
2. Use the algorithm of Theorem 1 to obtain a k -coloring of G^2 where $k \leq \min\{\Delta^2 + 1, n\}$.
3. Let C_1, C_2, \dots, C_k denote the color classes created in Step 2. (Color class C_j consists of all nodes assigned color j , $1 \leq j \leq k$.) Create the query sequence $M = \langle q_0, q_1, \dots, q_k \rangle$ with $k + 1$ queries as follows.
 - (a) Query q_0 is a bit vector where every element is 0.
 - (b) **for** $j = 1$ **to** k **do**
 - Create query q_j by choosing the value 1 for all the nodes in $C_1 \cup \dots \cup C_j$ and 0 for the other nodes.
4. Output the query sequence M .

We now present an algorithm to show that if the underlying graph G has n nodes, then there is a monotone complete query sequence M for \mathcal{S} with at most $\min\{\Delta^2 + 2, n + 1\}$ queries, where Δ is the maximum node degree of G . This sequence of queries can be submitted as a batch to learn all the symmetric local functions. To establish this result, we recall the following definitions.

Definition 3

- (a) Given an undirected graph $H(V_H, E_H)$ and an integer $k \geq 1$, a **k -coloring** of H assigns a color from the set $\{1, 2, \dots, k\}$ to each node of H such that for each edge $\{u, v\} \in E_H$, the colors assigned to u and v are different.
- (b) Given an undirected graph $G(V, E)$, the **square** of G , denoted by $G^2(V, E')$, is an undirected graph on the same vertex set V . The edge set E' is defined as: $\{u, v\} \in E'$ iff there is a path with at most 2 edges between u and v in G .

We will also use the following known result (West 2001).

Theorem 1 Let $H(V_H, E_H)$ be a graph with maximum node degree Δ_H . Then, H can be colored efficiently using at most $\Delta_H + 1$ colors. ■

Our algorithm ALG-MONOTONE-SEQ for generating a monotone complete query sequence M for the given SyDS \mathcal{S} is shown in Figure 2. It is easy to see that the algorithm runs in polynomial time. The following theorem shows its correctness and estimates the number of queries generated.

Theorem 2 Let \mathcal{S} be a symmetric SyDS whose graph $G(V, E)$ has n nodes and maximum node degree Δ . Algorithm ALG-MONOTONE-SEQ (Figure 2) produces a monotone complete query sequence M with at most $\min\{\Delta^2 + 2, n + 1\}$ queries. ■

For some graphs with maximum node degree Δ , Algorithm ALG-MONOTONE-SEQ may generate a query sequence with $\Omega(\Delta^2)$ queries but it guarantees that the resulting query sequence is complete for a symmetric SyDS. For graphs where $\Delta \geq (\log n)^2$, the number of queries can be reduced to $O(\Delta^{1.5} \log n)$, if we only need the query set to be complete with *high probability*. This result is stated below.

Theorem 3 Let \mathcal{S} be a symmetric SyDS with graph $G(V, E)$ where $|V| = n$ and maximum node degree = Δ . A query set Q of size $O(\Delta^{1.5} \log(n))$ which is complete with probability at least $(1 - \frac{1}{n})$ can be constructed for \mathcal{S} . ■

Generating Query Sets Under the Adaptive Mode

For threshold SyDSs, the adaptive query mode can reduce the number of queries significantly. To illustrate this, consider a SyDS whose underlying graph is a **star graph** with n nodes; that is, there is one node v_1 with degree $n - 1$ which is the root of the tree and each of the other nodes v_2 through v_n is child of the root. As will be shown in the section on lower bounds, in the batch mode, $n + 1$ queries are necessary even for the star graph to identify all the thresholds. However, under the adaptive mode, using the following method, $O(\log n)$ queries are sufficient.

The idea is simple: use *binary search* to identify the threshold of node v_1 whose degree is $n - 1$ using $O(\log n)$ queries. After this, the following 3 additional queries are sufficient to identify the thresholds of the remaining $n - 1$ nodes: a query with all 0's, a second query with all 1's and a third one in which v_1 has the value 1 and all the remaining nodes have the value 0. Thus, all the thresholds can be identified $O(\log n)$ queries under the adaptive mode.

The above idea can be applied to a more general class of graphs. Let a class of graphs with n nodes be called (α, β) -simple, if at most α nodes have degree $> \beta$ (the degree may be $\Omega(n)$) and all the remaining $n - \alpha$ nodes have a degree of at most β , with α and β being **constants** independent of n . Thus, each star graph belongs to the class of $(1, 1)$ -simple graphs. The following result shows the usefulness of the adaptive query mode for (α, β) -simple graph.

Theorem 4 For any threshold SyDS whose underlying graph G belongs to the class of (α, β) -simple graphs, $O(\log n)$ queries are sufficient in the adaptive mode to identify all the threshold values. ■

The above result can be used to establish a bound on the number of queries under the adaptive mode for scale-free graphs as stated below.

Theorem 5 For a threshold SyDS whose underlying graph $G(V, E)$ is scale-free with exponent $\gamma \geq 1$, the thresholds can be found using $O(n^{\frac{2}{\gamma+1}})$ queries under the adaptive query mode. ■

Lower Bounds on Sizes of Query Sets

Here, we present lower bounds under batch and adaptive query modes. We begin with a result that provides a lower bound for any symmetric SyDS under the batch mode.

Proposition 1 Let \mathcal{S} be a symmetric SyDS where the underlying graph $G(V, E)$ has a maximum node degree Δ . Under the batch query model, every complete query set must contain at least $\Delta + 2$ queries.

As a simple consequence of the above proposition, the following result points out that there are SyDSs with n nodes for which every complete query set must have $n + 1$ queries. This lower bound matches the upper bound of $n + 1$ given by Theorem 2 for all graphs.

Corollary 1 For a symmetric SyDSs whose underlying graph is a clique on n nodes, every complete query set under the batch mode must have at least $n + 1$ queries. ■

We now establish a lower bound under the adaptive query model to show that there are threshold SyDSs for which a large number of queries are needed even under the adaptive query mode. However, this result does not rule out the possibility of smaller query sets for special graph classes.

Theorem 6 For every $n \geq 1$, there is a threshold SyDS whose underlying graph is a clique on n nodes such that at least $n + 1$ queries are necessary under the adaptive query mode to correctly identify all the threshold values. ■

Complexity of Generating Small Monotone Complete Query Sequences

Here, we present a result that provides an indication of the difficulty of efficiently generating small query sets. In particular, we will show the NP-completeness of the following problem.

Short Monotone Complete Query Sequence (SMCQS)

Given: The underlying graph $G(V, E)$ of a SyDS \mathcal{S} where each local function is symmetric and a positive integer k .

Question: Is there a monotone complete query sequence Q with at most k queries for \mathcal{S} ?

Theorem 7 Problem SMCQS is NP-complete. ■

Results for Query Set Compaction

Under the batch mode, after generating a complete set of queries, it is useful to reduce the size of the set by eliminating redundant queries. We refer to this as the **Query Set Compaction** problem and its formulation is as follows.

Query Set Compaction (QSC)

Given: The underlying graph $G(V, E)$ of a symmetric SyDS \mathcal{S} , a complete query set Q and an integer $k \leq |Q|$.

Question: Is there a subset $Q' \subseteq Q$ such that (i) $|Q'| \leq k$ and (ii) Q' is also a complete query set for \mathcal{S} ?

The following result points out the intractability of QSC.

Theorem 8 (a) The problem QSC is NP-complete even when the underlying graph has no edges. (b) Unless $P = NP$, QSC cannot be approximated to within the factor $O(\log n)$, where n is the number of nodes in the underlying graph of the SyDS. ■

To complement the non-approximability result of the previous section, we present an efficient approximation algorithm with a performance guarantee of $O(\log n)$ for the QSC problem. The idea is to use a reduction from the QSC problem to the well known **Minimum Set Cover** (MSC) problem (Garey and Johnson 1979). A (greedy) approximation algorithm for MSC which provides a performance guarantee of $O(\log n)$ for the MSC problem is well known (Vazirani 2001).

The steps of our approximation algorithm Approx-QSC for QSC are shown in Figure 3. It can be seen that the approximation algorithm runs in polynomial time. The performance guarantee provided by Approx-QSC is indicated in the following theorem.

Input: The underlying graph $G(V, E)$ of a symmetric SyDS \mathcal{S} and a complete query set Q .

Output: A subset $Q' \subseteq Q$ such that Q' is also a complete query set and $|Q'|$ is as small as possible.

Steps:

1. To construct the base set X of the MSC instance, consider each node v_i ; let d_i denote the degree of v_i . Create a set A_i of $d_i + 1$ elements, given by $A_i = \{a_{ik} : 0 \leq k \leq d_i\}$, for v_i . The set X is given by $X = \cup_{i=1}^n A_i$.
2. From each query $q_j \in Q$, construct a subset Y_j of X as follows. Initially, Y_j is empty. For each $v_i \in V$, $1 \leq i \leq n$, if q_j sets k of the inputs to v_i to 1, then the element a_{ik} is added to the set Y_j .
3. Use the greedy algorithm (Vazirani 2001) to get an approximate solution Y' to the resulting MSC instance.
4. Construct the query set Q' by choosing the query corresponding to each subset in Y' and output Q' .

Figure 3: Details regarding Algorithm Approx-QSC

Theorem 9 Algorithm Approx-QSC provides a performance guarantee of $O(\log n)$ where n is the number of nodes in the underlying graph of the SyDS. ■

Experimental Results

We performed extensive experiments on more than 20 diverse real-world and synthetic networks. They are listed in Table 1 along with some of their properties. We present representative results for selected networks, with other networks exhibiting the same behavior unless stated otherwise.

We studied three approaches for inferring thresholds, two of which correspond to the batch mode and hence applicable to symmetric SyDSs as well, and one being an adaptive approach. The first batch mode approach is based on coloring G^2 and the other is a random query approach based on Theorem 3. In both these cases, we applied the compaction algorithm (Figure 3) on the complete sets that were constructed. Next, we propose a greedy algorithm for inferring thresholds in the adaptive mode and evaluate its performance.

Our theoretical results indicate that both network structure and the threshold assignments influence the number of queries required to infer the system. The experiments conducted were designed to further explore these aspects.

Method 1: G^2 Coloring Based Approach

We studied the performance of ALG-MONOTONE-SEQ (Figure 2). The results of are in Table 2. For most real world networks considered in this paper, it gives the best possible performance, i.e., $n_c(G^2)$ is equal to $\Delta + 1$, the lower bound on the size of complete set (Proposition 1). For synthetic networks (random regular and Erdős-Rényi graphs) though, $n_c(G^2)$ is significantly higher than $\Delta + 1$, yet much lower than $\Delta^2 + 1$. The reader should note that the observed performance is due to a combination of the structure of G^2 and the nature of the greedy coloring scheme. We observe that unlike the synthetic networks considered, most of the real-world networks are scale-free with maximum degree being much larger than average degree d_{avg} . This is

Table 1: Networks used in our experiments, their properties, and results of the different algorithms for inferring local functions or thresholds. The networks are grouped by type: social online, friendship, co-authorship (collaboration) (Leskovec and Krevl 2014) and synthetic networks. To conserve space, we have provided range of values for some network families.

Network (num. of instances)	Properties					Results	
	Type	n	avg. deg. d_{avg}	max. deg. Δ	Spec. rad.	Query set size Meth. 1 $n_c(G^2) + 1$	Query set size Meth. 3 $t(v) = \frac{d(v)+2}{2}$
FB	social media	43,953	8.30	223	39.7	225	53
p2p-gnutella04	hw connectivity	10,876	7.35	103	17.08	105	31
Enron	email	33,696	10.73	1383	118.4	1385	624
Epinions	online opinions	75,879	10.69	3044	246	3046	294
Slashdot0811	online	77,360	12.13	2539	250.3	2541	214
Slashdot0902	online	82,168	12.27	2552	252.6	2554	267
Wikipedia	online voting	7,115	28.32	1065	138.2	1067	114
ca-astroph	co-author	17,903	22.00	504	94.43	506	76
ca-condmat	co-author	21,363	8.55	279	37.89	281	67
ca-grqc	co-author	4,158	6.46	81	45.62	83	25
ca-hepph	co-author	11,204	21.00	491	244.9	619	72
ca-hepth	co-author	8,638	5.74	65	31.03	67	28
cit-hepph	co-author	34,401	24.46	846	76.58	848	78
Clique	synthetic	1000	999	999	999	1001	8
Rand. reg. A (10)*	synthetic	1000	10,800	10,800	10,800	34-36,1001	Fig. 4(a)(0.0)
Rand. reg. B (10)	synthetic	80,000	10,12	10,12	10,12	38	20 (avg)
Erdős-Rényi (10)	synthetic	80,000	10, 12	25-28,27-32	11.1,13.04-13.09	36-38, 46-47	—

* Degrees are 10, 50, 100, 200, 250, 400, 500, 700, 800. For $d_{\text{avg}} = 50, 100, n_c(G^2) + 1 = 348-358, 988-996$, and for greater $d_{\text{avg}}, n_c(G^2) = 1000$.

a possible reason for the superior performance of this approach. We also compared the results to the spectral radius bound, that is, the number of colors needed to color G^2 is at most $1 + \lambda_{\max}^2$ (Miao and Fan 2014). It is a well-known fact that $\sqrt{\Delta} \leq \lambda_{\max} \leq \Delta$, and for the real-world networks considered, λ_{\max} is indeed much less than Δ . However, despite this fact, we observe that $\lambda_{\max}^2 + 1$ is much larger than $n_c(G^2) + 1$ in these cases.

Compaction. We note that the query set generated by this approach is already compact, i.e., no subset of queries can be complete. We provide an intuitive explanation for this in the supplement.

Method 2: Randomized Algorithm

In this approach, we use the method of Theorem 3 to construct a complete set. The query set contains the configurations of all zeros, of all ones and $\ell\Delta$ random queries where ℓ queries are sampled from distributions $\mathbb{D}(i/\Delta)$ for $1 \leq i \leq \Delta$. Compared to Method 1, this is a very simple approach not requiring construction of G^2 or graph coloring. However, it is not guaranteed that the constructed query set is complete and therefore the process may have to be repeated a number of times. However, as discussed below, with ℓ sufficiently large, we can obtain a complete set in few repetitions. Further, the resulting set, even though large, can be compressed using the compaction algorithm.

We constructed 50 such query sets for three values of ℓ (2, 5 and 10) and checked if each of them is a complete set. For $\ell = 2$, out of the 50 sets none of them were complete. However, for $\ell = 10$, from 5 to 50 query sets turned out to be complete sets depending on the network. We applied the compaction algorithm on the complete sets generated by

Table 2: Results of Method 2.

Network	Query set size	% Compaction	Network	Query set size	% Compaction
FB	407	81	ca-grqc	153	81
p2p	159	84	ca-hepph	1201	75
Enron	2306	83	ca-hepth	140	78
Wikipedia	1420	86	cit-hepph	1240	85
ca-astroph	899	82	Rand. reg. A	$\approx 5\Delta$	40
ca-condmat	393	85	—	—	—

the randomized algorithm. The results for $\ell = 10$ are in Table 2. The compaction ratio depends on the size of complete set which was given as input. On an average, the combination of randomized algorithm and compaction gives query sets of size around 1.5 to 2 times that of Method 1 (Table 1 (Meth. 1)). However, comparatively these are much easier to generate.

Performance of compaction. We note that compaction of query sets generated by Method 2 consistently yields 80% reduction in the size of the query set (Table 2).

Method 3: Adaptive Algorithm

While the previous two methods correspond to the batch mode, here we develop an adaptive algorithm to infer the thresholds. We give an outline of the approach. The algorithm description is in the supplementary material. For every node, let $t_L(v)$ and $t_H(v)$ be the minimum and maximum possible values of threshold that v can be assigned. These values quantify the uncertainty about the threshold. The threshold is said to have been inferred when $t_H(v) = t_L(v)$. In a query q , if $\text{score}(q, v)$ falls in the range $[t_L(v), t_H(v) - 1]$, then, the uncertainty reduces to either $[\text{score}(q, v) +$

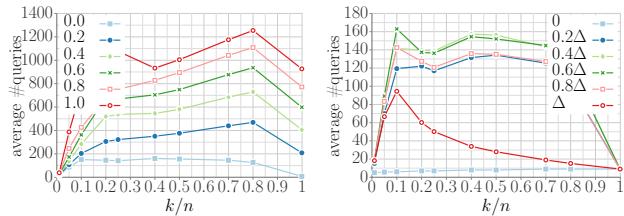


Figure 4: Experiments with 1000 node random k -regular graphs. (a) The threshold of a node is randomly assigned an integer in the interval $[(k+2)(1-\theta)/2, (k+2)(1+\theta)/2]$. The legend shows values of θ . (b) All nodes are assigned a fixed threshold relative to k . The legend shows values of t_i .

$1, t_H(v) - 1]$ or $[t_L(v), \text{score}(q, v)]$ depending on the state observed in the successor configuration. In this heuristic, we use a greedy adaptive approach where the current query is constructed iteratively in the following way. To begin with all nodes are in state 0. We first choose that vertex, say v_{\max} for which the threshold range is maximum. We set exactly $\lfloor(t_L(v) + t_H(v))/2\rfloor$ of nodes in its closed neighborhood to state 1. This guarantees a reduction in the range by half. In the next iteration, we ignore all nodes in G within distance-2 of v_{\max} and repeat this process. The query is fully constructed there are no more vertices to consider. After each query, the range $[t_L(v), t_H(v) - 1]$ for every v is updated based on its state in the successor. We terminate this process when for all v , $t_L(v) = t_H(v)$. The analysis of our experimental results follows.

Influence of threshold values and ranges. In general, the number of queries required is highly dependent on the possible threshold values the nodes can be assigned. We conducted experiments in the following manner. Let $0 \leq \theta \leq 1$ be a real number. For a fixed value of θ , each node v was assigned a threshold value uniformly at random from the interval $[(d(v) + 2)(1 - \theta)/2, (d(v) + 2)(1 + \theta)/2]$. Note that for $\theta = 0$, the interval corresponds to the fixed threshold of $(d(v) + 2)/2$ and for $\theta = 1$, any value from 0 to $d(v) + 2$ is possible. The results are in Figure 4(a) and 5(a) for random k -regular and real-world networks respectively. For the random-regular graphs, the number of queries (averaged over 10 instances of graphs for each k) increases from an order of $\log k$ to as high as n , the size of the graph. We note that for $k = n - 1$, this is in accordance with Theorem 6. For the real-world graphs, we see that increasing the range of threshold has the effect of gradually increasing the number of queries, but the number is less than 1.5Δ . In Figure 4(b), we investigate the influence that the threshold value on query set size. Again, we considered random k -regular graphs with varying k . Every node was assigned the same threshold. We see that the number of queries required is maximum when the threshold is around $\Delta/2$, and it decreases as the threshold approaches either 0 or Δ .

Influence of network structure. The theoretical bounds developed in the previous sections provide bounds with respect to size of the graph and maximum degree. Here, our objectives are two-fold. Firstly, we compare our adaptive approaches to the non-adaptive bounds, particularly the number of queries required relative to $\log \Delta$, Δ and Δ^2 . Secondly, we investigate the effect of graph density and degree

distribution on the performance of the heuristic.

We note that graph density plays an important role in the performance of the algorithm. First we will consider the synthetic networks. In Figure 4(b), we see that for low values of k the number of queries required is very small, but it increases rapidly (for higher values of thresholds). When the graph is sparse, for every node, the number of nodes within distance two ($k^2 + 1$ nodes) is small. Therefore, for every query constructed by the heuristic, the uncertainty range of around n/k^2 nodes (the “ v_{\max} ” vertices) is halved. However, as k increases, this number decreases drastically. Hence we see that the number of queries required increases. However, as the graph density increases, the intersection of neighborhoods of any two nodes is large which has the effect of reducing the variation in the scores of nodes. Therefore, particularly when the range of threshold values is limited, the thresholds can be inferred with far fewer queries than for sparser networks.

Progress towards inferring thresholds. In Figure 5(b), we plot the accumulated threshold ranges for all vertices as the algorithm moves from one query to the next. We note that within one-tenth of the total query size, the total accumulated threshold range decreases to 5% of its original value for all the studied networks.

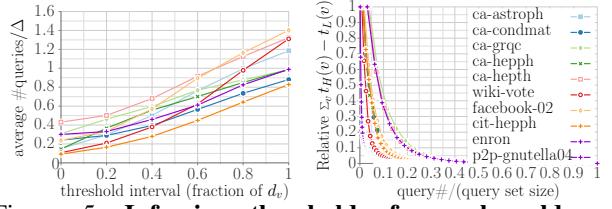


Figure 5: Inferring thresholds for real-world networks. (a) Adaptive heuristic for varying threshold ranges. (b) Progress made by the adaptive algorithm (Method 3) in each query.

Future Work

In this paper, our focus was on learning the threshold and symmetric functions of dynamical systems. One direction for future work is to investigate other classes of functions. Another direction is to explore the use of queries to infer other components of a dynamical system such as the network topology. A third direction is to extend the results for other classes of dynamical systems.

References

- Adiga, A.; Kuhlman, C. J.; Marathe, M. V.; Ravi, S. S.; Rosenkrantz, D. J.; and Stearns, R. E. 2017. Inferring local transition functions of discrete dynamical systems from observations of system behavior. *Theor. Comput. Sci.* 679:126–144.
- Berry, G., and Cameron, C. J. 2017. A new method to reduce overestimation of thresholds with observational network data. arXiv:1702.02700v1 [cs.SI].
- Centola, D., and Macy, M. 2007. Complex contagions and the weakness of long ties. *American Journal of Sociology* 113(3):702–734.
- Christakis, N. A., and Fowler, J. H. 2007. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine* 357(4):370–379.
- Easley, D., and Kleinberg, J. 2010. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*.
- Gai, P., and Kapadia, S. 2010. Contagion in financial networks. *Proceedings of the Royal Society A* 466:2401–2423.
- Garey, M. R., and Johnson, D. S. 1979. *Computers and Intractability: A Guide to the Theory of NP-completeness*. San Francisco: W. H. Freeman & Co.
- González-Bailón, S.; Borge-Holthoefer, J.; Rivero, A.; and Moreno, Y. 2011. The dynamics of protest recruitment through an online network. *Scientific Reports* 1:7 pages.
- Granovetter, M. 1978. Threshold models of collective behavior. *American Journal of Sociology* 1420–1443.
- He, X.; Xu, K.; Kempe, D.; and Liu, Y. 2016. Learning influence functions from incomplete observations. arXiv:1611.02305 [cs.SI].
- Juba, B. 2016. Learning abductive reasoning using random examples. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 999–1007.
- Karsai, M.; Iniguez, G.; Kaski, K.; and Kertesz, J. 2014. Complex contagion process in spreading of online innovation. *Journal of the Royal Society Interface* 11:20140694–1–20140694–8.
- Kearns, M. J., and Vazirani, U. V. 1994. *An Introduction to Computational Learning Theory*. Cambridge, MA: MIT Press.
- Kleinberg, J.; Mullainathan, S.; and Ugander, J. 2017. Comparison-based choices. arXiv:1705.05735v1 [cs.DS].
- Leskovec, J., and Krevl, A. 2014. SNAP Datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data>.
- McCormick, S. T. 1983. Optimal approximation of sparse Hessians and its equivalence to a graph coloring problem. *Math. Programming* 26(2):153–171.
- Miao, L., and Fan, Y. 2014. The distance coloring of graphs. *Acta Mathematica Sinica* 30(9):1579–1587.
- Murphy, K. P. 1996. Passively learning finite automata. Technical Report 96-04-017, Santa Fe Institute, Santa Fe, NM.
- Papadimitriou, C. H., and Roughgarden, T. 2003. Equilibria in symmetric games. Report, Stanford University.
- Romero, D. M.; Meeder, B.; and Kleinberg, J. 2011. Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web*, 695–704. ACM.
- Rosenthal, S. B.; Twomey, C. R.; Hartnett, A. T.; Wu, H. S.; and Couzin, I. D. 2015. Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proceedings of the National Academy of Sciences* 112(15):4690–4695.
- Schelling, T. C. 1971. Dynamic models of segregation. *Journal of Mathematical Sociology* 1:143–186.
- Schelling, T. C. 1978. *Micromotives and Macrobbehavior*.
- Sherman, S. G.; Ganna, D. S.; Tobin, K. E.; Latkin, C. A.; Welsh, C.; and Bielenson, P. 2009. The life they save may be mine: Diffusion of overdose prevention information from a city sponsored programme. *International Journal of Drug Policy* 20:137–142.
- Siettos, C.; Anastassopoulou, C.; Russo, L.; Grigoras, C.; and Mylonakis, E. 2015. Modeling the 2014 ebola virus epidemic—agent-based simulations, temporal analysis and future predictions for liberia and sierra leone. *PLOS Currents Outbreaks* 1–22.
- Ugander, J.; Backstrom, L.; Marlow, C.; and Kleinberg, J. 2012. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences* 109(16):5962–5966.
- Valente, T. W. 1996. Social network thresholds in the diffusion of innovations. *Social Networks* 18:69–89.
- Valente, T. W. 2010. *Social Networks and Health: Models, Methods, and Applications*.
- Valiant, L. G. 1984. A theory of the learnable. *Communications of the ACM* 18(11):1134–1142.
- Vazirani, V. V. 2001. *Approximation Algorithms*.
- Watts, D. J. 2002. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences* 99:5766–5771.
- West, D. B. 2001. *Introduction to Graph Theory*.
- Zhang, M.; Mathew, T.; and Juba, B. A. 2017. An improved algorithm for learning to perform exception-tolerant abduction. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 1257–1265.
- Zhu, K.; Chen, Z.; and Ying, L. 2017. Catch’em all: Locating multiple diffusion sources in networks with partial observations. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 1676–1683.

Learning the Behavior of a Dynamical System via a “20 Questions” Approach

(Supplementary Material)

This supplement provides proofs for many results mentioned in the main paper. (As some of the proofs involve long mathematical formulas, this supplement has been formatted using the single column mode.)

Statement and Proof of Theorem 2

Statement of Theorem 2: Let \mathcal{S} be a symmetric SyDS whose graph $G(V, E)$ has n nodes and maximum node degree Δ . Algorithm ALG-MONOTONE-SEQ (Figure 2) produces a monotone complete query sequence M with at most $\min\{\Delta^2 + 2, n + 1\}$ queries.

Proof: We first show that Step 2 of the algorithm can indeed color G^2 using at most $\min\{\Delta^2 + 1, n\}$ colors. Since the maximum node degree in G is Δ , each node v of G has at most Δ neighbors and at most $\Delta(\Delta - 1)$ nodes at a distance of 2 from v . Thus, the maximum node degree in G^2 is at most $\Delta(\Delta - 1) + \Delta = \Delta^2$. Hence, by Theorem 1, G^2 can be colored using at most $\Delta^2 + 1$ colors. Since G^2 has n nodes, n colors are sufficient. Thus, G^2 can be colored with at most $k = \min\{\Delta^2 + 1, n\}$ colors. Hence, the number of queries in $M = k + 1$ is at most $\min\{\Delta^2 + 2, n + 1\}$.

We now argue that the query sequence $M = \langle q_0, q_1, \dots, q_k \rangle$ is monotone. Query q_0 is the bit vector with all 0’s. For any $j \geq 1$, query q_j sets all the nodes in color classes C_1 through C_j to 1 and the remaining nodes to 0. Thus, each node that is set to 1 in query q_j remains 1 in all the subsequent queries q_{j+1}, \dots, q_k . In other words, the sequence is monotone.

Thus, we are left with the proof that M is complete; that is, for each node v with degree α in G and each value ℓ , $0 \leq \ell \leq \alpha + 1$, there is a query q in M such that $\text{score}(q, v_i) = \ell$. Query q_0 ensures that $\text{score}(q, v_i) = 0$. For the other values of ℓ , consider the closed neighborhood $N[v]$ of v in G . Note that $|N[v]| = \alpha + 1$. For each pair of nodes v_x and v_y in $N[v]$, there is a path consisting of at most two edges in G . Thus, the nodes in $N[v]$ form a clique in G^2 . In other words, each node in $N[v]$ must be in a different color class of G^2 . Let $C_{j_1}, C_{j_2}, \dots, C_{j_{\alpha+1}}$ denote the color classes of G^2 in which the nodes in $N[v]$ appear, and assume without loss of generality that $j_1 < j_2 < \dots < j_{\alpha+1}$. It is easy to see that for $1 \leq \ell \leq \alpha + 1$, query q_{j_ℓ} ensures that $\text{score}(q_{j_\ell}, v_i) = \ell$. This completes the proof of Theorem 2. ■

Generating Coloring from a Monotone Complete Query Sequence

Theorem 2 shows that a monotone complete query sequence can be constructed from the coloring of the graph G^2 . We now point out that this relationship is not accidental; indeed, from every monotone complete query sequence a valid coloring of G^2 can be generated.

Theorem 10 Let \mathcal{S} be a SyDS where each local function is symmetric. Let $G(V, E)$ be the underlying graph of \mathcal{S} . Suppose there is a monotone complete query sequence M with ℓ queries for \mathcal{S} . Then, G^2 can be colored using $\ell - 1$ colors.

Proof: Let $M = \langle q_1, q_2, \dots, q_\ell \rangle$ denote the given monotone complete query sequence for \mathcal{S} . Let C_i denote the set of nodes of G which have the value 0 in q_i and the value 1 in q_{i+1} , $1 \leq i \leq \ell - 1$. Assign color i to all the nodes in C_i , $1 \leq i \leq \ell - 1$. We now prove that this scheme assigns a color to each node and that this is a valid coloring of G^2 .

First, we prove that each node is assigned a color. To see this, note that since M is a monotone complete query sequence, query q_1 has all its bits set to 0 and q_ℓ has all its bits set to 1. Therefore, for each node v , there is an index r such that the value assigned to v in q_r is 0 and that in q_{r+1} is 1. Thus, v appears in set C_r and receives color r . The monotonicity of M ensures that v remains 1 in queries q_{r+1} through q_ℓ . In other words, the color assigned to v does not change subsequently.

We now prove by contradiction that the above method produces a valid coloring of G^2 . So, suppose that v_i and v_j are two nodes which receive the same color, say k , but G^2 has the edge $\{v_i, v_j\}$. By our coloring scheme, both v_i and v_j had the value 0 in q_k and the value 1 in q_{k+1} . There are two cases to consider.

Case 1: The edge $\{v_i, v_j\}$ is in G .

Note that both v_i and v_j have color k . Let $\text{score}(v_i, q_k) = \alpha$. Since both v_i and v_j changed from 0 in q_k to 1 in q_{k+1} and $\{v_i, v_j\}$ is an edge in G , $\text{score}(v_i, q_{k+1}) \geq \alpha + 2$. Because M is monotone, none of the other queries in M provides a score of $\alpha + 1$ to v_i . This contradicts the assumption that M is complete.

Case 2: The edge $\{v_i, v_j\}$ is not in G but in G^2 .

In this case, there is a node v_x such that the edges $\{v_i, v_x\}$ and $\{v_j, v_x\}$ are in G . Let $\text{score}(v_x, q_k) = \beta$. Since both v_i and v_j changed from 0 to 1 in q_{k+1} and both $\{v_i, v_x\}$ and $\{v_j, v_x\}$ are edges in G , $\text{score}(v_x, q_{k+1}) \geq \beta + 2$. Because M is monotone, none of the other queries in M provides a score of $\beta + 1$ to v_x . Again, this contradicts the assumption that M is complete, and this completes the proof of Theorem 10. ■

Generating a Complete Query Set with High Probability

Here, we present a proof of Theorem 3. A statement of the theorem is given below.

Statement of Theorem 3: Let \mathcal{S} be a symmetric SyDS with graph $G(V, E)$ where $|V| = n$ and maximum node degree $= \Delta$. A query set Q of size $O(\Delta^{1.5} \log(n))$ which is complete with probability at least $(1 - \frac{1}{n})$ can be constructed for \mathcal{S} .

In proving the above theorem, we will use the following notation. For any node v , let f_v denote the symmetric function at v and let $d(v)$ denote the degree of v . Note that each input to f_v is an integer that gives the number of 1's assigned to the closed neighborhood of v .

Proof of Theorem 3: Our method, discussed below, produces a query set with size at most $22\Delta\sqrt{\Delta+2}\log(n\Delta)$.

We first note that the all zeros and the all ones configurations can be used to query $f_v(0)$ and $f_v(d(v) + 1)$, respectively for all $v \in V$. This contributes the additive term 2. For a real number x , let $\langle x \rangle$ denote $\lfloor x + .5 \rfloor$, the integer closest to x . Let $Q = \{q_{ij} \sim \mathbb{D}(\frac{i}{\Delta+1}) \mid 1 \leq i \leq \Delta, 1 \leq j \leq 22\sqrt{\Delta+2}\log(n\Delta)\}$ be the query set. For any $v \in V, b \in \{1, \dots, d(v)\}$ and $q \sim \mathbb{D}(z)$, where $z = \left\langle \frac{b(\Delta+1)}{d(v)+1} \right\rangle$, we have $\Pr(f_v(b) \text{ is queried in } q) = \Pr(d_1[v, q] = b)$. Now, letting $p' = \Pr(d_1[v, q] = b)$, we have

$$\begin{aligned} p' &\geq \binom{d(v)+1}{b} \left(\frac{z}{\Delta+1} \right)^b \left(1 - \frac{z}{\Delta+1} \right)^{d(v)+1-b} \\ &\geq \frac{1}{11\sqrt{d(v)+2}} \\ &\geq \frac{1}{11\sqrt{\Delta+2}}, \end{aligned}$$

where the second inequality follows from Lemma 1 (below). Now, $\Pr(f_v(b) \text{ is not queried by } Q) \leq \Pr(f_v(b) \text{ is not queried by } q_{zj}, 1 \leq j \leq 22\sqrt{\Delta+2}\log(n\Delta))$.

The latter quantity is $\leq \left(1 - \frac{1}{11\sqrt{\Delta+2}}\right)^{22\sqrt{\Delta+2}\log(n\Delta)} < \frac{1}{(n\Delta)^2}$. By union bound, $\Pr(Q \text{ is not a complete set}) = \Pr(\exists v, b \text{ such that } f_v(b) \text{ is not queried by } Q)$. Now, the latter quantity is $\leq \sum_{v \in V} \sum_{b=1}^{\Delta} \Pr(f_v(b) \text{ is not queried by } Q) < \frac{1}{n\Delta}$. This completes the proof. ■

Lemma 1 For any three positive integers $b \leq d \leq D$ and $z = \left\langle \frac{bD}{d} \right\rangle$, $\binom{d}{b} \left(\frac{z}{D} \right)^b \left(1 - \frac{z}{D} \right)^{d-b} \geq \frac{1}{11\sqrt{d+1}}$.

We will first prove the following claims.

Claim 1 $\left(1 + \frac{1}{b}\right)^b$ is monotone increasing in b for positive integers.

Proof: Consider the collection of $(b+1)$ numbers $(1, \frac{b+1}{b}, \dots, \frac{b+1}{b})$. Using the fact that their arithmetic mean is \geq their geometric mean,

$$\begin{aligned} \frac{1+b(\frac{b+1}{b})}{b+1} &\geq \left(1^1 \left(\frac{b+1}{b} \right)^b \right)^{\frac{1}{b+1}} \\ \frac{(b+1)+1}{b+1} &\geq \left(\frac{b+1}{b} \right)^{\frac{b}{b+1}}. \end{aligned}$$

Claim 2 $\binom{d}{b} \left(\frac{b}{d} \right)^b \left(1 - \frac{b}{d} \right)^{d-b} \geq \frac{1}{\sqrt{2(d+1)}}$. ■

Proof: Let $h(b, d) = \binom{d}{b} \left(\frac{b}{d} \right)^b \left(1 - \frac{b}{d} \right)^{d-b}$. We will first show that for $b < \frac{d}{2}$, $h(b+1, d) \leq h(b, d)$ and for $b \geq \frac{d}{2}$, $h(b+1, d) > h(b, d)$, and hence, $h(\cdot)$ attains a minimum value at $b = \lfloor \frac{d}{2} \rfloor$.

$$\begin{aligned} \frac{h(b+1, d)}{h(b, d)} &= \frac{\binom{d}{b+1} \left(\frac{b+1}{d} \right)^{b+1} \left(1 - \frac{b+1}{d} \right)^{d-b-1}}{\binom{d}{b} \left(\frac{b}{d} \right)^b \left(1 - \frac{b}{d} \right)^{d-b}} \\ &= \frac{d-b}{b+1} \left(\frac{b+1}{b} \right)^b \frac{b+1}{d} \left(\frac{d-b-1}{d-b} \right)^{d-b} \frac{d}{d-b-1} \\ &= \left(\frac{b+1}{b} \right)^b \left(\frac{d-b-1}{d-b} \right)^{d-b-1} = \left(\frac{b+1}{b} \right)^b \left(\frac{b'}{b'+1} \right)^{b'}, \end{aligned}$$

where, $b' = d - b - 1$. When $b < \frac{d}{2}$, $b' \geq b$ and when $b \geq \frac{d}{2}$, $b' < b$. Applying Claim 1, we have for $b < \frac{d}{2}$, $h(b+1, d) \leq h(b, d)$ and for $b \geq \frac{d}{2}$, $h(b+1, d) > h(b, d)$.

When b is even, $h\left(\frac{d}{2}, d\right) \geq \frac{1}{\sqrt{2d}}$. Now we will show that when b is odd, $h\left(\frac{d-1}{2}, d\right) \geq \frac{1}{\sqrt{2(d+1)}}$. Let $b = 2k + 1$.

$$\begin{aligned} \frac{h(k, 2k+1)}{h(k, 2k+2)} &= \frac{\binom{2k+1}{k}}{\binom{2k+2}{k}} \frac{\left(\frac{k}{2k+1}\right)^k \left(1 - \frac{k}{2k+1}\right)^{k+1}}{\left(\frac{k}{2k+2}\right)^k \left(1 - \frac{k}{2k+2}\right)^{k+2}} \\ &= \frac{k+2}{2k+2} \left(\frac{2k+2}{2k+1}\right)^k \left(\frac{k+1}{k+2}\right)^{k+1} \left(\frac{2k+2}{2k+1}\right)^{k+1} \frac{2k+2}{k+2} \\ &= \left(1 + \frac{1}{2k+1}\right)^{2k+1} \left(1 + \frac{1}{k+1}\right)^{-(k+1)} > 1. \end{aligned}$$

The inequality follows from Claim 1. Therefore, when d is odd, $h\left(\frac{d-1}{2}, d\right) > h\left(\frac{d-1}{2}, d+1\right) \geq h\left(\frac{d+1}{2}, d+1\right) \geq \frac{1}{\sqrt{2(d+1)}}.$ ■

Claim 3 For any positive $x \leq \frac{1}{2}$, $1 - x \geq e^{-2x}$.

Proof: $e^{2x}(1-x) > (1+2x)(1-x) = 1+x(1-2x) \geq 1.$ ■

Proof of Lemma 1: We have two cases to consider: (a) $z \leq \frac{bD}{d}$ and (b) $z > \frac{bD}{d}$. But first we note that by definition, $|z - \frac{bD}{d}| \leq \frac{1}{2}$.

Case (a). $\frac{bD}{d} - \frac{1}{2} \leq z \leq \frac{bD}{d}$.

$$\begin{aligned} \binom{d}{b} \left(\frac{z}{D}\right)^b \left(1 - \frac{z}{D}\right)^{d-b} &\geq \binom{d}{b} \left(\frac{z}{D}\right)^b \left(1 - \frac{b}{d}\right)^{d-b} \\ &\geq \binom{d}{b} \left(\frac{b}{d} - \frac{1}{2D}\right)^b \left(1 - \frac{b}{d}\right)^{d-b} \\ &\geq \binom{d}{b} \left(\frac{b}{d}\right)^b \left(1 - \frac{b}{d}\right)^{d-b} \left(1 - \frac{d}{2bD}\right)^b \\ &\geq \frac{1}{2\sqrt{d}} \left(1 - \frac{d}{2bD}\right)^b \geq \frac{1}{e^2 \sqrt{2(d+1)}} \geq \frac{1}{11\sqrt{d+1}}. \end{aligned}$$

The last but one inequality follows from Claim 3.

Case (b). $\frac{bD}{d} \leq z \leq \frac{bD}{d} + \frac{1}{2}$.

$$\begin{aligned} \binom{d}{b} \left(\frac{z}{D}\right)^b \left(1 - \frac{z}{D}\right)^{d-b} &\geq \binom{d}{b} \left(\frac{b}{d}\right)^b \left(1 - \frac{z}{D}\right)^{d-b} \\ &\geq \binom{d}{b} \left(\frac{b}{d}\right)^b \left(1 - \frac{b}{d} - \frac{1}{2D}\right)^{d-b} \\ &\geq \binom{d}{b} \left(\frac{b}{d}\right)^b \left(1 - \frac{b}{d}\right)^{d-b} \left(1 - \frac{\frac{1}{2D}}{1 - \frac{b}{d}}\right)^{d-b} \\ &\geq \frac{1}{\sqrt{2(d+1)}} \left(1 - \frac{d}{2(d-b)D}\right)^{d-b} > \frac{1}{11\sqrt{d+1}}. \end{aligned}$$

This completes the proof. ■

Lower Bounds on Query Set Sizes

Statement of Proposition 1: Let \mathcal{S} be a symmetric SyDS where the underlying graph $G(V, E)$ has a maximum node degree Δ . Under the batch query model, every complete query set must contain least $\Delta + 2$ queries.

Proof: Under the batch mode, suppose a complete query set Q has less than $\Delta + 2$ queries. Since G has a node v of degree Δ , the number of 1's in the input to the symmetric function f_v at v varies from 0 to $\Delta + 1$, a total of $\Delta + 2$ values. Thus, there is

at least one value k such that none of the queries in Q has a score of k with respect to v . Hence, the query set cannot correctly determine the value of the function f_v when the number of 1's in the input to f_v is exactly ks . This contradicts the assumption that Q is a complete query set. The proposition follows. ■

Statement of Theorem 6: For every $n \geq 1$, there is a threshold SyDS whose underlying graph is a clique on n nodes such that at least $n + 1$ queries are necessary under the adaptive query mode to correctly identify all the threshold values.

Proof: Consider a threshold SyDS \mathcal{S} whose underlying graph $G(V, E)$ is a clique on n nodes. Let $V = \{v_1, v_2, \dots, v_n\}$. We will tentatively choose the threshold of node v_i to be i to answer queries under the adaptive model, $1 \leq i \leq n$. We will show that if the total number of queries is less than $n + 1$, the answers to the queries cannot distinguish between this tentative assignment and a slightly different assignment of threshold values.

Suppose Q be a sequence of queries under the adaptive model, with $|Q| \leq n$. We generate the responses to the queries using the chosen tentative assignment of threshold values to nodes. The threshold of any node of \mathcal{S} is in the range 0 through $n + 1$ (where the threshold value $n + 1$ indicates the function which is zero for every input). Since G is a clique, the closed neighborhood of each node is the node set V . Thus, each query $q \in Q$ provides the same score to each node of G . There are $n + 1$ scores in the range 0 through n . Since Q has at most n queries, there is at least one value k , $0 \leq k \leq n$ such that none of the queries in Q provides the score k . We have three cases depending on the value of k .

Case 1: $k = 0$. In this case, from the responses to the queries in Q , one cannot distinguish between the case where the threshold of node v_1 is 0 and the case where the threshold of v_1 is 1. (In both cases, the new state of v_1 is 1 in the response to each query in Q .)

Case 2: $k = n$. In this case, from the responses to the queries in Q , one cannot distinguish between the case where the threshold of node v_n is n and the case where the threshold of v_n is $n + 1$. (In both cases, the new state of v_n is 0 in the response to each query in Q .)

Case 3: $1 \leq k \leq n - 1$. In this case, from the responses to the queries in Q , one cannot distinguish between the case where the threshold of node v_k is k and the case where the threshold of v_k is $k + 1$. (In both cases, the responses have the following property. For any query $q \in Q$ where $\text{score}(q, v_k) \leq k - 1$, the new state of v_k is 0 in the response. For any query $q \in Q$ where $\text{score}(q, v_k) \leq k + 1$, the new state of v_k is 1 in the response.)

Thus, under the adaptive model, a query set with n or fewer queries cannot correctly identify all the thresholds for the chosen SyDS. This completes the proof of Theorem 6. ■

Complexity of Generating Small Monotone Complete Query Sequences

Statement of Theorem 7: Problem SMCQS is **NP**-complete.

Proof: It is easy to see that SMCQS is in **NP**. To prove **NP**-hardness, we use a reduction from the **Distance-2 Coloring** (D2C) problem defined as follows: given an undirected graph $G(V, E)$ and an integer r , is G^2 r -colorable? It is known that D2C is **NP**-complete (McCormick 1983).

The reduction is straightforward. Given an instance of the D2C problem consisting of graph G and integer r , we obtain an instance of the SMCQS problem where the graph is G itself and the length k of the query sequence is $r + 1$. It was shown in the proof of Theorem 2 that when G^2 is r -colorable, there is a monotone complete query sequence with $k = r + 1$ queries. Also, it was shown in the proof of Theorem 10 that from any monotone complete query sequence of length $r + 1$, one can obtain a valid coloring of G^2 with r colors. Thus, there is a solution to the SMCQS problem iff there is a solution to the D2C problem and this completes the proof. ■

Complexity of Query Set Compaction

Statement of Theorem 8: (a) The problem QSC is **NP**-complete even when the underlying graph has no edges. (b) Unless **P** = **NP**, QSC cannot be approximated to within the factor $o(\log n)$, where n is the number of nodes in the underlying graph of the SyDS.

Our proof of this result relies on known results for the **Minimum Set Cover** (MSC) problem which is defined as follows: given a base set $X = \{x_1, x_2, \dots, x_n\}$, a collection $Y = \{Y_1, Y_2, \dots, Y_m\}$, where each Y_j is a subset of X , $1 \leq j \leq m$, and an integer $\alpha \leq m$, is there a subcollection $Y' \subseteq Y$ such that (i) $|Y'| \leq \alpha$ and (ii) the union of all the sets in Y' is equal to X ? It is well known that MSC is **NP**-complete (Garey and Johnson 1979) and that unless **P** = **NP**, it cannot be approximated to within the factor $o(\log(n))$, where n is the size of the base set (Vazirani 2001).

Proof:

Part (a): It is easy to see that QSC is in **NP**. We prove **NP**-hardness through a reduction from MSC. Let the given instance of MSC consist of base set $X = \{x_1, x_2, \dots, x_n\}$, collection $Y = \{Y_1, Y_2, \dots, Y_m\}$ of nonempty subsets of X and integer $\alpha \leq m$. Without loss of generality, we may assume that each element of X appears in some subset in Y ; otherwise, there is no

solution to the MSC instance. We will construct the underlying graph $G(V, E)$ of a SyDS \mathcal{S} and a complete query set Q for \mathcal{S} as follows.

1. The node set V of G is given by $V = V_1 \cup V_2$, where $V_1 = \{v_1, v_2, \dots, v_n\}$ is in one-to-one correspondence with the base set $X = \{x_1, x_2, \dots, x_n\}$ of the MSC instance and $V_2 = \{v_{n+1}\}$ consists of just one node. (Thus, V has a total of $n + 1$ nodes.)
2. The edge set E of G is empty; that is, the degree of each node is 0. Thus, for each node $v \in V$, the number of 1's in the input to the local function f_v at v can only be either 0 or 1.
3. The query set Q consists of $m + 1$ queries (where $m = |Y|$) constructed as discussed below. Note that each query is an $(n + 1)$ -bit vector, where the i^{th} bit specifies the value of node v_i , $1 \leq i \leq n + 1$.
 - (a) For each $Y_j \in Y$, $1 \leq j \leq m$, Q contains a query q_j constructed as follows. Let $Y_j = \{x_{j_1}, x_{j_2}, \dots, x_{j_r}\}$. Then, in query q_j , the bits corresponding to the nodes $v_{j_1}, v_{j_2}, \dots, v_{j_r}$ are all 1 and the other bits are 0.
 - (b) We add one more query q_{m+1} to Q ; in query q_{m+1} , bits 1 through n are set to 0 and bit $n + 1$ is set to 1.
4. The upper bound on the size of the required subset Q' of queries is set to $\alpha + 1$.

This completes the construction of the QSC instance. It can be seen that the construction can be carried out in polynomial time. We now show that Q is a complete query set for \mathcal{S} .

Claim 1: The query set Q constructed above is a complete query set for \mathcal{S} .

Proof of Claim 1: We must show that for each node $v_i \in V$, Q contains two queries, say q_{i_0} and q_{i_1} , such that $\text{score}(q_{i_0}, v_i) = 0$ and that $\text{score}(q_{i_1}, v_i) = 1$. First, consider any node v_i , where $1 \leq i \leq n$. Query q_{m+1} sets the value of v_i to 0; thus $\text{score}(q_{m+1}, v_i) = 0$. Suppose the element x_i (corresponding to node v_i) appears in subset Y_j . By our construction, query q_j sets the value of v_i to 1; thus, $\text{score}(q_j, v_i) = 1$. For node v_{n+1} , each query q_j created from Y_j sets the value of v_{n+1} to 0; that is, $\text{score}(q_j, v_{n+1}) = 0$; Also, query q_{m+1} sets the value of v_{n+1} to 1; thus, $\text{score}(q_{m+1}, v_{n+1}) = 1$. The claim follows. \square

We now prove that there is a solution to the QSC instance if and only if there is a solution to the MSC instance.

Part 1: Suppose there is a solution Y' to the MSC instance consisting of sets $Y_{j_1}, Y_{j_2}, \dots, Y_{j_\ell}$, for some $\ell \leq \alpha$. Consider the query set $Q' = \{q_{j_1}, q_{j_2}, \dots, q_{j_\ell}, q_{m+1}\}$, which includes the queries corresponding to the sets in Y' along with query q_{m+1} . Note that $Q' \subseteq Q$. Also, since $\ell \leq \alpha$, $|Q'| \leq \alpha + 1$. Thus, we only need to show that Q' is a complete query set. Consider any node v_i , where $1 \leq i \leq n$. Query q_{m+1} sets the value of v_i to 0; thus, $\text{score}(q_{m+1}, v_i) = 0$. Further, Since Y' is a set cover, the element x_i (corresponding to node v_i) appears in some subset $Y_{j_z} \in Y'$. By our construction, in query q_{j_z} , the value of v_i is 1; thus, $\text{score}(q_{j_z}, v_i) = 1$. For node v_{n+1} , each query $q \in Q' - \{q_{m+1}\}$ sets the value of v_{n+1} to 0; thus, $\text{score}(q, v_{n+1}) = 0$. Further, query q_{m+1} sets the value of v_{n+1} to 1; in other words, $\text{score}(q_{m+1}, v_{n+1}) = 1$. Hence, Q' is a complete query set.

Part 2: Let Q' be a solution to the QSC instance with $|Q'| \leq \alpha + 1$. We claim that $q_{m+1} \in Q'$. This is because q_{m+1} is the only query for which $\text{score}(v_{n+1}) = 1$. Define $Q'' = Q' - \{q_{m+1}\}$. Let $|Q''| = \ell$ and note that $\ell \leq \alpha$. Further, let $Q'' = \{q_{j_1}, q_{j_2}, \dots, q_{j_\ell}\}$. Consider the following subcollection Y' of Y given by $Y' = \{Y_{j_1}, Y_{j_2}, \dots, Y_{j_\ell}\}$. We now show that Y' is a solution to the MSC instance. To see this consider any element $x_i \in X$, where $1 \leq i \leq n$. Query $q_{m+1} \in Q'$ sets node v_i to 0. Since Q' is a complete query set, some query $q_{j_z} \in Q''$ must set v_i to 1. By our construction, the subset Y_{j_z} contains x_i . Thus, Y' is a set cover. Since $|Y'| \leq \alpha$, Y' is a solution to the MSC instance, and this completes the **NP-hardness proof**.

We use the same reduction to prove the non-approximability result. Suppose \mathcal{A} is an approximation algorithm that provides a performance guarantee of $\rho = o(\log n)$ for the QSC problem, where n is the number of nodes in the underlying graph of the SyDS. We will show that \mathcal{A} can be used to construct an $2\rho = o(\log n)$ approximation algorithm for the MSC problem, contradicting the known non-approximability result for MSC. Towards this proof, consider any instance of the MSC optimization problem. Let $\text{OPT}(\text{MSC})$ denote the value of an optimal solution (i.e., the minimum size of a set cover) for the MSC instance and let $\text{OPT}(\text{QSC})$ denote the value of an optimal solution (i.e., the minimum size of a complete query subset) for the QSC instance constructed from the MSC instance. In the **NP-hardness proof**, we showed that

$$\text{OPT}(\text{QSC}) \leq \text{OPT}(\text{MSC}) + 1. \quad (1)$$

Suppose we run Algorithm \mathcal{A} on the resulting QSC instance. Since \mathcal{A} provides a ρ approximation, the solution produced by \mathcal{A} has at most $\rho \text{OPT}(\text{QSC})$ queries. From this query set, it was shown in the **NP-hardness proof** that a solution to the MSC instance with $\rho \text{OPT}(\text{QSC}) - 1$ subsets can be constructed. Letting $\text{APPROX}(\text{MSC})$ denote the resulting number of subsets, we have

$$\begin{aligned} \text{APPROX}(\text{MSC}) &\leq \rho \text{OPT}(\text{QSC}) - 1 \\ &\leq \rho [\text{OPT}(\text{MSC}) + 1] - 1 \quad (\text{using Equation (1)}) \\ &\leq 2\rho \text{OPT}(\text{MSC}) \quad (\text{since } \text{OPT}(\text{MSC}) \geq 1). \end{aligned}$$

Thus, if \mathcal{A} provides a performance guarantee of $\rho = o(\log n)$ for the QSC problem, then there is a $2\rho = o(\log n)$ approximation algorithm for the MSC problem as well. This completes the proof of Theorem 8. ■

Statement of Theorem 9: Algorithm Approx-QSC provides a performance guarantee of $O(\log n)$ for the QSC problem, where n is the number of nodes in the underlying graph of the SyDS.

To establish the above theorem, we need the following lemma.

Lemma 2 *The reduction from QSC to MSC used in Steps 1 and 2 of Approx-QSC (Figure 3) produces an instance of MSC such that any solution with r subsets to the MSC instance is a solution with r queries to the QSC instance and vice versa.*

Proof: First, consider any solution $Y' = \{Y_{j_1}, Y_{j_2}, \dots, Y_{j_r}\}$ with r subsets to the MSC instance. Let $Q' = \{q_{j_1}, q_{j_2}, \dots, q_{j_r}\}$ be the corresponding query set with r queries. We need to show that Q' is a complete query set; that is, for any node v_i ($1 \leq i \leq n$) and any integer k , $0 \leq k \leq d_i$, there is a query $q \in Q'$ such that the number of 1's in the input to the local function f_i at v_i due to q is exactly k . To see this, note that Y' is a set cover. Thus, there is a set $Y_{j_z} \in Y'$ such that the element $a_{ik} \in X$ appears in Y_{j_z} . By our construction, a_{ik} was added to Y_{j_z} because query q_{j_z} provides exactly k 1's to the local function f_i at v_i .

To prove the converse, let $Q' = \{q_{j_1}, q_{j_2}, \dots, q_{j_r}\}$ be a solution with r queries to the QSC instance. Consider the collection Y' of sets given by $Y' = \{Y_{j_1}, Y_{j_2}, \dots, Y_{j_r}\}$. We claim that Y' is a solution to the MSC instance. To see this, consider any element $a_{ik} \in X$. Since Q' is a complete query set, there is some query $q_{j_z} \in Q'$ such that the number of 1's in the input to the function f_i at node v_i due to q_{j_z} is exactly k . By our construction, set Y_{j_z} contains the element a_{ik} . In other words, Y' is a solution to the MSC instance with r sets. ■

The following is an immediate consequence of the above lemma.

Observation 2 *Let $OPT(QSC)$ denote the size of an optimal query set for a given QSC instance and let $OPT(MSC)$ denote the size of an optimal solution to the MSC instance obtained at the end of Step 2 of Approx-QSC. Then, $OPT(QSC) = OPT(MSC)$.* ■

We can now prove Theorem 9.

Proof of Theorem 9: Let $OPT(QSC)$ denote the size of an optimal query set for the QSC instance and let $OPT(MSC)$ denote the size of an optimal solution to the MSC instance. As mentioned earlier, the size of the base set X is $2|E| + n$. Since $|E| < n^2$, we note that $|X| < 3n^2$. The greedy algorithm for MSC provides a performance guarantee of $O(\log |X|)$. Since $|X| < 3n^2$, this performance guarantee is $O(\log n)$. Thus, the approximation algorithm for MSC produces a solution with at most $O(\log n)$ $OPT(MSC)$ sets. By Lemma 2, any solution to MSC with r subsets leads to a solution with r queries for QSC. Thus, the size of the resulting query set is at most $O(\log n)$ $OPT(MSC)$ which is equal to $O(\log n)$ $OPT(QSC)$ by Observation 2. Thus, Approx-QSC has a performance guarantee of $O(\log n)$. ■

Statement of Theorem 4: For any threshold SyDS whose underlying graph G belongs to the class of (α, β) -simple graphs, $O(\log n)$ queries are sufficient in the adaptive mode to identify all the threshold values.

Proof: We give an approach that uses $\alpha \lceil \log n \rceil + \beta^2 \lceil \log \beta \rceil$ queries under the adaptive model. Since α and β are constants independent of n , the number of queries is $O(\log n)$.

The idea is to first use a separate binary search for each of the α nodes of degree q ; this uses at most $\alpha \lceil \log n \rceil$ queries. Let V' denote the subset of nodes of G such that each node in V' has a degree of at most β . Now, if we construct G^2 , it can be seen that the subgraph G' of G^2 on V' , it can be seen that the maximum node degree in G' is at most β^2 . Therefore, G' can be colored using at most β^2 colors and a binary search can be done simultaneously for all the nodes in each color class. This will use at most $\beta^2 \lceil \log \beta \rceil$ queries, giving a total of $\alpha \lceil \log n \rceil + \beta^2 \lceil \log \beta \rceil$ queries. ■

Statement of Theorem 5: For a threshold SyDS whose underlying graph $G(V, E)$ is scale-free with exponent $\gamma \geq 1$, the thresholds can be found using $O(n^{\frac{2}{\gamma+1}})$ queries under the adaptive query mode.

Proof: Let $V_1 \subseteq V$ correspond to the set of nodes with degree at most β and V_2 be the remaining set. By Theorem 4, the number of queries required is at most $g(\beta) := |V_2| \log n + \beta^2 \log \beta$. Since $|V_2|$ is the number of nodes with degree $\geq \beta$, it is at most

$$\sum_{x=\beta}^{\Delta} c' \frac{n}{x^\gamma} \leq c'' \int_{\beta}^{\infty} \frac{n}{x^\gamma} dx = c \frac{n}{\beta^{\gamma-1}},$$

for some constants c , c' and c'' . Therefore, $g(\beta) = c \frac{n}{\beta^{\gamma-1}} \log n + \beta^2 \log \beta$. For $\beta > 0$, $g(\beta)$ is a convex function. Equating its first derivative to 0 and rearranging, we note that $g(\beta)$ attains a minimum value for β satisfying $\beta^{\gamma+1} \log \beta = \Theta(n \log n)$ or $\beta = \Theta(n^{\frac{1}{\gamma+1}})$. For this value, $g(\beta) = O(n^{\frac{2}{\gamma+1}})$. ■

Why Query Sets Generated by Method I are Difficult to Compress: As mentioned in the paper, our experimental results show that query sets generated by coloring G^2 are already compact; that is, i.e., no subset of these query sets can be complete. This can be explained intuitively as follows. If this set is not compact, there exists at least one query which is not required. We recall that by construction, the query set can be arranged as a monotone increasing sequence q_0, q_1, \dots , such that in query q_i all nodes of color i are set to switched to state 1. Suppose query q_k is not required for the set to be complete. Then, using the arguments in the proof of Theorem 2, we can show that if all vertices of color k were assigned color $k+1$, the coloring would still be valid. This means that in the greedy strategy, all nodes colored $k+1$ could actually have been assigned color k . Since the greedy strategy always gives the minimum color available to nodes, this situation cannot occur.

Method 3: Description of Our Adaptive Algorithm

Algorithm 1: Greedy heuristic to infer the thresholds.

Data: Network $G(V, E)$ and thresholds t_v for every node v .
Result: Complete query set Q

```

1 for  $v$  in  $V$  do
2   | Let  $t_L(v) = 0$  and  $t_H(v) = d(v) + 2$ ;
3 end
4 Let  $V_t = V$  denote the set of nodes for which threshold needs to be inferred;
5 Let  $Q = \emptyset$  be the query set;
6 while  $V_t \neq \emptyset$  do
7   | Let  $i = 1$ ;
8   | Let  $q_i$  be the current query. Let  $q_i[v] = 0$ ,  $\forall v \in V$ ;
9   | Let  $V_{\text{rem}} = V_t$ ;
10  | while  $V_{\text{rem}} \neq \emptyset$  do
11    |   | Let  $v_{\text{max}} = \arg \max_{v \in V_{\text{rem}}} t_H(v) - t_L(v)$ ;
12    |   | Set exactly  $\lfloor (t_H(v_{\text{max}}) - t_L(v_{\text{max}}))/2 \rfloor$  neighbors of  $v_{\text{max}}$  to state 1 in  $q_i$ ;
13    |   |  $V_{\text{rem}} \leftarrow V_{\text{rem}} \setminus \{N[v_{\text{max}}, G^2]\}$ ;
14  | end
15  | Compute the successor  $s$  of  $q_i$ ;
16  | //Update  $t_L(v)$  and  $t_H(v)$  for all  $v \in V_t$ //
17  | for  $v \in V_t$  do
18    |   | if  $s[v] = 0$  and  $t_L(v) \leq \text{score}(q_i, v)$  then
19    |   |   |  $t_L(v) \leftarrow \text{score}(q_i, v) + 1$ ;
20    |   | end
21    |   | else if  $s[v] = 1$  and  $t_H(v) > \text{score}(q_i, v)$  then
22    |   |   |  $t_H(v) \leftarrow \text{score}(q_i, v)$ ;
23    |   | end
24  | end
25  | Remove all nodes  $v$  from  $V_t$  such that  $t_L(v) = t_H(v)$ ;
26  |  $Q \leftarrow Q \cup \{q_i\}$ ;
27 end

```



A k -Median Based Online Algorithm for the Stochastic k -Server Problem

Abhijin Adiga¹ , Alexander D. Friedman², and Sharath Raghvendra³

¹ Biocomplexity Institute of Virginia Tech, Blacksburg, USA

abhijin@bi.vt.edu

² Department of Mathematics, Virginia Tech, Blacksburg, USA

adfriedm@vt.edu

³ Department of Computer Science, Virginia Tech, Blacksburg, USA

sharathr@vt.edu

Abstract. We consider the k -server problem in the random arrival model and present a simple k -median based algorithm for this problem. Let $\sigma = \langle r_1, \dots, r_n \rangle$ be the sequence of requests. Our algorithm will batch the requests into $\log n$ groups where the $(i+1)^{st}$ group contains requests $\langle r_{2^i+1}, \dots, r_{2^{i+1}} \rangle$. To process the requests of group $i+1$, our algorithm will place the k servers at the k -median centers of the first 2^i requests. When a new request of this group arrives, the algorithm will simply assign the server associated with the nearest k -median center to serve it. We show that this simple algorithm, in the random arrival model, has a competitive ratio of at most $O(\alpha)$ and an additive cost of $O(\Delta k \log n)$, where Δ is the diameter of the requests and α is a lower bound on the competitive ratio of any online algorithm in this model.

For our analysis, we use the following fact: In the random arrival model, the expected cost of serving the next request is minimized when servers are located at the k -median of the requests that have not yet arrived (unprocessed requests). But our algorithm instead uses the k -median of the requests seen so far as a proxy. Using existing analysis techniques, we obtain only a large bound on the difference between k -median of the unprocessed requests and that of the processed ones. In particular, in addition to $O(\alpha)$ times the optimal cost, for some $\epsilon > 0$, existing analysis techniques will also give an additive cost of ϵn for serving n requests. We present a new analysis to show that when the number of processed and unprocessed requests are of comparable sizes, the cost of serving n requests incurs only an additive cost of $k\Delta$ (independent of n and significantly better than the previous methods). We then apply this bound for serving each of the $\log n$ groups and obtain an overall bound which is $O(\alpha)$ times the optimal cost with an additive error of $O(k\Delta \log n)$.

Keywords: k -server problem · Random arrival model
 k -median clustering

A. D. Friedman and S. Raghvendra were supported under grant NSF-CCF 1464276.

A. Adiga was supported by the DTRA CNIMS Contract HDTRA1-11-D-0016-0001.

© Springer International Publishing AG, part of Springer Nature 2018

R. Solis-Oba and R. Fleischer (Eds.): WAOA 2017, LNCS 10787, pp. 176–189, 2018.

https://doi.org/10.1007/978-3-319-89441-6_14

1 Introduction

In this era of instant gratification, consumers demand access to goods and services in real-time. Business ventures, therefore, have to schedule their delivery of goods and services often without the complete knowledge of the future requests or their order of arrival. To assist with this, we need robust and competitive *online algorithms* that immediately and irrevocably allocate resources to requests in real-time with minimal cost.

In this paper, we consider the celebrated *k -server problem*. Initially, we are given a set of locations \mathcal{K} for the k servers. These are locations in a discrete metric space. Requests are generated in the same metric space by an adversary and are revealed one at a time. When a request $r \in \sigma$ is revealed, we have to immediately move one of the k servers to serve this request and incur a cost equal to the distance between their locations. The objective is to design an algorithm that allocates servers to requests that is competitive with respect to the minimum-cost offline solution.

For any algorithm \mathcal{A} , initial configuration of servers \mathcal{K} , and a sequence of requests σ , let $w_{\mathcal{A}}(\sigma, \mathcal{K})$ be the cost incurred when the algorithm \mathcal{A} assigns the requests to servers. Let $w_{\text{OPT}}(\sigma, \mathcal{K})$ be the smallest cost solution generated by an offline algorithm that has complete knowledge of the request sequence σ and assigns servers to requests based on their arrival order. We say that \mathcal{A} is α -competitive if, for a constant $\Phi_0 \geq 0$, the cost incurred by our algorithm satisfies,

$$w_{\mathcal{A}}(\sigma, \mathcal{K}) \leq \alpha w_{\text{OPT}}(\sigma, \mathcal{K}) + \Phi_0$$

for any request set and their arrival order.

In the *adversarial model*, there is an adversary who knows the server locations and the assignments made by the algorithm and generates a sequence to maximize α . In the *random arrival model* [1], the adversary chooses the locations of the requests in σ before the algorithm executes but their arrival order is a permutation chosen uniformly at random from the set of all possible permutations of the requests. In practical situations, it may be useful to assume that the requests are arriving i.i.d. from a known or an unknown distribution \mathcal{D} . Under these (known and unknown) models, the adversary is weaker than in the random arrival model and therefore, the competitive ratio for the random arrival model is an upper bound on its competitive ratio in the known and the unknown distribution models; see [2] for an algorithm in these models. We refer to the k -server problem under the known and unknown distribution as well as the random arrival model as the *stochastic k -server problem*.

For the stochastic k -server problem, the competitive ratio is expressed with respect to the expected costs. More specifically, \mathcal{A} is α -competitive if,

$$\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \leq \alpha \mathbb{E}[w_{\text{OPT}}(\sigma, \mathcal{K})] + \Phi_0.$$

Previous Work. The k -server problem is central to the theory of online algorithms. The problem was first posted by Manasse *et al.* [3]. In the adversarial

model, the best-known deterministic algorithm for this problem is the $(2k - 1)$ -competitive work function algorithm [4]. It is known that no deterministic algorithm can achieve a competitive ratio better than k and is conjectured that in fact there is a k -competitive algorithm for this problem. This conjecture is popularly called the *k -server conjecture*.

Bansal *et al.* [5] presented an $O(\log^{O(1)} n \log k)$ -competitive randomized algorithm for the k -server problem in the *oblivious adversary model*. This model is similar to the adversarial model; however, the adversary does not have any knowledge of the random choices made by the algorithm. There is also an online algorithm [6] for the closely related online metric matching problem. This algorithm is known to achieve an optimal competitive ratio of $2H_n - 1$ in the random arrival model. To the best of our knowledge, the stochastic k -server problem has not been studied before.

In the stochastic models, we can view any initial set of requests to be a sample that is chosen uniformly at random from σ . In this paper, using this sample, we approximate the k -median of the remaining requests leading to an improved algorithm for the stochastic k -server problem. Independent and uniform random samples have been used before in the context of sub-linear time algorithms for the k -median problem; see [7, 8]. In [7], it has been shown that a random sample of size $\tilde{O}(\Delta/\epsilon^2)$ can be used to approximate the k -median within a constant factor of the optimal k -median with an additional additive cost of ϵn . Meyerson *et al.* [8] show that if all the optimal k -median clusters are dense ($\geq \epsilon n/k$), then a very small random sample of size $\approx k/\epsilon$ can be used to approximate the k -median within a constant factor. In this paper, we will present and analyze a k -median based deterministic algorithm for the stochastic k -server problem.

Our Results. First, we present a simple algorithm which we refer to as the zoned algorithm for the k -server problem in the known distribution model, i.e., the request locations are i.i.d from a distribution \mathcal{D} on the discrete metric space. The zoned algorithm will associate one server to each of the centers of the k -median of \mathcal{D} and when a request arrives, the algorithm simply assigns the server associated with the closest k -median center to this request. The cost of serving any request is lower bounded by the average k -median cost of the distribution \mathcal{D} . Using triangle inequality, we can bound the cost incurred by zoned algorithm by twice the cost incurred by any optimal online algorithm for this problem.

Next, for the unknown distribution model and the random arrival model, we present an adaptive version of the zoned algorithm. Let $\sigma = \langle r_1, \dots, r_n \rangle$ be the request sequence. Our algorithm will batch the requests into $\log n$ groups where the $(i + 1)^{st}$ group (denoted by σ_{i+1}) contains requests $\langle r_{2^{i+1}}, \dots, r_{2^{i+1}} \rangle$. To process requests group $(i + 1)$, we apply the zoned algorithm using the k -median centers of the first 2^i requests.

In the random arrival model, the first t requests is a uniformly chosen random subset of size t . Using existing bounds [7], a random subset of size $\tilde{O}(\Delta/\epsilon^2)$ can be used to estimate the average k -median cost within a constant factor with an additional additive cost of ϵ . Unfortunately, despite having a large random

subset $(\sigma_0 \cup \sigma_1, \dots \cup \sigma_{\log n - 1})$ of $n/2$ requests (i.e., $\epsilon \approx \sqrt{\Delta/n}$) to estimate the k -median of $n/2$ requests of $\sigma_{\log n}$, we can only bound the average k -median cost within a constant factor with an additional additive cost of $\epsilon n/2 \approx \sqrt{\Delta n}$. Lower bounds on uniform random sample based estimation of k median seem to suggest that there is very little scope for improving this analysis for small-sized random subsets; see [7] for details on the upper and lower bound. In our case, since the sample size is large ($= n/2$), we present a different analysis to show that the k -median of this large random subset is a good proxy for the k -median of σ_{i+1} . Using this analysis, we show that the overall cost incurred in serving requests of σ_{i+1} has an additive cost of only $k\Delta$ (independent of n) and the total additive cost over all the $\log n$ groups to be $O(k\Delta \log n)$ leading to the following theorem (in Sect. 4.2):

Theorem 1. *Let σ be a sequence of n requests from a discrete metric space (X, d) . For any $\alpha > 1.5$, the expected cost of the adaptive zoned algorithm (Algorithm 2) for serving σ is upper bounded as follows.*

$$\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \leq 2\alpha n \text{medavg}(\sigma) + \left(\frac{8}{e} \left(\frac{2\alpha + 1}{2\alpha - 3} \right)^2 + 1 \right) k\Delta \log n + \Phi_0.$$

Here $\text{medavg}(\sigma)$ is the average k -median cost of all the requests in σ and is formally defined in Sect. 2.

In the random arrival model, the cost of serving the i th request can be lower bounded by the average k -median cost of the remaining (unprocessed) $n - i + 1$ requests. We show (in Sect. 4.3) that the lower bound on the cost of serving all the requests can still be related to the average k -median cost $\text{medavg}(\sigma)$ within an additive cost of $O(k\Delta \log n)$.

Theorem 2. *Let (X, d) be any discrete metric space and σ be a multi-set of n points from X . Let \mathcal{A} be any online algorithm to serve σ under the random arrival model, with initial configuration of servers \mathcal{K} . Let Δ be the diameter of X and for any $0 < \delta < 1$, the expected cost of \mathcal{A} is*

$$\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \geq \frac{n-1}{2}\delta \text{medavg}(\sigma) - \frac{2\delta}{(1-\delta)^2}(k+2)\Delta \log n.$$

Combining the two theorems, it follows that the adaptive zoned algorithm performs within a constant factor of the cost incurred by the best online algorithm along with an additional additive cost of $O(k\Delta \log n)$.

In Sect. 2, we present the basic terminology required for our algorithm and its analysis. In Sect. 3, we present the zoned algorithm for the known distribution model along with its analysis. In Sect. 4, we present the adaptive zoned algorithm in the random arrival model. We present our analysis of the upper bound of the cost in Sect. 4.2 and lower bound of the cost in Sect. 4.3.

2 Preliminaries

Let P be a multi-set of n points in a given discrete metric space (X, d) . For any point $p \in P$ and a set $K \subset X$, we define $d(p, K)$ to be the distance of p to its

nearest neighbor in K . We define the distance of the set P to the set K denoted by $d(P, K)$.

$$d(P, K) = \sum_{p \in P} d(p, K).$$

The average distance of P from K , denoted as $d_{\text{avg}}(P, K)$ is $d_{\text{avg}}(P, K) = \frac{1}{|P|} \sum_{x \in P} d(x, K)$. We define the k -median of P to be a set of k points $K^* \subseteq X$, given by

$$K^* = \arg \min_{K \subset X, |K|=k} d(P, K).$$

We refer to K^* as the *k -median centers* of P . The cost of the k -median K^* denoted by $\text{med}(P)$ is $\text{med}(P) = d(P, K^*)$, and the average cost of this k -median, denoted by $\text{medavg}(P)$, is given by

$$\text{medavg}(P) = \frac{\text{med}(P)}{|P|}.$$

In several instances, we denote the k -median of a set A by K^A .

The definition of k -median K^* extends easily to the case where we are given a probability distribution $\mathcal{D}(\cdot)$ on the discrete metric space X :

$$K^* = \arg \min_{K \subset X, |K|=k} \sum_{x \in X} \mathcal{D}(x) d(x, K),$$

and let $\text{medavg}(\mathcal{D}, X) = \sum_{x \in X} \mathcal{D}(x) d(x, K^*)$.

Theorem 3 (Chernoff bounds). *Suppose X_1, \dots, X_n are independent random binary variables, X denotes their sum, and $\mu = \mathbb{E}[X]$. Then*

$$\mathbb{P}[X \geq (1 + \delta)\mu] \leq e^{-\delta^2 \mu / 3}, \quad 0 < \delta < 1, \tag{1a}$$

$$\mathbb{P}[X \geq (1 + \delta)\mu] \leq e^{-\delta \mu / 3}, \quad 1 < \delta, \tag{1b}$$

$$\mathbb{P}[X \leq (1 - \delta)\mu] \leq e^{-\delta^2 \mu / 2}, \quad 0 < \delta < 1. \tag{1c}$$

3 Zoned Algorithm

We begin by introducing the zoned algorithm for the k -server problem in the known distribution (Algorithm 1) and the random arrival model (Algorithm 2). The core idea behind the algorithm is that the discrete metric space (X, d) can be partitioned into k zones each with a single server. For any request in a given zone, the corresponding server of this zone will serve it. For the known distribution, the partition is induced by the k -median of the distribution and is presented below.

Algorithm 1. Zoned algorithm for a known distribution

Data: Metric space (X, d) , probability distribution \mathcal{D} , a sequence of requests $\sigma = (r_1, \dots, r_n)$, k -servers $(\kappa_1, \dots, \kappa_k)$ and their initial positions
Result: Sequence of servers assigned in an online manner

- 1 Compute the k -median centers K^* of the distribution \mathcal{D} ;
- 2 $\phi : K^* \rightarrow (\kappa_1, \dots, \kappa_k)$, minimum-cost bipartite matching from K^* to the initial locations of k -servers;
- 3 **for** r in σ **do**
- 4 Find k -median center c with distance $d(r, K^*)$;
- 5 Move server $\phi(c)$ to r ;
- 6 **end**

The following lemma bounds the cost of any online algorithm with the k -median cost.

Lemma 1. *Given a discrete metric space (X, d) , and any request sequence σ of n locations chosen i.i.d. from a known distribution \mathcal{D} on X , the expected cost of any online algorithm \mathbb{A} is*

$$\mathbb{E}[w_{\mathbb{A}}(\sigma, \mathcal{K})] \geq n \cdot \text{medavg}(\mathcal{D}, X).$$

Proof. When a request arrives the algorithm must assign a server to it, so the cost of the algorithm must be at least the distance from the closest server to the request. The expected cost is therefore bounded below by the expected distance of a request to its nearest server. This is minimized if the servers have the configuration of the k -median centers K^* with the expected distance of $\text{medavg}(\mathcal{D}, X)$. The result then follows by linearity of expectation.

Theorem 4. *The zoned algorithm \mathcal{A} has an expected cost that is at most twice the cost incurred by any optimal online algorithm in the known distribution model.*

Proof. Let the initial configuration of the k servers be \mathcal{K} . For every request $r \in \sigma$, let $k \in K^*$ be its closest center. The zoned algorithm will move the server $\phi(k)$ to the request point r . By the triangle inequality this distance is less than if we had moved the server to c first, and then to r after. Every request under such a modification therefore incurs at most two costs, movement from c and movement to c . The expected distance of any request to its closest center is $\text{medavg}(\mathcal{D}, X)$, so using the modification, by linearity of expectation, and by Lemma 1:

$$\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \leq 2n \cdot \text{medavg}(\mathcal{D}, X) + \Phi_0 \leq 2w_{\mathbb{A}}(\sigma, \mathcal{K}) + \Phi_0$$

for n requests, any online algorithm \mathbb{A} , and where Φ_0 is the cost of the matching of \mathcal{K} and K^* .

4 Random Arrival Model

4.1 Adaptive Zoned Algorithm

We present and analyze a slightly modified zoned algorithm for the random arrival model. We partition the request sequence σ into $\log n$ groups $\sigma = \sigma_0 \cap \sigma_1, \dots, \cap \sigma_{\log n}$ where $|\sigma_0| = 1$ and $|\sigma_i| = 2^{i-1}$. For the request in σ_i we apply the zoned algorithm by using the k -median of the requests in $\sigma_0 \cup \sigma_1, \dots \cup \sigma_{i-1}$. Note that after serving requests in σ_i , we need to recompute the k -median of all the requests seen so far and move the k -servers to these locations (implicitly through the mapping ϕ). This results in a reconfiguration cost which is bounded by $O(k\Delta \log n)$. The algorithm is presented next.

Algorithm 2. Adaptive zoned algorithm A that runs in the Random Arrival Model

Data: Metric (X, d) , a sequence of requests $\sigma = (r_1, \dots, r_n)$, k -server $(\kappa_1, \dots, \kappa_k)$ and their initial positions

Result: Sequence of servers assigned in an online manner

```

1 1 → i;
2 1 → j;
3  $K^{(1)}$  is the initial positions of the  $k$ -servers  $(\kappa_1, \dots, \kappa_k)$ ;
4  $\phi$  is the map of locations in  $K^{(1)}$  to the servers that are located there;
5 for  $r$  in  $\sigma$  do
6   Find center  $c \in K^{(j)}$  that minimizes  $d(r, K^{(j)})$ ;
7   Move server  $\phi(c)$  to  $\sigma$ ;
8   if  $i = 2^{j-1}$  then
9     Compute the  $k$ -median centers  $K^{(j)}$  on  $\{r_1, \dots, r_i\}$ ;
10     $\phi : K^{(j)} \rightarrow (\kappa_1, \dots, \kappa_k)$ , minimum-cost bipartite matching
11      from  $K^{(j)}$  to the servers at their current locations;
12     $j \leftarrow j + 1$ ;
13  end
14   $i \leftarrow i + 1$ ;
15 end
```

4.2 Upper Bound

We use the following simple lemma in bounding the cost of the adaptive zoned algorithm.

Lemma 2. *Let P be any set of n points and let Q be a subset of P that is chosen uniformly at random from all possible subsets of size t . Then,*

$$\mathbb{E}[d(Q, K^P)] = t \cdot \text{medavg}(P).$$

Proof. Let \mathcal{S}_t be the set of all subsets of P with cardinality exactly t . The number of such subsets is given by $|\mathcal{S}_t| = \binom{n}{t}$. Therefore, the expected value of $d(Q, K^P)$ can be bounded by

$$\mathbb{E}[d(Q, K^P)] = \sum_{Q \in \mathcal{S}_t} \frac{1}{\binom{n}{t}} \sum_{q \in Q} d(q, K^P).$$

Every point $q \in P$ appears in exactly $\binom{n-1}{t-1}$ subsets of \mathcal{S}_t . Therefore, we can rewrite the expected value as

$$\mathbb{E}[d(Q, K^P)] = \frac{\binom{n-1}{t-1}}{\binom{n}{t}} \sum_{q \in P} d(q, K^P) = \frac{t}{n} \sum_{q \in P} d(q, K^P) = t \cdot \text{medavg}(P).$$

Lemma 3. *Let P be a set of n points with diameter Δ where n is a power of 2. For a random permutation of the points of P , let A and B correspond to the first $n/2$ and the last $n/2$ points of this permutation. Let K^P be the k -median centers of P . Given that A has been observed and K^A is the k -median centers of A , and for any $\alpha > 1.5$, the expected value of distance $d(B, K^A)$ is at most $\alpha n \text{medavg}(P) + \frac{4}{e} \left(\frac{2\alpha+1}{2\alpha-3} \right)^2 k\Delta$.*

Proof. Since A is the first $n/2$ points of a random permutation of P , A can be considered to be a subset chosen uniformly at random from the set of all subsets of size $n/2$. For the optimal k -median centers of A , namely K^A , we will bound the expected cost of $d(B, K^A)$. To help with the analysis, we partition P into k clusters based on assigning each point $p \in P$ to its closest k -median center in K^P (the optimal k -median of P). Let C_j^P denote the j th cluster with $k_j^P \in K^P$ as its median center and let $s_j = |C_j^P|$ be the size of this cluster. Let $A_j = A \cap C_j^P$ and $B_j = B \cap C_j^P$. By using triangle inequality (as in [8]), we can bound the distance of median center of cluster C_j^P , i.e., k_j^P to its closest median center in K^A

$$d(k_j^P, K^A) \leq \min_{x \in A_j} (d(k_j^P, x) + d(x, K^A)) \leq \frac{1}{|A_j|} \sum_{x \in A_j} (d(k_j^P, x) + d(x, K^A)).$$

The last inequality follows from the fact that the minimum of a set of numbers is bounded from above by its average. Using this bound, we can bound the distance from any point $y \in C_j^P$ to its closest median center in K^A by

$$d(y, K^A) \leq d(y, k_j^P) + d(k_j^P, K^A) \leq d(y, k_j^P) + \frac{1}{|A_j|} \sum_{x \in A_j} (d(k_j^P, x) + d(x, K^A)).$$

Therefore, the total distance of points in B_j to their closest center in K^A is bounded by

$$d(B_j, K^A) \leq \sum_{y \in B_j} d(y, k_j^P) + \frac{|B_j|}{|A_j|} \sum_{x \in A_j} (d(k_j^P, x) + d(x, K^A)). \quad (2)$$

Next, we will bound the expected value of $d(B_j, K^A)$. Note that the expected value $\mathbb{E}[|A_j|] = \frac{s_j}{2} = \mathbb{E}[|B_j|]$. Consider the event $\mathcal{E} : (|A_j| > (1 - \delta)\mathbb{E}[|A_j|])$, where $0 < \delta < 1$. Applying Chernoff bound (Theorem 3, inequality (1c)), event \mathcal{E} occurs with probability at least $1 - \exp\left(-\frac{\delta^2 s_j}{4}\right)$. When event \mathcal{E} occurs,

$$\frac{|B_j|}{|A_j|} = \frac{s_j - |A_j|}{|A_j|} < \frac{s_j(1 - \frac{1-\delta}{2})}{(1 - \delta)\frac{s_j}{2}} = \frac{1 + \delta}{1 - \delta}.$$

and we can bound $d(B_j, K^A)$ by

$$\sum_{y \in B_j} d(y, k_j^P) + \frac{1 + \delta}{1 - \delta} \sum_{x \in A_j} (d(k_j^P, x) + d(x, K^A)).$$

When \mathcal{E} does not occur (which has a probability of at most $\exp\left(-\frac{\delta^2 s_j}{4}\right)$), i.e., $|A_j| \leq (1 - \delta)\mathbb{E}[|A_j|]$, we use a trivial upper bound of $d(B_j, K^A) \leq s_j \Delta$. Applying this, we have the following upper bound for $\mathbb{E}[d(B_j, K^A)]$, where the expectation is over all possible permutations of P .

$$\begin{aligned} \mathbb{E}[d(B_j, K^A)] &= \Pr(\mathcal{E})\mathbb{E}[d(B_j, K^A) | \mathcal{E}] + \Pr(\bar{\mathcal{E}})\mathbb{E}[d(B_j, K^A) | \bar{\mathcal{E}}] \\ &< \mathbb{E}\left[\sum_{y \in B_j} d(y, k_j^P) + \frac{1 + \delta}{1 - \delta} \sum_{x \in A_j} (d(k_j^P, x) + d(x, K^A))\right] \\ &\quad + \exp\left(-\frac{\delta^2 s_j}{4}\right) s_j \Delta. \end{aligned} \tag{3}$$

We set $\delta := 2\sqrt{\frac{\log(s_j/\tau)}{s_j}}$. When $s_j \geq \tau$, we can reduce the last term in (3) to $\exp\left(-\frac{\delta^2 s_j}{4}\right) s_j \Delta \leq \tau \Delta$. When $s_j < \tau$, we can simply bound $d(B_j, K^A) \leq |B_j| \Delta < \tau \Delta$. Also, we note that $\frac{1+\delta}{1-\delta}$ is a monotonically increasing function of δ for $0 < \delta < 1$. For $x > 0$, $\delta = 2\sqrt{\frac{\log(s_j/\tau)}{s_j}}$ attains the maximum value of $\frac{2}{\sqrt{e\tau}}$ at $s_j = e\tau$. Therefore, $\frac{1+\delta}{1-\delta} \leq \frac{\sqrt{e\tau}+2}{\sqrt{e\tau}-2}$. Using these bounds, we rewrite (3) as

$$\mathbb{E}[d(B_j, K^A)] \leq \mathbb{E}\left[\sum_{y \in B_j} d(y, k_j^P) + \frac{\sqrt{e\tau}+2}{\sqrt{e\tau}-2} \sum_{x \in A_j} (d(k_j^P, x) + d(x, K^A))\right] + \tau \Delta.$$

Summing over all clusters,

$$\begin{aligned} \mathbb{E}[d(B, K^A)] &\leq \mathbb{E}\left[\sum_{j=1}^k \sum_{y \in B_j} d(y, k_j^P) + \frac{\sqrt{e\tau}+2}{\sqrt{e\tau}-2} \sum_{j=1}^k \sum_{x \in A_j} (d(k_j^P, x) + d(x, K^A))\right] \\ &\quad + k\tau \Delta \\ &\leq \mathbb{E}\left[\sum_{y \in B} d(y, K^P) + \frac{\sqrt{e\tau}+2}{\sqrt{e\tau}-2} \sum_{x \in A} (d(K^P, x) + d(x, K^A))\right] + k\tau \Delta. \end{aligned}$$

Since $\sum_{x \in A} d(x, K^A) \leq \sum_{x \in A} d(x, K^P)$, we have

$$\mathbb{E}[d(B, K^A)] \leq \mathbb{E}\left[\sum_{y \in B} d(y, K^P) + 2 \frac{\sqrt{e\tau} + 2}{\sqrt{e\tau} - 2} \sum_{x \in A} d(x, K^P)\right] + k\tau\Delta.$$

Since A and B are subsets chosen uniformly at random from all possible subsets of P of size $n/2$, by Lemma 2, we have $\mathbb{E}[\sum_{y \in B} d(y, K^P)] = |B| \text{medavg}(P)$ and $\mathbb{E}[\sum_{y \in A} d(y, K^P)] = |A| \text{medavg}(P)$. Therefore,

$$\mathbb{E}[d(B, K^A)] \leq |B| \text{medavg}(P) + 2 \frac{\sqrt{e\tau} + 2}{\sqrt{e\tau} - 2} |A| \text{medavg}(P) + k\tau\Delta$$

and,

$$\mathbb{E}[d(B, K^A)] \leq \left(\frac{\sqrt{e\tau} + 2}{\sqrt{e\tau} - 2} + \frac{1}{2}\right) n \text{medavg}(P) + k\tau\Delta.$$

Setting $\alpha = \frac{\sqrt{e\tau} + 2}{\sqrt{e\tau} - 2} + \frac{1}{2}$ we arrive at the final expression.

Proof (Proof of Theorem 1). Let $\sigma = \sigma^{(0)} \frown \sigma^{(1)} \frown \sigma^{(2)} \frown \dots$, where \frown denotes concatenation, so that $|\sigma^{(0)}| = 1$ and $|\sigma^{(i)}| = 2^{i-1}$. Let $T^{(i)} = \sigma^{(0)} \frown \sigma^{(1)} \frown \dots \frown \sigma^{(i)}$ and $K^{(i)}$ denote the k -median centers of $T^{(i)}$. Here, we use the terms multi-set and sequence interchangeably. By algorithm \mathcal{A} , each sequence $\sigma^{(i)}$ is served by $K^{(i-1)}$. Noting that $T^{(i)} = T^{(i-1)} \frown \sigma^{(i)}$ and $|\sigma^{(i)}| = |T^{(i-1)}| = 2^{i-1}$, we apply Lemma 3 with $P = T^{(i)}$, $A = T^{(i-1)}$ and $B = \sigma^{(i)}$. The expected distance $d(\sigma^{(i)}, K^{(i-1)}) \leq \alpha |T^{(i)}| \text{medavg}(T^{(i)}) + \frac{4}{e} \left(\frac{2\alpha+1}{2\alpha-3}\right)^2 k\Delta$, and the cost to serve $\sigma^{(i)}$ is at most twice this cost. In addition, for each $\sigma^{(i)}$, the cost to move the servers from their position at the end of $\sigma^{(i-1)}$ to $K^{(i)}$ is at most $k\Delta$. Therefore, $\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})]$ can be bounded as follows.

$$\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \leq 2 \sum_i \alpha |T^{(i)}| \text{medavg}(T^{(i)}) + \sum_i \left(\frac{8}{e} \left(\frac{2\alpha+1}{2\alpha-3}\right)^2 k\Delta + k\Delta\right) + \Phi_0.$$

Since, $\text{medavg}(T^{(i)}) = \frac{1}{|T^{(i)}|} \sum_{x \in T^{(i)}} d(x, K^{(i)}) \leq \frac{1}{|T^{(i)}|} \sum_{x \in T^{(i)}} d(x, K^*)$,

$$\begin{aligned} \mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] &\leq 2\alpha \sum_i \sum_{x \in T^{(i)}} d(x, K^*) + \left(\frac{8}{e} \left(\frac{2\alpha+1}{2\alpha-3}\right)^2 + 1\right) k\Delta \log n + \Phi_0 \\ &= 2\alpha \sum_{x \in \sigma} d(x, K^*) + \left(\frac{8}{e} \left(\frac{2\alpha+1}{2\alpha-3}\right)^2 + 1\right) k\Delta \log n + \Phi_0 \\ &= 2\alpha n \text{medavg}(\sigma) + \left(\frac{8}{e} \left(\frac{2\alpha+1}{2\alpha-3}\right)^2 + 1\right) k\Delta \log n + \Phi_0. \end{aligned}$$

Hence proved.

4.3 Lower Bound

Let \mathcal{S}_j denote the set of all possible subsets of σ of size j . The expected value of $\text{medavg}(A)$ for a set A of size m chosen uniformly at random from \mathcal{S}_m is denoted by $\text{medavg}_{\mathbb{E}}(m)$, and defined as

$$\text{medavg}_{\mathbb{E}}(m) = \frac{1}{|\mathcal{S}_m|} \sum_{A \in \mathcal{S}_m} \text{medavg}(A).$$

We will first prove the following lemmas which bounds the cost of any online algorithm from below by $\sum_{i=1}^j \text{medavg}_{\mathbb{E}}(i)$.

Lemma 4. $\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \geq \sum_{i=1}^n \text{medavg}_{\mathbb{E}}(i)$.

Proof. Supposing $i - 1$ requests have been served, let A denote the subset of S containing the points yet to be served, i.e., $|A| = n - i + 1$. Let K be the current configuration of the servers. Since every element of A has the same probability of being picked, for any k -set K , the expected distance of the i^{th} request from K is

$$\frac{1}{|A|} \sum_{x \in A} d(x, K) \geq \text{medavg}(A).$$

Since every set $A \in \mathcal{S}_{n-i+1}$ is equiprobable, the cost of serving the i th request is $\text{medavg}_{\mathbb{E}}(n - i + 1)$ and the result follows.

Lemma 5. For any $m > m'$, $\text{medavg}_{\mathbb{E}}(m) \geq \text{medavg}_{\mathbb{E}}(m')$.

Proof. Let $A \in \mathcal{S}_m$ and $B \subset A$ of size m' . Noting that every element of A occurs in exactly $\binom{m-1}{m'-1}$ subsets B , we have

$$\begin{aligned} \text{medavg}(A) &= \frac{1}{|A|} \sum_{x \in A} d(x, K^A) = \frac{1}{m} \frac{1}{\binom{m-1}{m'-1}} \sum_{B \subset A, |B|=m'} d(x, K^A) \\ &\geq \frac{1}{m} \frac{1}{\binom{m-1}{m'-1}} \sum_{B \subset A, |B|=m'} d(x, K^B) \\ &= \frac{m'}{m} \frac{1}{\binom{m-1}{m'-1}} \sum_{B \subset A, |B|=m'} \text{medavg}(B). \end{aligned}$$

Since, every B is a subset of exactly $\binom{n-m'}{m-m'}$ sets of \mathcal{S}_m ,

$$\begin{aligned} \sum_{A \in \mathcal{S}_m} \text{medavg}(A) &\geq \binom{n-m'}{m-m'} \frac{m'}{m} \frac{1}{\binom{m-1}{m'-1}} \sum_{B \in \mathcal{S}_{m'}} \text{medavg}(B) \\ &= \frac{|\mathcal{S}_m|}{|\mathcal{S}_{m'}|} \sum_{B \in \mathcal{S}_{m'}} \text{medavg}(B). \end{aligned}$$

Lemma 6. $\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \geq \lceil \frac{n}{2} \rceil \text{medavg}_{\mathbb{E}}(\lceil \frac{n}{2} \rceil)$.

Proof. From Lemma 5, for any $m > \frac{n}{2}$, $\text{medavg}_{\mathbb{E}}(m) \geq \text{medavg}_{\mathbb{E}}(\lceil \frac{n}{2} \rceil)$. From Lemma 4,

$$\mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] \geq \sum_{i=1}^n \text{medavg}_{\mathbb{E}}(i) \geq \sum_{i=1}^{\lceil \frac{n}{2} \rceil} \text{medavg}_{\mathbb{E}}(n-i+1) \geq \lceil \frac{n}{2} \rceil \text{medavg}_{\mathbb{E}}(\lceil \frac{n}{2} \rceil).$$

Following Meyerson *et al.* [8], for any k -set K of σ , let $\beta(K, b) = (B_1, B_2, \dots, B_b)$ denote the partition of σ induced by K as follows: We order points of σ by their distance to the closest point in K (low to high), and divide this sequence into b bins each containing equal number of points $\frac{n}{b}$. Henceforth, let $b = \frac{n(1-\delta)^2}{4(k+2)\log n}$ for $0 < \delta < 1$.

Lemma 7. *Let A be any random subset of σ of size $\lceil \frac{n}{2} \rceil$. With probability at least $1 - \frac{1}{n}$, every bin of $\beta(K^A, b)$ has at most $\frac{\delta n}{2b}$ points from A .*

Proof. We will show that with high probability the above statement is satisfied for all k -sets K . Since K^A is one among these sets, the bound follows. Let K be any k -set and $\beta(K, b)$ be the induced partition. The expected number of points of A belonging to each bin is $\frac{n}{2b}$. Using Chernoff bound (1c), for $0 < \delta < 1$, the probability that a bin has at most $\frac{\delta n}{2b}$ points is at most $e^{-\frac{n(1-\delta)^2}{4b}}$. Since there are at most n^k sets K and b bins, applying union bound the proof follows.

Proof (Proof of Theorem 2). Let $A \subset \sigma$ of size $\frac{n}{2}$, and $\beta(K^A, b)$ be the partition of σ induced by K^A . Let d_j^{\min} and d_j^{\max} denote the minimum and maximum distances respectively of points from bin B_j . From Lemma 7, with probability at least $1 - \frac{1}{n}$, every bin of $\beta(K^A, b)$ has at most $\frac{\delta n}{2b}$ points of A . Therefore,

$$\sum_{x \in A} d(x, K^A) \geq \sum_{j=1}^b \frac{\delta n}{2b} d_j^{\min} \geq \frac{\delta}{2} \sum_{j=1}^{b-1} \frac{n}{b} d_j^{\max}.$$

The last inequality follows from the fact that $d_j^{\min} \geq d_{j-1}^{\max}$. For each j , since d_j^{\max} is the maximum assigned distance, $\frac{n}{b} d_j^{\max} \geq \sum_{y \in B_j} d(y, K^A)$. Finally, we note that $d_b^{\max} \leq \Delta$, the diameter. Combining,

$$\begin{aligned} \sum_{x \in A} d(x, K^A) &\geq \frac{\delta}{2} \sum_{j=1}^{b-1} \sum_{y \in B_j} d(y, K^A) + \frac{\delta}{2} \sum_{y \in B_b} d(y, K^A) - \frac{\delta n}{2b} \Delta \\ &= \frac{\delta}{2} \sum_{j=1}^b \sum_{y \in B_j} d(y, K^A) - \frac{\delta n}{2b} \Delta \\ &= \frac{\delta}{2} \sum_{y \in \sigma} d(y, K^A) - \frac{\delta n}{2b} \Delta. \end{aligned}$$

Therefore, with probability $(1 - 1/n)$

$$\text{medavg}(A) \geq \frac{\delta}{n} \sum_{y \in \sigma} d(y, K^A) - \frac{\delta\Delta}{b} \geq \delta \text{medavg}(\sigma) - \frac{\delta\Delta}{b}.$$

Hence,

$$\text{medavg}_{\mathbb{E}}\left(\left\lceil \frac{n}{2} \right\rceil\right) \geq \left(1 - \frac{1}{n}\right)\left(\frac{\delta}{n} \sum_{y \in \sigma} d(y, K^A) - \frac{\delta\Delta}{b}\right) + \frac{1}{n} \cdot 0$$

From Lemma 6 and the choice of $b = \frac{n(1-\delta)^2}{4(k+2)\log n}$,

$$\begin{aligned} \mathbb{E}[w_{\mathcal{A}}(\sigma, \mathcal{K})] &\geq \left\lceil \frac{n}{2} \right\rceil \text{medavg}_{\mathbb{E}}\left(\left\lceil \frac{n}{2} \right\rceil\right) \geq \left\lceil \frac{n}{2} \right\rceil \left(1 - \frac{1}{n}\right) \left(\delta \text{medavg}(\sigma) - \frac{\delta\Delta}{b}\right) \\ &\geq \frac{n-1}{2} \left(\delta \text{medavg}(\sigma) - \frac{\delta\Delta}{b}\right) \\ &= \frac{n-1}{2} \delta \text{medavg}(\sigma) - \frac{2\delta}{(1-\delta)^2} (k+2)\Delta \log n. \end{aligned}$$

Hence proved.

5 Conclusion

In this paper, we presented and analyzed a simple k -median based algorithm for the stochastic k -server problem. Our result is based on proving a new and sharper approximation bound of the k -median of a large random subset of a point set with respect to the k -median of entire point set.

In the random arrival model, the cost of serving the next request is lower bounded by the average k -median cost of the requests that have not yet been processed. Clearly, the k servers cannot always be in the optimal k -median configuration even for the best online algorithm. Therefore, it is conceivable that one can prove a stronger lower bound. In particular, can we prove a stronger lower or upper bound and reduce the additive error from $O(k\Delta \log n)$ to $O(k\Delta)$ (completely independent of n)?

References

1. Mahdian, M., Yan, Q.: Online bipartite matching with random arrivals: an approach based on strongly factor-revealing LPs. In: Proceedings of the 43rd Annual ACM Symposium on Theory of Computing, STOC 2011, pp. 597–606 (2011)
2. Karande, C., Mehta, A., Tripathi, P.: Online bipartite matching with unknown distributions. In: Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing, STOC 2011, pp. 587–596. ACM, New York (2011)
3. Manasse, M.S., McGeoch, L.A., Sleator, D.D.: Competitive algorithms for server problems. J. Algorithms **11**(2), 208–230 (1990)

4. Koutsoupias, E., Papadimitriou, C.H.: On the k -server conjecture. *J. ACM* **42**(5), 971–983 (1995)
5. Bansal, N., Buchbinder, N., Madry, A., Naor, J.: A polylogarithmic-competitive algorithm for the k -server problem. In: Proceedings of the IEEE 52nd Annual Symposium on Foundations of Computer Science (FOCS), pp. 267–276, October 2011
6. Raghvendra, S.: A Robust and optimal online algorithm for minimum metric bipartite matching. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2016, vol. 60, pp. 18:1–18:16 (2016)
7. Czumaj, A., Sohler, C.: Sublinear-time approximation algorithms for clustering via random sampling. *Random Struct. Algorithms* **30**(1–2), 226–256 (2007)
8. Meyerson, A., O’callaghan, L., Plotkin, S.: A k -median algorithm with running time independent of data size. *Mach. Learn.* **56**(1–3), 61–87 (2004)

Towards Robust Models of Food Flows and Their Role in Invasive Species Spread

Srinivasan Venkatramanan^{1,*}, Sichao Wu¹, Bowen Shi¹, Achla Marathe^{1,2}, Madhav Marathe¹, Stephen Eubank¹, Lalit P. Sah^{3,4}, A. P. Giri^{3,4}, Luke A. Colavito^{3,4}, K. S. Nitin⁵, V. Sridhar⁵, R. Asokan⁵, Rangaswamy Muniappan³, G. Norton², Abhijin Adiga¹

¹ Biocomplexity Institute of Virginia Tech

² Department of Agricultural and Applied Economics, Virginia Tech

³ Feed the Future Integrated Pest Management Innovation Lab, Virginia Tech

⁴ International Development Enterprises, Nepal

⁵ Indian Institute of Horticultural Research

* email:vsrini@vt.edu

Abstract—We develop a general data-driven methodology that yields network representations of agricultural flows pertaining to the spread of invasive species. The methodology synthesizes sparse, diverse, noisy and incomplete data that is typically available to build realistic spatio-temporal network representations. We illustrate the methodology by modeling the seasonal flow of the tomato crop in Nepal between major domestic markets. Through dynamical analysis of the network, we study its role in the spread of a major pest of tomato, *Tuta absoluta*, an emerging outbreak in this country. In the absence of high-resolution pest distribution data, we apply a novel ranking-based inference approach to establish that tomato trade is a driving factor in the rapid spread of this pest.

I. INTRODUCTION

Food security is an increasingly important societal problem. Increased globalization, climate change, population growth, scarce per capita resources, international trade, travel and invasive species are important factors contributing to the issue of global food security. In this paper, we will focus on commodity flows, a quintessential component of our food systems. Production and consumption of agricultural produce is no longer a local phenomenon – agro products travel thousands of miles over global supply chain networks. While economically attractive in the short term, global trade increases the risk of rapid spread of invasive species and bio-terrorism. The situation is quite similar to spread of infectious diseases in human and animal populations.

In this paper, we study the seasonal flow of agricultural commodities, focusing on their role in the spread of invasive species. The spread of pests

and pathogens is driven by various natural and anthropogenic factors. An in-depth understanding of the biology and climatic conditions is essential to assess establishment risk and devise sustainable management strategies and has been the focus of ecologists for a long time. In contrast, not much is understood as regards to the role of human-mediated pathways (including trade and travel) in preventing introduction and mitigating immediate impact [2, 4, 6, 10]. See [7, 8, 13, 17] for further discussion on this important subject.

Our contributions: Here we develop an integrated methodology that combines data science, algorithmics, machine learning and ecological modeling that allow us to address important factors that affect the human mediated pathways contributing to invasive species spread. Our key contributions in this paper are as follows:

(i) We develop an integrated data-driven methodology for synthesizing realistic spatio-temporal networks of seasonal agro-products between major markets. The methodology is outlined in Figure 1a. It combines diverse multi-type, noisy, misaligned and sparse datasets with detailed context specific domain knowledge provided by local experts. A particular challenge we address is data sparsity. The methodology is generic and can be adapted to other agro products and regions.

(ii) We illustrate the methodology by developing a spatio-temporal domestic tomato trade network in Nepal and investigate its role in the spread of *Tuta absoluta*, a devastating pest of the tomato crop [3] and an emerging pest in Nepal [1].

(iii) We analyze the spatio-temporal properties of the

flow networks. Further, through dynamical analysis of the networks and a novel rank-based inference approach, we assess the role of trade in the spread of the pest.

(iv) We conduct an in-depth sensitivity analysis to quantify the role of input parameters. This analysis is used in validating our synthesized networks; furthermore the analysis provides improved understanding of the pest dynamics.

Challenges. Agro-trade networks for moving agricultural products is a complex system. The networks depend on varied factors, including seasonal production, population distribution, cultural factors, economic activity, storage and transport infrastructure. Furthermore, data needed to develop agro-trade networks is often sparse, noisy and is not openly available. For instance, even standard information such as region-level production is unavailable for many countries. Even if available, these datasets vary in format, they are misaligned in reporting time and vary in spatial and temporal resolution. Apart from quantitative datasets, there is also need for qualitative information pertaining to the study region such as cultural practices, seasonal production cycles, etc. Interpreting this data and integrating it into model design requires local knowledge. Similar challenges exist in obtaining high-resolution pest distribution data.

Another challenge is validating the network representations. While international trade data is available at the commodity level, domestic data is hard to come by. Even in data-rich regions such as the US, the available sample data (e.g. Freight Analysis Framework¹) is aggregated at the commodity category level. Secondly, the role of these networks in the study of invasive species requires one to understand the ecological contagion processes. Also, monitoring is a resource intensive task: the placement of traps is largely determined by accessibility and availability of trained personnel. The pest might not be detected during off season due to host unavailability. In the absence of monitoring, its presence will become apparent only during the growing season. But its reporting might be delayed by farmers due to lack of awareness or fear of quarantining. Given these constraints, there may be several months of delay in reporting.

Related work: In recent years, there has been a lot of interest in studying the role of international

trade and travel in invasive species spread. Ercsey-Ravasz et al. [8] analyze the International Agro-Food Trade Network to identify countries of importance in the context of food safety. Early et al. [7] study the terrestrial threat from invasive species and evaluate national capacities to prevent and manage invasions. Tatem [17] showed that the world-wide airline network increases the risks of establishment by providing busy transport links between spatially distant, but climatically similar regions of the world.

There has been some work on domestic commodity flow and its role in pest spread. Nopsa et al. [13] evaluated the structure of rail networks in the US and Australia for pest and mycotoxin dispersal. Colunga-Garcia et al. [5] use the regional freight transport information to characterize risk of urban and periurban areas to exotic forest insect pests in the US. In [15] provides a survey of recent modeling efforts.

T. absoluta. There is general consensus that vegetable and seedling trade is a primary driver of *T. absoluta* spread [3]. However, previous modeling efforts have only focused on establishment potential [18] and spatial dispersion [9]. This is the first work that analyzes human-mediated pathways in the context of *T. absoluta*. Nepal's vegetable production and trade has been extensively studied from a socio-economic perspective ([19] for example), but, to the best of our knowledge, there is no such work in the context of invasive species spread with focus on this region.

II. MODELING FRAMEWORK

Figure 1a outlines the different components that constitute the framework. As we describe each component, we will also discuss the associated data challenges and key modeling assumptions that allowed us to integrate them. The symbols and abbreviations used henceforth are summarized in Table II.

A. Data

Table I lists the datasets used in our framework. We link several open-source datasets along with qualitative inputs from local experts in order to model the seasonal trade of tomato crop as well as pest dynamics. Some of the challenges arise from the fact that the datasets vary in their spatial and temporal resolution and their year of release (see Table I). Owing to the unique geography of Nepal, the vegetable production cycle varies with altitude (see Figure 1c). The annual production data was combined with the knowledge of production cycle to model the spatio-temporal variation in

¹https://ops.fhwa.dot.gov/freight/freight_analysis/faf/

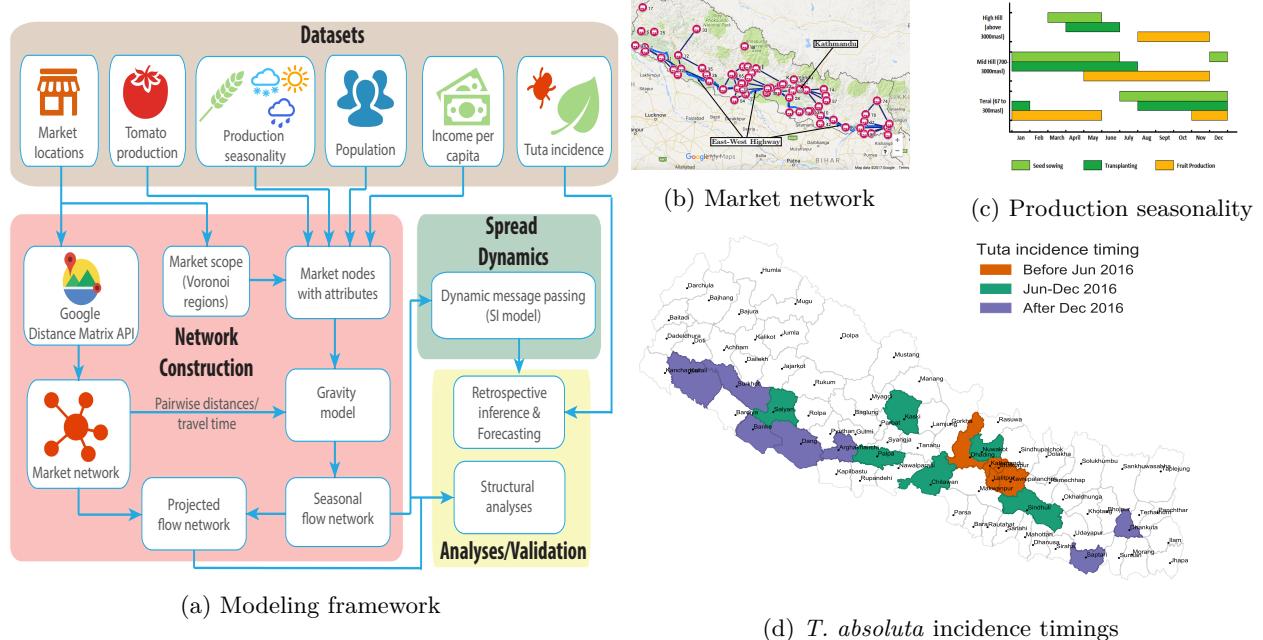


Figure 1: Modeling framework and some datasets

production across seasons. Major vegetable markets were geolocated using Google Maps, and Google Distance Matrix API was used to construct the road network and compute travel times. Several organizations have been involved in the monitoring of *T. absoluta* spread in Nepal: NARC, USAID, iDE Nepal, ENBAITA and Agricare Pvt. Ltd. The pest is monitored using pheromone traps that have been installed in several Village Development Committees. In May 2016, *T. absoluta* was officially reported by NARC’s entomology division in Lalitpur (near Kathmandu).

B. Network construction

Regional markets serve as key locations facilitating agricultural commodity flow, hence it makes sense to model the flow network with markets represented as nodes. We model the flow of agricultural produce among markets based on the following premise: The total outflow from a market depends on the amount of produce in its surrounding regions, and the total inflow is a function of the population it caters to and the corresponding per capita income. The main assumptions in this model are: (i) imports and exports are not significant enough to influence domestic trade: For instance, in 2014, Nepal exported only 1% of its tomatoes and imported about 6-7% of its total consumption (<http://www.fao.org/faostat>); (ii) Fresh tomatoes are mainly traded for consump-

tion: The tomato processing industry in Nepal is not well developed [19]. This motivates the use of population and per capita income as indicators of tomato consumption in a given district. and (iii) the higher the per capita income, the greater the consumption: Tomato is among the top two vegetables which provide highest profit to farmers (expensive for a typical consumer), and unlike cauliflower and cabbage, tomato is not considered a staple vegetable in the Nepalese household [19].

The flows are estimated using a doubly constrained gravity model [11]. The flow F_{ij} from location i to j is given by

$$F_{ij} = a_i b_j O_i I_j f(d_{ij}) \quad (1)$$

where, O_i is the total outflow of the commodity from i , I_j is the total inflow to j , d_{ij} is the time taken to travel from i to j , $f(\cdot)$ is the *distance deterrence function*, and coefficients a_i and b_j are computed through an iterative process to ensure flow balance.

However, as seen in Table I, data pertaining to these quantities are available at different spatial and temporal resolutions. Thus, before we apply (1), we need to synthesize these datasets to capture the seasonal commodity flow at the level of markets. The steps involved are described as follows:

Seasonality of production: Based on the physiognomy, districts of Nepal are partitioned into three regions, namely Terai, Mid Hills and High Hills

Table I: Datasets.

Description	Source	Resolution	Year
Population	Nepal Central Bureau of Statistics (http://cbs.gov.np/)	District/Town	2011
Per Capita Income	Nepal Central Bureau of Statistics (http://cbs.gov.np/)	District	2011
Tomato production	Nepal Ministry of Agricultural Development (MOAD) (http://moad.gov.np/)	District, Annual	2015
Production seasonality	iDE Nepal (http://idenepal.org/) and MOAD	Region, Monthly	2016
Major vegetable markets	MOAD Marketing Information System (http://www.agrimis.gov.np/)	Town	2017
Market distances	Google Maps, Distance matrix API	Market	2017
Tomato import,exports	Food and Agriculture Organization (FAOSTAT) (www.fao.org/faostat/)	Country, Annual	2013
Tomato consumption	FAOSTAT, MOAD	Country, Annual	2013
Flows to Kalimati market	Official website (kalimatimarket.gov.np/)	District, Annual	2015
<i>T. absoluta</i> incidence reports	Nepal National Agriculture Research Council(http://narc.gov.np/)	District/town	2017
	USAID IPM Innovation Lab, iDE Nepal		

Table II: Notation and abbreviations.

Variables	Description
F_{ij}	Commodity flow from node i to j
O_i	Total outflow of commodity from node i
I_i	Total inflow of commodity into node i
d_{ij}	Distance between nodes i and j
$f(\cdot)$	deterrence function
β	Power-law exponent of gravity model
κ	Cutoff time of gravity model
γ	Per capita income parameter
σ	Gaussian parameter for spatial seeding
t	Time step for the spread model

(see Figure 3e) Due to altitude and temperature variations, the tomato production season varies among these regions (see Figure 1c). Production in the Mid Hills and High Hills is largely restricted to the summer months of June to November (referred to as season S1), while Terai region produces during the winter months of December to May (referred to as season S2). As a result, we have two distinct flow networks, one for each season. We partitioned the districts into two groups: Mid Hills and High Hills belong to group 1, while the Terai districts belong to group 2. All districts belonging to group i were assigned their respective annual production for season S_i and zero for the other season.

Market scope definition: The nodes of the flow network are the major markets, 69 in all, after merging markets that belong to the same town. Recall that the amount of production is specified at the district level. In order to obtain the production estimates at market level, we defined *market scope* as follows: The country’s map was overlaid by a grid cell of size $5\text{km} \times 5\text{km}$ and we constructed a Voronoi partition of these cells using market locations as centroids. This is under the assumption that tomato sellers and buyers will seek out the nearest market. We assumed uniform spatial distribution of production within each district. Each grid cell was assigned a value of production in a particular season proportional to the fraction of the area of the

district covered by the cell. The total outflow from the market is the sum of production of the grid cells assigned to it for a particular season.

Modeling consumption: We modeled the total inflow I_i into a market as a product of the population catered to by the market and a function of the average per capita income associated with the market η_i, η_i^γ , where γ is a tunable parameter. The population catered to by the market, was derived from district level population data and the market scope as defined for production redistribution.

Inter-market travel time: Owing to the diverse landscape of Nepal and varying road conditions we used travel time by road instead of the geodesic or road distance between the markets. We begin with list of major vegetable markets in Nepal (see Table I), and geolocate them using Google Maps. We then manually embedded the market locations onto Nepal road network, and constructed a planar network by connecting the markets which have a direct route (without going through other markets) between them. We also removed markets which were completely inaccessible by road. We used Google Distance Matrix API² to compute travel times by road along the edges of this planar network. This in turn, yields a road network among the markets, where the edges are weighted by their travel time. Distance between any two markets is then obtained as the shortest travel time on the road network. The distance deterrence function $f(d_{ij}) = d_{ij}^{-\beta} \exp(-d_{ij}/\kappa)$ combines power-law and exponential decay with d_{ij} which can be controlled by the tunable parameters β , the power-law exponent, and κ , the cutoff time.

C. Spread Dynamics

We develop a discrete-time SI (Susceptible-Infected) epidemic model on directed weighted networks [14] to model pest dispersal. Each node is

²<https://developers.google.com/maps/documentation/distance-matrix/>

either susceptible (free from pest) or infected (pest is present). Henceforth, we use the term “infected” for a node or a region frequently to imply *T. absoluta* infestation at that location. A node i in state I infects each of its out-neighbors j in the network with probability proportional to the flow F_{ij} at each time step t . The infection probabilities are obtained by normalizing flows globally: $\lambda_{ij} = \frac{F_{ij}}{\max_{i,j} F_{ij}}$. The model is based on two assumptions: (i) an infected node remains infected and continues to infect its neighbors and (ii) the chance of infection is directly proportional to the volume traded. Considering the fact that Nepal was ill-prepared for this invasion and the lack of effective intervention methods, (i) is a fair assumption. Historically, *T. absoluta* has spread rapidly in regions where tomato trade has been the highest (parts of Europe and Middle-East for example) thus motivating assumption (ii).

Let $P_S(i, t, f_0)$ denote the probability that node i remains uninfected (i.e., susceptible) by time t given the initial condition f_0 which assigns probability of infection at time step $t = 0$ to each node. In general, computing P_S is hard. Efficient methods have been proposed to estimate this probability. Here, we adopt the *dynamic message passing algorithm* by Lokhov et al. [12], summarized by the following equations.

$$\begin{aligned} P_S^{i \rightarrow j}(t+1) &= P_S(i, 0, f_0) \Pi_{k \in \delta i \setminus j} \theta^{k \rightarrow i}(t+1) \\ \theta^{k \rightarrow i}(t+1) &= \theta^{k \rightarrow i}(t) - \lambda_{ki} \phi^{k \rightarrow i}(t) \\ \phi^{k \rightarrow i}(t) &= (1 - \lambda_{ki}) \phi^{k \rightarrow i}(t-1) \\ &\quad - [P_S^{k \rightarrow i}(t) - P_S^{k \rightarrow i}(t-1)] \end{aligned} \quad (2)$$

In the above equations, λ_{ki} is the infection probability across edge (k, i) , and θ, ϕ are intermediate messages used to update the node states. Finally, the quantity of interest $P_S(i, t, f_0)$, the probability that node i remains uninfected (i.e., susceptible) till time t is given as:

$$P_S(i, t+1, f_0) = P_S(i, 0, f_0) \Pi_{k \in \delta i} \theta^{k \rightarrow i}(t+1)$$

Note that for any given t , $P_S(i, t, f_0) + P_I(i, t, f_0) = 1$, and hence the entire evolution of the epidemic on the network is captured by $P_S(i, t, f_0), \forall i, t$ given the initial condition f_0 .

The initial configuration f_0 is chosen to mimic a spatially dispersed seeding scenario. We first select a *central* seed node, and then use a Gaussian kernel with parameter σ around the seed node to assign initial infection probabilities for neighboring markets. A market at a geodesic distance d from the seed,

is assigned the infection probability $e^{-\frac{d^2}{2\sigma^2}}$. The kernel accounts for factors such as uncertainty in determining the pest location, the possibility of spread of the pest through natural means as well as interactions between these markets.

III. ANALYSES AND RESULTS

Flow validation: The unavailability of sample data on seasonal trade of tomato crop makes it challenging to calibrate and validate the flow network model. In fact, to the best of our knowledge, even information on annual flow of vegetables between markets is not available. However, for the largest wholesale market of Nepal, Kalimati (located in Kathmandu), yearly data on volume of tomato arriving from each district is available (Table I). In Figures 2d–2f, we compare this data with the network flows. Given a set of network parameters (β, κ, γ) , we obtained the inflow from a particular district to Kathmandu as follows: We combined the weights of all edges of the corresponding network with destination node “Kathmandu” and source nodes belonging to that district.

As seen in Figure 2d, for γ values between 0.5 and 1, the flows from the networks are comparable to the Kalimati data except for two districts: Dhading (the top contributor) and Sarlahi (third highest). Upon further investigation we find that Dhading, which is a major producer west of Kathmandu, serves the Mid Hills and Terai regions of the Central Development Region in the flow networks (Figure 2e). While the gravity model predicts that these flows will be directly delivered to these regions, in reality, it is possible that Dhading’s produce is routed through Kalimati market as there are several traders from Dhading registered in the Kalimati market³. As for Sarlahi, even though there is little inflow to Kalimati market in the flow networks, other markets in the Kathmandu valley (belonging to Bhaktapur and Lalitpur districts) receive significant flows from Sarlahi (Figure 2f), which could, as in the previous case be routed through Kalimati market. These issues highlight some of the limitations of the gravity model, which do not account for real-world trader dynamics.

A. Structural properties

For each set of network parameters (β, κ, γ) , there are two networks, one for each season. Both networks have 69 nodes. The cumulative distribution of flows with respect to travel time are plotted in Figures 2a–2c for different values of network parameters for

³<http://mrsmp.gov.np/files/download/tomato%20book.pdf>

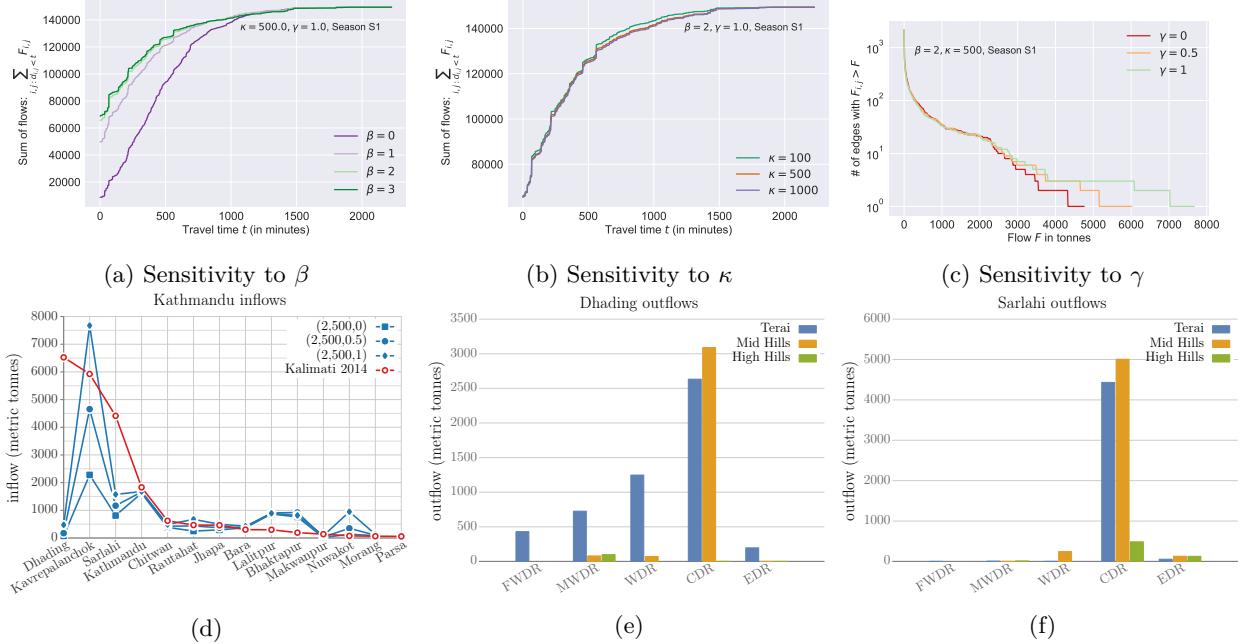


Figure 2: Sensitivity analysis and flow validation

season S1 (the network corresponding to season S2 have similar properties). Except for $\beta = 1$, the flow plateaus for $t > 500$ minutes, which corresponds to ≈ 8 hours of travel time.

For further analysis of the flow network we describe the different regions within Nepal. Nepal has significant altitude variations along the North-South axis, and is divided into three major physiographic regions namely: Terai, Mid-hills and High hills (Figure 3e). For administrative reasons, Nepal has been divided along the East-West axis (Figure 3a) into five major development regions. Kathmandu, for instance, belongs to Mid-hills and Central Development Region. It is useful to remember that the Central Development Region is by far the most economically prosperous, while the population density is high along the Terai region and Kathmandu valley (Table I).

The general trends of tomato trade between markets is depicted in Figure 3 (generated for $\beta = 2$, $\kappa = 500$ and $\gamma = 1.0$). We recall that our model accounts for the fact that the Hills/Mid Hills and the Terai are the primary sources of tomato during seasons S1 and S2 respectively. This is clearly reflected in the net flow diagram between geographic regions: north (Hills/Mid Hills) to south (Terai) in S1 and south to north in S2. However, an interesting pattern to be noted is the significant flow from east to west during S1 as observed in the net flow diagram between the Development Regions. These could be

due to the variability in vegetable production, and the presence of an arterial East-West highway that almost covers the entire breadth of the country.

Comparison with the annual flow network: To evaluate the importance of seasons, we constructed the annual flow network by using the gravity model with annual production for each district. The resulting flows are shown in Figures 3d and 3h. Compared to the seasonal flows we see that annual flows are of shorter distance and thus there is not much flow between regions (either between east and west or south and north).

Sensitivity analysis of the flow network: Figures 2a–2c show the sensitivity of edge weight distribution of season S1 network to β , κ and γ . We find that for $\beta \geq 2$ and $\kappa \geq 500$ the weight distribution is relatively stable. A similar behavior was observed for the season S2 flow network with respect to β and κ . Increasing γ tends to redistribute flows towards high income regions (in this case, regions around Kathmandu in the Mid Hills, Central Development Region, see Figure 3a), and leads to higher maximum flows in the network in season S1, and lower maximum flows in season S2 (not shown here). However, changing γ had minimal effect on most of the low weight edges in the network.

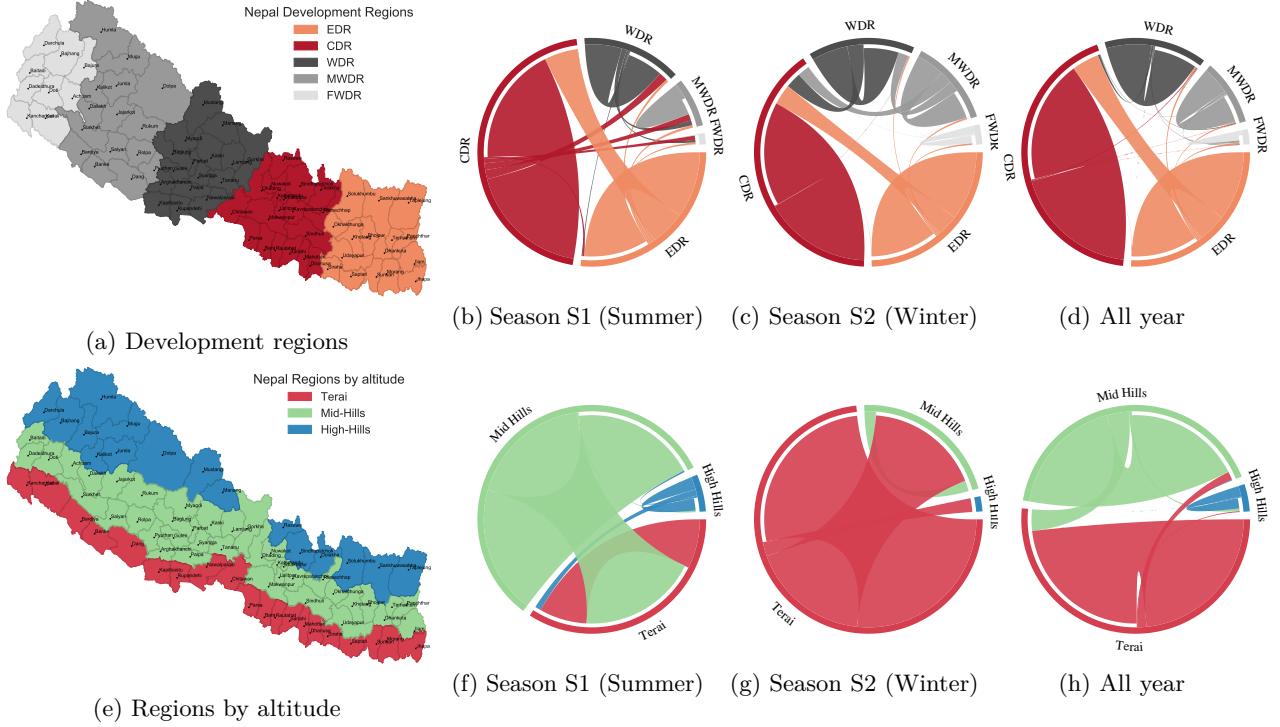


Figure 3: **The spatio-temporal structure of the flow network.** The first row shows the flow from east to west between development regions of Nepal. The second row depicts the flow from north to south between regions of different altitudes. While the second and third columns correspond to seasonal flow, the last column corresponds to the flows generated from annual data.

B. Role of trade network in pest spread

We applied the network diffusion model described in Section II-C to study the role of the flow networks in the spread of *T. absoluta* in Nepal. To interpret the spread model’s output, in terms of incidence reports, we need to translate t of the SI model to a real-world equivalent temporal unit (e.g., month). Validating this requires pest reports at a high spatio-temporal resolution. Since this is absent in the case of *T. absoluta*, to circumvent this problem, we make use of SI model’s monotonic property: For any $t' > t$, $P_S(i, t', f_0) \leq P_S(s, t, f_0)$, and thus the ranking of relative vulnerabilities of market nodes could inform how the process unfolds. We also observed that the rank list is stable (or changes slowly) with respect to t with other parameters fixed (see Table III).

The experiment was setup under the following premise: *T. absoluta* was first introduced to the Kathmandu valley. Ground experts have high confidence in this assumption since the pest was not discovered in the previous growing season in other parts of Nepal. Given the pest reports till December 2016 (Figure 1d), we evaluate our model based on

the following backward inference problem: for an observation of node states at time t , what is the most likely origin of invasion? (also known as the source detection problem [16]). We examine the likelihood of markets or regions being the source nodes, and in particular, we compare this with the likelihood of the region around Kathmandu being the source (see Figure 4). Suppose \mathcal{O} is the observation criteria; it consists of pairs (v, X) where v is a node and $X \in \{S, I\}$ is a state. For each candidate initial condition f_0 , we estimate the joint probability of \mathcal{O} at a time step t , as a product of the marginal probability estimates from the message passing algorithm and define an *energy function* for each tuple (f_0, t) as

$$\phi(\mathcal{O}|f_0, t) = -\log \left(\prod_{(v, X) \in \mathcal{O}} P_X(i, t, f_0) \right).$$

The lower the value of ϕ , the higher the likelihood of f_0 being the initial condition. Secondly, recalling the uncertainty in interpreting time step t , we examined the relative likelihoods of each f_0 and the stability of the ranking across a range of model parameters.

We consider the spread during June–November (season S1) for model evaluation. Using the S1 flow network, our objective was to rank various starting configurations f_0 based on $\phi(\mathcal{O}|f_0, t)$ given \mathcal{O}, t . For a given σ , we evaluated the likelihood of each node being the central node. We considered two criteria based on which the likelihood of each f_0 as the starting configuration was computed: (i) \mathcal{O}_G : this is the set of all pairs (v, I) where v is a market node that belongs to a district that reported pest presence by December 2016. (ii) \mathcal{O}_B : this is the set of (v, I) for all nodes v . This is the baseline which assumes no observational data.

The results are shown in Figure 4. Firstly, we observed that for both criteria \mathcal{O}_G and \mathcal{O}_B , the top few ranks are relatively robust to varying network and model parameters. Also, for both criteria, markets from the Central Development Region (CDR) that belong to Kathmandu and its adjacent districts are among the top ranked nodes. Interestingly, for the criterion \mathcal{O}_G , Dhankuta (EDR), with the highest assigned production has a very low rank (Figure 4a) and a low ϕ value compared to the top market in \mathcal{O}_G . However, for \mathcal{O}_B , it is ranked second (Figure 4b). This clearly shows that while Dhankuta has the potential to infect a large number of areas, given what has been observed, it is very unlikely that it was the source of infection. Dhankuta reported presence of the pest only towards the end of 2016 (see Figure 1d).

Spread in season S2: To study the spread from November 2016 to May 2017, we considered the dynamics on season S2 network. To set the initial conditions, we used the results of our inference study, and chose Kathmandu with $\sigma = 10$ as the seed distribution. For this initial condition, we obtained the probability of infection for all nodes in S1 for T1 time steps. This distribution is used as initial condition for the S2 network spread. Figure 4c shows the infection probabilities for a particular combination of (T_1, T_2) . As seen in Figure 4c, our model suggests that most Terai and Mid Hills regions of CDR, WDR would be affected by the end of May 2017, and subsequent seasons are only going to see increasing incidence of the pest throughout the country. From Figure 1d, we see that regions belonging to Terai in CDR and Mid Hills of WDR and MWDR have already reported pest presence (marked in Figure 4c).

While the intended usage of the origin inference formulation is to determine the source of infection,

we have adapted it to compare expected spread in the model with observed data. Our results demonstrate that this framework is in general very useful in finding the likely pathways of introduction of the pest.

Sensitivity analyses: A full factorial design was performed with levels for the parameters of interest as given in Table III, and analysis of variance (ANOVA) was used to evaluate single parameter effect. It is worth noting that assessment of parameter sensitivity depends on the choice of quantity of interest. Since the outcome of origin inference is a ranking on markets, we used Spearman’s rho to test its stability across the parameter space. The experiment was set up within the GENEUS framework [20], a general computational environment for experimental design, uncertainty quantification and sensitivity analysis.

We studied the sensitivity of individual market ranks as well as rank lists to network parameters (β, κ, γ) , and diffusion model parameters (σ, t) . We found that the market ranks are more sensitive to spatial seeding parameter σ and distance exponent β than other parameters. In particular, we observed that the sensitivity was highest when $\sigma = 0$ was included in the analysis. In this case (and in general for very low values of σ), substantial spread occurs only when the seed node is a source. Even if a node is in close proximity to several sources (such as Kathmandu), there is hardly any spread. This is unrealistic in the context of pest and pathogen dispersal. Hence, we restricted σ to be greater than 0 in our analysis. Also, we observe that the variance in rank is small for higher ranked nodes. This can be seen in Figure 4, and is more pronounced in the single parameter analyses. This property gives higher confidence in interpreting the results on top markets.

We used Spearman’s rank correlation coefficient to analyze the rank stability. Here we use the rank list that results from configuration $(\beta = 2, \kappa = 500, \gamma = 0, \sigma = 5, T = 10)$ as the reference and calculate the Spearman’s rho value with respect to it for rank lists induced by other parameter settings. Table III gives the Analysis of Variance (ANOVA) results. Under 95% confidence level, p -value < 0.05 means that the particular parameter has a significant effect. Therefore, we see that β and σ have significant effects, while others do not. Here, we note that this is despite not considering $\sigma = 0$ in the analysis.

IV. CONCLUSION AND FUTURE WORK

We have described a first-principles based commodity modeling framework that integrates easily available datasets on population, production, etc.

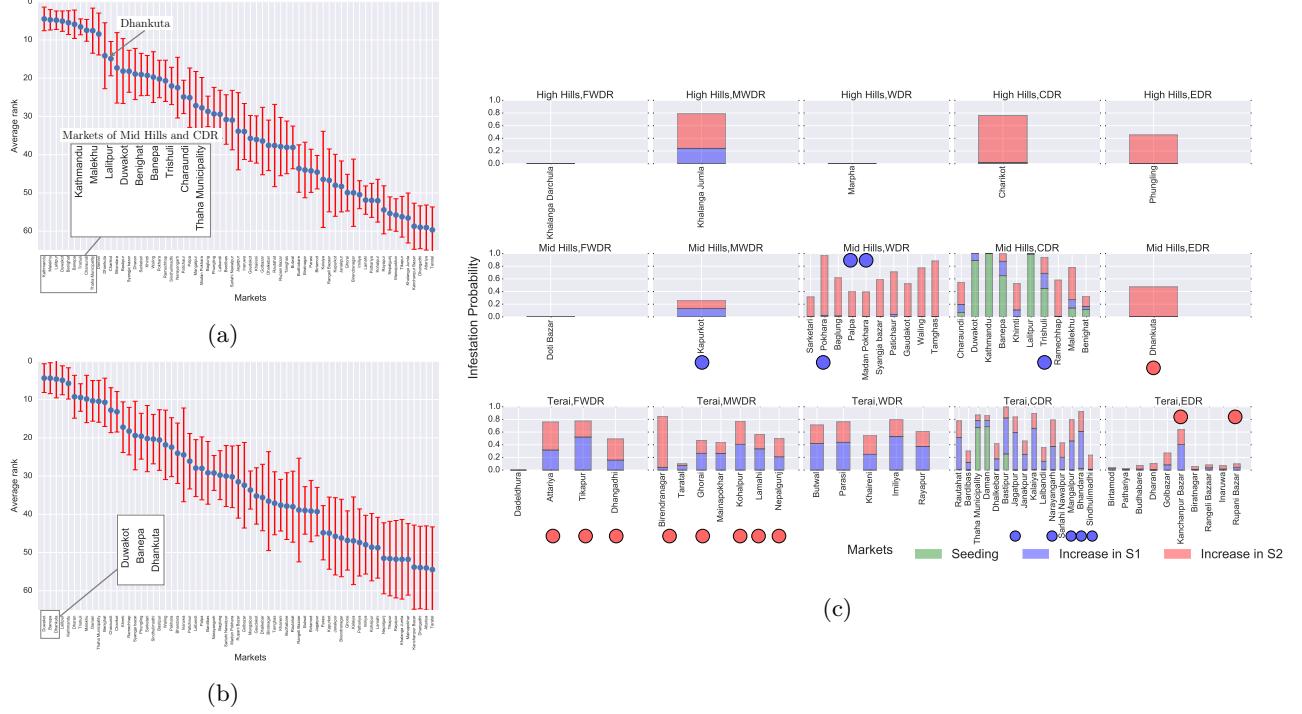


Figure 4: **Evaluating the spread model using epidemic source inference framework.** (a) The average rank of each market based on the likelihood for the criterion \mathcal{O}_G for a range of model parameters (see Table III). (b) Same as (a), but for criterion \mathcal{O}_B . (c) Spread in S2: The parameters used were $\beta = 2$, $\kappa = 500$, $\sigma = 15$, $\gamma = 1$, $T_1 = T_2 = 10$ with Kathmandu as the seed node. The blue dots correspond to markets whose districts reported *T. absoluta* presence before December 2016 (season S1), while the red dots correspond to markets which reported later.

Param.	Levels	t-ratio	F-value	p-value
β	[0, 1, 2]	-5.16	26.6059	< 0.0001
σ	[5, 10, 15, 20]	-3.29	10.8424	< 0.0001
κ	[100, 500, 1000]	-0.42	0.1758	0.6753
γ	[0, 0.5, 1.0]	0.89	0.7970	0.3727
T	[5, 10, 20]	1.14	1.2976	0.2556

Table III: **Analyzing sensitivity to model parameters using ANOVA.**

to model the flow of agricultural produce. We have demonstrated the validity of the constructed networks, and have used it to understand the impact of commodity flow on pest spread. Despite being limited by the availability of quality validation datasets, a bare bones framework such as ours can be quickly extended to other vegetables, pests and regions with minimal effort. Our approach provides a modular framework for integration of other models that can be refined with increased availability of data and sophisticated methods.

Since our study is one of the first to consider regional commodity flow analysis in the context of pest spread, especially *T. absoluta*, there are

several avenues for improvement. While some of the limitations arise from lack of refined data, others are due to the limited understanding of the underlying complexity of pest invasions. The former may be the norm for emerging contagions in a data-poor region, whereas the latter will need several iterations of model development and validation by the scientific community. Our model predominantly focuses on commodity flow, and does not explicitly account for natural or other modes of spread (infected seedlings from nurseries for example). A more comprehensive model will need to integrate ecological suitability and biology directly in the diffusion process.

ACKNOWLEDGMENTS

This work was supported in part by the United States Agency for International Development under the Cooperative Agreement NO. AID-OAA-L-15-00001 Feed the Future Innovation Lab for Integrated Pest Management, DTRA CNIMS Contract HDTRA1-11-D-0016-0001, NSF BIG DATA Grant IIS-1633028, NSF DIBBS Grant ACI-1443054, NIH

Grant 1R01GM109718 and NSF NRT-DESE Grant DGE-154362. G.N. was also partly supported by Virginia Agricultural Experiment Station project VA-136324.

REFERENCES

- [1] A. S. R. Bajracharya, R. P. Mainali, B. Bhat, S. Bista, P. Shashank, and N. Meshram. The first record of South American tomato leaf miner, *Tuta absoluta* (Meyrick 1917)(Lepidoptera: Gelechiidae) in Nepal. *J. Entomol. Zool. Stud.*, 4:1359–1363, 2016.
- [2] N. C. Banks, D. R. Paini, K. L. Bayliss, and M. Hodda. The role of global trade and transport network topology in the human-mediated dispersal of alien species. *Ecology letters*, 18(2):188–199, 2015.
- [3] M. R. Campos, A. Biondi, A. Adiga, R. N. Guedes, and N. Desneux. From the western palaearctic region to beyond: *Tuta absoluta* 10 years after invading europe. *Journal of Pest Science*, pages 1–10, 2017.
- [4] L. Carrasco, J. Mumford, A. MacLeod, T. Harwood, G. Grabenweger, A. Leach, J. Knight, and R. Baker. Unveiling human-assisted dispersal mechanisms in invasive alien insects: integration of spatial stochastic simulation and phenology models. *Ecological Modelling*, 221(17):2068–2075, 2010.
- [5] M. Colunga-Garcia, R. A. Haack, and A. O. Adelaja. Freight transportation and the potential for invasions of exotic insects in urban and periurban forests of the united states. *Journal of Economic Entomology*, 102(1):237–246, 2009.
- [6] N. J. Cunniffe, B. Koskella, C. J. E. Metcalf, S. Parnell, T. R. Gottwald, and C. A. Gilligan. Thirteen challenges in modelling plant diseases. *Epidemics*, 10:6–10, 2015.
- [7] R. Early, B. A. Bradley, J. S. Dukes, J. J. Lawler, J. D. Olden, D. M. Blumenthal, P. Gonzalez, E. D. Grosholz, I. Ibañez, L. P. Miller, et al. Global threats from invasive alien species in the twenty-first century and national response capacities. *Nature Communications*, 7, 2016.
- [8] M. Ercsey-Ravasz, Z. Toroczkai, Z. Lakner, and J. Baranyi. Complexity of the international agro-food trade network and its impact on food safety. *PloS one*, 7(5):e37810, 2012.
- [9] R. Y. Guimapi, S. A. Mohamed, G. O. Okeyo, F. T. Ndjomatchoua, S. Ekesi, and H. E. Tonnang. Modeling the risk of invasion and spread of *Tuta absoluta* in Africa. *Ecological Complexity*, 28:77–93, 2016.
- [10] P. E. Hulme. Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology*, 46(1):10–18, 2009.
- [11] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius. The complex network of global cargo ship movements. *Journal of the Royal Society Interface*, 7(48):1093–1103, 2010.
- [12] A. Y. Lokhov, M. Mézard, H. Ohta, and L. Zdeborová. Inferring the origin of an epidemic with a dynamic message-passing algorithm. *Physical Review E*, 90(1):012801, 2014.
- [13] J. F. H. Nopsa, G. J. Daglish, D. W. Hagstrum, J. F. Leslie, T. W. Phillips, C. Scoglio, S. Thomas-Sharma, G. H. Walter, and K. A. Garrett. Ecological networks in stored grain: Key postharvest nodes for emerging pests, pathogens, and mycotoxins. *BioScience*, page biv122, 2015.
- [14] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015.
- [15] C. Robinet, H. Kehlenbeck, D. J. Kriticos, R. H. Baker, A. Battisti, S. Brunel, M. Dupin, D. Eyre, M. Faccoli, Z. Ilieva, et al. A suite of models to support the quantitative assessment of spread in pest risk analysis. *PLoS One*, 7(10):e43366, 2012.
- [16] D. Shah and T. Zaman. Rumors in a network: Who’s the culprit? *IEEE Transactions on information theory*, 57(8):5163–5181, 2011.
- [17] A. J. Tatem. The worldwide airline network and the dispersal of exotic species: 2007–2010. *Ecography*, 32(1):94–102, 2009.
- [18] H. E. Tonnang, S. F. Mohamed, F. Khamis, and S. Ekesi. Identification and risk assessment for worldwide invasion and spread of *Tuta absoluta* with a focus on Sub-Saharan Africa: implications for phytosanitary measures and management. *PloS one*, 10(8):e0135283, 2015.
- [19] USAID/Nepal. Value Chain/Market Analysis of the vegetable Sub-Sector in Nepal. 2011.
- [20] S. Wu, H. Mortveit, and S. Gupta. A Framework for Validation of Network-based Simulation Models: an Application to Modeling Interventions of Pandemics. In *Proceedings of ACM SIGSIM Conference on Principles of Advanced Discrete Simulation*. ACM, 2017.



From the Western Palaearctic region to beyond: *Tuta absoluta* 10 years after invading Europe

Mateus R. Campos^{1,2} · Antonio Biondi³ · Abhijin Adiga⁴ · Raul N. C. Guedes² · Nicolas Desneux¹

Received: 7 February 2017 / Revised: 12 April 2017 / Accepted: 18 April 2017 / Published online: 5 May 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract The South American tomato pinworm, *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae), is a devastating pest currently threatening the global tomato industry worldwide. In the last 10 years, it has spread and expanded to most of Europe, Africa and Asia, causing extensive damage to the crop itself and to the international tomato trade. With the aim of providing an overview of the current knowledge on this pest, we have briefly reviewed the available literature relying on its spread, quarantine, modeling and management. Finally, we have underlined the gaps in knowledge and provided several recommendations on how to achieve sustainable control as well as how to prevent further spread into unaffected areas.

Communicated by M. Traugott.

Electronic supplementary material The online version of this article (doi:[10.1007/s10340-017-0867-7](https://doi.org/10.1007/s10340-017-0867-7)) contains supplementary material, which is available to authorized users.

✉ Antonio Biondi
antonio.biondi@unitn.it

Nicolas Desneux
nicolas.desneux@sophia.inra.fr

¹ INRA (French National Institute for Agricultural Research), Université Côte d'Azur, CNRS, UMR 1355-7254, Institut Sophia Agrobiotech, 06903 Sophia Antipolis, France

² Departamento de Entomologia, Universidade Federal de Viçosa, Viçosa, MG 36570-900, Brazil

³ Department of Agriculture, Food and Environment, University of Catania, via Santa Sofia 100, 95123 Catania, Italy

⁴ Biocomplexity Institute, Virginia Tech, Blacksburg, VA 24061, USA

Keywords South American tomato pinworm · Invasive pest · Biological control · Integrated pest management · Quarantine

Key message

- We have summarized the available information on *Tuta absoluta* 10 years after it was documented outside of its native range
- *Tuta absoluta* has increased its range radius by 800 km per year, infesting almost 60% of the tomato cultivated land worldwide, causing extensive direct and indirect damage
- Various phytosanitary measures, quarantine restrictions and early warning actions are being implemented in unaffected areas
- In the invaded regions, various non-chemical control packages have been set up and need to be prioritized over reiterative applications of broad-spectrum insecticides

Introduction

Trade and introduction of alien species have been intertwined since the dawn of the Industrial Revolution (Kolar and Lodge 2001; Schmitz and Simberloff 1997). However, with the advent of globalization, the current levels of international trade and human travel have not only brought the risk of introduction of invasive species to unprecedented levels (Essl et al. 2011) but also act as conduits for their rapid spread across the landscape (Hulme 2009; Lowry et al. 2013; Meyerson and Mooney 2007). While such typical anthropogenic activities favor biological

invasion, this phenomenon is further aggravated by current global climate changes. These factors often determine successful invasions while affecting the patterns of species distribution and resource dynamics in the environment (Battisti and Larsson 2015; Dukes and Mooney 1999; Hellmann et al. 2008). As a consequence, the concern about invasive species, i.e., particularly invasive arthropod pests, has increased in recent years (Biondi et al. 2016a; Haye et al. 2015; Ragsdale et al. 2011; Keller et al. 2011). Among the invasive arthropod pest species drawing attention, the South American tomato pinworm *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) is one of the most harmful pests of solanaceous crops and represents a major threat to tomato production worldwide (Desneux et al. 2010, 2011; Guedes and Picanço 2012).

The tomato pinworm was observed for the first time outside South America, in eastern Spain, a little over 10 years ago. Currently, it can be found throughout Europe, Africa, the Middle East and parts of Asia with risks of outbreaks in bordering countries (Fig. 1). The economic and ecological problems caused by the tomato pinworm, its rapid dispersion and the difficulty in controlling it, especially in these past 10 years, has raised the level of concern of the principal producing countries, exporters and importers of tomatoes even further afield (Desneux et al. 2010, 2011). Some of these countries are still free from the tomato pinworm, thus requiring quarantine initiatives and further phytosanitary measures to minimize the risk of

introduction. The need to determine the genotype of the spreading of *T. absoluta* is also relevant for gaining knowledge on the genetic basis of traits that enables successful invasion (Cifuentes et al. 2011). Detailed records on the spread of invasive pests and pathways of introduction are essential in order to understand the potential range and economic importance of invasive species (Guillemaud et al. 2015), as well as to be able to predict future trends and to identify management options (Asplen et al. 2015; Haye et al. 2016; Tonnang et al. 2015). In this paper, we have summarized and updated the information available about the tomato pinworm, including its economic impact, the recently invaded regions, current management efforts and the need for further future research.

Spread dynamics in invaded areas

Knowledge of the movement pathways of specific pest species is fundamental in reducing the likelihood of their introduction into new environments, forecasting trends and in identifying the most suitable and sustainable management options (Cesari et al. 2015; Roques et al. 2016). Recently, molecular evidence about the Old-World invasion by the tomato pinworm has been gained, suggesting that central Chile is the sole and very likely origin of the introduction of this species into Europe (Guillemaud et al. 2015). Spreading from its native range in the Peruvian

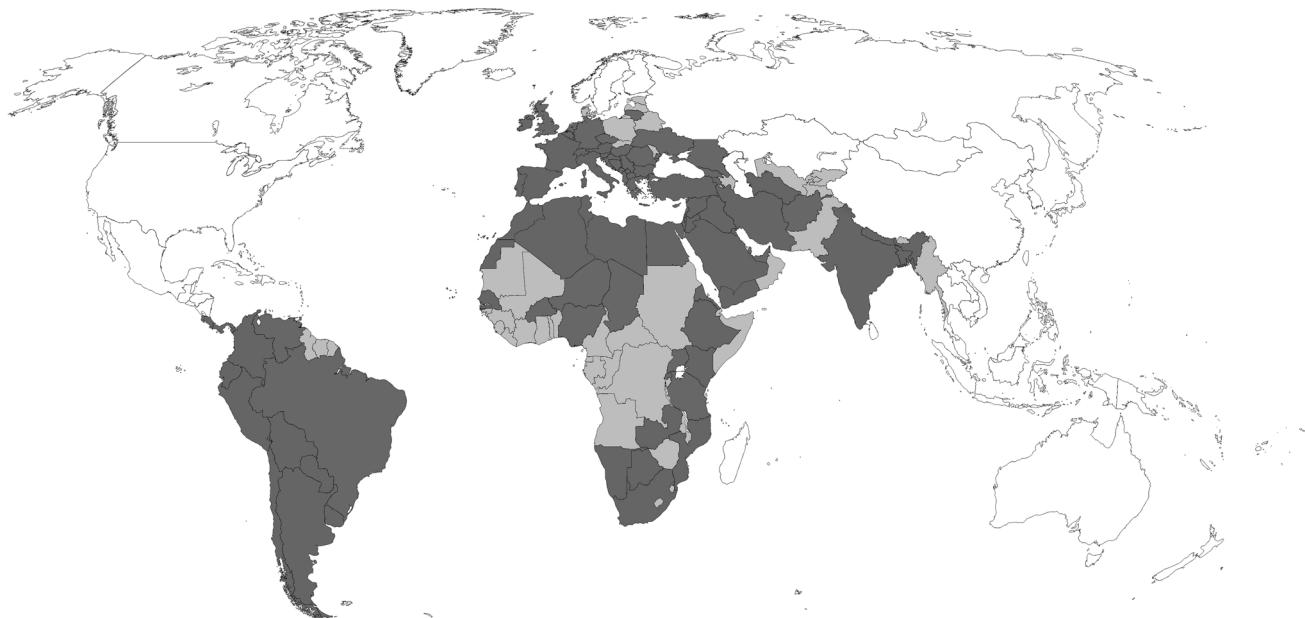


Fig. 1 Current worldwide *Tuta absoluta* distribution map (as of April 2017). Countries are indicated as follows: (1) *T. absoluta* presence has been confirmed (dark gray), and (2) *T. absoluta* could be considered present because of geographical and ecological proximity, or because the presence has not yet been confirmed after an initial

record (light gray). The information provided is based on a compilation of reports from plant protection services and extension specialists (e.g., US Federal Order July 2014, EPPO reporting services), on published scientific articles and on personal communications (see also Table 1S for more details)

central highlands, the tomato pinworm expanded into other regions and countries of Latin America countries between the 1960s and 1980s, including Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Panama, Paraguay, Peru, Uruguay and Venezuela (USDA–APHIS 2014).

Invasive species usually spread very fast on most continents generally around 5 and 10 years after having been recorded in the invaded area. A recent evaluation on the spread of the tomato pinworm in Europe has shown that *T. absoluta* after first being recorded in the invaded area increased its radius by around 600 km per year over a period of 9 years (Roques et al. 2016). The rapid dispersion and the ecological and economic impacts on the countries after infestation by the species has increased the level of concern in tomato-producing countries free from this species (Fig. 1). This pest was first recorded outside Neotropical America in Spain in 2006, and within a few years it spread into most countries of the Mediterranean basin, besides North and Central Europe (Desneux et al. 2011) (Table S1). More recently, *T. absoluta* has invaded several Middle East countries, e.g., Afghanistan, and some sub-Saharan areas, from Gambia to South Africa, threatening tomato crop sustainability in these regions (Brévault et al. 2014; Pfeiffer et al. 2013; Tonnang et al. 2015).

Upon reaching the northern coast of Africa and the Sub-Saharan region, the pest started moving swiftly southwards, thanks to the spatial continuity of vegetable cultivation across political borders, the absence of effective surveillance mechanisms, the lack of specific phytosanitary expertise to intercept infested vegetables, ever-growing tourism and increasing intra-continental trade (Tonnang et al. 2015). The most recent African records are from southern countries, such as Zambia, Mozambique, Botswana and South Africa (Table S1). Nevertheless, records from several other African countries still need to be checked (Fig. 1; Table 1S). By 2017, it has been estimated the whole African continent will be totally invaded (Guimapi et al. 2016). By early 2014, *T. absoluta* was reported in southern India (Kallewaraswamy et al. 2015; Sridhar et al. 2014). In early 2016, the moth was found infesting tomato farms in the Kathmandu valley of Nepal (NARC 2016). In addition, reports have confirmed the pest north of Bangladesh (Hossain et al. 2016), thus showing how far east the tomato pinworm has so far spread.

Damage and economic impact

The economic impact of invasive pest species is directly reflected in the rising costs of crop production, namely additional costs for pest management, in the decrease of

marketable products, and in the potential loss of trading partners through restrictions on export to non-infested countries. Moreover, as observed in many countries, it is very difficult to control and limit the spread of the invasive pest (Garnas et al. 2016; Keller et al. 2011). The tomato pinworm is a multivoltine species, showing high reproductive potential because of its adaptability which allows the pest population to increase very quickly (Tropea Garzia et al. 2012). The high reduction of tomato production is mainly due to the severe degree of tomato pinworm attacks, which mine the leaves, affecting the tomato plant photosynthetic capacity. Indeed, wandering larvae can bore through the stems, thus killing young plants which in turn compromise yield, or occasionally feed on fruits, increasing the costs for post-harvest selection (Galdino et al. 2015). Indirect damage can also be caused by secondary infections, with pathogens developing on the infested plant and fruit tissues (Tropea Garzia et al. 2012). The economic impact based on crop losses has been estimated at between €5 and 25 million per year in the Netherlands (Potting et al. 2013).

Control measures against the tomato pinworm are thus often used and chemical control is the most common tool used to suppress populations that suddenly grow in the open-field tomato crops (Guedes and Picanço 2012). In a worst-case scenario, an estimated average increase of 13–15 extra insecticide applications are necessary to fully control *T. absoluta* in the Netherlands, with an estimated increase in pest control-related costs of €4 million per year (Potting et al. 2013). In Spain and Tunisia, 15 and 18 active insecticide molecules, respectively, were specifically introduced to target *T. absoluta* between 2009 and 2011 (Abbes et al. 2012; Desneux et al. 2011). In South America, the introduction of *T. absoluta* in countries bordering the origin country has led to sudden increases in insecticide use in tomato fields from 12 to over 30 applications per cultivation period (Guedes and Picanço 2012; Gontijo et al. 2013). Such a huge insecticide selection pressure is obviously not compatible with most of the integrated pest management (IPM) programs implemented in the tomato systems. The increasing number of insecticide applications is mainly due to the problems linked to insecticide resistance in *T. absoluta* populations (Campos et al. 2014; Roditakis et al. 2015). It is also the consequence of potential disturbances of ecological services provided by pollinators and natural enemies exposed to the sublethal exposure of pesticides (Abbes et al. 2015; Barbosa et al. 2015; Biondi et al. 2012a, 2013a; Tomé et al. 2012, 2015). What is more, secondary pest chronic exposure to sublethal concentrations of pesticides can trigger unforeseeable consequences, such as pest outbreaks (Desneux et al. 2007; Guedes et al. 2016).

Current threat to global tomato industry

After 10 years outside Neotropical America, the tomato pinworm has already infested 2.8 million ha of tomato crop, i.e., more than half the land devoted to cultivating tomatoes (FAOSTAT 2012). Among the ten largest world tomato producers, six countries, namely India, Turkey, Egypt, Iran, Italy and Spain, have already been infested by this pest, and around 50% of the world tomato production is marketed among infested countries (FAOSTAT 2012). Three countries, China, India and Turkey, account for almost half of the land area covered worldwide with tomato crops i.e., 31, 11 and 7%, respectively. India is the second largest tomato producer in the world, 17.5 M tons, and exhibiting a tomato cultivated area of 870 M ha (FAOSTAT 2012). Despite efforts to prevent the *T. absoluta* invasion in China, the risk of introduction into this country within the next few years has increased considerably, as previously reported (Desneux et al. 2011). In addition to China's neighboring countries, Afghanistan and India, newly infested countries bordering China, such as Tajikistan and Kyrgyzstan, are currently under high risk of tomato pinworm outbreaks (Fig. 1).

The tomato acreage in Europe is half a million hectares, 11% of the global tomato area cultivated (FAOSTAT 2012). Most of Europe has recorded the presence of *T. absoluta*, except Moldova and central and eastern Russia, which currently remain free. Russia has the largest tomato cultivated area in Europe (117 M ha) (FAOSTAT 2012), but the presence of tomato pinworm has been restricted to western Russia since 2010 (Izhevskya et al. 2011). Although the tomato pinworm does not survive low temperatures, it can successfully overwinter in warmer environments, such as protected crops and indoor warehouses (Tonnang et al. 2015; USDA-APHIS 2011).

The African continent produces 18 M tons of tomatoes, which corresponds to 12% of the world production as of 2012. The biggest tomato producers in Africa are Nigeria, Egypt, Morocco and Tunisia; these four countries alone produced more than 70% of African tomato in 2012 (FAOSTAT 2012). Recently, the governments of northern Nigerian states declared an emergency after the tomato pinworm destroyed more than 80% of the tomato fields in the region, causing a 20-fold increase in the tomato price (FAO 2015). Due to the pattern of climate similarity with South America and the rapid spread of the tomato pinworm in recent years, all countries in the African continent are expected to record the presence of this pest species within 2018 (Tonnang et al. 2015).

Central America produces 4 M tons of tomatoes accounting for more than 16% of the total production in the Americas covering an area of 113 M ha. The spread of the

tomato pinworm is still taking place in the continent, and its presence in Costa Rica was reported in 2014 (EPPO 2014), increasing the quarantine concern in Central and North America (USA and Canada). North and Central America, mainly Mexico, are seeking data about the dispersion dynamics and sustainable management options for the tomato pinworm (NAPPO 2013). The invasions of this pest in new regions appears linked to the import of tomato fruits, as in confirmed cases in Eurasia after 2010, e.g., the Netherlands and Russia (Desneux et al. 2011; Karadjova et al. 2013). The United States accounts for 25% of the world tomato imports of the tomato, and in 2015 such imports were received from different countries reaching a total of 1.5 M tons (Trade Map 2016). A fact worth mentioning is that in 2014 few of the US tomato imports came from tomato pinworm-infested countries, i.e., 0.01% or 245 tons, including tomatoes from the Netherlands, Spain, Ecuador, Italy and Colombia (Trade Map 2016).

In 2005 and 2006, Spain imported 23 and 29 tons of fresh tomatoes, respectively, from Chile, whose problems with tomato pinworm infestation date back from the 1960s. This, together with recent molecular evidence (Guillemaud et al. 2015), suggest that exported Chilean tomatoes are the most likely source responsible for the European invasion. Chile was also the source of the earliest report on insecticide resistance in this species, followed by Brazil and Argentina (Guedes and Siqueira 2012). Thus, the introduction of insecticide-resistant genotypes of the tomato pinworm into Europe probably occurred at the outset. By contrast, tomato trading could not be the sole cause for the Indian invasion, which took place in late 2014, because this country stopped importing tomatoes by late 2013 (Trade Map 2016). Therefore, the tomato pinworm presence may have remained undetected in India for a few years. The pest could have entered the country through (1) the importation of other commercial goods, including alternative host plant species or used packing materials, or (2) due to workers commuting between Middle Eastern countries and the Indian subcontinent.

Management efforts

Substantial economic and ecological damage was created by international trade and global environmental change which are significant components of the invasion of ecological systems by exotic species (Garnas et al. 2016). Policies against invasive species often involve unilateral defensive actions as opposed to coordinated international initiatives. However, an important step towards preventing threats from such infestation is the coordination and global cooperation in the management of phytosanitary risks

(Perrings et al. 2010). Trade patterns relative to pest species distribution and commodity production locations should be the main drivers of phytosanitary actions. However, pests vary among countries in their perceived risk to the agricultural industries and to the environment.

The International Plant Protection Convention of the United Nations Food and Agriculture Organization (UN/FAO) provides guiding principles applicable to phytosanitary importation and exportation activities (FAO 2006). Countries associated with the North America Plant Protection Organization (NAPPO) also selected appropriate phytosanitary measures that are important to minimize further spread and establishment of the tomato pinworm (USDA–APHIS 2014). The established measures were based on considerations followed by the International Standards for Phytosanitary Measures (ISPM) which guides the import and export of goods (NAPPO 2013). ISPM requirements are adapted to prevent spread of the organism via import and internal movement of plants and fruits of Solanaceae, especially tomato. In case of fruits originating from infested areas, procedures are required to sanitize them, thus preventing adult moths from escaping during the fruit transportation or at the time of its unloading; hermetic packaging conditions, refrigeration or netting are procedures used for the purpose (FAO 2006). The tomato pinworm is a major concern for the United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine (USDA–APHIS–PPQ), which has imposed quarantine restrictions for tomato import into the country. In response to this pest outbreaks in the EU, the US and Canadian authorities imposed strict import conditions for tomatoes from several EU countries and is also highly likely to impose quarantine restrictions on export of tomatoes from West and Central African countries (ECOWAS) (Pfeiffer et al. 2013; USDA–APHIS 2014; van der Straten et al. 2011).

Tomato fruit, either green, red or pink, produced in Chile and which are imported to US must be subjected to methyl bromide fumigation. However, this treatment can only be applied in a preclearance program. This implies conjunction of international services, preclearance inspections, treatments and/or other mitigation measures procedures which are designed to identify and/or mitigate the risk of invasive pest introductions through action taken in foreign countries under the supervision of APHIS personnel. Among the countries known to be infested with the tomato pinworm, only Chile has implemented a preclearance program (USDA–APHIS 2013). Nonetheless, fumigant applications and particularly methyl bromide use are currently facing phasing out due to its ozone-depletion characteristic. An alternative method in use which is internationally accepted for different fruits and vegetables is irradiation treatment

protocols (Hallman et al. 2013). Ionizing radiation use as a phytosanitary treatment has increased against invasive pest species of fruits and vegetables in various countries, including Australia, New Zealand, and Malaysia (Hallman et al. 2013).

The use of pheromone traps is considered the major tool for early detection of this pest in newly invaded areas. Between 2010 and 2012, the US State of Florida implemented a *T. absoluta* detection network in 29 cities using traps baited with the sexual pheromone (NAPPO 2013). In Bosnia and Herzegovina and Montenegro, a molecular protocol to detect *T. absoluta* DNA from trapped individuals and from infested fruits was successfully tested (Đurić et al. 2014). Besides the high cost to implement the use of pheromone traps, decision-making studies focusing on the control of *T. absoluta* in infested crops are still linked to plant sampling (Cocco et al. 2015). Therefore, more studies are still needed to enable proper sampling when pheromone traps are used next to the plants in order to take the necessary decisions. Nevertheless, the use of synthetic pheromones for male mass-trapping or for mating disruption gave unsatisfactory results. As a result of such shortcomings, the detection of parthenogenesis among specimens of the tomato pinworm has given rise to further concerns (Caparros Megido et al. 2012, 2013; Cocco et al. 2013).

The first years following detection of the tomato pinworm and the awareness of its impact on tomato crops in Europe and North Africa have led to the extensive use of insecticides by growers in these regions (Desneux et al. 2011). This is similar to what was previously observed in South America (Guedes and Picanço 2012; Guedes and Siqueira 2012). However, insecticide applications are not always as efficient as expected to control this species. Insecticide efficacy is usually compromised due to the cryptic nature of the larvae, the high reproduction potential and the long tomato reproductive stage where the fruits remain exposed to direct injury (Guedes and Picanço 2012). In addition, several cases of insecticide resistance have been reported, including resistance to organophosphates, pyrethroids, abamectin, cartap, anthranilic diamides and spinosad (Campos et al. 2014, 2015; Haddi et al. 2012; Roditakis et al. 2015, 2017; Silva et al. 2016; Siqueira et al. 2000).

The susceptibility of the tomato pinworm to insecticides has been monitored over the past few years in Europe (Roditakis et al. 2013), following previous South American studies (as reviewed by Guedes and Picanço 2012), and subsequent intensification of tomato pinworm resistance to different active ingredients (Campos et al. 2014, 2015; Silva et al. 2016). In Italy, after <5 years of the *T. absoluta* arrival, the pest population was more than 1000-fold resistant to chlorantraniliprole and flubendiamide, as well

as showing cross-resistance to both compounds (Roditakis et al. 2015, 2017). Besides the resistance of this species to pesticides, there are potential side effects on natural enemies that can induce multiple sublethal effects in individuals that survive pesticide exposure (Biondi et al. 2012b, 2013a; Desneux et al. 2007). Pollinators are also the unintentional targets of pesticide use in tomatoes, where pollination enhances fruit production and pesticide use can compromise such economic and environmental service (Barbosa et al. 2015; Biondi et al. 2012a; Lima et al. 2016; Tomé et al. 2012, 2015). As a consequence, non-chemical tools have to be prioritized.

IPM programs for the tomato pinworm management under development and/or implementation in several countries include tools for the pest prevention, monitoring and sustainable control. After the initial detection of the tomato pinworm, the elimination of symptomatic leaves and destruction of infested tomato plants are highly recommended against this pest, as well as environmental management practices such as tilling, manuring, irrigation, crop rotation and solarization (Terzidis et al. 2014). In greenhouses, one of the management practices used to reduce initial tomato pinworm infestations is to keep infested greenhouses closed after the harvest to prevent the migration of adults to open-field crops or to an alternative host (Derbalah et al. 2012; Desneux et al. 2011). Alternating host crops, mainly tomato and potato, with non-host cultures can ensure long-term reduction of pest pressure (Terzidis et al. 2014; Tropea Garzia et al. 2012). Nonetheless, area-wide tomato cultivation throughout the year compromises the potential effectiveness of this recommendation, which is also limited by the year-round existence of alternative hosts (Guedes and Picanço 2012).

Among the agronomic techniques experimented to control *T. absoluta*, several tomato accessions and cultivars (Oliveira et al. 2009; Sohrabi et al. 2016), as well as the nitrogen and water management (Han et al. 2014, 2016a), have been shown to impact the pinworm population dynamics. However, caution should be exercised in using IPM programs against *T. absoluta* when irrigation in tomato crops is performed with saline water (Han et al. 2016b). This knowledge has not been incorporated into *T. absoluta* IPM packages. By contrast, biological control against this pest has been widely implemented using indigenous natural enemies (Zappalà et al. 2013). Various predators and parasitoids have been reported naturally attacking the tomato pinworm in tomato crops in Europe and North Africa and more recently in India and South America (Biondi et al. 2013b; Calvo et al. 2016; Perdikis and Arvaniti 2016; Salehi et al. 2016). Some efforts have targeted the activity of the *T. absoluta* egg (Cabello et al. 2012; Chailleur et al. 2012; El-Arnaouty et al. 2014) and larval parasitoids (Biondi et al. 2013c; Chailleur et al.

2014). However, the most cost-effective and successful biological control programs are those based on the augmentation and/or conservation of mirid bugs, such as the omnivorous *Macrolophus pygmaeus* and *Nesidiocoris tenuis*, employed alone or in combination with parasitoids (Chailleur et al. 2013a, b; De Backer et al. 2015; Mollá et al. 2014; Naselli et al. 2017) or with selective pesticides (Zappalà et al. 2012).

Among the biopesticides used on tomato crops, commercial formulations based on *Bacillus thuringiensis* (*Bt*) and *Beauveria bassiana* are considered the most efficient and selective options, as they are highly compatible with the tomato pinworm natural enemies (Biondi et al. 2012b, 2013a; Klieber and Reineke 2016). Further studies should investigate possibilities for improving natural enemy ecosystem services via conservation and augmentation strategies, which would help their early establishment and persistence during the cropping season by providing additional food sources, and/or alternative plants and/or non-pest host sources (Biondi et al. 2016b; Urbaneja-Bernat et al. 2015). Nevertheless, in the pest native range, high natural enemy biodiversity has been recorded feeding and developing on this prey/host (Desneux et al. 2010; Salas Gervassio et al. 2016), but effective, coevolved and specialized natural enemy species have not yet been identified. Therefore, despite classical biological control programmes are being pursued for controlling several invasive insect pests in other agroecosystems (Daane et al. 2015, 2016; Desneux et al. 2012), this strategy is not currently considered as a priority for *T. absoluta* in Africa and Eurasia.

Modeling efforts

From the perspective of ecological models, there have been two CLIMEX-based approaches to determine regions of suitability for *T. absoluta* induced by current and future climatic conditions, one focusing on the Mediterranean basin (Desneux et al. 2010), and the other on the global distribution (Tonnang et al. 2015). The latter combines molecular characterization and geo-reference of incidence records in conjunction with ecological parameters to estimate potential spread of *T. absoluta*, mainly in sub-Saharan Africa. More recently, Guimapi et al. (2016) have developed a cellular automata model to study the dynamics of the pest in Africa. They have taken several factors into account such as: vegetation, temperature, humidity, production data, and have been able to estimate the invasion time and pattern for the whole of Africa. Integrated surveys are also the Physiologically Based Demographic Models (PBDM) as proposed by Ponti et al. (2015) to study the spread of *T. absoluta* in the Mediterranean area. Though

literature contains mechanistic models for tomato growth, knowledge gaps in *T. absoluta* biology that need to be filled have been identified, in order to develop an effective PBDM. In the case of *T. absoluta*, however, any approach that only takes into account the ecological factors is inadequate. While these models provide accurate predictions on the potential risk of invasion and establishment, they do not throw light on possible pathways of introduction. Particularly in the case of *T. absoluta*, anthropogenic factors such as trade, travel and the global proliferation of greenhouses (main overwintering sites in cooler areas) are the leading causes for the rapid spread and survival of the pest (Desneux et al. 2011). Researchers are increasingly acknowledging the need for more holistic approaches towards modeling invasive species dynamics (Venette 2015). However, unlike ecological models, research on integrated models, especially those representing human-mediated factors, are in process of being developed, although well-defined methods for data acquisition and model design are lacking.

Conclusion and future outlook

A decade or so ago, *T. absoluta* arrived in Europe from South America, and within this period it has spread to almost the whole of Africa and Eurasia. It remains a significant threat to tomato production because it has reached the borders of the main tomato producing, exporting and importing countries, thus increasing the odds of the approaching introduction and subsequent outbreaks, particularly in China, Mexico and the US. Given the aggressive nature of the pest and the ever-increasing commercial travel, the introduction of the tomato pinworm into unaffected countries is an enormous and constant threat. Nevertheless, quarantine and phytosanitary measures, e.g., ionizing radiation could serve as important deterrents, and should be adopted according to international trade recommendations.

The tomato pinworm might be controlled effectively by devising an integrated approach which keeps in mind the host and climatic conditions of the geographical area. The integrated approach for pest prevention, monitoring and sustainable control include: (1) inspection and careful installation of pheromone traps in countries bordering areas exhibiting high risk of introduction; (2) highly recommended methods are the elimination of symptomatic leaves and destruction of infested tomato plants; (3) the development of resistant tomato varieties to minimize tomato pinworm infestations and losses, and (4) biological control implementation through the scouting of effective and specialized natural enemies in the native range and by the conservation and augmentation of locally available

biocontrol agents in combination with selective insecticides. The knowledge and control methods studied are of great relevance not only to infested countries but also to the uninfested ones threatened by the imminent invasion of the tomato pinworm. The outcome would be enhanced human and environmental safety maintaining the affordability of tomatoes, the main horticultural product worldwide.

Author contribution statement

AB and ND conceived and designed the manuscript. MRC, AB and ND organized the manuscript based on written contributions from all authors. All authors wrote, read and approved the manuscript.

Acknowledgements Funding was provided by the EU FP7-IRSES ASCII project No. 318246 (to AB, ND and MRC), the H2020-SFS EUCLID project No. 633999 (to ND), by the EU FP7 for research, technological development and demonstration, ARIMNet2 No. 618127 (to AB and ND); the Italian Ministry of Education, University and Research (SIR project, No. RBSI14I02A) (to AB), the CAPES Foundation to (MRC and RNCG), the National Council of Scientific and Technological Development (CNPq) (to RNCG) and the USAID Cooperative Agreement No. AID-OAA-L-15-00001 (to AA) and IPM Innovation Lab Insect Modeling Project No. AID-OAA-14-000018 (to AA and ND). Authors are also grateful to R. Muniappan (Virginia Tech) and D. A. Muruvanda (USDA-APHIS-PPQ) for information on monitoring and quarantine efforts, and to four anonymous reviewers and the editor for their helpful comments on the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Abbes K, Harbi A, Chermiti B (2012) The tomato leafminer *Tuta absoluta* (Meyrick) in Tunisia: current status and management strategies. EPPO Bull 42:226–233
- Abbes K, Biondi A, Kurtulus A, Ricupero M, Russo A, Siscaro G, Chermiti B, Zappalà L (2015) Combined non-target effects of insecticide and high temperatures on the parasitoid *Braccon nigricans*. PLoS ONE 10(9):e0138411
- Asplen MK, Anfora G, Biondi A et al (2015) Invasion biology of spotted wing Drosophila (*Drosophila suzukii*): a global perspective and future priorities. J Pest Sci 88:469–494
- Barbosa WF, Smagghe G, Guedes RNC (2015) Pesticides and reduced-risk insecticides, native bees and pantropical stingless bees: pitfalls and perspectives. Pest Manag Sci 71:1049–1053
- Battisti A, Larsson S (2015) Climate change and insect pest distribution range. In: Björkman C, Niemelä P (eds) Climate change and insect pests. CABI, Padua, p 279
- Biondi A, Mommaerts V, Smagghe G, Viñuela E, Zappalà L, Desneux N (2012a) The non-target impact of spinosyns on beneficial arthropods. Pest Manag Sci 68:1523–1536

- Biondi A, Desneux N, Siscaro G, Zappalà L (2012b) Using organic-certified rather than synthetic pesticides may not be safer for biological control agents: selectivity and side effects of 14 pesticides on the predator *Orius laevigatus*. Chemosphere 87:803–812
- Biondi A, Zappalà L, Stark JD, Desneux N (2013a) Do biopesticides affect the demographic traits of a parasitoid wasp and its biocontrol services through sublethal effects? PLoS ONE 8:e76548
- Biondi A, Chailleur A, Lambion J, Han P, Zappalà L, Desneux N (2013b) Indigenous natural enemies attacking *Tuta absoluta* (Lepidoptera: Gelechiidae) in Southern France. Egypt J Biol Pest Control 23:117–121
- Biondi A, Desneux N, Amiens-Desneux E, Siscaro G, Zappalà L (2013c) Biology and developmental strategies of the Palaearctic parasitoid *Bracon nigricans* (Hymenoptera: Braconidae) on the Neotropical moth *Tuta absoluta* (Lepidoptera: Gelechiidae). J Econ Entomol 106:1638–1647
- Biondi A, Traugott M, Desneux N (2016a) Special issue on *Drosophila suzukii*—from global invasion to sustainable control. J Pest Sci 89:603–604
- Biondi A, Zappalà L, Di Mauro A, Tropea Garzia G, Russo A, Desneux N, Siscaro G (2016b) Can alternative host plant and prey affect phytophagy and biological control by the zoophytophagous mirid *Nesidiocoris tenuis*? Biocontrol 61:79–90
- Brévault T, Sylla S, Diatte M, Bernadas G, Diarra K (2014) *Tuta absoluta* Meyrick (Lepidoptera: Gelechiidae): a new threat to tomato production in Sub-Saharan Africa. Afr Entomol 22:441–444
- Cabello T, Gallego JR, Fernandez FJ, Gamez M, Vila E, Pino MD, Hernandez-Suarez E (2012) Biological control strategies for the South American tomato moth (Lepidoptera: Gelechiidae) in greenhouse tomatoes. J Econ Entomol 105:2085–2096
- Calvo FI, Soriano JD, Stansly PA, Belda JE (2016) Can the parasitoid *Necremnus tutae* (Hymenoptera: Eulophidae) improve existing biological control of the tomato leafminer *Tuta absoluta* (Lepidoptera: Gelechiidae)? Bull Entomol Res 106:502–511
- Campos MR, Rodrigues ARS, Silva WM, Silva TBM, Silva VRF, Guedes RNC, Siqueira HAA (2014) Spinosad and the tomato borer *Tuta absoluta*: a bioinsecticide, an invasive pest threat, and high insecticide resistance. PLoS ONE 9:e103235
- Campos MR, Silva TBM, Silva WM, Silva JE, Siqueira HAA (2015) Spinosyn resistance in the tomato borer *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae). J Pest Sci 88:405–412
- Caparros Megido R, Haubrige E, Verheggen FJ (2012) First evidence of deuterotokous parthenogenesis in the tomato leafminer, *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae). J Pest Sci 85:409–412
- Caparros Megido R, Haubrige É, Verheggen FJ (2013) Pheromone-based management strategies to control the tomato leafminer, *Tuta absoluta* (Lepidoptera: Gelechiidae). A review. Biotechnol Agron Soc Environ 17:475–482
- Cesari M, Maistrello L, Ganzerli F, Dioli P, Rebecchi L, Guidetti R (2015) A pest alien invasion in progress: potential pathways of origin of the brown marmorated stink bug *Halyomorpha halys* populations in Italy. J Pest Sci 88:1–7
- Chailleur A, Desneux N, Seguret J, Do Thi Khanh H, Maignet P, Tabone E (2012) Assessing European egg parasitoids as a mean of controlling the invasive South American tomato pinworm *Tuta absoluta*. PLoS ONE 7:e48068
- Chailleur A, Bearez P, Pizzol J, Amiens-Desneux E, Ramirez-Romero R, Desneux N (2013a) Potential for combined use of parasitoids and generalist predators for biological control of the key invasive tomato pest *Tuta absoluta*. J Pest Sci 86:533–541
- Chailleur A, Biondi A, Han P, Tabone E, Desneux N (2013b) Suitability of the pest—plant system *Tuta absoluta* (Lepidoptera: Gelechiidae)—tomato for Trichogramma (Hymenoptera: Trichogrammatidae) parasitoids and insights for biological control. J Econ Entomol 106:2310–2321
- Chailleur A, Desneux N, Arnó J, Gabarra R (2014) Biology of two key Palaearctic larval ectoparasitoids when parasitizing the invasive pest *Tuta absoluta*. J Pest Sci 87:441–448
- Cifuentes D, Chynoweth R, Bielza P (2011) Genetic study of Mediterranean and South American populations of tomato leafminer *Tuta absoluta* (Povolny, 1994) (Lepidoptera: Gelechiidae) using ribosomal and mitochondrial markers. Pest Manag Sci 67:1155–1162
- Cocco A, Deliperi S, Delrio G (2013) Control of *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) in greenhouse tomato crops using the mating disruption technique. J Appl Entomol 137:16–28
- Cocco A, Serra G, Lentini A, Deliperi S, Delrio G (2015) Spatial distribution and sequential sampling plans for *Tuta absoluta* (Lepidoptera: Gelechiidae) in greenhouse tomato crops. Pest Manag Sci 71:1311–1323
- Daane KM, Wang XG, Nieto DJ et al (2015) Classic biological control of olive fruit fly in California, USA: release and recovery of introduced parasitoids. Biocontrol 60:317–330
- Daane KM, Wang XG, Biondi A et al (2016) First exploration of parasitoids of *Drosophila suzukii* in South Korea as potential classical biological agents. J Pest Sci 89:823–835
- De Backer L, Megido RC, Fauconnier M-L, Brostaux Y, Francis F, Verheggen F (2015) *Tuta absoluta*-induced plant volatiles: attractiveness towards the generalist predator *Macrolophus pygmaeus*. Arthropod-Plant Interact 9:465–476
- Derbalah AS, Morsey SZ, El-Samahy M (2012) Some recent approaches to control *Tuta absoluta* in tomato under greenhouse conditions. Afr Entomol 20:27–34
- Desneux N, Decourtye A, Delpuech J (2007) The sublethal effects of pesticides on beneficial arthropods. Annu Rev Entomol 52:81–106
- Desneux N, Wajnberg E, Wyckhuys KAG et al (2010) Biological invasion of European tomato crops by *Tuta absoluta*: ecology, geographic expansion and prospects for biological control. J Pest Sci 83:197–215
- Desneux N, Luna MG, Guillemaud T, Urbaneja A (2011) The invasive South American tomato pinworm, *Tuta absoluta*, continues to spread in Afro-Eurasia and beyond: the new threat to tomato world production. J Pest Sci 84:403–408
- Desneux N, Blahnik R, Delebecque CJ, Heimpel GE (2012) Host phylogeny and specialisation in parasitoids. Ecol Lett 15:453–460
- Dukes JS, Mooney HA (1999) Does global change increase the success of biological invaders? Trends Ecol Evol 14:135–139
- Durić Z, Delić D, Hrnčić S, Radonjić S (2014) Distribution and molecular identification of *Tuta absoluta* (Meyrick 1917) (Lepidoptera, Gelechiidae) populations in Bosnia and Herzegovina and Montenegro. Pol J Entomol 83:121–129
- El-Arnaouty SA, Pizzol J, Galal HH et al (2014) Assessment of Two Trichogramma species for the control of *Tuta absoluta* in North African tomato greenhouses. Afr Entomol 22:801–809
- EPPO (2014) New data on quarantine pests and pests of the EPPO Alert List (Costa Rica). EPPO Global Database. <https://gd.eppo.int>. Accessed 20 Dec 2016
- Essl F, Dullinger S, Rabitsch W et al (2011) Socioeconomic legacy yields an invasion debt. Proc Natl Acad Sci USA 108:203–207
- FAO (2006) International standards for phytosanitary measures 1 to 24, 2005. Secretariat of the International Plant Protection Convention, FAO, Rome
- FAO (2015) Quarterly early warning Bulletin of Food and Agriculture. FAO, Rome

- FAOSTAT (2012) FAO Statistical Database (FAOSTAT). ICTUpdate, FAO, Rome. <http://www.fao.org/faostat>. Accessed 20 Dec 2016
- Galdino TVS, Picanço MC, Ferreira DO, Silva GAR, de Souza TC, Silva GA (2015) Is the performance of a specialist herbivore affected by female choices and the adaptability of the offspring? *PLoS ONE* 10:e0143389
- Garnas JR, Auger-Rozenberg MA, Roques A et al (2016) Complex patterns of global spread in invasive insects: eco-evolutionary and management consequences. *Biol Invasions* 18:935–952
- Gontijo PC, Picanço MC, Pereira EJG, Martins JC, Chediak M, Guedes RNC (2013) Spatial and temporal variation in the control failure likelihood of the tomato leaf miner, *Tuta absoluta*. *Ann Appl Biol* 162:50–59
- Guedes RNC, Picanço MC (2012) The tomato borer *Tuta absoluta* in South America: pest status, management and insecticide resistance. *EPPO Bull* 42:211–216
- Guedes RNC, Siqueira HAA (2012) The tomato borer *Tuta absoluta*: insecticide resistance and control failure. *CAB Rev Perspect Agric Vet Sci Nutr Nat Resour* 7:1–7
- Guedes RNC, Smagghe G, Stark JD, Desneux N (2016) Pesticide-induced stress in arthropod pests for optimized integrated pest management programs. *Annu Rev Entomol* 61:43–62
- Guillemaud T, Blin A, Le Goff I et al (2015) The tomato borer, *Tuta absoluta*, invading the Mediterranean Basin, originates from a single introduction from Central Chile. *Sci Rep* 5:8371
- Guimapi RY, Mohamed SA, Okeyo GO et al (2016) Modeling the risk of invasion and spread of *Tuta absoluta* in Africa. *Ecol Complex* 28:77–93
- Haddi K, Berger M, Bielza P et al (2012) Identification of mutations associated with pyrethroid resistance in the voltage-gated sodium channel of the tomato leaf miner (*Tuta absoluta*). *Insect Biochem Mol Biol* 42:506–513
- Hallman GJ, Arthur V, Blackburn CM, Parker AG (2013) The case for a generic phytosanitary irradiation dose of 250 Gy for Lepidoptera eggs and larvae. *Radiat Phys Chem* 89:70–75
- Han P, Lavois A-V, Le Bot J, Amiens-Desneux E, Desneux N (2014) Nitrogen and water availability to tomato plants triggers bottom-up effects on the leafminer *Tuta absoluta*. *Sci Rep* 4:4455
- Han P, Desneux N, Michel T et al (2016a) Does plant cultivar difference modify the bottom-up effects of resource limitation on plant-insect herbivore interactions? *J Chem Ecol* 42:1293–1303
- Han P, Wang Z-j, Lavois A-V et al (2016b) Increased water salinity applied to tomato plants accelerates the development of the leaf miner *Tuta absoluta* through bottom-up effects. *Sci Rep* 6:32403
- Haye T, Fischer S, Zhang J, Gariepy T (2015) Can native egg parasitoids adopt the invasive brown marmorated stink bug, *Halymorpha halys* (Heteroptera: Pentatomidae), in Europe? *J Pest Sci* 88:693–705
- Haye T, Girod P, Cuthbertson AGS et al (2016) Current SWD IPM tactics and their practical implementation in fruit crops across different regions around the world. *J Pest Sci* 89:643–651
- Hellmann JJ, Byers JE, Bierwagen BG, Dukes JS (2008) Five potential consequences of climate change for invasive species Cinco Consecuencias Potenciales del Cambio Climático para Especies Invasoras. *Conserv Biol* 22:534–543
- Hossain MS, Mian MY, Muniappan R (2016) First record of *Tuta absoluta* (Lepidoptera: Gelechiidae) from Bangladesh. *J Agric Urban Entomol* 32:101–105
- Hulme PE (2009) Trade, transport and trouble: managing invasive species pathways in an era of globalization. *J Appl Entomol* 46:10–18
- Izhevskaya SS, Akhatov AK, Sinyov SY (2011) *Tuta absoluta* has been detected in Russia. *Zashchita i Karantin Rastenii* 3:40–44
- Kalleshwaraswamy CM, Murthy MS, Viraktamath CA, Kumar NKK (2015) Occurrence of *Tuta absoluta* (Lepidoptera: Gelechiidae) in the Malnad and Hyderabad-Karnataka Regions of Karnataka, India. *Fla Entomol* 98:970–971
- Karadjova O, Ilieva Z, Krumov V, Petrova E, Vensislavov V (2013) *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae): potential for entry, establishment and spread in Bulgaria. *Bulg J Agric Sci* 19:563–571
- Keller RP, Geist J, Jeschke JM, Kühn I (2011) Invasive species in Europe: ecology, status, and policy. *Environ Sci Eur* 23:23
- Klieber J, Reineke A (2016) The entomopathogen *Beauveria bassiana* has epiphytic and endophytic activity against the tomato leaf miner *Tuta absoluta*. *J Appl Entomol* 140:580–589
- Kolar CS, Lodge DM (2001) Progress in invasion biology: predicting invaders. *Trends Ecol Evol* 16:199–204
- Lima MAP, Martins GF, Oliveira EE, Guedes RNC (2016) Agrochemical-induced stress in stingless bees: peculiarities, underlying basis, and challenges. *J Comp Physiol A* 202:733–747
- Lowry E, Rollinson EJ, Laybourn AJ et al (2013) Biological invasions: a field synopsis, systematic review, and database of the literature. *Ecol Evol* 3:182–196
- Meyerson LA, Mooney HA (2007) Invasive alien species in an era of globalization. *Front Ecol Environ* 5:199–208
- Mollá O, Biondi A, Alonso-Valiente M, Urbaneja A (2014) A comparative life history study of two mirid bugs preying on *Tuta absoluta* and *Ephestia kuhniella* eggs on tomato crops: implications for biological control. *Biocontrol* 59:175–183
- NAPPO - North American Plant Protection Organization (2013) Surveillance protocol for the tomato leaf miner, *Tuta absoluta*, for NAPPO member countries. USDA. https://www.aphis.usda.gov/import_export/plants/plant_exports/downloads/Tuta_absoluta_surveillanceprotocol_08-06-2012-e.pdf. Accessed 20 Dec 2016
- NARC (2016) The First record of *Tuta absoluta* (Meyrick 1917) (Lepidoptera: Gelechiidae) in Nepal. Nepal Agricultural Research Council (NARC), Kathmandu
- Naselli M, Biondi A, Tropea Garzia G, Desneux N, Russo A, Siscaro G, Zappalà L (2017) Insights on food webs associated with the South American Tomato Pinworm. *Pest Manag Sci*. doi:10.1002/ps.4562
- Oliveira FA, da Silva DJH, Leite GLD, Jham GN, Picanço M (2009) Resistance of 57 greenhouse-grown accessions of *Lycopersicon esculentum* and three cultivars to *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae). *Sci Hortic* 119:182–187
- Perdikis D, Arvaniti K (2016) Nymphal development on plant vs. leaf with and without prey for two omnivorous predators: *Nesidiocoris tenuis* (Reuter, 1895) (Hemiptera: Miridae) and *Dicyphus errans* (Wolff, 1804) (Hemiptera: Miridae). *Entomol Gen* 35:297–306
- Perrings C, Burgiel S, Lonsdale M, Mooney H, Williamson M (2010) International cooperation in the solution to trade-related invasive species risks. *Ann N Y Acad Sci* 1195:198–212
- Pfeiffer DG, Muniappan R, Sall D, Diatta P, Diongue A, Dieng EO (2013) First record of *Tuta absoluta* (Lepidoptera: Gelechiidae) in Senegal. *Fla Entomol* 96:661–662
- Ponti L, Gutierrez AP, Altieri MA (2015) Holistic approach in invasive species research: the case of the tomato leaf miner in the Mediterranean Basin. *Agroecol Sust Food* 39:436–468
- Potting RPJ, van der Gaag DJ, Loomans A, van der Straten M, Anderson H, MacLeod A et al (2013) *Tuta absoluta*, Tomato leaf miner moth or South American tomato moth. Ministry of Agriculture, Nature and Food Quality, Plant Protection Service of the Netherlands, Utrecht, The Netherlands
- Ragsdale DW, Landis DA, Brodeur J, Heimpel GE, Desneux N (2011) Ecology and management of the soybean aphid in North America. *Annu Rev Entomol* 56:375–399
- Roditakis E, Skarmoutsou C, Staurakaki M (2013) Determination of baseline susceptibility of European populations of *Tuta absoluta*

- (Meyrick) to indoxacarb and chlorantraniliprole using a novel dip bioassay method. Pest Manag Sci 69:217–227
- Roditakis E, Vasakis E, Grispou M, Stavrakaki M, Nauen R, Gravouil M, Bassi A (2015) First report of *Tuta absoluta* resistance to diamide insecticides. J Pest Sci 88:9–16
- Roditakis E, Steinbach D, Moritz G et al (2017) Ryanodine receptor point mutations confer diamide insecticide resistance in tomato leafminer, *Tuta absoluta* (Lepidoptera: Gelechiidae). Insect Biochem Mol Biol 80:11–20
- Roques A, Auger-Rozenberg MA, Blackburn TM et al (2016) Temporal and interspecific variation in rates of spread for insect species invading Europe during the last 200 years. Biol Invasions 18:907–920
- Salas Gervassio NG, Luna MG, Lee S, Salvo A, Sánchez NE (2016) Trophic web associated with the South American tomato moth *Tuta absoluta*: implications for its conservation biological control in Argentina. Agric For Entomol 18:137–144
- Salehi Z, Yarahmadi F, Rasekh A et al (2016) Functional responses of *Orius albidipennis* Reuter (Hemiptera, Anthocoridae) to *Tuta absoluta* Meyrick (Lepidoptera, Gelechiidae) on two tomato cultivars with different leaf morphological characteristics. Entomol Gen 36:127–136
- Schmitz DC, Simberloff D (1997) Biological invasions: a growing threat. Issues Sci Technol 13:33–40
- Silva JE, Assis CPO, Ribeiro LMS, Siqueira HAA (2016) Field-evolved resistance and cross-resistance of Brazilian *Tuta absoluta* (Lepidoptera: Gelechiidae) populations to diamide insecticides. J Econ Entomol 109:2190–2195
- Siqueira HAA, Guedes RNC, Picanço MC (2000) Insecticide resistance in populations of *Tuta absoluta* (Lepidoptera: Gelechiidae). Agric For Entomol 2:147–153
- Sohrabi F, Nooryazdan H, Gharati B et al (2016) Evaluation of ten tomato cultivars for resistance against tomato leaf miner, *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) under field infestation conditions. Entomol Gen 36:163–175
- Sridhar V, Chakravarthy A, Asokan R, Vinesh L, Rebijith D, Vennila S (2014) New record of the invasive South American tomato leaf miner, *Tuta absoluta* (Meyrick) (Lepidoptera: Gelechiidae) in India. Pest Manag Hortic Ecosyst 20:148–154
- Terzidis AN, Wilcockson S, Leifert C (2014) The tomato leaf miner (*Tuta absoluta*): conventional pest problem, organic management solutions? Organ Agric 4:43–61
- Tomé HVV, Martins GF, Lima MAP, Campos LAO, Guedes RNC (2012) Imidacloprid-induced impairment of mushroom bodies and behavior of the native stingless bee *Melipona quadrifasciata anthidioides*. PLoS ONE 7:e38406
- Tomé HVV, Barbosa WF, Martins GF, Guedes RNC (2015) Spinosad in the native stingless bee *Melipona quadrifasciata*: regrettable non-target toxicity of a bioinsecticide. Chemosphere 124:103–109
- Tonnang HEZ, Mohamed SF, Khamis F, Ekesi S (2015) Identification and risk assessment for worldwide invasion and spread of *Tuta absoluta* with a focus on Sub-Saharan Africa: implications for phytosanitary measures and management. PLoS ONE 10:e0135283
- Trade Map (2016) Market Access Map, Investment Map and Standards Map. International Trade Centre. www.intracen.org/marketanalysis. Accessed 20 Dec 2016
- Tropea Garzia G, Siscaro G, Biondi A, Zappalà L (2012) *Tuta absoluta*, a South American pest of tomato now in the EPPO region: biology, distribution and damage. EPPO Bull 42:205–210
- Urbaneja-Bernat P, Mollá O, Alonso M, Bolckmans K, Urbaneja A, Tena A (2015) Sugars as complementary alternative food for the establishment of *Nesidiocoris tenuis* in greenhouse tomato. J Appl Entomol 139:161–167
- USDA-APHIS - United States Department of Agriculture, Animal and Plant Health Inspection Service (2011) New pest response guidelines: tomato leafminer (*Tuta absoluta*). USDA. https://www.aphis.usda.gov/import_export/plants/manuals/emergency/downloads/Tuta-absoluta.pdf. Accessed 20 Dec 2016
- USDA-APHIS - United States Department of Agriculture, Animal and Plant Health Inspection Service (2013) Treatment manual. USDA. https://www.aphis.usda.gov/import_export/plants/manuals/ports/downloads/treatment.pdf. Accessed 20 Dec 2016
- USDA-APHIS - United States Department of Agriculture, Animal and Plant Health Inspection Service (2014) Federal Order for U.S. Imports of host materials of tomato leaf miner (*Tuta absoluta*). USDA. <https://www.aphis.usda.gov/aphis/ourfocus/planthealth/import-information/federal-import-orders>. Accessed 20 Dec 2016
- van der Straten MJJ, Potting RPJ, Linden AVD (2011) Introduction of the tomato leafminer *Tuta absoluta* into Europe. Proc Neth Entomol Soc Meet 22:23–30
- Venette RC (2015) The challenge of modelling and mapping the future distribution and impact of invasive alien species. In: Venette RC (ed) Pest risk modelling and mapping for invasive Alien Species, CABI Invasives Series. CABI, St. Paul, p 252
- Zappalà L, Siscaro G, Biondi A, Mollá O, González-Cabrera J, Urbaneja A (2012) Efficacy of sulphur on *Tuta absoluta* and its side effects on the predator *Nesidiocoris tenuis*. J Appl Entomol 136:401–409
- Zappalà L, Biondi A, Alma A et al (2013) Natural enemies of the South American moth, *Tuta absoluta*, in Europe, North Africa and Middle East, and their potential use in pest control strategies. J Pest Sci 86:635–647



Activity in Boolean networks

Abhijin Adiga¹ · Hilton Galyean² · Chris J. Kuhlman¹ · Michael Levet⁴ ·
Henning S. Mortveit^{1,3} · Sichao Wu¹

Published online: 14 October 2016
© Springer Science+Business Media Dordrecht 2016

Abstract In this paper we extend the notion of activity for Boolean networks introduced by Shmulevich and Kauffman (Phys Rev Lett 93(4):48701:1–4, 2004). In contrast to existing theory, we take into account the actual graph structure of the Boolean network. The notion of activity measures the probability that a perturbation in an initial state produces a different successor state than that of the original unperturbed state. It captures the notion of sensitive dependence on initial conditions, and provides a way to rank vertices in terms of how they may impact predictions. We give basic results that aid in the computation of activity and apply this to Boolean networks with

threshold functions and nor functions for elementary cellular automata, d -regular trees, square lattices, triangular lattices, and the Erdős–Renyi random graph model. We conclude with some open questions and thoughts on directions for future research related to activity, including long-term activity.

Keywords Boolean networks · Finite dynamical system · Activity · Sensitivity · Network · Sensitive dependence on initial conditions

1 Introduction

A Boolean network (*BN*) is a map of the form

$$F = (f_1, \dots, f_n) : \{0, 1\}^n \longrightarrow \{0, 1\}^n. \quad (1)$$

BNs were originally proposed as a model for many biological phenomena (Kauffman 1969, 1993), but have now been used to capture and analyze a range of complex systems and their dynamics (Aldana et al. 2003). Associated to F we have the *dependency graph* of F whose vertex set is $\{1, 2, \dots, n\}$ and with edges all (i, j) for which the function f_i depends non-trivially on the variable x_j . See (Goles and Martinez 1990; Robert 1986; Mortveit and Reidys 2007) for more general discussions of maps F .

The study of stability and the response to perturbations of Boolean networks is central to increased understanding of their dynamical properties. Perturbations may take many forms, with examples including perturbations of the dependency graph (Adiga et al. 2013; Luo and Turner 2012; Pomerance et al. 2009), the vertex states (Shmulevich and Kauffman 2004; Ghanbarnejad and Klemm 2012; Serra et al. 2010), the vertex functions (Shmulevich et al. 2003; Xiao and Dougherty 2007), or combinations of these.

✉ Henning S. Mortveit
henning.mortveit@vt.edu

Abhijin Adiga
adiga@vbi.vt.edu

Hilton Galyean
hgalyean@vt.edu

Chris J. Kuhlman
ckuhlman@vbi.vt.edu

Michael Levet
mlevet@vt.edu

Sichao Wu
sichao@vbi.vt.edu

¹ Network Dynamics and Simulation Science Laboratory, VBI, Virginia Tech, 1015 Life Science Circle, Blacksburg, VA 24061, USA

² Department of Physics, Virginia Tech, Blacksburg, VA 24061, USA

³ Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, USA

⁴ Department of Computer Science, University of South Carolina, Columbia, SC, USA

This paper is concerned with noise applied to vertex states. Specifically, we want to know the following:

What is the probability that $F(x)$ and $F(x + e_i)$ are different?

Here i is a vertex while e_i is the i th unit vector with the usual addition modulo 2. In Shmulevich and Kauffman (2004), Shmulevich and Kauffman considered Boolean networks over regular graphs with $f_j = f$ for all vertices j , that is, those induced by a common function. They defined the notion of *activity* of f with respect to its i th argument as the expected value of the Boolean derivative of f with respect to its i th variable. Under their assumptions, this may give a reasonable indication of the expected impact of perturbations to the i th variable under the evolution of F . However, this approach does not consider the impact of the dependency graph structure. We remark that Layne et al. (2012) compute the activities of nested canalyzing functions given their canalyzing depth, extending results in Shmulevich and Kauffman (2004) on canalyzing functions.

This question of sensitivity has also been studied when x is restricted to attractors in order to assess stability of long-term dynamics under state noise. The notion of threshold ergodic sets (TESs) is introduced in Ribeiro and Kauffman (2007) and studied further in Luo and Turner (2012), and Kuhlman and Mortveit (2014). The structure of TESs capture long-term stability under state perturbations of periodic orbits and the resulting mixing between attractors that may happen as a result.

Other tools for analyzing sensitivity of vertex noise includes Lyapunov exponents, see for example (Baetens and Baets 2010; Baetens et al. 2012), although this is perhaps mostly relevant or suited for the infinite case such as cellular automata over (infinite) regular lattices. Also, the notion of *Derrida diagrams* has been used to quantify how *Hamming classes* of states separate on average (Derrida and Pomeau 1986; Fretter et al. 2009) after one or ℓ transitions under F . Derrida diagrams, however, are mainly analyzed through numerical experiments via sampling. Moreover, analyzing how Hamming classes of large distance separate under F may not be so insightful—it seems more relevant to limit oneself to the case of nearby classes of vertex states.

Returning to the original question, we note that $F(x)$ and $F(x + e_i)$ may only differ in the components j for which f_j depends on x_i . The answer to the question therefore depends on the vertex functions in the 1-neighborhood of i in the dependency graph X , and therefore the structure of the induced subgraph of the 2-neighborhood of i in X with the omission of edges connecting pairs of vertices both of distance 2 from i . We denote this subgraph by $X(i; 2)$, see Fig. 1.

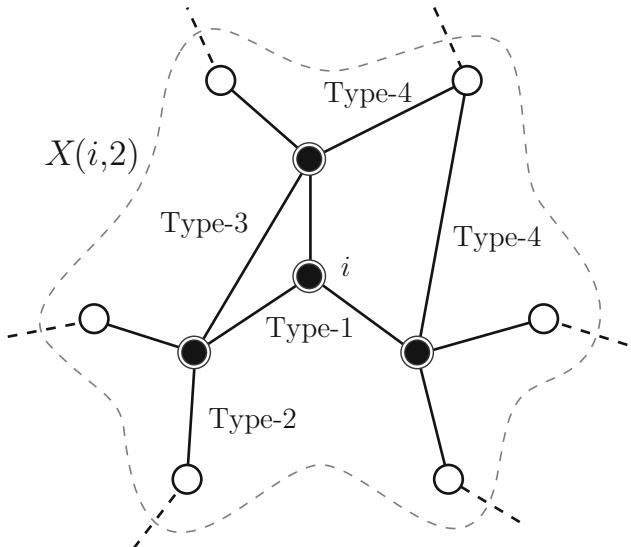


Fig. 1 The subgraph $X(i; 2)$ of X induced by vertex i and its distance ≤ 2 neighbors. Vertices belonging to the 1-neighborhood of i (note that i is included) are marked using *filled circles*, remaining vertices are marked using *open circles*. Edges connecting vertices of distance 2 from i in X do not belong to $X(i; 2)$. Type-1 edges are edges incident to i ; Type-2 edges are edges of $X(i; 2)$ not incident to i and not part of any 3- or 4-cycle. Type-3 (resp. type-4) edges are edges in 3-cycles (resp. 4-cycles) containing i that are not incident to i . For additional terminology we refer to Sect. 2

To determine the activity of vertex i one will in general have to evaluate F over all possible states of $X(i; 2)$, a problem which, in the general case, is computationally intractable. We will address three cases in this paper: the first case is when $X(i; 2)$ is a tree, the second case is that of elementary cellular automata (a special case of the former case), and the case where X is a regular, square, 2-dimensional lattice. Work for other graph classes is in progress and we comment on some of the challenges in the Summary section. Here we remark that a major source of challenges for analytic computations is the introduction of *type-3* and *type-4* edges as illustrated in Fig. 1. Lack of symmetry adds additional challenges. The activity of a vertex can be evaluated analytically using the inclusion-exclusion principle, and type-3 and type-4 edges impact the complexity of the combinatorics.

A goal of the work on activity started here is as follows: when given a network X and vertex functions $(f_i)_i$, we would like to rank the vertices by activity in decreasing order. Just being able to identify for example the ten (say) vertices of highest activity would also be very useful. This information would allow one to identify the vertices for which state perturbations are most likely to produce different outcomes, at least in the short term. In a biological experiment for which there is a BN model of the form (1), one may then be able to allocate more resources to the measurement of such states, or perhaps ensure that these

states are carefully controlled and locked at their intended values.

1.1 Paper organization

After basic definitions and terminology in Sect. 2 we carefully define a new notion of activity denoted by $\bar{\alpha}_{F,i}$. Basic results to help in analytic evaluations of $\bar{\alpha}_{F,i}$ are given in Sect. 3 followed by specific results for the elementary cellular automata in Sect. 4, d -regular trees in Sect. 5, and then nor- and threshold-BNs over square and triangular lattices in Sect. 6. In Sect. 8 we consider activity for the $G_{n,p}$ random graph model. A central goal is to relate the structure of $X(i; 2)$ and the functions $(f_j)_j$ to $\bar{\alpha}_{F,i}$. We conclude with open questions and possible directions for followup work in Sect. 9. We remark that this paper is an extended version of work that we presented at Automata 2015 and that appeared in its proceedings.

2 Background, definitions and terminology

In this paper we consider the discrete dynamical systems of the form (1) where each map f_i is of the form $f_i : \{0, 1\}^n \rightarrow \{0, 1\}$. However, f_i will in general depend nontrivially only on some subset of the variables x_1, x_2, \dots, x_n , a fact that is captured by the dependency graph defined in the introduction and denoted by X_F or simply X when F is implied. The graph X is generally directed and will contain a loop at every vertex. We will, however, limit ourselves to undirected graphs. A particular graph that we will refer to is the *circle graph on n vertices* denoted by Circle_n and defined by vertex and edge sets

$$v[\text{Circle}_n] = \{1, 2, \dots, n\}$$

$$e[\text{Circle}_n] = \{\{i, i+1\} \mid 1 \leq i \leq n\},$$

respectively, and where indices are taken modulo 2.

Each vertex i has a *vertex state* $x_i \in K = \{0, 1\}$ and a *vertex function* of the form $f_i : \{0, 1\}^{d(i)+1} \rightarrow \{0, 1\}$ taking as arguments the states of vertices in the 1-neighborhood of i in X . Here $d(i)$ is the degree of vertex i . We write $n[i]$ for the ordered sequence of vertices contained in the 1-neighborhood of i with i included and set $n(i) = n[i] \setminus \{i\}$. Similarly, we write $x[i]$ for the subsequence of vertex states corresponding to $n[i]$. Finally, we denote the *system state* by $x = (x_1, \dots, x_n) \in \{0, 1\}^n$.

We will write the evaluation of F in (1) as

$$F(x) = (f_1(x[1]), f_2(x[2]), \dots, f_n(x[n])).$$

The phase space of the map F in (1) is the directed graph $\Gamma(F)$ with vertex set $\{0, 1\}^n$ and directed edges all pairs $(x, F(x))$. A state on a cycle in $\Gamma(F)$ is called a *periodic*

point and a state on a cycle of length one is a *fixed point*. The sets of all such points are denoted by $\text{Per}(F)$ and $\text{Fix}(F)$ respectively. All other states are *transient states*. Since $\{0, 1\}^n$ is finite, the phase space of F consists of a collection of oriented cycles (called periodic orbits), possibly with directed trees attached at states contained on cycles.

In this paper we analyze *short-term stability of dynamics* through the function $\alpha_{F,i} : K^n \rightarrow \{0, 1\}$ defined by

$$\alpha_{F,i}(x) = \mathbb{I}[F(x + e_i) \neq F(x)] \quad (2)$$

where \mathbb{I} is the indicator function and e_i is the i th unit vector with $1 \leq i \leq n$. In other words, $\alpha_{F,i}(x)$ measures whether perturbing x by e_i results in a different successor state under F than $F(x)$.

Definition 1 The *activity of F with respect to vertex i* is the expected value of $\alpha_{F,i}$ using the uniform measure on K^n :

$$\bar{\alpha}_{F,i} = \mathbb{E}[\alpha_{F,i}] . \quad (3)$$

The *activity of F* is the vector

$$\bar{\alpha}_F = (\bar{\alpha}_{F,1}, \bar{\alpha}_{F,2}, \dots, \bar{\alpha}_{F,n}) , \quad (4)$$

while the *sensitivity of F* is the average activity $\bar{\alpha} = \sum_{i=1}^n \bar{\alpha}_{F,i}/n$.

For a randomly chosen state $x \in K^n$, the value $\bar{\alpha}_{F,i}$ may be interpreted as the probability that perturbing x_i will cause $F(x + e_i) \neq F(x)$ to hold. This activity notion may naturally be regarded as a measure of sensitivity with respect to initial conditions.

From (2) it is clear that $\bar{\alpha}_{F,i}$ depends on the functions f_j with $j \in n[i]$ and the structure of the distance-2 subgraph $X(i; 2)$, see Fig. 1. The literature (see, e.g. Shmulevich and Kauffman 2004; Layne et al. 2012) has focused on a very special case when considering activity. Rather than considering the general case and Eq. (2), they have focused on the case where X is a regular graph where each vertex has degree d and all vertex functions are induced by a common function $f : K^{d+1} \rightarrow K$ through $f_v(x[v]) = f(x[v])$. In this setting, activity is defined with respect to f and its i th argument, that is, as the expectation value of the function $\mathbb{I}[f(x + e_i) \neq f(x)]$ where $x \in K^{d+1}$. Clearly, this measure of activity is always less than or equal to $\mathbb{E}[\alpha_{F,i}]$. Again, we note that this simpler notion of activity does not account for the network structure of $X(i; 2)$.

3 Preliminary results

For the evaluation of $\bar{\alpha}_{F,i}$ we introduce some notation. In the following we set $K = \{0, 1\}$, write N_i for the size of $X(i; 2)$, and $K(i) = K^{N_i}$ for the projection of K^n onto the set

of vertex states associated to $X(i; 2)$. For $j \in n[i]$, define the sets $A_j(i) \subset K(i)$ by

$$A_j(i) = \{x \in K(i) \mid F(x + e_i)_j \neq F(x)_j\}.$$

These sets appear in the evaluation of $\bar{\alpha}_{F,i}$, see Proposition 1. For convenience, we also set

$$A_j^m(i) = \{x \in A_j(i) \mid x_i = m\},$$

for $m = 0, 1$. We write $\bar{A}_j(i) = A_j^0(i)$ and $\binom{n}{k}$ for binomial coefficients using the convention that it evaluates to zero if either $k < 0$ or $n - k < 0$.

The following proposition provides a somewhat simplified approach for evaluating $\bar{\alpha}_{F,i}$ in the general case.

Proposition 1 *Let X be a graph and F a map over X as in (1). The activity of F with respect to vertex i is*

$$\bar{\alpha}_{F,i} = \Pr \left(\bigcup_{j \in n[i]} A_j \mid x_i = 0 \right) = \Pr \left(\bigcup_{j \in n[i]} \bar{A}_j \right). \quad (5)$$

Proof As stated earlier, we can write

$$\begin{aligned} \bar{\alpha}_{F,i} &= \mathbb{E}[\alpha_{F,i}] = \sum_{x \in K^n} \alpha_{F,i}(x) \Pr(x) = \sum_{x \in K(i)} \alpha_{F,i}(x) \Pr(x) \\ &= \Pr \left[\bigcup_{j \in n[i]} A_j \right], \end{aligned}$$

where probabilities in the first and second sum are in K^n and $K(i)$, respectively. This follows by considering the relevant cylinder sets. Equation (5) follows by conditioning on the possible states for x_i . Since we are in the Boolean case, we have a bijection between each pair of sets A_j^0 and A_j^1 for $j \in n[i]$ which immediately allows us to deduce Eq. (5). \square

As an example, consider the complete graph $X = K_n$ with threshold functions at every vertex. Recall that the standard Boolean *threshold function*, denoted by $\tau_{k,n} : K^n \rightarrow K$, is defined by

$$\tau_{k,n}(x_1, \dots, x_n) = \begin{cases} 1, & \text{if } \sum_{j=1}^n x_j \geq k, \\ 0, & \text{otherwise,} \end{cases} \quad (6)$$

where k is the threshold. Here we have

$$\bar{\alpha}_{F,i} = \binom{n-1}{k-1} / 2^{n-1}.$$

To see this, note first that all the sets $\bar{A}_j(i)$ are identical. For a state x with $x_i = 0$ to satisfy $\tau_k(x) \neq \tau_k(x + e_i)$ it is necessary and sufficient that x belong to Hamming class $k-1$. Since $x_i = 0$, it follows from Proposition 1 that $\Pr(\bar{A}_j(i)) = \frac{1}{2^{n-1}} |\bar{A}_j| = \binom{n-1}{k-1} / 2^{n-1}$ as stated.

Next, consider the Boolean nor-function $\text{nor}_m : K^m \rightarrow K$ defined by

$$\text{nor}_m(x_1, \dots, x_m) = (1 + x_1) \cdots (1 + x_m), \quad (7)$$

with arithmetic operations modulo 2. If we use the nor-function over K_n we obtain

$$\bar{\alpha}_{F,i} = 1/2^{n-1}.$$

Again, $\bar{A}_j = \bar{A}_k$ for all $j, k \in n[i]$ and we have $F(x + e_i)_i \neq F(x)_i$ precisely when $x_j = 0$ for all $j \in n(i)$ leading to $\Pr(\bar{A}_i) = \frac{1}{2^{n-1}}$. We record the previous two results as a proposition:

Proposition 2 *If F is the Boolean network induced by the nor-function over K_n , then*

$$\bar{\alpha}_{F,i} = \frac{1}{2^{n-1}},$$

and if F is induced by the k -threshold function over K_n , then

$$\bar{\alpha}_{F,i} = \binom{n-1}{k-1} / 2^{n-1}.$$

In the computations to follow, we will frequently need to evaluate the probability of the union of the A_j 's. For this, let B denote the union of A_j 's for all $j \neq i$. We then have

$$\begin{aligned} \Pr \left[\bigcup_{j \in n[i]} A_j \right] &= \Pr(A_i \cup B) = \Pr(B) + \Pr(A_i \cap B^c) \\ &= 1 - \Pr(B^c) + \Pr(A_i \cap B^c), \end{aligned} \quad (8)$$

where B^c denotes the complement of B .

4 Activity of elementary cellular automata

The evaluation of Eq. (5) can often be done through the inclusion–exclusion principle. We demonstrate this in the context of elementary cellular automata (ECA). This also makes it clear how the structure of X comes into play for the evaluation of $\bar{\alpha}_{F,i}$.

Let F be the ECA map over $X = \text{Circle}_n$ where each vertex function is given by $f : \{0, 1\}^3 \rightarrow \{0, 1\}$. Here we will assume that $n \geq 5$; the case $n = 3$ corresponds to the complete graph K_3 and the case $n = 4$ can be handled quite easily. Here we have

$$\bar{\alpha}_{F,i} = \Pr(\bar{A}_{i-1}(i) \cup \bar{A}_i(i) \cup \bar{A}_{i+1}(i)).$$

Applying the definitions,

$$\begin{aligned} \bar{A}_j(i) &= \{x = (x_{i-2}, x_{i-1}, x_i = 0, x_{i+1}, x_{i+2}) \mid \\ &\quad f(x[j]) \neq f((x + e_i)[j])\} \end{aligned} \quad (9)$$

where $j \in n[i] = \{i-1, i, i+1\}$ and with indices modulo n .

Proposition 3 *The activity for a k -threshold ECA is*

$$\bar{\alpha}_{F,i} = \begin{cases} 0, & \text{if } k = 0 \text{ or } k > 3, \\ 1/2, & \text{if } k = 1 \text{ or } k = 3, \\ 7/8, & \text{if } k = 2. \end{cases}$$

Proof Clearly, for $k = 0$ and $k > 3$ we always have $F(x + e_i) = F(x)$ so in these cases it follows that $\bar{\alpha}_{F,i} = 0$. The cases $k = 1$ and $k = 3$ are symmetric (rule r and $255 - r$ have identical activity, see below), and, using $k = 1$, we have

$$\bar{A}_{i-1} = \{(0, 0, 0, x_{i+1}, x_{i+2})\}, \quad \bar{A}_i = \{(x_{i-2}, 0, 0, 0, x_{i+2})\}, \\ \text{and } \bar{A}_{i+1} = \{(x_{i-2}, x_{i-1}, 0, 0, 0)\}.$$

By the inclusion–exclusion principle, it follows that

$$|\bar{A}_{i-1} \cup \bar{A}_i \cup \bar{A}_{i+1}| = |\bar{A}_{i-1}| + |\bar{A}_i| + |\bar{A}_{i+1}| \\ - |\bar{A}_{i-1} \cap \bar{A}_i| - |\bar{A}_{i-1} \cap \bar{A}_{i+1}| - |\bar{A}_i \cap \bar{A}_{i+1}| + |\bar{A}_{i-1} \cap \bar{A}_i \cap \bar{A}_{i+1}| \\ = 3 \times 4 - 2 - 1 - 2 + 1 = 8.$$

This yields $\bar{\alpha}_{F,i} = 8/2^4 = 1/2$. The proof for the case $k = 2$ is similar to that of $k = 1$ so we leave this to the reader, but see also Table 1. \square

Remark 1 For the ECA induced by the nor-function, $\bar{\alpha}_{F,i} = 1/2$ for $n \geq 5$.

More generally, an ECA rule $f : \{0, 1\}^3 \rightarrow \{0, 1\}$ may be represented by an integer $0 \leq r(f) \leq 255$, or simply

Table 1 The activity of all ECA rules for $n \geq 5$

0 0	1 8	2 9	3 12	4 11	5 12	6 14	7 15
8 9	9 14	10 12	11 13	12 12	13 13	14 14	15 16
16 9	17 12	18 12	19 12	20 14	21 15	22 15	23 14
24 14	25 15	26 15	27 14	28 16	29 16	30 16	31 15
32 11	33 16	34 12	35 14	36 14	37 16	38 14	39 14
40 14	41 16	42 14	43 14	44 14	45 16	46 12	47 13
48 12	49 14	50 15	51 16	52 14	53 14	54 16	55 12
56 16	57 16	58 14	59 14	60 16	61 15	62 16	63 12
64 9	65 14	66 14	67 15	68 12	69 13	70 16	71 16
72 12	73 16	74 15	75 16	76 15	77 14	78 14	79 13
80 12	81 13	82 15	83 14	84 14	85 16	86 16	87 15
88 15	89 16	90 16	91 16	92 14	93 13	94 15	95 12
96 14	97 16	98 16	99 16	100 14	101 16	102 16	103 15
104 15	105 16	106 16	107 16	108 16	109 16	110 16	111 14
112 14	113 14	114 14	115 14	116 12	117 13	118 16	119 12
120 16	121 16	122 15	123 16	124 16	125 14	126 14	127 8

In the table, rule number and activity number alternate. For example, in the lower right corner, “127 8” encodes rule 127 and activity number 8. Here, all activity numbers must be divided by 16 to obtain the actual activity value. Rules in the range $128 \leq r \leq 255$ are omitted since the activity of rule r and rule $255 - r$ coincide. The reader will observe that the possible activity numbers are $\{0, 8, 9, 11, 12, 13, 14, 15, 16\}$

r. Each triple (x_{i-1}, x_i, x_{i+1}) can be viewed as a binary number $0 \leq j \leq 7$ where x_{i-1} is taken as the most significant digit. Let x^j denote the triple corresponding to the integer and set $a_j = f(x^j)j$. We can then represent f as the 8-tuple $a = (a_0, a_1, \dots, a_7)$. The associated rule number is $r(f) = \sum_{0 \leq j \leq 7} a_j \cdot 2^j$. From this and through either inclusion–exclusion or exhaustive enumeration (perhaps easier), one can obtain the activity value of all ECA through conditions on the a_i values. The frequency distribution of activity values for the ECA is derived from the data in Table 1 and is shown in Fig. 2.

Remark 2 In Shmulevich and Kauffman (2004), activity is defined with respect to the function f instead of F and its i th argument. With this context,

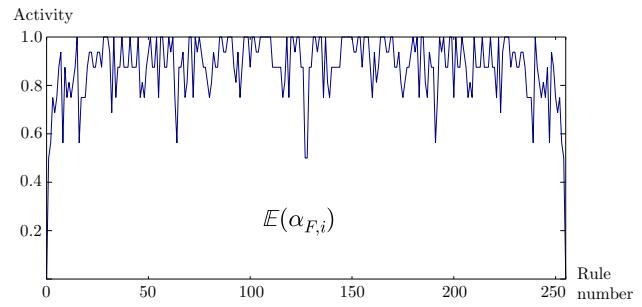


Fig. 2 Activity of ECA by rule number. With the exception of the constant rules of activity 0, all rules have activity $\geq 1/2$ with most rules exceeding 3/4

$$\bar{\alpha}_{f,i}(x) = \mathbb{E}[\mathbb{I}[f(x + e_i) \neq f(x)]] .$$

Using their definition of activity for ECA with a function of three Boolean variables, the only possible values for activity of ECA are 0, 1/4, 1/2, 3/4, and 1. The fact that our new notion of activity gives $\bar{\alpha} = 7/8$ for threshold-2 Boolean networks demonstrates that network structure does indeed impact results.

Additionally, we always have $\bar{\alpha}_{f,i} \leq \bar{\alpha}_{F,i}$. As an example, for a circle graph of girth ≥ 5 and threshold-1 functions (i.e., $k = 1$) it can be verified that $\bar{\alpha}_{f,i} = 1/4$. It was shown above that in this case $\bar{\alpha}_{F,i} = 1/2$.

5 Activity over d -regular trees

A natural starting point for analyzing activity is the case where the graph X is a d -regular tree. The reason is that in this case the sets A_j (or more precisely, the sets $n[j]$) with $j \neq i$ only have vertex i in common. As a result, when we condition on $x_i = m$, the resulting sets are independent. This fact will hold if the girth of the graph is at least 5, so the results we give are applicable to this broader graph class. We remark that a d -regular tree is an infinite tree where all vertices have degree d . Alternatively, one may consider this to be finite trees where all vertices either have degree d or degree 1. In the latter case, our result only applies to vertices whose neighbors all have degree d .

The Boolean bi-threshold function $\tau_{i,k_{01},k_{10},n} : K^n \rightarrow K$ generalizes standard threshold functions and is defined by

$$\tau_{i,k_{01},k_{10},n}(x_1, \dots, x_n) = \begin{cases} 1, & \text{if } x_i = 0 \text{ and } \sum_{j=1}^n x_j \geq k_{01} \\ 0, & \text{if } x_i = 1 \text{ and } \sum_{j=1}^n x_j < k_{10} \\ x_i, & \text{otherwise,} \end{cases} \quad (10)$$

where the integers k_{01} and k_{10} are the up- and down-thresholds, respectively. Here i is a designated vertex – it will be the index of a vertex function. If the up- and down-thresholds are the same we get the standard threshold systems.

Proposition 4 Let X be a d -regular graph of girth ≥ 5 , and F the Boolean network over X with the bi-threshold vertex functions as in Eq. (10). Then the activity of F with respect to vertex i is given by

$$\begin{aligned} \bar{\alpha}_{F,i} = 1 - \frac{1}{2^{d^2}} & \left[2^d - \binom{d-1}{k_{01}-1} - \binom{d-1}{k_{10}-2} \right]^d \\ & + \frac{1}{2^{d^2+1}} \sum_{k=k_{01},k_{10}} \binom{d}{k-1} \left[2^{d-1} - \binom{d-1}{k_{10}-2} \right]^{k-1} \\ & \left[2^{d-1} - \binom{d-1}{k_{01}-1} \right]^{d-(k-1)}. \end{aligned} \quad (11)$$

Proof Let B denote the union of A_j 's for all $j \neq i$ as in Eq. (8) and note that B^c is the event that none of the A_j 's occur for $j \neq i$. Using girth ≥ 5 and the resulting independence from conditioning on x_i we can write

$$\Pr(B^c) = \Pr\left(\bigcap_{j \in n(i)} A_j^c \mid x_i = 0\right) = \prod_{j \in n(i)} \Pr(A_j^c \mid x_i = 0). \quad (12)$$

Next we have

$$\begin{aligned} \Pr(A_j^c \mid x_i = 0) &= \frac{1}{2} \sum_{r \in \{0,1\}} \Pr(A_j^c \mid x_i = 0 \text{ and } x_j = r) \\ &= \frac{1}{2} \cdot \frac{2^{d-1} - \binom{d-1}{k_{01}-1}}{2^{d-1}} \\ &+ \frac{1}{2} \cdot \frac{2^{d-1} - \binom{d-1}{k_{10}-2}}{2^{d-1}} \\ &= \left[2^d - \binom{d-1}{k_{01}-1} - \binom{d-1}{k_{10}-2} \right] / 2^d, \end{aligned}$$

which substituted into Eq. (12) yields

$$\Pr(B^c) = \left[2^d - \binom{d-1}{k_{01}-1} - \binom{d-1}{k_{10}-2} \right]^d / 2^{d^2}. \quad (13)$$

Next, we compute the probability of $A_i \cap B^c$ as:

$$\begin{aligned} 2^{d^2+1} \times \Pr(A_i \cap B^c) &= 2^{d^2} \sum_{r=0,1} \Pr(A_i \cap B^c \mid x_i = r) \\ &= 2^{d^2} \sum_{r=0,1} \Pr(A_i \mid x_i = r) \cdot \Pr(B^c \mid A_i \text{ and } x_i = r) \\ &= \binom{d}{k_{01}-1} \left[2^{d-1} - \binom{d-1}{k_{10}-2} \right]^{k_{01}-1} \\ &\quad \left[2^{d-1} - \binom{d-1}{k_{01}-1} \right]^{d-(k_{01}-1)} + \binom{d}{k_{10}-1} \\ &\quad \left[2^{d-1} - \binom{d-1}{k_{10}-2} \right]^{k_{10}-1} \left[2^{d-1} - \binom{d-1}{k_{01}-1} \right]^{d-(k_{10}-1)} \\ &= \sum_{k=k_{01},k_{10}} \binom{d}{k-1} \left[2^{d-1} - \binom{d-1}{k_{10}-2} \right]^{k-1} \\ &\quad \times \left[2^{d-1} - \binom{d-1}{k_{01}-1} \right]^{d-(k-1)} \end{aligned} \quad (14)$$

Substituting Eqs. (13) and (14) into Eq. (8) leads to Eq. (11), which ends the proof. \square

Since the standard threshold function is a special case of the bi-threshold function when k_{01} and k_{10} coincide, it is straightforward to obtain the following corollary.

Corollary 1 If F is the Boolean network with k -threshold vertex functions over a d -regular graph of girth ≥ 5 , then the activity of F with respect to vertex i is

$$\begin{aligned}\bar{\alpha}_{F,i} = 1 - & \left[1 - \binom{d}{k-1} / 2^d \right]^d \\ & + \frac{\binom{d}{k-1}}{2^d} \left[1 - \binom{d-1}{k-2} / 2^{d-1} \right]^{k-1} \\ & \times \left[1 - \binom{d-1}{k-1} / 2^{d-1} \right]^{d-(k-1)}.\end{aligned}\quad (15)$$

We next consider the nor-function.

Proposition 5 Let X be a d -regular graph of girth ≥ 5 and F the Boolean network over X induced by the nor-function. Then the activity of F with respect to i is given by

$$\bar{\alpha}_{F,i} = 1 - \left(1 - \frac{1}{2^d} \right)^d + \left(\frac{1}{2} - \frac{1}{2^d} \right)^d. \quad (16)$$

Proof Conditioning on $x_i = 0$ and using independence we have

$$\Pr(B^c) = \Pr\left(\bigcap_{j \in n(i)} A_j^c\right) = \prod_{j \in n(i)} \Pr(A_j^c). \quad (17)$$

For a state x with $x_i = 0$ to be in \bar{A}_j , all remaining d states of \bar{A}_j must be zero, leading to $\Pr(\bar{A}_j^c) = 1 - \frac{1}{2^d}$, and

$$\Pr(\bar{B}^c) = \left(1 - \frac{1}{2^d} \right)^d. \quad (18)$$

In order to calculate $\Pr(A_i \cap B^c)$ we note that $\Pr(A_i \cap B^c) = \Pr(A_i)\Pr(B^c|A_i)$ and again use independence to obtain

$$\Pr(\bar{B}^c|\bar{A}_i) = \prod_{j \in n(i)} \Pr(\bar{A}_j^c|\bar{A}_i). \quad (19)$$

Note that $\Pr(\bar{A}_j|\bar{A}_i) = 1/2^{d-1}$ so that $\Pr(\bar{A}_j^c|\bar{A}_i) = 1 - \Pr(\bar{A}_j|\bar{A}_i) = 1 - 1/2^{d-1}$ which substituted into (19) gives

$$\Pr(\bar{B}^c|\bar{A}_i) = \left(1 - \frac{1}{2^{d-1}} \right)^d.$$

Noting that $\Pr(\bar{A}_i) = 1/2^d$ we obtain the third term in (15), finishing the proof. \square

6 Activity over square lattices

In this section we consider graphs with girth-4 edges or girth 4. Here the 1-neighborhoods $n[j]$ with $j \neq i$ may intersect, the key aspect we want to address here. As an example, the reader may verify that for threshold-2

functions and Circle4, the activity of any vertex is $3/4$ and not $7/8$ as when $n \geq 5$ in Proposition 3 (Fig. 3).

To be specific we take the graph X to be a regular, square 2-dimensional lattice, see Fig. 4. It may be either infinite or with periodic boundary conditions. In the latter case, we assume for simplicity that its two dimensions are at least 5. This graph differs from the 4-regular tree by the introduction of type-4 edges: sets A_j and A_{j+1} of $n(i)$, when conditioned on the state of vertex i , are no longer independent. The graph has cycles of size 4 containing i . We first illustrate this case using nor-functions as these allow for a somewhat simplified evaluation as compared to what happens for threshold functions.

Proposition 6 Let X be the 2-dimensional lattice as above where every vertex has degree 4, and let F be the Boolean network over X induced by nor-functions. Then the activity of F for every vertex i is $\bar{\alpha}_{F,i} = 1040/2^{12} = 0.254$.

Proof The neighborhoods of i involved are illustrated in Fig. 4. We rewrite Eq. (8) as follows

$$\Pr\left(\bigcup_{j \in n[i]} A_j\right) = \Pr(B) + \Pr(A_i)(1 - \Pr(B|A_i)). \quad (20)$$

To determine $|B|$, refer to Fig. 4. With indices $p, q, r, s \in \{1, 2, 3, 4\}$ we have

$$\begin{aligned}|A_1 \cup A_2 \cup A_3 \cup A_4| = & \sum_{p=1}^4 |A_p| - \sum_{p < q} |A_p \cap A_q| \\ & + \sum_{p < q < r} |A_p \cap A_q \cap A_r| \\ & - \sum_{p < q < r < s} |A_p \cap A_q \cap A_r \cap A_s|,\end{aligned}\quad (21)$$

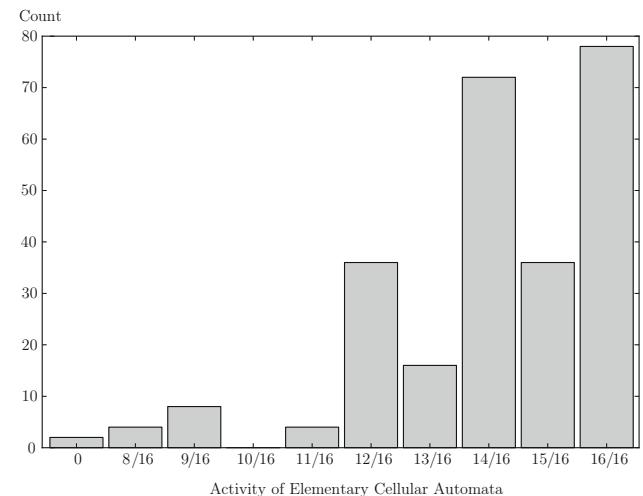


Fig. 3 A diagram showing the possible activity values of ECA (horizontal axis) and their frequency (vertical axis)

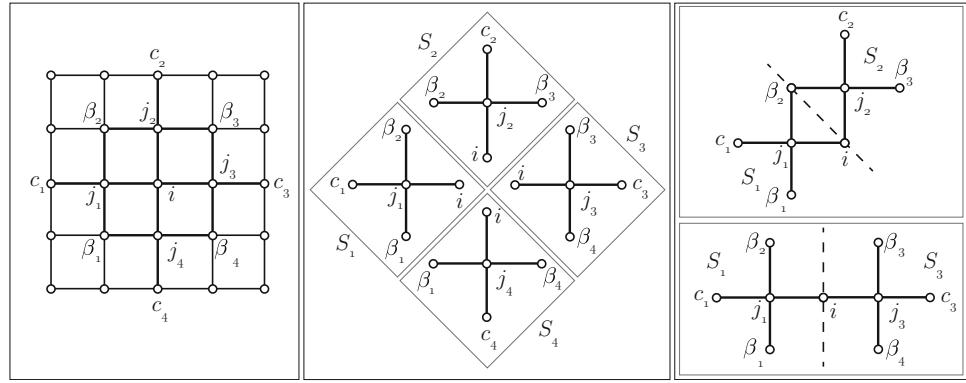


Fig. 4 A subgraph of the square lattice including $X(i; 2)$ at center (left). The four adjacent sets of A_{j_ℓ} correspond to the subgraphs S_ℓ that are illustrated in the middle as templates. Finally, on the right, the two

possible intersections of 2 adjacent sets/templates are illustrated. For one set/template (middle) or the intersection of 3 sets (not shown), there is only one possible configuration modulo rotational symmetry

which, along with Eq. (20), is also valid if we condition on $x_i = 0$ in every set. We take $i \neq j_\ell$ with $\ell \in \{1, 2, 3, 4\}$ and let j_ℓ denote the vertex at the center of $n[j_\ell]$ and i the vertex at the center of $n[i]$. Here $|X(i; 2)| = 13$. Using the symmetry of the sets \bar{A}_{j_ℓ} and the fact that a state x is in the set \bar{A}_{j_ℓ} precisely when $x[j_\ell] = 0$ leads to

$$\sum_{\ell=1}^4 |\bar{A}_{j_\ell}| = 4 \cdot 2^8 .$$

For the second term on the right in (21) we have two cases for intersections: (i) two adjacent vertices $j_\ell, j_{\ell+1} \in n(i)$ with all indices modulo 4, of which there are four instances, and (ii) two non-adjacent vertices $j_\ell, j_{\ell+2} \in n(i)$, of which there are two instances. In the first case we obtain $|\bar{A}_{j_\ell} \cap \bar{A}_{j_{\ell+1}}| = 2^5$, while in the second case we have $|\bar{A}_{j_\ell} \cap \bar{A}_{j_{\ell+2}}| = 2^4$, making the second term on the right in (21)

$$\sum_{p < q} |\bar{A}_{j_p} \cap \bar{A}_{j_q}| = 4 \cdot 2^5 + 2 \cdot 2^4 . \quad (22)$$

The third sum appearing on the right in (21) contains four terms all corresponding to an intersection of three sets A_{j_ℓ} , and we obtain

$$\sum_{p < q < r} |\bar{A}_{j_p} \cap \bar{A}_{j_q} \cap \bar{A}_{j_r}| = 4 \cdot 2^2 . \quad (23)$$

Finally, the intersection of all four sets \bar{A}_{j_ℓ} has size 1, and we get

$$\Pr(\bar{B}) = |\bar{B}|/2^{12} = [4 \cdot 2^8 - (4 \cdot 2^5 + 2 \cdot 2^4) + 4 \cdot 2^2 - 1]/2^{12} \quad (24)$$

We next determine $\Pr(B|A_i)$. For a state x to be in A_i we must have $x[i] = 0$ so that $|A_i| = 2^8$. Conditioning on this, we can calculate $|\bigcup_{\ell=1}^4 A'_{j_\ell}|$ as above. The reader can verify that

$$\begin{aligned} \sum_{\ell=1}^4 |A'_{j_\ell}| &= 4 \cdot 2^5 , \\ \sum_{p < q} |A'_{j_p} \cap A'_{j_q}| &= 4 \cdot 2^3 + 2 \cdot 2^2 , \\ \sum_{p < q < r} |A'_{j_p} \cap A'_{j_q} \cap A'_{j_r}| &= 4 \cdot 2 , \\ |A'_{j_1} \cap A'_{j_2} \cap A'_{j_3} \cap A'_{j_4}| &= 1 , \end{aligned} \quad (25)$$

leading to the expression

$$\begin{aligned} \Pr(\bar{A}_i \cap \bar{B}^c) &= \Pr(\bar{A}_i)(1 - \Pr(\bar{B}|\bar{A}_i)) \\ &= \frac{1}{2^4} \left(1 - \frac{1}{2^8} [2^7 - 2^5 - 2^3 + 2^3 - 1] \right) , \end{aligned}$$

which, together with (24) in (20) gives the stated result of $\bar{\alpha}_{F,i} = 1040/2^{12}$. \square

6.1 Activity of threshold functions over square lattices

We next consider the square lattice where each vertex function is the k -threshold function. For this case, the analytic derivations resemble those of the nor-case just covered, but the expressions appearing are more involved.

Proposition 7 *Let X be a rectangular grid (torus) in which every vertex has degree 4, and F the map over X with the threshold vertex functions. Then the activity of F with respect to vertex i is given by*

$$\cdot \alpha_{F,i} = \left(\frac{1}{2^{12}} \right) [\gamma_1 - \gamma_2 + \gamma_3 - \gamma_4 + \gamma_5] , \quad (26)$$

where

$$\gamma_1 = 2^{10} \cdot \binom{4}{k-1} , \quad (27)$$

$$\gamma_2 = 2^7 \left[\binom{3}{k-1}^2 + \binom{3}{k-2}^2 \right] + 2^5 \binom{4}{k-1}^2, \quad (28)$$

$$\begin{aligned} \gamma_3 = & 2^4 \left[\binom{3}{k-1}^2 \binom{2}{k-1} \right. \\ & \left. + 2 \binom{3}{k-2} \binom{2}{k-2} \binom{3}{k-1} + \binom{3}{k-2}^2 \binom{2}{k-3} \right], \end{aligned} \quad (29)$$

$$\begin{aligned} \gamma_4 = & \left[\binom{2}{k-1}^4 + 4 \binom{2}{k-1}^2 \binom{2}{k-2}^2 \right. \\ & + 4 \binom{2}{k-1} \binom{2}{k-2}^2 \binom{2}{k-3} \\ & \left. + 2 \binom{2}{k-2}^4 + 4 \binom{2}{k-2}^2 \binom{2}{k-3}^2 + \binom{2}{k-3}^4 \right], \end{aligned} \quad (30)$$

and

$$\gamma_5 = \sum_{x \in M} \mathbb{I}[x_{j_\ell} + x_{\beta_\ell} + x_{\beta_{\ell+1}} + x_{c_\ell} \neq k-1], \quad (31)$$

where M is the subset of states associated to $X(i; 2)$ for which the Hamming norm of its elements projected to the components j_ℓ with $1 \leq \ell \leq 4$ is $k-1$, that is, $|(x_{j_1}, x_{j_2}, x_{j_3}, x_{j_4})|_H = k-1$, and where x_{j_ℓ} , x_{β_ℓ} and x_{c_ℓ} are the states of the vertices j_ℓ , β_ℓ and c_ℓ as given in Fig. 4.

Proof We begin with Eq. (8), and by the symmetry of thresholds, we can condition on $x_i = 0$, to write

$$\bar{\alpha}_{F,i} = \Pr(B | x_i = 0) + \Pr(A_i \cap B^c | x_i = 0). \quad (32)$$

Here, B is the union of A_{j_ℓ} s, and we take $i \neq j_\ell$ for all $\ell \in \{1, 2, 3, 4\}$. We have A_{j_ℓ} as a particular A_j . Let j_ℓ denote the vertex at the “center” of subgraph S_ℓ with $d = 4$ (per Fig. 4), as vertex i is the “center” of S_i , and let x_{j_ℓ} denote the state of j_ℓ . There is a one-to-one correspondence between S_ℓ and A_{j_ℓ} since the vertices associated with each are the same subset of those of $n[j_\ell]$. We evaluate $\Pr(B | x_i = 0)$ and $\Pr(A_i \cap B^c | x_i = 0)$ in turn, using the “templates” formed from the S_ℓ .

The first four terms γ_r with $1 \leq r \leq 4$ in Eq. (26), come from the first term on the right hand side (RHS) of Eq. (32). We have, by symmetry in the $0 \rightarrow 1$ and $1 \rightarrow 0$ transitions for x_i ,

$$\begin{aligned} \Pr(B | x_i = 0) &= \Pr \left[\bigcup_{\substack{j_\ell \in n[i] \\ j_\ell \neq i}} (A_{j_\ell} | x_i = 0) \right] \\ &= |A_1 \cup A_2 \cup A_3 \cup A_4| / 2^{12}, \end{aligned}$$

where the 12 vertices in the template of Fig. 4 can be in either state 0 or 1 ($x_i = 0$ is fixed).

The inclusion/exclusion principle can be used to evaluate the numerator in the RHS of this last expression, which is independent of the girth of the graph. We also note that the last expression of the union of the A_{j_ℓ} s is subject to the condition $x_i = 0$. We use the templates in Fig. 4, without loss of generality, to compute the value of the last equation. Each of the expressions for γ_1 through γ_4 take the same form, which is a product of the multiplicity of the particular template in $X(i; 2)$; the number z of the 12 vertices whose states can be either zero or one because they do not play a role in the particular template being evaluated (giving rise to a term 2^z); and an expression in binomial coefficients that represent the number of sets of vertex states in the template whose value must be 1. In this case, the first summation on the RHS of (21) evaluates to

$$\gamma_1 = \sum_{p=1}^4 |A_p| = 4 \cdot 2^{12-4} \cdot \binom{4}{k-1}, \quad (33)$$

using symmetry across A_{j_ℓ} with $\ell \in \{1, 2, 3, 4\}$ of Fig. 4, center diagram. The multiplicity is 4 because there are four A_{j_ℓ} instances. The 2^{12-4} comes from the fact that there are 12 vertices in \bar{K}^{N_i} whose states can be chosen, and we specify 4 vertex states in each template S_ℓ . The 4 in the binomial comes from the fact that there are four vertices in each S_ℓ that can be assigned a value in $\{0, 1\}$ independently of the state assignments to the other vertices (we call these free vertices), noting $x_i = 0$, and $k-1$ comes from the fact that when x_i changes from state 0 to 1 (i.e., when its threshold is satisfied), A_{j_ℓ} is satisfied.

For the second term on the RHS of Eq. (21), we have two cases (see the right-most drawings of Fig. 4): (i) four instances with A_{j_ℓ} and $A_{j_{\ell+1}}$ (all indices are modulo 4) with a total of 7 vertices in each instance that can take states 0 and 1 (so that $z = 12 - 7$); and (ii) two instances with A_{j_ℓ} and $A_{j_{\ell+2}}$ where each instance has 8 vertices that can take states 0 and 1, giving $z = 12 - 8$. Hence, the β_ℓ are not free vertices for case (i), but they are for case (ii). In case (i), there is a common vertex, $\beta_{\ell+1}$, and hence there are three free vertices in each S_ℓ . Fixing $x_{\beta_{\ell+1}}$ in turn, to be 0 and then 1, we get $|A_{j_\ell} \cap A_{j_{\ell+1}}| = 4 \cdot 2^{12-7} \left[\binom{3}{k-1}^2 + \binom{3}{k-2}^2 \right]$. For case (ii), there is no common vertex beyond i , so we have four free vertices in each of A_{j_ℓ} and $A_{j_{\ell+2}}$. We get $|A_{j_\ell} \cap A_{j_{\ell+2}}| = 2 \cdot 2^{12-8} \binom{4}{k-1}^2$. In total, the second summation on the RHS of Eq. (21) is $\sum_{p < q} |A_p \cap A_q| = \gamma_2$, with γ_2 given in Eq. (28).

The third summation in the RHS of Eq. (21) has a single case: the intersection of three of the four A_{j_ℓ} s. There are four instances, with each instance composed of 10 vertices to which states may be assigned, giving a multiplier of $4 \cdot 2^{12-10} = 2^4$ in γ_3 in Eq. (29). Without loss of generality, we focus on A_{j_ℓ} , $\ell \in \{1, 2, 3\}$ of Fig. 4. Here, A_{j_1} has three free vertices (β_{j_2} is common to A_{j_1} and A_{j_2}); A_{j_2} has two free vertices (β_{j_2} and β_{j_3} are common to A_{j_1} and A_{j_3} , respectively); and A_{j_3} has three free vertices. The three terms in $[\cdot]$ in Eq. (29) are produced by assigning the state pairs $(x_{\beta_2}, x_{\beta_3})$ equal to, in turn, (a) (0,0); (b) (0,1) and then (1,0); and (c) (1,1).

The final summation in the RHS of Eq. (21) also has a single case: the intersection of all four A_{j_ℓ} s. There is one instance. All four β_ℓ are considered, so we take as subcases the number of β_ℓ where $x_{\beta_\ell} = 1$, is, in turn, 0, 1, 2, 3, and 4. It can be shown, following analogous procedures to those for the expression for γ_3 , that $\sum_{p < q < r < s} |A_p \cap A_q \cap A_r \cap A_s| = \gamma_4$; see Eq. (30).

To evaluate $\Pr(A_i \cap B^c \mid x_i = 0)$ in Eq. (32), we note that $B^c = \left[\bigcap_{j_\ell \in n(i)} A_{j_\ell}^c \right]$. The states contained in $A_i \cap B^c$ are therefore those contained in A_i but none of the four sets A_{j_ℓ} .

For a given k , consider the state x . We see that $x \in A_i$ if and only if exactly $k - 1$ of the adjacent vertices of i have state $x_{j_\ell} = 1$ and the remaining $4 - (k - 1)$ adjacent states satisfy $x_{j_\ell} = 0$. For $x \notin A_{j_\ell}$ to hold, it is necessary and sufficient that

$$x_{j_\ell} + x_{\beta_\ell} + x_{\beta_{\ell+1}} + x_{\ell} \neq k - 1, \quad (34)$$

for $1 \leq \ell \leq 4$. We can partition these cases as follows: First, consider all 16 elements $x_\beta = (x_{\beta_1}, x_{\beta_2}, x_{\beta_3}, x_{\beta_4}) \in \{0, 1\}^4$ (the four overlapping vertices $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)$ in Fig. 4). For each x_β , consider all $\hat{x}_j = (x_{j_1}, x_{j_2}, x_{j_3}, x_{j_4}) \in \{0, 1\}^4$ of adjacent vertex states of i of Hamming norm $|\hat{x}_j| = k - 1$. Finally, we consider all $x_c = (x_{c_1}, x_{c_2}, x_{c_3}, x_{c_4}) \in \{0, 1\}^4$. We thus count the the number of elements $x \in M$ for which Eq. (34) holds, producing γ_5 in Eq. (31) and completing the proof. \square

We have shown the graph of the activity as a function of the threshold value k in Fig. 5. Note that for thresholds in the range $1 < k < 5$, the activity for the threshold function is considerably larger than that for the nor-function given immediately above (Table 2; Fig. 5).

Table 2 Activity value (rounded to four decimal places) as a function of the threshold k for the 4-regular square lattice and the 4-regular tree. Thresholds 0 and 6 omitted

Graph	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
4-lattice	0.2539	0.7095	0.9009	0.7095	0.2539
4-regular tree	0.2641	0.7370	0.9046	0.7370	0.2641

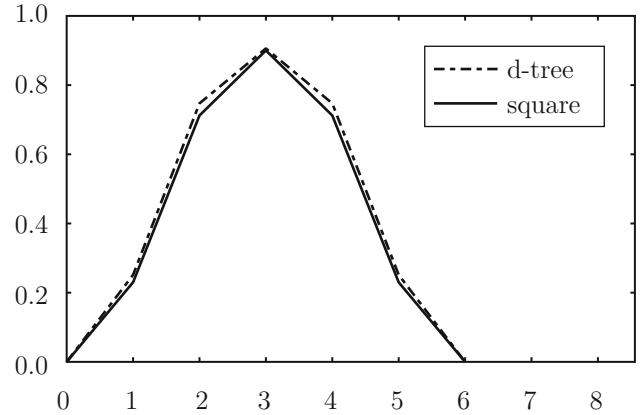


Fig. 5 Activity as a function of the threshold value k for the square lattice of Fig. 4 (solid line). For comparison, the activity for the 4-regular tree is included (dashed line)

7 Activity over triangular lattices

A triangular lattice is a regular graph with type-3 edges and a girth of 3 as indicated in the left of Fig. 6. In this case, the structure of the intersections among sets A_j (with $j \neq i$) is more involved than for the square lattice case, and type-3 edges are introduced. In principle, the derivations are similar to those of the square lattice case. Rather than giving a similar argument to the proof of Proposition 7, we limit ourselves to showing the graph of the activity as a function of the threshold value k , see Fig. 6 right side, but see also Table 3 for the approximate values.

8 Activity over the Erdős–Rényi random graphs

Here, we provide an upper bound for the expected activity in $G_{n,p}$ for the uniform k -threshold function using the union bound. While bound obtained this way may not turn out to strong, the work in this section takes the first steps towards analyzing activity over ER random graphs.

First, let $B[n, p]$ denote the binomial random variable and let

$$f(t; n, p) = \Pr(B[n, p] = t) = \binom{n}{t} p^t (1-p)^{n-t}.$$

Since we are considering a family of graphs, we find it necessary to redefine some of the terminology developed earlier. Let i be the vertex under consideration. Let $\bar{\alpha}_{F,i}(G)$ denote the activity for $G \in G_{n,p}$. The expected activity $\mathbb{E}[\bar{\alpha}_{F,i}] := \mathbb{E}_{G(n,p)}[\bar{\alpha}_{F,i}]$.

Note that, by symmetry $\bar{\alpha}_{F,i}$ is the same for all $i \in \{1, \dots, n\}$. Next, recall the definition of $\bar{A}_j(i)$ (or \bar{A}_j in short) from Section 3. Here, we modify it as $\bar{A}_j(G)$ to denote the set of states x which satisfy $x_i = 0$ and $F(x + e_i)_j \neq F(x)_j$ in G .

Fig. 6 Left: the triangular lattice. Right: activity as a function of threshold for the triangular lattice (solid line). For comparison, the activity for the 6-regular tree is included (dashed line)

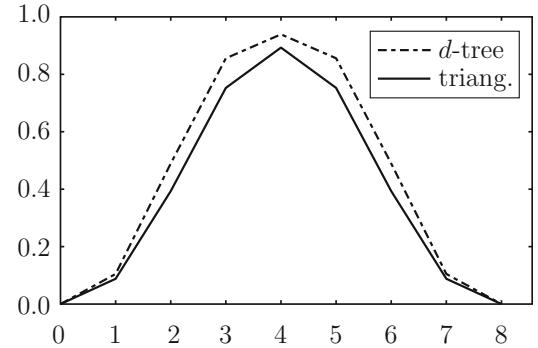
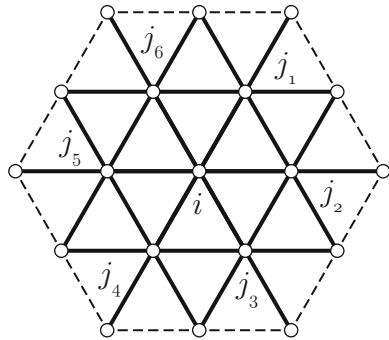


Table 3 Activity value (rounded to four decimal places) as a function of the threshold k for the triangular lattice. Thresholds 0 and 8 omitted

Graph k	1	2	3	4	5	6	7
Triangular lattice	0.0881	0.3963	0.7603	0.9019	0.7603	0.3963	0.0881

Proposition 8 If F is the Boolean network induced by the k -threshold function over $G(n, p)$, then $\mathbb{E}[\bar{x}_{F,i}] \leq f(k-1; n-1, p/2) + \frac{(n-1)p}{2}(f(k-1; n-1, p/2) + f(k-2; n-2, p/2))$.

Proof Let $n(i, G)$ and $n[i, G]$ denote the open and closed neighborhoods of i in G .

Using the union bound on (5), we have

$$\mathbb{E}[\bar{x}_{F,i}] = \sum_{G \in G_{n,p}} \Pr(G) \Pr\left(\bigcup_{j \in n[i,G]} \bar{A}_j(G)\right) \quad (35)$$

$$\leq \sum_{G \in G_{n,p}} \Pr(G) \Pr(\bar{A}_i(G)) + \sum_{G \in G_{n,p}} \Pr(G) \Pr\left(\bigcup_{j \in n(i,G)} \bar{A}_j(G)\right) \quad (36)$$

$$\leq \sum_{G \in G_{n,p}} \Pr(G) \Pr(\bar{A}_i(G)) + \sum_{G \in G_{n,p}} \Pr(G) \sum_{j \in n(i,G)} \Pr(\bar{A}_j(G)). \quad (37)$$

Now we will compute the two summands separately. First, we note that the term $\sum_{G \in G_{n,p}} \Pr(G) \Pr(\bar{A}_i(G))$ is precisely the probability that a randomly chosen pair (x, G) satisfies $x \in \bar{A}_i(G)$. By randomly chosen, we mean that (x, G) is sampled by choosing independently x from K^n and G from $G_{n,p}$. This probability can be computed easily as follows. For any $x \in \bar{A}_i(G)$, i must have exactly $k-1$ neighbors in state 1. For any vertex $j \neq i$, the probability that it is a neighbor of i and is in state 1 is $p/2$. Therefore, the number of neighbors of i in state 1 is binomially distributed with parameters $n-1$ and $p/2$, i.e., $B[n-1, p/2]$. Hence,

$$\begin{aligned} \sum_{G \in G_{n,p}} \Pr(G) \Pr(\bar{A}_i(G)) &= \Pr(B[n-1, p/2] = k-1) \\ &= f(k-1; n-1, p/2). \end{aligned} \quad (38)$$

Next, we will compute $\sum_{G \in G_{n,p}} \Pr(G) \sum_{j \in n(i,G)} \Pr(\bar{A}_j(G))$. Rearranging the summation,

$$\sum_{G \in G_{n,p}} \Pr(G) \sum_{j \in n(i,G)} \Pr(\bar{A}_j(G)) = \sum_{j \neq i} \sum_{G \in G_{n,p}} \Pr(G) \Pr(\bar{A}_j(G)).$$

As in the previous case, the term $\sum_{G \in G_{n,p}} \Pr(G) \Pr(\bar{A}_j(G))$ corresponds to the probability that a randomly chosen pair (x, G) satisfies $x \in \bar{A}_j(G)$. This probability can be computed as follows: For (x, G) to satisfy the above condition, j must be a neighbor of i and satisfy one of these conditions: (i) if $x_j = 0$, then, j must have exactly $k-1$ neighbors excluding i in state 1, and (ii) if $x_j = 1$, then, j must have exactly $k-2$ neighbors, excluding i , in state 1. Let Y_j^1 be the number of neighbors of j excluding i at state 1 in x . Note that this is binomially distributed with parameters $n-2$ and $p/2$. Therefore,

$$\begin{aligned} \sum_{G \in G_{n,p}} \Pr(G) \sum_{j \in n(i,G)} \Pr(\bar{A}_j(G)) &= (\Pr(x_j = 0) \Pr(x \in \bar{A}_j | x_j = 0) \\ &\quad + \Pr(x_j = 1) \Pr(x \in \bar{A}_j | x_j = 1)) \\ &= \frac{p}{2} (\Pr(Y_j^1 = k-1) + \Pr(Y_j^1 = k-2)) \\ &= \frac{p}{2} (f(k-1; n-2, p/2) + f(k-2; n-2, p/2)) \end{aligned} \quad (39)$$

Finally, applying (38) and (39) in (37), we have

$$\begin{aligned}\mathbb{E}[\bar{x}_{F,i}] &\leq f(k-1; n-1, p/2) + \sum_{j \neq i}^p \frac{p}{2} (f(k-1; n-1, p/2) \\ &\quad + f(k-2; n-2, p/2)) = f(k-1; n-1, p/2) \\ &\quad + \frac{(n-1)p}{2} (f(k-1; n-1, p/2) \\ &\quad + f(k-2; n-2, p/2)).\end{aligned}$$

Hence proved. \square

9 Summary and research directions

In this extended version of Adiga et al. (2015) we have incorporated new and additional results. Our notion of activity extends what was proposed by Shmulevich and Kauffman Shmulevich and Kauffman (2004). This extension takes into account the impact of the network structure when studying $\mathbb{E}[\mathbb{I}[F(x) \neq F(x + e_i)]]$, which estimates how likely the perturbation e_i will cause successor states to diverge after one time step.

Naturally, orbits that initially separate may later converge, reflecting that x and $x + e_i$ may belong to the same attractor basin. Nonetheless, this notion of activity provides a measure for sensitivity with respect to initial conditions that accounts for network structure. Investigating *long-term activity*, that is, the probability that perturbing a state in the i th coordinate will cause x and $x + e_i$ to have different ω -limit sets, is an interesting direction for future research. Involving the attractors of F , we naturally expect this to be challenging work, even in most special cases.

In this paper we have started analysis for activity over random graphs. Future research could include extensions of this work to provide stronger bounds. Analysis of dynamics of Boolean networks, as well as the more general class of graph dynamical systems, in the setting of random graphs also seem to hold much potential for interesting insight. Our results also show that the activity for the d -regular trees using threshold functions is larger than those of the corresponding square and triangular grids. Clearly, the combinatorial arguments become more involved with the presence of type-3 and type-4 edges. In general, what can be said about the impact of these classes of edges on activity?

For efficient computations of activity one may consider using model counting SAT solvers. This was not explored by us, however, it was commented by a reviewer. For networks where $n_i := |X(i; 2)|$ is “not too large”, one may consider developing a tool to map from vertex functions and the expressions involved in the definitions of the sets A_k to suitable formats for this class of SAT solvers. For

additional directions for future work we refer to Adiga et al. (2015).

Acknowledgments We thank our external collaborators and members of the Network Dynamics and Simulation Science Laboratory (NDSSL) for their suggestions and comments. We also thank two anonymous reviewers for valuable comments. This work has been partially supported by DTRA Grant HDTRA1-11-1-0016.

References

- Adiga A, Kuhlman CJ, Mortveit HS, Vullikanti AKS (2013) Sensitivity of diffusion dynamics to network uncertainty. In: Proceedings of the twenty-seventh AAAI conference on artificial intelligence (AAAI-13), July 14–18, 2013. Bellevue, Washington, USA, pp 2–8
- Adiga A, Galyean H, Kuhlman CJ, Levet M, Mortveit HS, Wu S (2015) Network structure and activity in Boolean networks. In: Kari J (ed) Cellular automata and discrete complex systems: proceedings of AUTOMATA 2015, Turku, Finland, June 8–10, 2015, Lecture Notes in Computer Science, vol 9099, pp 210–223, doi:[10.1007/978-3-662-47221-7_16](https://doi.org/10.1007/978-3-662-47221-7_16)
- Aldana M, Coppersmith S, Kadanoff LP (2003) Boolean dynamics with random couplings. In: Perspectives and problems in nonlinear science, Springer, pp 23–89
- Baetens JM, De Baets B (2010) Phenomenological study of irregular cellular automata based on Lyapunov exponents and Jacobians. Chaos 20:1–15. doi:[10.1063/1.3460362](https://doi.org/10.1063/1.3460362)
- Baetens JM, Van der Weeën P, De Baets B (2012) Effect of asynchronous updating on the stability of cellular automata. Chaos Solitons Fractals 45:383–394. doi:[10.1016/j.chaos.2012.01.002](https://doi.org/10.1016/j.chaos.2012.01.002)
- Derrida B, Pomeau Y (1986) Random networks of automata: a simple annealed approximation. Europhys Lett 1:45–49
- Fretter C, Szekja A, Drossel B (2009) Perturbation propagation in random and evolved Boolean networks. N J Phys 11:1–13. doi:[10.1088/1367-2630/11/3/033005](https://doi.org/10.1088/1367-2630/11/3/033005)
- Ghanbarnejad F, Klemm K (2012) Impact of individual nodes in Boolean network dynamics. EPL (Europhys Lett) 99(5):58,006
- Goles E, Martinez S (1990) Neural and automata networks: dynamical behaviour and applications. Kluwer Academic Publishers, Berlin
- Kauffman SA (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. J Theor Biol 22:437–467
- Kauffman SA (1993) The origins of order: self-organization and selection in evolution. Oxford University Press, Oxford
- Kuhlman CJ, Mortveit HS (2014) Attractor stability in nonuniform Boolean networks. Theor Comput Sci 559:20–33. doi:[10.1016/j.tcs.2014.08.010](https://doi.org/10.1016/j.tcs.2014.08.010) special volume: Non-uniform Cellular Automata
- Layne L, Dimitrova E, Matthew M (2012) Nested canalizing depth and network stability. Bull Math Biol. doi:[10.1007/s11538-011-9692-y](https://doi.org/10.1007/s11538-011-9692-y)
- Luo JX, Turner MS (2012) Evolving sensitivity balances Boolean networks. PLoS One 7(e36):010. doi:[10.1371/journal.pone.0036010](https://doi.org/10.1371/journal.pone.0036010)
- Mortveit HS, Reidys CM (2007) An introduction to sequential dynamical systems. Universitext, Springer. doi:[10.1007/978-0-387-49879-9](https://doi.org/10.1007/978-0-387-49879-9)
- Pomerance A, Ott E, Girvan M, Losert W (2009) The effect of network topology on the stability of discrete state models of genetic control. Proc Nat Acad Sci 106(20):8209–8214
- Ribeiro AS, Kauffman SA (2007) Noisy attractors and ergodic sets in models of gene regulatory networks. J Theor Biol 247:743–755

- Robert F (1986) Discrete iterations. A Metric Study. No. 6 in Springer Series in Computational Mathematics, Springer
- Serra R, Villani M, Barbieri A, Kauffman S, Colacci A (2010) On the dynamics of random Boolean networks subject to noise: attractors, ergodic sets and cell types. *J Theor Biol* 265(2):185–193
- Shmulevich I, Kauffman SA (2004) Activities and sensitivities in Boolean network models. *Phys Rev Lett* 93(4):048701:1–4
- Shmulevich I, Lähdesmäki H, Dougherty ER, Astola J, Zhang W (2003) The role of certain post classes in Boolean network models of genetic networks. *Proc Nat Acad Sci* 100(19):10,734–10,739
- Xiao Y, Dougherty ER (2007) The impact of function perturbations in Boolean networks. *Bioinformatics* 23(10):1265–1273



Inferring local transition functions of discrete dynamical systems from observations of system behavior



Abhijin Adiga^a, Chris J. Kuhlman^a, Madhav V. Marathe^a, S.S. Ravi^{b,*}, Daniel J. Rosenkrantz^b, Richard E. Stearns^b

^a Network Dynamics and Simulation Science Laboratory, VBI, Virginia Tech, USA

^b Computer Science Department, University at Albany – SUNY, USA

ARTICLE INFO

Article history:

Received 30 November 2015

Received in revised form 7 June 2016

Accepted 6 July 2016

Available online 16 July 2016

Keywords:

Threshold inference

Computational complexity

Synchronous dynamical systems

Threshold functions

Phase space

Trajectory

Stable and unstable configurations

Fixed parameter tractability

ABSTRACT

We consider the problem of inferring the local transition functions of discrete dynamical systems from observed behavior. Our focus is on synchronous systems whose local transition functions are threshold functions. We assume that the topology of the system is known and that the goal is to infer a threshold value for each node so that the system produces the observed behavior. We show that some of these inference problems are efficiently solvable while others are NP-complete, even when the underlying graph of the dynamical system is a simple path. We identify a fixed parameter tractable problem in this context. We also consider constrained versions of threshold inference problems where the input includes a set of equality or inequality constraints (which specify pairs of nodes which must have the same threshold value or different threshold values). We present algorithmic and complexity results for several constrained threshold inference problems.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

1.1. Motivation

Many studies have demonstrated the efficacy of using discrete dynamical systems to model population dynamics. Among these are studies of incarceration [1], epidemics and infectious diseases [2,3], and information spread through blogs [4]. It has also been demonstrated that dynamical system models can produce dramatically different results, depending on the parameter values used. For example, in certain disease models, small changes in infection and recovery rates, as quantified by the epidemic threshold, can change the size of an outbreak from large to small [5]. As another example, in complex contagion models, small changes in threshold values can produce large changes in contagion spread [6,7]. These considerations are among the issues within the realm of model verification and validation.

Methods that use observations of systems to calibrate or infer models and model properties arise in many contexts. These properties help explain system behaviors, and parameterized models may be transferable to other contexts [8]. A case in point is a protest in Spain in 2011, in which people demonstrated against economic austerity measures [9]. In that work on information spread via Twitter, threshold behavior was used to model tweeting about the protest. If a person v received

* Corresponding author.

E-mail addresses: abhijin@vbi.vt.edu (A. Adiga), ckuhlman@vbi.vt.edu (C.J. Kuhlman), mmarathe@vbi.vt.edu (M.V. Marathe), ssravi0@gmail.com (S.S. Ravi), drosenkrantz@gmail.com (D.J. Rosenkrantz), thestearns2@gmail.com (R.E. Stearns).

t_v tweets (from unique users that she followed) about the event, and then tweeted for the first time before v received the $(t_v + 1)$ th tweet, that person's threshold for participation was chosen as t_v . The inference of thresholds enabled the full specification of a model for the underlying socio-technical phenomenon. Several other studies along the same lines have also been carried out (e.g., [10,11]). The threshold model for social networks, introduced in [6], is used in many applications (see e.g., [12]). Discrete dynamical systems [13,14], which generalize cellular automata, represent a rigorous and convenient abstract model to study socio-technical phenomena.

In experimental evaluations of contagion processes, different subpopulations behave differently, suggesting that these subgroups have different thresholds. There is evidence for differences in contagion transmission based on such factors as ethnicity, gender, and academic standing. For example, in [15], it is shown that the probability of students dropping out of school is a function of ethnicity. Since dropping out of school can be modeled as a contagion [15,16], this suggests that a threshold model of dropping out would assign different thresholds to different ethnic groups. Similarly, graduate students have a much higher drop out rate than other students at selective institutions [17], indicating that threshold-based contagion models of dropping out would assign a lesser threshold (to drop out) to graduate students than to undergraduates. As a final example, treating depression as a contagion [18], women are more likely to become depressed from their friends' depression than men [19]. This implies that for a threshold model of depression, females would have a lesser threshold than males. These studies motivate the use of different thresholds for different sets of individuals.

The problem of inferring the components of a system from observed behavior has also received a lot of attention in the theoretical computer science literature. For example, many researchers have studied the problem of learning automata from sets of accepted strings (see e.g., [20–22]). Likewise, the problem of learning CNF and DNF formulas has also been studied extensively in the literature (e.g., [23]). Additional information on these topics will be provided in Section 1.3.

1.2. Problems considered

We consider inference problems that arise in the context of discrete dynamical systems. In particular, we focus on one such model, namely *synchronous* discrete dynamical systems (SyDSs). We provide an informal description of a SyDS here; a formal description is given later. A SyDS consists of an undirected graph whose vertices represent entities (agents) and edges represent local interactions among entities. Each vertex has a state value chosen from a finite domain (e.g., $\{0, 1\}$). In addition, each vertex v also has a local transition function whose inputs are the current state of v and those of its neighbors; the output of this function is the next state of v . The vector consisting of the state values of all the nodes at each time instant is referred to as the **configuration** of the system at that instant. In each time step, all nodes of a SyDS compute and update their states *synchronously*. Starting from a (given) initial configuration, the time evolution of a SyDS consists of a sequence of successive configurations. Models similar to SyDSs have been used in several applications, including the propagation of diseases and social phenomena (e.g., [12,24]).

Several researchers have studied the complexity of various **analysis problems** for discrete dynamical systems (e.g., [25–29]). Informally, the goal of an analysis problem is to predict the behavior of a system given its static description. An example of an analysis problem is that of **reachability**: given a SyDS S and two configurations C_1 and C_2 , will S reach C_2 starting from C_1 ? Such analysis questions are studied by considering the **phase space** of the SyDS, which is a directed graph with one vertex for each possible configuration and a directed edge (x, y) from a vertex x to vertex y if the SyDS can transition from the configuration corresponding to x to the one corresponding to y in one time step. In such a case, y is the **successor** of x and x is a **predecessor** of y . Each self loop in the phase space of a SyDS represents a **stable** configuration or a **fixed point** of the system¹ (i.e., a configuration in which the system will stay forever). Any configuration C whose successor is different from C is called an **unstable configuration**. A **trajectory** in phase space is a directed path with one or more edges. Also, any vertex in the phase space with no incoming edges represents a **Garden-of-Eden** (GE) configuration which can arise only as the initial configuration of the system.

Here, our focus is on a problem which may be considered as an *inverse* of the analysis problem. In such a problem, the goal is to infer some aspect of the structure of a partially specified system given a description of its observed behavior. In particular, we focus on inferring the local transition functions of a SyDS, given the underlying graph and appropriate behavior patterns. We consider several different behavior patterns (e.g., collection of stable configurations, collection of trajectories, collection of unstable configurations) and study the complexity of identifying the local transition functions that can explain the observed behavior. In particular, we assume that each local transition function f_i is a t_i -threshold function for some non-negative integer t_i (see Section 2 for the definition of threshold functions).

We present results for threshold inference problems for two categories of observed behavior, namely homogeneous collections (e.g., a set of stable configurations, a set of unstable configurations) and heterogeneous collections (e.g., a collection consisting of a set of stable configurations and a set of unstable configurations). Our results establish the complexity of the corresponding decision and optimization problems (see Sections 3 and 4). Motivated by the considerations mentioned in Section 1.1, we also present algorithmic and complexity results for constrained versions of threshold inference problems (Section 5).

¹ We will use the term "stable configuration" instead of "fixed point" throughout this paper since we use the word "fixed" in the context of fixed parameter tractability.

SyDSs are generalizations of cellular automata (CA) since the underlying graphs of SyDSs can be arbitrary while those of CA are generally restricted to geometric structures such as lines and lattices. Many of the results presented in this paper apply directly to CA with threshold functions at the nodes. For example, since all our efficient algorithms for threshold inference for SyDSs hold for arbitrary graphs, they also hold for CA. Further, our hardness results in Sections 3 and 4 which hold for SyDSs when the underlying graph is a simple path can be thought of as results for 1-dimensional CA. Other problems where hardness results for discrete dynamical systems directly imply similar results for CA are discussed in [27].

1.3. Related work

Several inference problems have been explored in the context of disease, information and meme spread. One direction of work considers estimating model parameters given the traces of a diffusion process, network and a class of models (e.g., Bailon et al. [9]). Abraho et al. [30], Gomez et al. [31], Soundarajan and Hopcroft [32] consider the problem of inferring the network structure given the model. Recently, there has been a lot of work on source detection, where the goal is to find the source of infection given limited information about the network, diffusion model and the set of infected nodes (e.g., [33]). Most of these problems turn out to be hard even for simple models (such as progressive systems [34]) and networks.

Learning finite automata [20,21] and Boolean functions [23] are two rich areas which consider problems with a similar flavor. In the case of learning finite automata, the general problem is to infer a finite (stochastic) automaton given a set of strings labeled as either in the language or not [20–22]. Similarly, in concept learning (or learning Boolean functions), the general task is to infer a Boolean function given information about its values for some inputs, together with the knowledge that it belongs to a particular class of functions [23].

As mentioned earlier, many researchers have addressed various questions regarding phase space properties of discrete dynamical systems. Goles and Martínez [35] provide bounds on the lengths of transients (i.e., trajectories that end in stable configurations) and cycles in threshold dynamical systems. The complexity of determining whether a given configuration y of a deterministic SyDS has a predecessor has been studied in [25]. Problems similar to predecessor existence have also been considered in the context of CA [26,36]. Researchers have also studied various questions for dynamical systems under the sequential update model, where the vertex functions are applied according to a specified order [37,13,28].

Many researchers have explored dynamics on networks. For example, variants of the influence maximization problem initially proposed in [38] are studied in [39], and several fixed parameter tractability results are established. The idea of “harmless individuals” (nodes) in a social network is investigated in [40]. Formally, a set S of nodes is harmless if every node v of the input graph has less than its threshold number of neighbors in S . Hence, a contagion where each initially infected node is in S cannot spread through the network. This is also related to the target set selection problem (see [40]). The firefighting problem, where a maximum number of blocking nodes is specified for each time step to stop the contagion spread, is studied in [41].

Several problems in graph theory, often referred to as re-configuration or re-optimization problems, are related to those we study here. These problem formulations rely on suitable definitions of distances between a pair of instances of a problem and a pair of solutions. Given two problem instances which are within a specified distance and a solution to the first instance, the goal is to determine a solution to the second instance that is within a specified distance of the first solution. In this context, two fixed parameter tractable problems—one for dominating sets and one for vertex covers—are provided in [42]. Other problems have also been solved under essentially this same setting (e.g., [43,44]). Dominating sets, alliances, and a generalizing framework are discussed in [45]. Łacki et al. [46] provide algorithms for maintaining approximations of a minimum Steiner tree of a predefined graph when the vertices of the graph may be deleted or (re)inserted in time. A survey of problems where the goal is to transform one solution into another by a series of intermediate steps, each of which is also a solution, is given in [47].

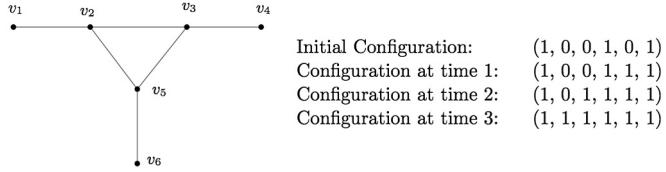
Beyond the graph theoretic problems discussed above (e.g., [42–44]), fixed parameter tractability and associated results have also been established in other contexts such as grammars and parsing [48,49], logic-based model checking [50], and construction of deterministic finite automata (DFA) [51]. In particular, the problem of finding a smallest deterministic finite automaton consistent with a set of positive and negative examples of the intended language, addressed in [51], has some similarity with our work.

2. Definitions, problem formulations, and summary of results

2.1. Formal definition of the SyDS model

Let \mathbb{B} denote the Boolean domain $\{0, 1\}$. A **Synchronous Dynamical System** (SyDS) S over \mathbb{B} is specified as a pair $S = (G, \mathcal{F})$, where

- (a) $G(V, E)$, an undirected graph with $|V| = n$, represents the underlying graph of the SyDS, with node set V and edge set E , and
- (b) $\mathcal{F} = \{f_1, f_2, \dots, f_n\}$ is a collection of functions in the system, with f_i denoting the **local transition function** associated with node v_i , $1 \leq i \leq n$.



Note: Each configuration has the form $(s_1^t, s_2^t, s_3^t, s_4^t, s_5^t, s_6^t)$, where s_i^t is the state of node v_i at time t , $1 \leq i \leq 6$. The configuration at time 3 is a stable configuration.

Fig. 1. An Example of a SyDS.

Each node of G has a state value from \mathbb{B} . Each function f_i specifies the local interaction between node v_i and its neighbors in G . The inputs to function f_i are the state of v_i and those of the neighbors of v_i in G ; function f_i maps each combination of inputs to a value in \mathbb{B} . This value becomes the next state of node v_i . It is assumed that each local function can be computed efficiently. In a SyDS, all nodes compute and update their next state *synchronously*. Other update disciplines (e.g., sequential updates) for discrete dynamical systems have also been considered in the literature (e.g., [13,27]). At any time t , the **configuration** \mathcal{C} of a SyDS is the n -vector $(s_1^t, s_2^t, \dots, s_n^t)$, where $s_i^t \in \mathbb{B}$ is the state of node v_i at time t ($1 \leq i \leq n$).

The local function f_v associated with node v of a SyDS \mathcal{S} is a t_v -**threshold** function for some integer $t_v \geq 0$ if the following condition holds: the value of f_v is 1 if the number of 1's in the input to f_v is *at least* t_v ; otherwise, the value of the function is 0. Thus, the state of a node may change from 0 to 1 or from 1 to 0. Throughout this paper, we assume that the local transition function for each node is a threshold function.

We let d_v denote the degree of node v , and t_v denote the threshold of node v . The number of inputs to the function f_v is $d_v + 1$. We assume without loss of generality that $0 \leq t_v \leq d_v + 2$. (The threshold values 0 and $d_v + 2$ allow us to realize functions that always output 1 and 0 respectively.)

Example. Consider the graph shown in Fig. 1. Suppose the local transition functions at each of the nodes v_1, v_4, v_5, v_6 is the **1-threshold function**, the function at v_3 is the **2-threshold function** and that at v_2 is the **3-threshold function**. Assume that initially, v_1, v_4 and v_6 are in state 1 and all other nodes are in state 0. During the first time step, the state of node v_5 changes to 1 since its neighbor v_6 is in state 1; the states of other nodes do not change. The configurations at subsequent time steps are shown in the figure. The configuration $(1, 1, 1, 1, 1, 1)$ at time step 3 is a stable configuration for this system. \square

2.2. Additional terminology and notation

A **trajectory** of a SyDS \mathcal{S} is a sequence of configurations $\langle C_1, C_2, \dots, C_r \rangle$, with $r \geq 2$, such that C_{i+1} is the successor of C_i , for $1 \leq i \leq r - 1$. When the last two configurations in a trajectory are identical, the trajectory ends in a stable configuration. A stable configuration is a special trajectory consisting of two identical configurations. Given a configuration \mathcal{C} and a node v , we let $\mathcal{C}(v)$ denote the value of v in \mathcal{C} , and \mathcal{C}^v denote the number of 1's in the input to f_v in \mathcal{C} .

A problem is **fixed parameter tractable** (FPT) with respect to a parameter k if there is an algorithm for the problem with a running time of $O(h(k)N^{O(1)})$, where N is the size of the problem instance and the function $h(k)$ depends only on k (see e.g., [52]). In particular, the function h does not depend on N . Many combinatorial problems are known to be fixed parameter tractable [52]. In Sections 4 and 5 we show that two of the inference problems considered in this paper are fixed parameter tractable.

A problem is solvable in **quasi-polynomial time** if there is an algorithm for the problem with a running time of $2^{O((\log N)^c)}$ where N is the size of the problem instance and c is a constant. In Section 5, we present quasi-polynomial time algorithms for two constrained versions of threshold inference problems.

Given a bipartite graph $G(V_1, V_2, E)$, a **matching** in G is a subset M of edges such that no two edges of M share an end point. A matching of largest cardinality is called a **maximum matching**. It is well known that for any bipartite graph with n nodes and m edges, a maximum matching can be found in $O(m\sqrt{n})$ time [53]. We will use this result in Section 4.2.

2.3. Problem formulations and summary of results

2.3.1. Overview

In all of the threshold inference problems considered in this paper, it is assumed that the underlying graph of the SyDS is given and that each local function is a threshold function with an unknown threshold value. We first define unconstrained versions of threshold inference problems considered in this paper. Subsequently, we define constrained versions of threshold inference problems.

Table 1

Table showing results for homogeneous and heterogeneous behavior specifications.

Problem	Results	
ITSC	Efficiently solvable	Theorem 3
ITUC	Efficiently solvable	Theorem 5
ITT	Efficiently solvable	Theorem 7
ITGE	Efficiently solvable	Theorem 9
Max-ITSC	$W[1]$ -complete and no $O(n^{1-\epsilon})$ approximation unless $P = NP$	Theorem 4
Max-ITUC	Efficiently solvable	Corollary 6
Max-ITT	$W[1]$ -complete and no $O(n^{1-\epsilon})$ approximation unless $P = NP$	Corollary 8
Max-ITGE	Efficiently solvable	Corollary 10
ITSUC	NP -complete but fixed parameter tractable with respect to the number of unstable configurations	Theorems 11 and 14

2.3.2. Unconstrained inference problems

We start with the definitions of threshold inference problems with *homogeneous* behavior specifications. In this case, the input representing behavior is a set Q along with a type tag; all elements of Q are of the type specified by the tag. For example, a set Q with type tag “stable configuration” indicates that each element of Q is a stable configuration. Similar interpretations can be given for the type tags “unstable configuration”, “trajectory” and “GE configuration”. We say that a SyDS S **exhibits the behavior specified by** Q if S satisfies the property specified by the type tag of Q for each element of Q . For example, if the type tag of Q is “stable configuration”, then S exhibits the behavior specified by Q if each element of Q is a stable configuration of S . The general problem formulation for the homogeneous case can now be stated as follows.

Inferring Thresholds from Homogeneous Behavior Specifications

Given: A partially specified SyDS S over $\{0, 1\}$ consisting of the underlying graph G and a set Q with a type tag.

Question: Is there an assignment of threshold values to the nodes of S such that the resulting fully specified SyDS S exhibits the behavior corresponding to the type tag of Q ?

A number of specific problems can be derived from the above general definition. For example, when the type tag of Q is “stable configuration”, we refer to the resulting problem as **Inferring Thresholds from Stable Configurations** (ITSC). Likewise, when the type tag of Q is “trajectory” or “unstable configuration” or “GE configuration”, we refer to the corresponding problems as Inferring Thresholds from Trajectories (ITT), Unstable Configurations (ITUC) and GE Configurations (ITGE) respectively.

When the answer to an instance of a decision problem such as ITSC is “no”, it is natural to consider a maximization version where the goal is to find a threshold assignment that makes a largest subset of Q to be stable configurations. We use the prefix “Max” in naming these problems. Thus, we denote the maximization versions of ITSC, ITT, ITUC and ITGE by Max-ITSC, Max-ITT, Max-ITUC and Max-ITGE respectively. At the end of Section 3, we briefly comment on the complementary versions of these problems where the goal is to delete the minimum number of observations so that there is a solution for the remaining set of observations.

When the behavior specification consists of two or more sets, each with a different type tag, we obtain inference problems for *heterogeneous* behavior specifications. Many such inference problems can be formulated by considering combinations of type tags. Here, we focus on the problem where the observed behavior is specified by two sets Q_1 and Q_2 with type tags “stable configuration” and “unstable configuration” respectively. We say that a SyDS S exhibits the behavior specified by Q_1 and Q_2 if S exhibits the behavior specified by Q_1 as well as Q_2 . We refer to the corresponding problem as **Inferring Thresholds from Stable and Unstable Configurations** (ITSUC). Table 1 shows our results for the above problems.

2.3.3. Inference problems with constraints

As explained in Section 1.1, several application scenarios impose constraints on the threshold values of nodes. In such cases, the solution for a threshold inference problem must also satisfy the specified constraints. This section defines the constrained versions of the threshold inference problems studied in this paper.

We consider two forms of constraints, namely “equal” (denoted by EQ) and “not equal” (denoted by NE) constraints. The constraint EQ(v_i, v_j) requires that nodes v_i and v_j must have the same threshold value in the solution to the corresponding inference problem. Similarly, the constraint NE(v_i, v_j) requires that nodes v_i and v_j must have different threshold values.

We consider these constraints in conjunction with two of the inference problems, namely ITSC (Inferring Thresholds from Stable Configurations) and ITUC (Inferring Thresholds from Unstable Configurations), defined in Section 2.3.2. The ITSC problem with equality constraints (denoted by ITSC-EQ) is similar to the ITSC problem except that the input includes a set Γ_{EQ} of equality constraints, and the solution to the ITSC problem must also satisfy those constraints. The ITSC problem with

Table 2

Table showing results for constrained versions of inference problems.

Problem	Results	
ITSC-EQ	Efficiently solvable	Theorem 15
ITSC-NE	NP -complete but fixed parameter tractable with respect to the number of constraints	Proposition 16 and Theorem 17
ITUC-EQ	Solvable in quasi-polynomial time	Theorem 18
ITUC-NE	NP -complete but solvable in quasi-polynomial time when the number of inequality constraints is fixed	Proposition 19 and Theorem 21

inequality constraints is denoted by ITSC-NE. The corresponding problems derived from ITUC are denoted by ITUC-EQ and ITUC-NE respectively. Table 2 shows our results for four constrained threshold inference problems.

2.4. Preliminary results

Here, we present some preliminary results which will be used throughout this paper.

Lemma 1. Every SyDS \mathcal{S} where each local transition function is a threshold function has at least one stable configuration. Furthermore, a stable configuration can be found in polynomial time.

Proof. Given the SyDS, we say that a given node v is **forced** if either $t_v = 0$ or $t_v = d_v + 2$. Note that if $t_v = 0$, then in every stable configuration, the value (i.e., state) of node v is 1; and if $t_v = d_v + 2$, then in every stable configuration, the value of node v is 0.

We can construct a stable configuration D of the given SyDS \mathcal{S} as follows. As long as there is at least one forced node, we repeatedly delete a forced node, modifying the SyDS each time, as described below. Suppose we select node v as the forced node to be deleted.

1. If $t_v = 0$, we set the value of v in D to 1. We then delete node v from the SyDS. Every neighbor w of v such that $t_w > 0$ has its threshold reduced by one.
2. If $t_v = d_v + 2$, we set the value of v in D to 0. We then delete node v from the SyDS. Every neighbor w of v such that $t_w = d_w + 2$ (where d_w is the degree of node w before the deletion of v) has its threshold reduced by one.

Note that the deletion of a node can cause a neighboring node that was previously unforced to become forced.

If there are no nodes left, a stable configuration has been constructed. If there are any nodes left, they are all unforced. We can extend the partial configuration constructed so far by making all the remaining nodes 0 (or all the remaining nodes 1). The result is a stable configuration. \square

Lemma 2. Given the underlying graph G of a SyDS \mathcal{S} over $\{0, 1\}$ and a configuration D of \mathcal{S} , there is a linear time algorithm for constructing an assignment of threshold values to the nodes of \mathcal{S} such that D is the successor of every configuration of \mathcal{S} .

Proof. The thresholds are set as follows.

1. If node v has value 1 in D , we set t_v to be 0, so that the successor for every configuration has the value 1 for v .
2. If node v has value 0 in D , we set t_v to be $d_v + 2$, so that the successor for every configuration has the value 0 for v . \square

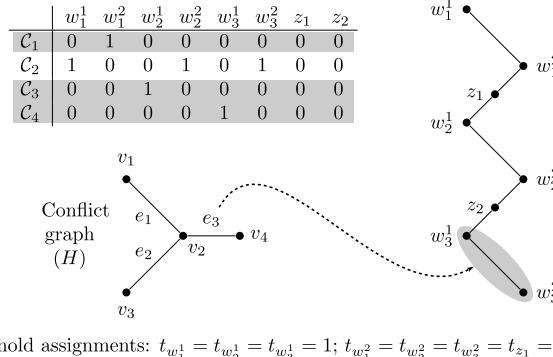
3. Threshold inference from homogeneous behavior specifications

3.1. Inferring thresholds from stable configurations

In this section, we present an efficient algorithm for the problem of inferring thresholds from stable configurations (ITSC). We also present complexity and non-approximability results for Max-ITSC.

Theorem 3. The ITSC problem can be solved efficiently. When there is a solution, an assignment of threshold values to the nodes can also be obtained efficiently.

Proof. Let $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_r\}$ be the given set with type tag “stable configuration”. We can consider the threshold assignment separately for each node. For any node v , let Q_v^0 (Q_v^1) be the subset of \mathcal{C} such that for each configuration in Q_v^0 (Q_v^1) the value of v is 0 (1). We use f_v to denote the threshold function at node v . Recall that \mathcal{C}^v denotes the number of 1’s in the input to f_v in \mathcal{C} .



Threshold assignments: $t_{w_1^1} = t_{w_2^1} = t_{w_3^1} = 1$; $t_{w_1^2} = t_{w_2^2} = t_{w_3^2} = t_{z_1} = t_{z_2} > 1$.

Fig. 2. An example of Max-ITSC instance and corresponding conflict graph. The shaded configurations are satisfied by the threshold assignments.

- If Q_v^0 is nonempty, we consider each configuration $C_i \in Q_v^0$. Since C_i is a stable configuration, the threshold t_v must satisfy the condition $t_v > C_i^v$, i.e., $t_v \geq C_i^v + 1$. Let $t_v^{low} = 1 + \max_{C_i \in Q_v^0} C_i^v$. If Q_v^0 is empty, we let $t_v^{low} = 0$.
- If Q_v^1 is nonempty, we consider each configuration $C_j \in Q_v^1$. Since C_j is a stable configuration, the threshold t_v must satisfy the condition $t_v \leq C_j^v$. Let $t_v^{high} = \min_{C_j \in Q_v^1} C_j^v$. If Q_v^1 is empty, we let $t_v^{high} = d_v + 2$.
- The two steps above provide a collection of constraints that can be satisfied provided $t_v^{low} \leq t_v^{high}$. When there is a solution, any value for t_v can be chosen such that $t_v^{low} \leq t_v \leq t_v^{high}$.

Clearly, the above computations can be done in polynomial time. \square

Also note that for a given node v , if Q_v^0 is nonempty, then $t_v^{low} \geq 1$; and if Q_v^1 is nonempty, then $t_v^{high} \leq d_v + 1$. Thus, if there is a solution, then there is a solution where none of the local functions are constant functions; that is, there is a solution where for each node v , $1 \leq t_v \leq d_v + 1$.

Recall that in the Max-ITSC problem, the goal is to choose threshold values so that a maximum number of elements in the set Q are stable points of \mathcal{S} . We now establish a hardness result for this problem which holds even when the underlying graph of a SyDS is a simple path.

The idea behind the proof is the following. Let us say that two configurations C_i and C_j **conflict** if there is a node v such that $C_i(v) = 0$, $C_j(v) = 1$, and $C_i^v \geq C_j^v$. Note that if C_i and C_j conflict, then there is no assignment of thresholds under which both C_i and C_j are stable. On the other hand, if a set of configurations is conflict-free (i.e., no pair of configurations in the set conflict), then thresholds can be assigned so that all the configurations in the set are stable. Given an ITSC problem instance, define the **conflict graph** for the instance to be the undirected graph with a node for each configuration in Q and an edge between each pair of nodes whose corresponding configurations conflict. Each independent set [54] of size r in the conflict graph gives a subset Q' of Q with r configurations that can be made stable. Since the Maximum Independent Set (MIS) problem is in $W[1]$ [55], this reduction from the Max-ITCS problem to the MIS problem shows that Max-ITCS is in $W[1]$, with the parameter being the number of configurations that must be made stable.

Theorem 4. *Max-ITSC is $W[1]$ -complete, where the parameter is the number of configurations that must be made stable. Further, for any $\epsilon > 0$, there is no polynomial time $O(n^{1-\epsilon})$ approximation algorithm for Max-ITSC, unless $P = NP$. (Here n is the number of configurations in the given set Q .) These results hold even when the underlying graph of the SyDS is a simple path.*

Proof. The membership of Max-ITSC in $W[1]$ was discussed just before the statement of the theorem. To establish $W[1]$ -hardness, we use a reduction from the MIS problem: given an undirected graph $H(V_H, E_H)$ and an integer $K \leq |V_H|$, does H have an independent set of size at least K ? MIS is known to be $W[1]$ -complete [55].

The key to the $W[1]$ -hardness reduction is to construct an ITSC problem instance from the MIS in such a way that the conflict graph (defined above) for the constructed Max-ITSC problem instance is identical to the graph in the given MIS problem instance. We do this as follows. (Fig. 2 illustrates the construction.) There is a configuration for each node of H , with configuration C_i corresponding to node $v_i \in V_H$. We let $Q = \{C_1, C_2, \dots, C_n\}$, where $n = |V_H|$. Let $E_H = \{e_1, e_2, \dots, e_m\}$. For each edge $e_k \in E_H$, the underlying graph G for SyDS \mathcal{S} has two nodes denoted by w_k^1 and w_k^2 along with the edge $\{w_k^1, w_k^2\}$. Suppose the edge $e_k \in E_H$ joins nodes v_i and v_j of H . We set $C_i(w_k^1) = 0$, $C_i(w_k^2) = 1$, $C_j(w_k^1) = 1$, and $C_j(w_k^2) = 0$. All other configurations in Q have value 0 for both w_k^1 and w_k^2 . Nodes w_k^1 and w_k^2 induce a conflict between C_i and C_j , but do not induce a conflict between any other pair of configurations. Thus, any independent set of V_H corresponds to a conflict-free subset of configurations of \mathcal{S} and vice versa.

To ensure that the underlying graph of SyDS \mathcal{S} is a simple path, we add $m - 1$ new nodes denoted by z_1, z_2, \dots, z_{m-1} . For each j , $1 \leq j \leq m - 1$, node z_j is adjacent to w_j^2 and w_{j+1}^1 . In each of the n configurations constructed above, the values of the nodes z_1, z_2, \dots, z_{m-1} are all 0.

It can be verified that H has an independent set of size K if and only if there is a conflict-free subset of Q with K configurations. This proves the $W[1]$ -hardness of Max-ITSC.

It is well known that for any $\epsilon > 0$, there is no polynomial time $O(n^{1-\epsilon})$ -approximation algorithm for the MIS problem, unless $P = NP$ [56]. Since our construction preserves approximations (i.e., for each r , any independent set of size r in H leads to a subset of r conflict-free configurations of \mathcal{S} and vice versa), the same negative result holds for Max-ITS as well. \square

3.2. Inferring thresholds from unstable configurations

Here, we consider the ITUC problem where the goal is to infer thresholds from a given set Q of unstable configurations. In this case, we show that both ITUC and Max-ITUC can be solved efficiently.

Theorem 5. *The ITUC problem can be solved efficiently. When there is a solution, an assignment of threshold values to the nodes can also be obtained efficiently.*

Proof. If Q contains every possible configuration of \mathcal{S} (i.e., $|Q| = 2^n$, where n is the number of nodes in the underlying graph of \mathcal{S}), then, by Lemma 1, there is no solution to the problem.

Suppose that Q excludes at least one configuration of \mathcal{S} . Then there is always a solution to the problem, as outlined below. Let \mathcal{C} be a configuration that is not in Q . From Lemma 2, we can set the thresholds so that the successor of every configuration is the configuration \mathcal{C} ; that is, every configuration other than \mathcal{C} is an unstable configuration. \square

The following result for Max-ITUC is a simple consequence of the above theorem.

Corollary 6. *Max-ITUC is efficiently solvable.*

3.3. Inferring thresholds from trajectories

Here, we present results for the ITT problem, where the goal is to infer thresholds given a set Q of trajectories.

Theorem 7. *The ITT problem can be solved efficiently. When there is a solution, an assignment of threshold values to the nodes can also be obtained efficiently.*

Proof. The main idea is similar to the one used in establishing the efficient solvability of ITSC (Theorem 3). As before, we can consider the threshold assignment separately for each node v of the underlying graph. Let $T_i = \langle \mathcal{C}_{i,1}, \mathcal{C}_{i,2}, \dots, \mathcal{C}_{i,r} \rangle$ denote one of the trajectories. Consider any transition $\mathcal{C}_{i,j}$ to $\mathcal{C}_{i,j+1}$ in T_i . We now observe that this transition gives rise to an inequality for the threshold t_v of v , depending on the values of node v in $\mathcal{C}_{i,j}$ and $\mathcal{C}_{i,j+1}$. There are four cases to consider.

Case 1: $\mathcal{C}_{i,j}(v) = 0$ and $\mathcal{C}_{i,j+1}(v) = 0$. Here, $\mathcal{C}_{i,j}^v$ (i.e., the number of 1's in $\mathcal{C}_{i,j}$ which are inputs to the function f_v) is too small to change the value of v from 0 to 1. In other words, we have the inequality $t_v > \mathcal{C}_{i,j}^v$ in this case.

Case 2: $\mathcal{C}_{i,j}(v) = 0$ and $\mathcal{C}_{i,j+1}(v) = 1$. Here, $\mathcal{C}_{i,j}^v$ is large enough to change the value of v from 0 to 1. Thus, we get the inequality $t_v \leq \mathcal{C}_{i,j}^v$.

Case 3: $\mathcal{C}_{i,j}(v) = 1$ and $\mathcal{C}_{i,j+1}(v) = 0$. Here, the corresponding inequality is $t_v > \mathcal{C}_{i,j}^v$.

Case 4: $\mathcal{C}_{i,j}(v) = 1$ and $\mathcal{C}_{i,j+1}(v) = 1$. Here, the corresponding inequality is $t_v \leq \mathcal{C}_{i,j}^v$.

Thus, each transition in each trajectory gives rise to an inequality for t_v . The number of inequalities for any node v is at most the total number of transitions over all the trajectories. There is a solution to the given ITT instance if and only if for each node v , all the inequalities can be satisfied. Clearly, these computations can be done in polynomial time. \square

Since each stable configuration is a special case of a trajectory (containing two identical configurations), the following result is a direct consequence of Theorem 4.

Corollary 8. *Max-ITT is $W[1]$ -hard. Further, for any $\epsilon > 0$, there is no polynomial time $O(n^{1-\epsilon})$ approximation algorithm for Max-ITT, unless $P = NP$. (Here n is the number of trajectories in the given set Q .) These results hold even when underlying graph of the SyDS is a simple path.*

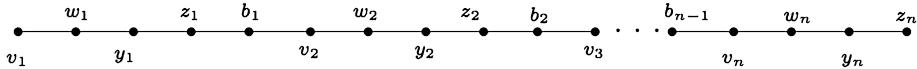


Fig. 3. Reduction used in the Proof of [Theorem 11](#).

3.4. Inferring thresholds from Garden of Eden configurations

We now present results for the ITGE problem, where the goal is to infer thresholds given a set Q of Garden of Eden configurations.

Theorem 9. *The ITGE problem can be solved efficiently. When there is a solution, an assignment of threshold values to the nodes can also be obtained efficiently.*

Proof. Suppose that Q contains every possible configuration of \mathcal{S} (i.e., $|Q| = 2^n$, where n is the number of nodes in the underlying graph of \mathcal{S}). Since every configuration has a successor, there is at least one configuration with a predecessor, so there is no solution to the problem instance.

Suppose that Q excludes at least one configuration of \mathcal{S} . Then there is always a solution to the problem instance, as outlined below. Let C be a configuration that is not in Q . From [Lemma 2](#), we can set the thresholds so that the successor of every configuration is configuration C , so that every configuration other than C has no predecessor. In other words, every element of Q is a GE configuration. \square

The following result for Max-ITGE is a simple consequence of the above theorem.

Corollary 10. *Max-ITGE is efficiently solvable.* \square

3.5. Some observations regarding minimization versions

When there is no solution to a given threshold inference problem, we considered the problem of finding a maximum subset of configurations for which there is a solution. One can also consider the complementary problem, where the goal is to delete the *minimum* number of configurations so that there is a solution to the remaining set of configurations. Since Max-ITUC and Max-ITGE are efficiently solvable ([Corollaries 6 and 10](#)), the corresponding minimum deletion versions are also efficiently solvable. By starting with the minimum vertex cover (MVC) problem [54] and carrying out the same reduction used in the hardness proofs for Max-ITSC and Max-ITT ([Theorem 4](#) and [Corollary 8](#)), the corresponding minimum deletion versions can be seen to be **NP**-complete. Also, the minimization versions of these two problems can be reduced to the MVC problem for the corresponding conflict graphs along the lines of the discussion preceding the statement of [Theorem 4](#). Since MVC is fixed parameter tractable [52], this direct connection to MVC points out that the minimization versions of ITSC and ITT are also fixed parameter tractable. Further, all known approximation results for MVC [57] also apply to the minimization versions of ITSC and ITT.

4. Inference from heterogeneous collections of behavior

4.1. The complexity of ITSUC

We now consider the ITSUC problem, where there are two sets Q_1 and Q_2 of configurations, and the requirement is to find a threshold value for each node of \mathcal{S} such that each configuration in Q_1 is stable and each configuration in Q_2 is unstable. The following theorem shows the complexity of this decision problem.

Theorem 11. *The ITSUC problem is **NP**-complete even when the underlying graph of the given SyDS is a simple path.*

Proof. It is readily seen that ITSUC is in **NP**. We show **NP**-hardness via a reduction from 3SAT.

Suppose the given 3SAT formula has n variables and m clauses. The reduction constructs a SyDS \mathcal{S} whose underlying graph G is a *simple path* containing $5n - 1$ nodes as follows. For each variable x_i of the formula, graph G contains four nodes, which we denote as v_i , w_i , y_i , and z_i , and the three edges $\{v_i, w_i\}$, $\{w_i, y_i\}$, and $\{y_i, z_i\}$. We also have $n - 1$ additional nodes b_1, b_2, \dots, b_{n-1} (referred to as “buffer nodes”). For each i , $1 \leq i \leq n - 1$, we add the two edges $\{z_i, b_i\}$ and $\{b_i, v_{i+1}\}$. Thus, the buffer nodes connect together the subgraphs created from the variables so that the underlying graph of \mathcal{S} is a simple path on $5n - 1$ nodes. (See [Fig. 3](#).)

The set $Q_1 = \{C_1, C_2, C_3, C_4, C_5\}$ consists five configurations which are required to be stable. These five configurations are constructed as follows.

1. In \mathcal{C}_1 , every node v_i and w_i has value 0, every node y_i and z_i has value 1 and all buffer nodes have value 0.
2. In \mathcal{C}_2 every node v_i and w_i has value 1, every node y_i and z_i has value 0 and all buffer nodes have value 0.
3. In \mathcal{C}_3 , every node v_i , w_i , and y_i has value 0, every node z_i has value 1 and all buffer nodes have value 0.
4. In \mathcal{C}_4 , every node v_i , w_i , y_i and z_i has value 1, and all buffer nodes have value 0.
5. In \mathcal{C}_5 , all the nodes of \mathcal{S} have value 1.

The set $Q_2 = \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_m\}$ contains m configurations, with configuration \mathcal{D}_j corresponding to clause c_j of the 3SAT instance, $1 \leq j \leq m$. Each of these m configurations is required to be unstable. Consider a given clause c_j . If a given variable x_i occurs as a positive literal in c_j , then in \mathcal{D}_j , node v_i has value 1, w_i has value 0, y_i has value 0, and z_i has value 0. If variable x_i occurs as a negative literal (i.e., as \bar{x}_i) in c_j , then in \mathcal{D}_j , node v_i has value 0, w_i has value 1, y_i has value 1, and z_i has value 1. If no literal for variable x_i occurs in c_j , then in \mathcal{D}_j , node v_i has value 0, w_i has value 0, y_i has value 1, and z_i has value 1. In all the m configurations, each buffer node has value 0.

Now consider the thresholds of the nodes in G . For each i , $1 \leq i \leq n$, configuration \mathcal{C}_1 imposes the constraints that $t_{v_i} > 0$, $t_{w_i} > 1$, $t_{y_i} \leq 2$, and $t_{z_i} \leq 2$. For each i , $1 \leq i \leq n$, configuration \mathcal{C}_2 imposes the constraints that $t_{v_i} \leq 2$, $t_{w_i} \leq 2$, $t_{y_i} > 1$, and $t_{z_i} > 0$. For each i , $1 \leq i \leq n$, configuration \mathcal{C}_3 imposes the constraints that $t_{v_i} > 0$, $t_{w_i} > 0$, $t_{y_i} > 1$, and $t_{z_i} \leq 1$. Together, these constraints require that t_{v_i} is 1 or 2, $t_{w_i} = 2$, $t_{y_i} = 2$, and $t_{z_i} = 1$. We view setting $t_{v_i} = 2$ as corresponding to setting Boolean variable x_i to 1, and setting $t_{v_i} = 1$ as corresponding to setting x_i to 0. For each i , $1 \leq i \leq n - 1$, configuration \mathcal{C}_4 imposes the constraint that $t_{b_i} > 2$. For each i , $1 \leq i \leq n - 1$, configuration \mathcal{C}_5 imposes the constraint that $t_{b_i} \leq 3$. Thus, \mathcal{C}_4 and \mathcal{C}_5 together ensure that $t_{b_i} = 3$ for $1 \leq i \leq n - 1$.

Suppose the given 3SAT formula is satisfiable. Consider a given satisfying assignment α . We construct an assignment of threshold values to the nodes of G , as follows. For each i , $1 \leq i \leq n$, we set $t_{w_i} = 2$, $t_{y_i} = 2$, and $t_{z_i} = 1$. If a given variable x_i has value 1 in α , we set $t_{v_i} = 2$; and if x_i has value 0 in α , we set $t_{v_i} = 1$. Each buffer node is assigned a threshold value of 3. We now argue that these thresholds ensure that none of the configurations in Q_2 is stable. Suppose x_i has value 1 in α , and a given clause c_j contains the positive literal x_i . Then, node v_i has value 1 in \mathcal{D}_j , its only neighbor w_i has value 0 in \mathcal{D}_j , and $t_{v_i} = 2$, so configuration \mathcal{D}_j is unstable. On the other hand, suppose x_i has value 0 in α , and a given clause c_j contains the negative literal \bar{x}_i . Then, node v_i has value 0 in \mathcal{D}_j , its only neighbor w_i has value 1 in \mathcal{D}_j , and $t_{v_i} = 1$, so configuration \mathcal{D}_j is unstable. Thus, it can be seen that the answer to the ITSUC problem instance is “Yes”.

Now suppose that the answer to the ITSUC problem instance is “Yes”. Consider a given assignment of threshold values that makes the answer “Yes”. We can construct an assignment α of Boolean values to the variables of the given 3SAT formula, as follows. For $1 \leq i \leq n$, if $t_{v_i} = 2$, we set x_i to the value 1; and if $t_{v_i} = 1$, we set x_i to the value 0. It can be seen that the constructed assignment α is a satisfying assignment.

Thus, the given 3SAT formula is satisfiable iff the answer to the constructed ITSUC problem instance is “Yes”. \square

4.2. Fixed parameter tractability of ITSUC

We now show that ITSUC is fixed parameter tractable with respect to the number of unstable configurations specified in the problem instance, with no restrictions on the underlying graph of the given SyDS. Given a set $A = \{a_1, a_2, \dots, a_r\}$, let $P(A)$ denote a **partition** of A into nonempty subsets. Let each subset in $P(A)$ be called a **block**. We use $\pi(A)$ to denote the collection of all partitions of A . For a set A with r elements, it is known that $|\pi(A)| = O((r/\log r)^r)$ [58].

Let Q_1 and Q_2 denote the set of stable and unstable configurations respectively in the given ITSUC instance. Let $q = |Q_2|$ and let n be the number of nodes in the given SyDS \mathcal{S} . From the proof of [Theorem 3](#), it can be seen that the configurations in Q_1 impose constraints on the threshold value of each node of \mathcal{S} . Given any configuration $\mathcal{C} = (s_1, s_2, \dots, s_n)$ in Q_2 , we can try to make \mathcal{C} an unstable configuration by choosing a threshold t_{v_i} for node v_i so that in the successor \mathcal{C}' of \mathcal{C} , the state of v_i is different from s_i , $1 \leq i \leq n$. Such a choice must also satisfy the constraints imposed on t_{v_i} by the configurations of Q_1 . Given an instance of ITSUC, we say that a node v of \mathcal{S} is **compatible** with a configuration $\mathcal{C} \in Q_2$ if a value for t_v can be chosen so that (i) t_v satisfies all the constraints imposed by the collection Q_1 and (ii) this choice makes \mathcal{C} an unstable configuration (regardless of the threshold values assigned to the other nodes). Extending this definition, we say that a node v is **compatible with a subset** R of Q_2 if v is compatible with every configuration in R .

The above definitions are used in our algorithm (Alg-ITSUC) shown in [Fig. 4](#). The algorithm considers each possible partition of the set Q_2 of unstable configurations. For each partition P , it constructs a bipartite graph H_P with the set V of nodes of the SyDS \mathcal{S} on one side and a set V_P of nodes corresponding to the blocks of P on the other side. Each edge $\{x, y\}$ of H_P , where $x \in V$ and $y \in V_P$ indicates that node x of \mathcal{S} is compatible with the block of P corresponding to node y . (The construction of this bipartite graph is depicted in [Fig. 5](#).) If the algorithm encounters a partition P for which the corresponding bipartite graph H_P contains a matching whose size is equal to the number of blocks in P , the algorithm outputs “Yes”; otherwise, the algorithm outputs “No”. We now establish the correctness of the algorithm and its running time.

Lemma 12. Algorithm Alg-ITSUC given in [Fig. 4](#) correctly decides whether the ITSUC instance has a solution.

Proof. We will show that the ITSUC instance has a solution iff Alg-ITSUC outputs “Yes”.

Input: Graph $G(V, E)$ of a SyDS \mathcal{S} , and configuration sets Q_1 and Q_2 .
Requirement: Output “Yes” if there is a threshold value t_v for each $v \in V$ such that in the resulting SyDS, all the configurations in Q_1 are stable and all the configurations in Q_2 are unstable. Otherwise, output “No”.

Steps:

1. **for** each partition P in $\pi(Q_2)$ **do**
 - (a) Let k denote the number of blocks in P and let B_1, B_2, \dots, B_k denote the blocks themselves.
 - (b) Construct the bipartite graph $H_P(V, V_P, E_P)$ where V_P has one node for each block in P and $E_P = \{(x, y) : x \in V, y \in V_P \text{ and node } x \text{ of } \mathcal{S} \text{ is compatible with the block of } P \text{ represented by node } y\}$.
 - (c) **if** H_P has a matching with k edges **then output** “Yes” and **stop**.
2. **Output** “No”.

Fig. 4. Algorithm Alg-ITSUC to show the fixed parameter tractability of ITSUC.

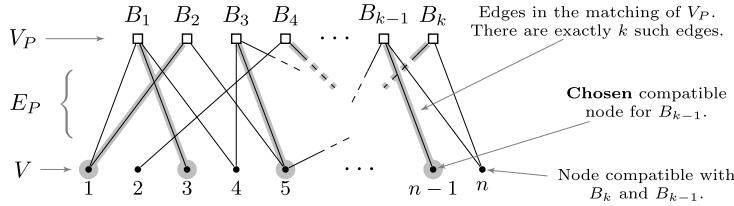


Fig. 5. The bipartite graph $H_P(V, V_P, E_P)$.

Part 1 (if): Suppose the algorithm outputs the answer “Yes”. We will prove that there is a solution to the ITSUC instance.

From the description of the algorithm, we note that in this case, there is a partition P of Q_2 with k (nonempty) blocks such that there is a matching with k edges in the corresponding bipartite graph H_P . In such a matching, each node y that corresponds to block B_y of P is matched to some node x of the SyDS \mathcal{S} . By our construction, node x is compatible with block B_y . By the definition of compatibility, there is a threshold value t_x for node x such that t_x satisfies all the constraints imposed by the configurations in Q_1 , and further, this value of t_x for x makes all the configurations in B_y unstable. Since every block of P is matched to a distinct node of \mathcal{S} , it follows that threshold values can be chosen for each node of \mathcal{S} independently to satisfy the required conditions. In other words, there is a solution to the ITSUC instance.

Part 2 (only if): Suppose there is a solution to the ITSUC instance. We will prove that the algorithm outputs the answer “Yes”. The essence of the argument is the following: each solution to the ITSUC instance induces a partition P of the set Q_2 of unstable configurations so that the bipartite graph H_P constructed in Step 1(b) of Alg-ITSUC has a matching whose size is the number of blocks in P .

Consider a solution Γ to the ITSUC instance which assigns the threshold value t_{v_i} for each node v_i of \mathcal{S} . Let $Q_2 = \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_m\}$. Since Γ is a solution to the ITSUC instance, for each $\mathcal{D}_j \in Q_2$, there is a node v_r which is compatible with \mathcal{D}_j , $1 \leq j \leq m$; that is, the threshold assignment t_{v_r} to v_r ensures that \mathcal{D}_j is an unstable configuration. (When there are two or more nodes which are compatible with a given configuration $\mathcal{D}_j \in Q_2$, we choose a node arbitrarily.) Thus, the given solution Γ allows us to assign each configuration in Q_2 to some node of \mathcal{S} . By this method, we assign a (possibly empty) subset, say D_i , of Q_2 to each node v_i of \mathcal{S} so that v_i is compatible with each configuration in D_i . Since each configuration of Q_2 is assigned to only one node of \mathcal{S} , the collection of sets D_1, \dots, D_n is pairwise disjoint. Thus, the non-empty sets in this collection are the blocks of a partition P of Q_2 . Let k denote the number of such blocks in P .

Since Alg-ITSUC considers all the partitions of Q_2 , it will surely encounter the partition P . Further, as argued above, for every block B_y of P , there is a node x of \mathcal{S} such that x is compatible with B_y . Moreover, for the partition P , there is an assignment of blocks of P to the nodes of \mathcal{S} such that no two blocks are assigned to the same node of \mathcal{S} . In other words, the bipartite graph H_P constructed by the algorithm from P has a matching with k edges, where k is the number of blocks of P . Thus, the algorithm will output the answer “Yes”, and this completes the proof. \square

Lemma 13. Algorithm Alg-ITSUC can be implemented to run in time $O(h(q)N^{O(1)})$, where q is the number of unstable configurations, N is the size of the problem instance and the function h depends only on q .

Proof. Let Q_1 (the set of configurations to be made stable) contain r configurations. We first discuss some preprocessing steps. For any node v and any configuration C in Q_1 , the constraint on the threshold t_v of v imposed by C can be found in $O(n)$ time (since each configuration has n state values). Thus, the constraints on t_v imposed by all r configurations in Q_1 can be found in $O(nr)$ time. As indicated in the proof of Theorem 3, all of these constraints can be combined into a single constraint of the form $t_v^{\text{low}} \leq t_v \leq t_v^{\text{high}}$ for appropriate integers t_v^{low} and t_v^{high} in $O(nr)$ time. Thus, obtaining such a single

constraint for each of the n nodes can be done in $O(n^2r)$ time. These preprocessing steps can be done before starting the execution of the **for** loop in Step 1 of the algorithm.

We now estimate the time needed to check whether a node v is compatible with a configuration \mathcal{C} in Q_2 . For a node v , the constraint on t_v needed to make \mathcal{C} an unstable configuration (regardless of the threshold values of the other nodes) can be determined in $O(n)$ time. Checking whether this constraint also satisfies the constraint on t_v obtained by considering the configurations in Q_1 can be done in $O(1)$ time. Thus, checking whether a node v is compatible with a configuration in Q_2 can be done in $O(n)$ time. The results from these compatibility checks can be stored in an $n \times q$ Boolean matrix M so that during the execution of the **for** loop of the algorithm, we can determine whether a node v is compatible with a configuration \mathcal{C} of Q_2 in $O(1)$ time.

We now estimate the time used for each iteration of the **for** loop of Fig. 4. For each partition P in $\pi(Q_2)$, the algorithm constructs an appropriate bipartite graph H_P and checks whether the graph has a matching whose size is equal to the number of blocks of P . If P has k blocks, the number of nodes in H_P is $n+k$ and the number of edges is at most nk . For each node $x \in V$, the blocks with which x is compatible can be found in $O(q)$ time from the precomputed matrix M . Thus, constructing the graph H_P can be done in time $O(nkq) = O(nq^2)$ since $k \leq q$. Since H_P has $n+k \leq n+q$ nodes and at most $nk \leq nq$ edges, as mentioned in Section 2.2, a maximum matching in H_P can be found in $O(nq\sqrt{n+q})$ time. Thus, each iteration of the **for** loop can be implemented to run in time $O(nq^2 + nq\sqrt{n+q})$ time.

Since the number of iterations of the **for** loop is at most $|\pi(Q_2)|$, the running time of Step 1 is $O(|\pi(Q_2)|(nq^2 + nq\sqrt{n+q}))$. The overall running time of the algorithm, including the preprocessing steps, is $O(n^2(r+q) + |\pi(Q_2)|(nq^2 + nq\sqrt{n+q}))$. As mentioned earlier, $|\pi(Q_2)| = O((q/\log q)^q)$. Since $n+r+q \leq N$, where N is the size of the given ITSUC instance, it follows that the running time of our algorithm for ITSUC has the form $O(h(q)N^{O(1)})$, where $h(q) = (q/\log q)^q$ depends only on q . \square

The following theorem is a direct consequence of Lemmas 12 and 13.

Theorem 14. *The ITSUC problem is fixed parameter tractable where the parameter is the number of unstable configurations.* \square

5. Threshold inference under constraints

5.1. Overview

Here, we consider threshold inference problems with constraints. Section 5.2 considers constrained inference problems for stable configurations while Section 5.3 considers the corresponding problems for unstable configurations.

Throughout this section, Q denotes the set of configurations (which must be made stable or unstable depending on the problem). For problems involving equality constraints, Γ_{EQ} denotes the given set of constraints. For problems involving inequality constraints, the corresponding set is denoted by Γ_{NE} .

We recall an assumption from Section 2: when a node v of a given SyDS S has degree d_v , any threshold value t_v assigned to the node satisfies the condition $0 \leq t_v \leq d_v + 2$. This assumption is used throughout this section.

For any node v , choosing $t_v = d_v + 2$ ensures that the local transition function f_v at v always outputs the value 0. In other words, choosing $d_v + 2$ as the threshold value of v is akin to choosing the threshold value of “infinity” for v . For two nodes u and v whose degrees are different, when we choose $t_u = d_u + 2$ and $t_v = d_v + 2$, the threshold values t_u and t_v are numerically different; however, the values effectively represent the same threshold, namely “infinity”. So, in such cases, we regard u and v as having the same threshold value; that is, such a threshold assignment satisfies the condition $t_u = t_v$. Also, to satisfy the constraint $t_u \neq t_v$, at least one of u and v must be assigned a threshold different from “infinity”.

For simplicity, we present our algorithms for decision versions of the various problems. It is straightforward to modify the algorithms to find a threshold assignment when one exists.

5.2. Inferring thresholds with stable configurations and constraints

Here, we consider the constrained versions of the ITSC problem, namely ITSC-EQ and ITSC-NE. We first show that ITSC-EQ can be solved in polynomial time. Then we show that ITSC-NE is **NP**-complete and also that it is fixed parameter tractable with respect to the number of inequality constraints.

5.2.1. A polynomial time algorithm for ITSC-EQ

Our polynomial time algorithm for the ITSC-EQ problem relies on the ideas developed in the algorithm for the ITSC problem (Theorem 3). In particular, we recall that for each node $v \in V$, the set Q gives rise to lower and upper bounds t_v^{low} and t_v^{high} respectively on the threshold value of v . These bounds have the property that all configurations in Q can be made stable if and only if for each node $v \in V$, a threshold value t_v can be chosen so that $t_v^{low} \leq t_v \leq t_v^{high}$. Thus, if there is any node v for which $t_v^{low} > t_v^{high}$, then there is no solution to the ITSC-EQ instance. Therefore, for the remainder of this section, we assume that for each node v , $t_v^{low} \leq t_v^{high}$.

Input: Graph $G(V, E)$ of a SyDS \mathcal{S} , a set Q of configurations \mathcal{S} and a collection Γ_{EQ} of constraints of the form $t_{v_i} = t_{v_j}$.

Requirement: Output “Yes” if there is a threshold value t_v for each $v \in V$ such that in the resulting SyDS, all the configurations in Q are stable and all the constraints in Γ_{EQ} are satisfied. Otherwise, output “No”.

Steps:

1. Using the configurations in Q , for each node $v \in V$, find lower (t_v^{low}) and upper (t_v^{high}) bounds on the threshold value t_v of v using the method given in the proof of [Theorem 3](#).
2. If there is a node $v \in V$ such that $t_v^{low} > t_v^{high}$ then **output** “No” and **stop**.
3. Find the partition $P = \{V_1, V_2, \dots, V_r\}$ of V using the equivalence relation Γ_{EQ} . (Nodes which do not appear in any constraint form singleton blocks in P .)
4. **for** $i = 1$ **to** r **do**
 - (a) Determine whether there is a value α_i such that assigning the threshold value α_i to each node in V_i satisfies all the bounds on the threshold values of each node in V_i . (Note that $0 \leq \alpha_i \leq n + 1$.)
 - (b) If there is no such α_i , then **output** “No” and **stop**.
5. **Output** “Yes”.

Fig. 6. A polynomial time algorithm for ITSC-EQ.

Let $V' \subseteq V$ be the set of nodes which appear in one or more constraints of Γ_{EQ} . Since equality of threshold values is an equivalence relation, the set Γ_{EQ} induces a partition of V' which can be constructed in polynomial time. We augment this partition by creating $|V - V'|$ additional blocks, with each such block containing just one node of $V - V'$. Let V_1, V_2, \dots, V_r denote the blocks in the resulting partition of V . Thus, in any solution to the given ITSC-EQ instance, all the nodes in block V_i must have the same threshold value, $1 \leq i \leq r$. Recall that for each node v , the threshold values must be in the range $[0 .. d_v + 2]$; since $d_v \leq n - 1$, the number of possible threshold values for any node is $O(n)$. So, by trying each such threshold value, it is possible to determine in polynomial time whether there is a single threshold value α_i that satisfies all the bounds (determined by Q) for each node in block V_i , $1 \leq i \leq r$. Clearly, there is a solution to the ITSC-EQ instance if and only if the answer is “Yes” for each block.

An outline of our algorithm for ITSC-EQ appears in [Fig. 6](#). The correctness of the algorithm and the fact that it runs in polynomial time readily follow from the above discussion. The following theorem summarizes our result for the ITSC-EQ problem.

Theorem 15. *The ITSC-EQ problem can be solved in polynomial time.* \square

5.2.2. Results for ITSC-NE

We begin by showing that ITSC-NE is **NP**-complete. The proof is by a straightforward reduction from the 3-Coloring problem for undirected graphs: given a graph $H(V_H, E_H)$, can the nodes of H be colored using at most three colors so that no pair of adjacent nodes receives the same color? The 3-Coloring problem is known to be **NP**-complete [54]. In this problem, we may assume without loss of generality that the input graph H requires at least 3 colors since the 2-coloring problem can be solved in polynomial time [54]. The main idea used in our reduction is that each edge of the graph representing an instance of 3-Coloring becomes an inequality constraint for the resulting ITSC-NE instance.

Proposition 16. *The ITSC-NE problem is **NP**-complete.*

Proof. It is easy to see that ITSC-NE is in **NP**. The proof of **NP**-hardness is through a reduction from the 3-Coloring problem. Let the graph $H(V_H, E_H)$ represent the given instance of the 3-Coloring problem. Let $V_H = \{w_1, w_2, \dots, w_n\}$ so that $|V_H| = n$. We construct an instance of the ITSC-NE problem as follows. For each node $w_i \in V_H$, the underlying graph $G(V, E)$ of the SyDS \mathcal{S} has two nodes denoted by v_i and x_i , $1 \leq i \leq n$. Let $V = \{v_1, v_2, \dots, v_n\}$ and $X = \{x_1, x_2, \dots, x_n\}$. Node v_i corresponds to node w_i of H , $1 \leq i \leq n$. The edge set E of G consists of n edges, namely $\{v_1, x_1\}, \{v_2, x_2\}, \dots, \{v_n, x_n\}$. The only configuration \mathcal{C} of \mathcal{S} that must be made stable has the state value 1 for all the nodes of G . For each edge $\{w_i, w_j\} \in E_H$, the constraint $t_{v_i} \neq t_{v_j}$ is added to the set Γ_{NE} of inequality constraints. (There are no constraints on the threshold value of the nodes in X .) This completes the construction, which can obviously be carried out in polynomial time.

Suppose H is 3-colorable. Let the three colors be denoted by 0, 1 and 2. Further, let V_H^i denote the color class i (i.e., the set of nodes of H assigned color i), $i = 0, 1, 2$. Consider the following threshold assignment: all the nodes of V corresponding to the color class V_H^i are assigned the threshold value i ($i = 0, 1, 2$), and each node in X is assigned the threshold value 1. It can be verified that under this assignment, \mathcal{C} is a stable configuration and that all the constraints in Γ_{NE} are satisfied.

Now, suppose there is a solution to the ITSC-NE instance. Since the degree of each node in G is 1, the solution must assign the threshold value 0, 1, 2 or 3 to each node of G . Also, since the configuration \mathcal{C} (in which the state of every node is 1) is required to be stable, no node of G can be assigned the threshold of 3; that is, each node of G must be assigned a

Input: Graph $G(V, E)$ of a SyDS \mathcal{S} , a set Q of configurations and a collection Γ_{NE} of constraints of the form $t_{v_i} \neq t_{v_j}$. Let $k = |\Gamma_{NE}|$.

Requirement: Output “Yes” if there is a threshold value t_v for each $v \in V$ such that in the resulting SyDS, all the configurations in Q are stable and all the constraints in Γ_{NE} are satisfied. Otherwise, output “No”.

Steps:

1. Using the configurations in Q , for each node $v \in V$, find lower (t_v^{low}) and upper (t_v^{high}) bounds on the threshold value t_v of v using the method given in the proof of [Theorem 3](#).
 2. If there is a node $v \in V$ such that $t_v^{low} > t_v^{high}$ then **output** “No” and **stop**.
 3. Construct graph $G_\Gamma(V_1, E_1)$: $V_1 \subseteq V$ is the set of nodes involved in one or more constraints in Γ_{NE} and for each constraint $t_{v_i} \neq t_{v_j}$ in Γ_{NE} , E_1 contains the edge $\{v_i, v_j\}$.
 4. For each node $v \in V_1$, construct a list L_v containing each integer value α such that $t_v^{low} \leq \alpha \leq t_v^{high}$.
 5. Let $V'_1 \subseteq V_1$ be the set of nodes such that for each node $w \in V'_1$, the list L_w has at least $k + 1$ values. Let G_Γ^1 be the subgraph of G_Γ induced on the node set $V_1 - V'_1$.
 6. Consider each combination β of values in the lists for the nodes of G_Γ^1 . If some combination β yields a valid list coloring of G_Γ^1 , then **output** “Yes”; otherwise, **output** “No”.
-

Fig. 7. Algorithm to show the fixed parameter tractability of ITSC-NE.

threshold value of 0, 1 or 2. Consider the following assignment of colors to the nodes of H : if node v_i is assigned threshold α (for some $\alpha \in \{0, 1, 2\}$), assign color α to the corresponding node w_i , $1 \leq i \leq n$. It can be verified that this is a valid 3-coloring of H , and this completes the proof. \square

Our next result shows that the ITSC-NE problem is fixed parameter tractable with respect to the number of inequality constraints. Let $k = |\Gamma_{NE}|$ and let V_1 denote the set of nodes which appear in at least one constraint in Γ_{NE} . Since the number of constraints is k , we have $|V_1| \leq 2k$. The nodes in $V - V_1$ are not involved in any of the constraints in Γ_{NE} .

As in the algorithm for ITSC-EQ, the set of configurations Q (which must be made stable by the chosen threshold assignment) can be used to construct lower (t_v^{low}) and upper (t_v^{high}) bounds on the threshold value of each node $v \in V$. If there is a node v for which $t_v^{low} > t_v^{high}$, then there is no solution to the ITSC-NE instance. So, for the remainder of this discussion, we assume that for each node v , $t_v^{low} \leq t_v^{high}$.

To solve the ITSC-NE problem, we need to determine whether a threshold value for each node $v \in V_1$ can be chosen so that the constraints in Γ_{NE} are also satisfied. Consider the graph $G_\Gamma(V_1, E_1)$ whose node set is V_1 and whose edge set E_1 represents the constraints in Γ_{NE} ; that is, for each constraint $t_{v_i} \neq t_{v_j}$ in Γ_{NE} , E_1 has the edge $\{v_i, v_j\}$. For each node $v \in V_1$, let L_v denote the list of integer values in the (closed) interval $[t_v^{low} \dots t_v^{high}]$; thus, L_v is the list of all allowable threshold values for v , as determined by the set of configurations Q . Thus, our goal is to determine whether for each node $v \in V_1$, a value can be chosen from its list L_v so that no pair of adjacent nodes in G_Γ has the same value. This is the well known **list coloring problem** in graph theory [[59](#)].

Since $|E_1| = k$, the maximum degree of any node in G_Γ is k . Let V'_1 denote the set of all nodes v of G_Γ such that L_v contains $k + 1$ or more values. We note that each node $v \in V'_1$ (and the edges incident on v) can be removed from G_Γ . This is because after obtaining a valid list coloring for the nodes in $V_1 - V'_1$, for each node $v \in V'_1$, a threshold value α from L_v can be chosen so that α differs from the values chosen for the (at most k) neighbors of v in G_Γ . Let G_Γ^1 denote the subgraph of G_Γ induced on $V_1 - V'_1$. Thus, the list for each node in G_Γ^1 has at most k values. Since G_Γ^1 has at most $2k$ nodes and each list has at most k values, we need to try only k^{2k} combinations of values in an exhaustive search to determine whether there is a valid list coloring of G_Γ^1 ; if none, there is no solution to the ITSC-NE instance. Otherwise, any valid list coloring of G_Γ^1 can be extended to a valid list coloring of G_Γ . This coloring of G_Γ along with a choice of threshold values for nodes in $V - V_1$ (i.e., nodes which do not appear in any constraint) constitutes a solution to the ITSC-NE instance. The above discussion leads to the outline of our algorithm for the ITSC-NE problem presented in [Fig. 7](#).

Theorem 17. The ITSC-NE problem is fixed parameter tractable with respect to the number of inequality constraints.

Proof. The correctness of the algorithm for the ITSC-NE problem presented in [Fig. 7](#) follows from the above discussion. We now show that the running time has the form $O(f(k) + N^{O(1)})$, where N is the size of the problem instance and the function f depends only on k , the number of inequality constraints. In the ensuing discussion, we refer to the steps of the algorithm presented in [Fig. 7](#).

Step 1 can be done in $O(n^2|Q|)$ time by going through each configuration in Q and finding the number of appropriate input values for each of the n nodes of the SyDS \mathcal{S} (as discussed in the proof of [Theorem 3](#)). Step 2 can be done in $O(n)$

Input: Graph $G(V, E)$ of a SyDS \mathcal{S} , a set Q of configurations and a collection Γ_{EQ} of constraints of the form $t_{v_i} = t_{v_j}$. Let $q = |Q|$.

Requirement: Output “Yes” if there is a threshold value t_v for each $v \in V$ such that in the resulting SyDS, all the configurations in Q are unstable and all the constraints in Γ_{EQ} are satisfied. Otherwise, output “No”.

Steps:

1. Find the partition $P = \{V_1, V_2, \dots, V_r\}$ of V using the equivalence relation Γ_{EQ} . (Nodes which do not appear in any constraint form singleton blocks in P .)
2. **if** $q < 2^r$ **then output** “Yes” and **stop**.
3. **Comment:** Here, $q \geq 2^r$.
 - (a) Consider each combination $\beta = (\alpha_1, \alpha_2, \dots, \alpha_r)$ of threshold values to the r blocks, where α_i is the value for block i , $1 \leq i \leq r$.
 - (b) If some combination β makes each configuration in Q unstable, **output** “Yes”; otherwise, **output** “No”.

Fig. 8. A quasi-polynomial time algorithm for ITUC-EQ.

time. Step 3 can be done in $O(k)$ time by going through each constraint in Γ_{NE} . Since $k = O(n^2)$, the time for Step 3 is $O(n^2)$. Since the size of the list L_v for each node v is $O(n)$, Step 4 and Step 5 can be done in time $O(nk) = O(n^3)$. So, the time for Steps 1 through 5 is $O(n^3 + n^2|Q|) = O(N^3)$, where N is the size of the ITSC-NE instance.

Step 6 checks all combinations of values in the lists for the nodes of G_Γ^1 . Since G_Γ^1 has at most $2k$ nodes and each list has at most k values, the number of possible combinations is k^{2k} . For each combination, we can determine whether the resulting coloring of G_Γ^1 is valid in $O(k)$ time since G_Γ^1 has at most k edges. Therefore, Step 6 can be implemented to run in $O(k^{2k+1})$ time.

Hence, the overall running time of the algorithm is $O(k^{2k+1} + N^3)$; that is, the ITSC-NE problem is fixed parameter tractable with respect to k . \square

5.3. Inferring thresholds with unstable configurations and constraints

In this section, we focus on threshold inference problems given a set of unstable configurations along with constraints. We show that the ITUC-EQ problem can be solved in quasi-polynomial time. We also show that, in general, ITUC-NE is **NP**-complete. However, when the number of inequality constraints is *fixed*, we again show that ITUC-NE can be solved in quasi-polynomial time.

5.3.1. An algorithm for ITUC-EQ

We begin by discussing the ideas behind the algorithm. As in the case of the algorithm for ITSC-EQ (Section 5.2), we first construct a partition $\{V_1, V_2, \dots, V_r\}$ of the node set V using the constraints in Γ_{EQ} . In any solution to the ITUC-EQ instance, all nodes in block V_i must have the same threshold value, $1 \leq i \leq r$.

We say that a configuration \mathcal{C} of \mathcal{S} is **block uniform** if the following condition holds: for each block V_i , each node in V_i has the same state value in \mathcal{C} , $1 \leq i \leq r$. Since there are r blocks, the number of block uniform configurations of \mathcal{S} is 2^r .

Let $Q = \{C_1, C_2, \dots, C_q\}$, where $q = |Q|$. We first consider the case where $q < 2^r$. In this case, at least one block uniform configuration, say C , does *not* appear in Q . We note that such a configuration can be found in polynomial time by first sorting the configurations of Q using radix sort [53] and carrying out a simple linear scan of the resulting sorted list of configurations. If we make C the successor of each configuration in Q , then each configuration in Q is rendered unstable. Since C is block uniform, we can achieve this and ensure that all the nodes in each block V_i have the same threshold value by the following method (which is similar to the one used in the proof of Lemma 2 in Section 2): for each block V_i , if all the nodes in V_i have state value 1 in C , set the threshold of each node in V_i to 0; otherwise, set the threshold of each node $v \in V_i$ to $d_v + 2$ (which corresponds to the value “infinity”). Thus, when $q < 2^r$, there is always a solution to the ITUC-EQ instance.

Now, consider the other case where $|Q| = q \geq 2^r$. Since Q is part of the input, in this case, the size N of the ITUC-EQ instance is at least 2^r . For any block V_i , the possible threshold values are in the interval $[0 .. n + 1]$; thus, the number of possible values is at most $n + 2$. Since there are r blocks, the number of combinations of threshold values for all the blocks is at most $(n + 2)^r$. Our algorithm carries out an exhaustive search over all these combinations of threshold values. If some combination makes each configuration in Q unstable, then there is a solution to the ITUC-EQ instance; otherwise, there is no solution. In the proof of Theorem 18, we will show that the running time for this case is $O(N^c \log N)$ for some constant c .

An outline of our algorithm for ITUC-EQ appears in Fig. 8. Correctness of the algorithm follows from the above discussion. The following theorem establishes the running time of the algorithm.

Theorem 18. *The ITUC-EQ problem can be solved in quasi-polynomial time. More specifically, the running time of the algorithm in Fig. 8 is $O(N^c \log N)$ where N is the size of the problem instance and c is a constant.*

Proof. When $|Q| < 2^r$, the running time of the algorithm in Fig. 8 can be seen to be a polynomial in N , the size of the problem instance. So, we need to consider only the case when $|Q| \geq 2^r$. We show that the running time in this case is $O(N^{c \log N})$ for some constant c .

There are r blocks, and the number of threshold values to be considered for each block is at most $n + 2$. Thus, the number of combinations of threshold values to be considered is at most $(n + 2)^r = O(N^r)$. For each combination, we can determine in $O(|Q|n^2) = O(N^3)$ time whether the combination causes each configuration in Q to be unstable. Therefore, the total time spent in Step 3(b) of the algorithm is $O(N^{r+3})$. Since $|Q| \geq 2^r$, we have $N \geq 2^r$ or $r \leq \log N$. Therefore, the overall running time is $O(N^{\log N+3})$ which is $O(N^{c \log N})$ for a constant c . Since $O(N^{c \log N}) = 2^{O((\log N)^2)}$, the running time is quasi-polynomial. \square

5.4. Results for ITUC-NE

We first show that ITUC-NE is **NP**-complete through a reduction similar to the one used in the proof of Proposition 16.

Proposition 19. *The ITUC-NE problem is **NP**-complete.*

Proof. It is easy to verify that ITUC-NE is in **NP**. We establish **NP**-hardness using a reduction from 3-Coloring. Let the graph $H(V_H, E_H)$ represent the given instance of the 3-Coloring problem. Let $V_H = \{w_1, w_2, \dots, w_n\}$ so that $|V_H| = n$. We construct an instance of the ITUC-NE problem as follows. For each node $w_i \in V_H$, the underlying graph $G(V, E)$ of the SyDS \mathcal{S} has one corresponding node denoted by v_i , $1 \leq i \leq n$. Thus, $V = \{v_1, v_2, \dots, v_n\}$. The edge set E of G is empty. The only configuration C' of \mathcal{S} that must be made unstable has the state value 0 for all the nodes of G . For each edge $\{w_i, w_j\} \in E_H$, the constraint $t_{v_i} \neq t_{v_j}$ is added to the set Γ_{NE} of inequality constraints. This completes the construction, which can obviously be carried out in polynomial time.

Suppose H is 3-colorable. Let the three colors be denoted by 0, 1 and 2. Further, let V_H^i denote the color class i (i.e., the set of nodes of H assigned the color i), $i = 0, 1, 2$. Since H is assumed to require at least three colors, each of these color classes is nonempty. Consider the following threshold assignment: all the nodes of V corresponding to the color class V_H^i are assigned the threshold value i , $i = 0, 1, 2$. It is easy to see that this assignment satisfies all the constraints in Γ_{NE} . Further, since at least one node of G is assigned the threshold value 0, C' is an unstable configuration. Thus, we have a solution to the ITUC-NE instance.

Now, suppose there is a solution to the ITUC-NE instance. Since the degree of each node in G is zero, the solution must assign the threshold value 0, 1 or 2 to each node of G . Consider the following assignment of colors to the nodes of H : if node v_i is assigned threshold α (for some $\alpha \in \{0, 1, 2\}$), assign color α to the corresponding node w_i , $1 \leq i \leq n$. It can be verified that this is a valid 3-coloring of H , and this completes the proof. \square

We now show that when the number of inequality constraints is *fixed*, the ITUC-NE problem can be solved in quasi-polynomial time. The approach is similar to the one used for the ITUC-EQ problem. Let $k = |\Gamma_{NE}|$ and let V_1 denote the subset of nodes which appear in one or more constraints in Γ_{NE} . Let $V_2 = V - V_1$ be the set of nodes which are not involved in any constraint. Let $p = |V_1|$; thus, $|V_2| = n - p$. As before, we know that $p \leq 2k$. As a first step towards our algorithm for the ITUC-NE problem, we have the following simple lemma.

Lemma 20. *The problem of determining whether there is a threshold assignment to the nodes of V_1 that satisfies all the constraints in Γ_{NE} is in the complexity class **XP**, with $k = |\Gamma_{NE}|$ as the parameter. In particular, the problem can be solved in $O(kn^{2k})$ time, which is polynomial when k is fixed.*

Proof. A straightforward algorithm with a running time of $O(kn^{2k})$ can be obtained as follows. For each node in V_1 , there are only $O(n)$ possible threshold values. Since $|V_1| \leq 2k$, the number of combinations of threshold values to the nodes in V_1 is $O(n^{2k})$. For each combination, checking whether all the constraints in Γ_{NE} are satisfied can be done in $O(k)$ time. Therefore, the overall time is $O(kn^{2k})$, which is a polynomial since k is fixed. Since $O(kn^{2k}) = O(n^{2k+1})$, the running time is of the form $O(n^{f(k)})$, where $f(k)$ depends only on k . Thus, the problem is in **XP** [55]. \square

Our algorithm proceeds by considering two cases. For the first case, we let $q = |Q| < 2^{n-p}$. For each configuration $C_j \in Q$, let C'_j be the subconfiguration obtained by considering only the state values of nodes in V_2 , $1 \leq j \leq q$. Let $Q' = \{C'_1, C'_2, \dots, C'_q\}$. Let Π denote the set of all 2^{n-p} subconfigurations of \mathcal{S} obtained by assigning state values 0 or 1 to the $n - p$ nodes of V_2 . Since $|Q| < |\Pi|$, there is at least one subconfiguration C' in Π which does not appear in Q' . As in the algorithm for ITUC-EQ, such a subconfiguration can be found in polynomial time by first constructing Q' and sorting the subconfigurations in Q' . The subconfiguration C' can be extended into a configuration C of \mathcal{S} by choosing arbitrary values for the nodes in V_1 . As in the algorithm for ITUC-EQ, we can make C the successor for all the configurations in Q by choosing appropriate threshold values for the nodes in V_2 (which are not involved in any constraint). This ensures that each configuration in Q is unstable. Therefore, there is a solution to the ITUC-EQ instance if and only if there is a threshold

Input: Graph $G(V, E)$ of a SyDS \mathcal{S} , a set Q of configurations and a collection Γ_{NE} of constraints of the form $t_{v_i} \neq t_{v_j}$. Let $q = |Q|$.

Requirement: Output “Yes” if there is a threshold value t_v for each $v \in V$ such that in the resulting SyDS, all the configurations in Q are unstable and all the constraints in Γ_{NE} are satisfied. Otherwise, output “No”.

Steps:

1. Let V_1 be the set of nodes that appear in one or more constraints of Γ_{NE} . Let $p = |V_1|$ and let $V_2 = V - V_1$.

2. Case 1: $q < 2^{n-p}$.

- (a) Use the algorithm presented in the proof of [Lemma 20](#) to check whether there is a threshold assignment to the nodes in V_1 that satisfies all the constraints in Γ_{NE} .
- (b) If a threshold assignment is found in Step 2(a), then **output** “Yes” and **stop**; otherwise, **output** “No” and **stop**.

3. Case 2: Here, $q \geq 2^{n-p}$.

- (a) Consider each combination $\beta = (\alpha_1, \alpha_2, \dots, \alpha_n)$ of threshold values to the n nodes, where α_i is the value for node v_i , $1 \leq i \leq n$. (Note that $0 \leq \alpha_i \leq n+1$, $1 \leq i \leq n$.)
 - (b) If some combination β makes each configuration in Q unstable and satisfies all the constraints in Γ_{NE} , **output** “Yes”; otherwise, **output** “No”.
-

Fig. 9. A quasi-polynomial time algorithm for ITUC-NE when the number of constraints is fixed.

assignment for the nodes in V_1 that satisfies all the constraints in Γ_{NE} . From [Lemma 20](#), the latter task can be done in polynomial time. Therefore, when $|Q| < 2^{n-p}$, the algorithm runs in polynomial time when k is fixed.

For the second case, we have $|Q| \geq 2^{n-p}$. Therefore, in this case, the size of the problem instance N is at least 2^{n-p} . The algorithm uses a simple exhaustive search over all the $O(n^n)$ possible assignments of threshold values. As in the case of the algorithm for ITSC-EQ, this running time can be shown to be quasi-polynomial in the size of the problem instance.

An outline of our algorithm for ITUC-NE for the case when $|\Gamma_{NE}|$ is fixed appears in [Fig. 9](#). Since the correctness of the algorithm follows from the above discussion, the following theorem establishes that the running time is quasi-polynomial.

Theorem 21. *When the number of inequality constraints is fixed, the ITUC-NE problem can be solved in time $O(N^{c \log N})$ time, where N is the size of the problem instance and c is a constant.*

Proof. Let $k = |\Gamma_{NE}|$ and let p be the number of nodes appearing in one or more constraints. We note that $p \leq 2k$.

From the preceding discussion, when $|Q| < 2^{n-p}$, the running time of the algorithm in [Fig. 9](#) can be seen to be a polynomial in N , the size of the problem instance. When $|Q| \geq 2^{n-p}$, we show that the running time is $O(N^{c \log N})$ for some constant c .

There are n nodes, and the number of threshold values to be considered for each node is at most $n+2$. Thus, the number of combinations of threshold values to be considered is at most $(n+2)^n = O(N^n)$. For each combination, we can determine in $O(|Q|n^2) = O(N^3)$ time whether the combination causes each configuration in Q to be unstable. Therefore, the total time spent in Step 3 of the algorithm is $O(N^{n+3})$. Since $|Q| \geq 2^{n-p}$, we have $N \geq 2^{n-p}$ or $n \leq \log_2 N + p$. Since $p \leq 2k$, we have $n \leq \log_2 N + 2k$. Therefore, the overall running time is $O(N^{\log_2 N + 3 + 2k})$ which is $O(N^{c \log N})$ for a constant c since k is fixed. \square

6. Summary and future research directions

We considered many versions of threshold inference problems for deterministic SyDSs and presented complexity results, fixed parameter tractability results as well as polynomial and quasi-polynomial time algorithms. We conclude by mentioning some specific open problems and a few general directions for future work.

It is of interest to investigate whether the running times of our algorithms that show the fixed parameter tractability of some inference problems can be improved. Alternatively, it may also be of interest to investigate whether lower bounds can be established on the running times under some complexity theoretic assumptions such as strong exponential time hypothesis [60,61]. When considering constrained versions of threshold inference problems, we obtained quasi-polynomial time algorithms for two problems, namely ITUC-EQ and ITUC-NE when the number of constraints is fixed. It is of interest to investigate whether there are polynomial time algorithms for these problems.

Our work also suggests several general directions for future work. One direction is to consider inference problems for other forms of observed behavior such as a collection of snapshots of the system, where each snapshot specifies a time and the configuration of the system at that time. A second direction is to consider inference problems assuming more powerful local functions. Finally, it is also of interest to consider inference problems for stochastic SyDSs, whose local transition functions are probabilistic threshold functions.

Acknowledgements

It is our pleasure to thank the reviewer for carefully reading the manuscript and providing valuable suggestions. This work has been partially supported by DTRA Grant HDTRA1-11-1-0016 and DTRA CNIMS Contract HDTRA1-11-D-0016-0010, NSF NetSE Grant CNS-1011769, NSF SDCI Grant OCI-1032677 and NIH MIDAS Grant 5U01GM070694-11.

References

- [1] K. Lum, S. Swarup, S. Eubank, J. Hawdon, The contagious nature of imprisonment: an agent-based model to explain racial disparities in incarceration rates, *J. R. Soc. Interface* 11 (98) (2014), arXiv:2014.0409.
- [2] M.E. Halloran, N.M. Ferguson, S. Eubank, I.M. Longini, D.A. Cummings, B. Lewis, S. Xu, C. Fraser, A. Vullikanti, T.C. Germann, et al., Modeling targeted layered containment of an influenza pandemic in the United States, *Proc. Natl. Acad. Sci. USA* 105 (12) (2008) 4639–4644.
- [3] O. Diekmann, H. Heesterbeek, T. Britton, *Mathematical Tools for Understanding Infectious Disease Dynamics*, Princeton University Press, 2012.
- [4] D. Gruhl, R. Guha, D. Liben-Nowell, A. Tomkins, Information diffusion through blogspace, in: *Proceedings of the 13th International Conference on World Wide Web*, ACM, 2004, pp. 491–501.
- [5] B.A. Prakash, D. Chakrabarti, N.C. Valler, M. Faloutsos, C. Faloutsos, Threshold conditions for arbitrary cascade models on arbitrary networks, *Knowl. Inf. Syst.* 33 (3) (2012) 549–575.
- [6] M. Granovetter, Threshold models of collective behavior, *Amer. J. Sociol.* (1978) 1420–1443.
- [7] C.J. Kuhlman, V.A. Kumar, M.V. Marathe, S. Ravi, D.J. Rosenkrantz, Inhibiting diffusion of complex contagions in social networks: theoretical and experimental results, *Data Min. Knowl. Discov.* 29 (2) (2015) 423–465.
- [8] T.G. Trucano, L.P. Swiler, T. Igusa, W.L. Oberkampf, M. Pilch, Calibration, validation, and sensitivity analysis: what's what, *Reliab. Eng. Syst. Saf.* 91 (10) (2006) 1331–1357.
- [9] S. González-Bailón, J. Borge-Holthoefer, A. Rivero, Y. Moreno, The dynamics of protest recruitment through an online network, *Sci. Rep.* 1 (2011), 7 pages.
- [10] D.M. Romero, B. Meeder, J. Kleinberg, Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter, in: *Proceedings of the 20th International Conference on World Wide Web*, ACM, 2011, pp. 695–704.
- [11] J. Ugander, L. Backstrom, C. Marlow, J. Kleinberg, Structural diversity in social contagion, *Proc. Natl. Acad. Sci. USA* 109 (16) (2012) 5962–5966.
- [12] D. Easley, J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*, Cambridge University Press, 2010.
- [13] H. Mortveit, C. Reidys, *An Introduction to Sequential Dynamical Systems*, Springer Science & Business Media, New York, NY, 2007.
- [14] C. Barrett, H.B. Hunt, M.V. Marathe, S. Ravi, D.J. Rosenkrantz, R.E. Stearns, Modeling and analyzing social network dynamics using stochastic discrete graphical dynamical systems, *Theoret. Comput. Sci.* 412 (30) (2011) 3932–3946.
- [15] J. Crane, The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing, *Amer. J. Sociol.* (1991) 1226–1259.
- [16] A. Gaviria, S. Raphael, School-based peer effects and juvenile behavior, *Rev. Econ. Stat.* 83 (2) (2001) 257–268.
- [17] Y.J. Xu, Advance to and persistence in graduate school: identifying the influential factors and major-based differences, *J. Coll. Stud. Ret., Res. Theory Pract.* 16 (3) (2014) 391–417.
- [18] M.H.W. v. Zalk, M. Kerr, S.J. Branje, H. Stattin, W.H. Meeus, Peer contagion and adolescent depression: the role of failure anticipation, *J. Clin. Child Adolesc. Psychol.* 39 (6) (2010) 837–848.
- [19] E.A. Stevens, M.J. Prinstein, Peer contagion of depressogenic attributional styles among adolescents: a longitudinal study, *J. Abnorm. Child Psychol.* 33 (1) (2005) 25–37.
- [20] C. De la Higuera, *Grammatical Inference: Learning Automata and Grammars*, Cambridge University Press, 2010.
- [21] J. Heinz, C. De la Higuera, M. van Zaanen, Grammatical inference for computational linguistics, *Synth. Lect. Hum. Lang. Technol.* 8 (4) (2015) 1–139.
- [22] K.P. Murphy, Passively learning finite automata, *Tech. rep. 96-04-017*, Santa Fe Institute, Santa Fe, NM, 1996.
- [23] M.J. Kearns, U.V. Vazirani, *An Introduction to Computational Learning Theory*, MIT Press, Cambridge, MA, 1994.
- [24] M.W. Macy, R. Willer, From factors to actors: computational sociology and agent-based modeling, *Annu. Rev. Sociol.* 28 (2002) 143–166.
- [25] C. Barrett, H.B. Hunt, M.V. Marathe, S. Ravi, D.J. Rosenkrantz, R.E. Stearns, M. Thakur, Predecessor existence problems for finite discrete dynamical systems, *Theoret. Comput. Sci.* 386 (1) (2007) 3–37.
- [26] F. Green, NP-complete problems in cellular automata, *Complex Systems* 1 (3) (1987) 453–474.
- [27] C.L. Barrett, H.B. Hunt, M.V. Marathe, S. Ravi, D.J. Rosenkrantz, R.E. Stearns, Complexity of reachability problems for finite discrete dynamical systems, *J. Comput. System Sci.* 72 (8) (2006) 1317–1345.
- [28] S. Kosub, C.M. Homan, Dichotomy results for fixed point counting in Boolean dynamical systems, in: *Proceedings of the 10th Italian Conference on Theoretical Computer Science*, 2007, pp. 163–174.
- [29] K. Sutner, Computational classification of cellular automata, *Int. J. Gen. Syst.* 41 (6) (2012) 595–607.
- [30] B. Abrahao, F. Chierichetti, R. Kleinberg, A. Panconesi, Trace complexity of network inference, in: *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2013, pp. 491–499.
- [31] M. Gomez Rodriguez, J. Leskovec, A. Krause, Inferring networks of diffusion and influence, in: *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2010, pp. 1019–1028.
- [32] S. Soundarajan, J.E. Hopcroft, Recovering social networks from contagion information, in: *Proceedings of the 7th Annual Conference on Theory and Models of Computation*, Springer, 2010, pp. 419–430.
- [33] D. Shah, T. Zaman, Rumors in a network: who's the culprit?, *IEEE Trans. Inform. Theory* 57 (8) (2011) 5163–5181.
- [34] J. Kleinberg, Cascading behavior in networks: algorithmic and economic issues, in: *Algorithmic Game Theory*, Cambridge University Press, UK, 2007, pp. 613–632, Ch. 24.
- [35] E. Goles, S. Martínez, *Neural and Automata Networks: Dynamical Behavior and Applications*, Kluwer, Dordrecht, The Netherlands, 1990.
- [36] B. Durand, A random NP-complete problem for inversion of 2D cellular automata, *Theoret. Comput. Sci.* 148 (1) (1995) 19–32.
- [37] C.L. Barrett, H.B. Hunt III, M.V. Marathe, S. Ravi, D.J. Rosenkrantz, R.E. Stearns, P.T. Tomic, Gardens of Eden and fixed points in sequential dynamical systems, in: *Proceedings of the Discrete Mathematics and Theoretical Computer Science Conference*, 2001, pp. 95–110.
- [38] D. Kempe, J. Kleinberg, E. Tardos, Maximizing the spread of influence through a social network, in: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003, pp. 137–146.
- [39] C. Bazgan, M. Chopin, A. Nichterlein, F. Sikora, Parameterized approximability of maximizing the spread of influence in networks, *J. Discrete Algorithms* 27 (2014) 54–65.
- [40] C. Bazgan, M. Chopin, The complexity of finding harmless individuals in social networks, *Discrete Optim.* 14 (2014) 170–182.
- [41] C. Bazgan, M. Chopin, M. Cygan, M.R. Fellows, F.V. Fomin, E.J. van Leeuwen, Parameterized complexity of firefighting, *J. Comput. System Sci.* 80 (7) (2014) 1285–1297.

- [42] F.N. Abu-Khzam, J. Egan, M.R. Fellows, F.A. Rosamond, P. Shaw, On the parameterized complexity of dynamic problems with connectivity constraints, in: Combinatorial Optimization and Applications, Springer, 2014, pp. 625–636.
- [43] N. Boria, J. Monnot, V.T. Paschos, Reoptimization under vertex insertion: max P_k -free subgraph and max planar subgraph, Discrete Math. Algorithms Appl. 5 (2) (2013), 25 pages.
- [44] T. Ito, E.D. Demaine, N.J. Harvey, C.H. Papadimitriou, M. Sideri, R. Uehara, Y. Uno, On the complexity of reconfiguration problems, in: Proceedings of the 19th International Symposium on Algorithms and Computation, Springer, 2008, pp. 28–39.
- [45] H. Fernau, J.A. Rodriguez-Velazquez, A survey on alliances and related parameters in graphs, Electron. J. Graph Theory Appl. 2 (1) (2014) 70–86.
- [46] J. Łacki, J. Oćwieja, M. Pilipczuk, P. Sankowski, A. Zych, The power of dynamic distance oracles: efficient dynamic algorithms for the Steiner tree, in: Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, ACM, 2015, pp. 11–20.
- [47] J. van den Heuvel, The complexity of change, in: S.R. Blackburn, S. Gerke, M. Wildon (Eds.), Surveys in Combinatorics, in: London Math. Soc. Lecture Note Ser., vol. 409, Cambridge University Press, 2013, pp. 127–160.
- [48] M. Berglund, H. Björklund, F. Drewes, On the parameterized complexity of linear context-free rewriting systems, in: Proceedings of the 13th Meeting on the Mathematics of Language, Association for Computational Linguistics, 2013, pp. 21–29.
- [49] C.C. Florêncio, H. Fernau, On families of categorical grammars of bounded value, their learnability and related complexity questions, Theoret. Comput. Sci. 452 (2012) 21–38.
- [50] R.G. Downey, M.R. Fellows, B.M. Kapron, M.T. Hallett, H.T. Wareham, The parameterized complexity of some problems in logic and linguistics, in: Proceedings of the Third International Symposium on Logical Foundations of Computer Science, in: Lecture Notes in Comput. Sci., vol. 813, Springer, 1994, pp. 89–100.
- [51] H. Fernau, P. Heggernes, Y. Villanger, A multi-parameter analysis of hard problems on deterministic finite automata, J. Comput. System Sci. 81 (4) (2015) 747–765.
- [52] R. Niedermeier, Invitation to Fixed Parameter Algorithms, Oxford University Press, New York, NY, 2006.
- [53] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, Introduction to Algorithms, second edition, MIT Press and McGraw-Hill, Cambridge, MA, 2009.
- [54] M.R. Garey, D.S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, W.H. Freeman & Co., San Francisco, 1979.
- [55] J. Flum, M. Grohe, Parameterized Complexity Theory, Springer, Heidelberg, Germany, 2006.
- [56] J. Hästads, Clique is hard to approximate within $n^{1-\epsilon}$, Acta Math. 182 (1999) 105–142.
- [57] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, M. Protasi, Complexity and Approximation: Combinatorial Problems and Their Approximability Properties, Springer, Berlin, Germany, 1999.
- [58] R. Graham, D. Knuth, O. Patashnik, Concrete Mathematics, Addison-Wesley, Reading, MA, 1994.
- [59] D.B. West, Introduction to Graph Theory, Prentice-Hall, Inc., Englewood Cliffs, NJ, 2003.
- [60] R. Impagliazzo, R. Paturi, F. Zane, Which problems have strongly exponential complexity?, in: Proceedings of the 39th Annual Symposium on Foundations of Computer Science, IEEE, 1998, pp. 653–662.
- [61] D. Lokshtanov, D. Marx, S. Saurabh, Lower bounds based on the exponential time hypothesis, Bull. Eur. Assoc. Theor. Comput. Sci. EATCS 105 (2011) 41–72.

Near-Optimal Algorithms for Controlling Propagation at Group Scale on Networks

Yao Zhang, Abhijin Adiga, Sudip Saha, Anil Vullikanti, and B. Aditya Prakash

Abstract—Given a network with groups, such as a contact-network grouped by ages, which are the best groups to immunize to control the epidemic? Equivalently, how to choose best communities in social media like Facebook to stop rumors from spreading? Immunization is an important problem in multiple different domains like epidemiology, public health, cyber security, and social media. Additionally, clearly immunization at group scale (like schools and communities) is more realistic due to constraints in implementations and compliance (e.g., it is hard to ensure specific individuals take the adequate vaccine). Hence, efficient algorithms for such a “group-based” problem can help public-health experts take more practical decisions. However, most prior work has looked into individual-scale immunization. In this paper, we study the problem of controlling propagation at group scale. We formulate a set of novel Group Immunization problems for multiple natural settings (for both threshold and cascade-based contagion models under both node-level and edge-level interventions) and develop multiple efficient algorithms, including provably approximate solutions. Finally, we show the effectiveness of our methods via extensive experiments on real and synthetic datasets.

Index Terms—Graph mining, social networks, immunization, diffusion, groups

1 INTRODUCTION

INFECTIOUS diseases account for a large fraction of deaths worldwide. The main public health response to containing epidemic outbreaks is by vaccination and social distancing, e.g., [1], [2]. These interventions have resource constraints (e.g., limited supply of vaccines and the high cost of social distancing), and therefore, designing optimal control strategies is an active area of research in public health policy planning, e.g., [1], [3], [4], [5], [6], [7]. However, optimal strategies based on node level characteristics, such as the degree or spectral properties [6], [7] cannot be easily turned into implementable policies, because such targeted immunization of specific individuals raises significant social and moral issues. As a result, vaccination policies, such as those specified by CDC are at the level of groups (e.g., based on demographics), and almost all the efforts in epidemiology are focused on developing group level strategies, even though this may lead to sub-optimal solutions compared to the individual level policies. For instance, Medlock et al. [1] develop an optimal vaccine allocation for different age groups. Even so, all prior work on optimal group level immunization has focused on differential equation based models, and has not been studied on network models of epidemic spread. Implementing such interventions is challenging because people “comply” with them based on their

individual utility. We model such limited compliance by random vaccine allocation within each group, which motivates our paper. Our focus in this paper is on developing interventions that can be implemented before the start of the epidemic. Further, interventions can be of two kinds: vaccination (which can be modeled in terms of node removals) and social distancing (which can be modeled in terms of edge removal, e.g., reducing contacts between certain sub-populations). We consider two kinds of metrics: (1) maximizing the expected number of people who do not get infected, and (2) minimizing the time for the epidemic to die out. These are both commonly studied metrics in public health (see, e.g., [8], [9]). Most of the work in mathematical epidemiology has been formalized in terms of reducing the reproductive number. However, these methods do not extend to network based models. In this paper, we develop algorithms for optimizing these metrics in two different models of diffusion.

Similar diffusion processes arise in other domains such as social media, e.g., the spread of spam/rumors on Facebook, Twitter, LiveJournal or Friendster. These are also commonly modeled by models such as the Linear Threshold (LT) model [10]. Analogous to the public-health case, we can control such processes by ‘immunization’ via blocking users or preventing some interactions, such that the expected number of users who adopt spam/rumors is minimal. Past work has studied individual-level based immunization algorithms for the LT model [11]. However, it is more realistic to issue a warning bulletin on group pages, and some members within those groups comply with the warning to stop disseminating rumors. Similarly, Twitter can warn a group of accounts to control the spread of the malicious tweets. The same holds true for user groups in Friendster and LiveJournal.

In this paper, we present a unified approach to study strategies for controlling the spread of diffusion processes through group level interventions, capturing both uncertainty and lack of control at high resolution within groups. The main contributions of our paper are:

- Y. Zhang and B.A. Prakash are with the Department of Computer Science, Virginia Tech, Blacksburg, VA 24061.
E-mail: {yaozhang, badityap}@cs.vt.edu.
- A. Adiga is with NDSSL, Biocomplexity Institute of Virginia Tech, Blacksburg, VA 24061. E-mail: abhijin@vbi.vt.edu.
- S. Saha and A. Vullikanti are with the Department of Computer Science, Virginia Tech, Blacksburg, VA 24061, and NDSSL, Biocomplexity Institute of Virginia Tech, Blacksburg, VA 24061.
E-mail: {ssaha, akumar}@vbi.vt.edu.

Manuscript received 29 Dec. 2015; revised 9 July 2016; accepted 24 Aug. 2016. Date of publication 1 Sept. 2016; date of current version 2 Nov. 2016.

Recommended for acceptance by A. Gionis.

For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TKDE.2016.2605088

- 1) *Problem Formulation:* We develop group level intervention problems in both the LT model, and the SIS/SIR models, for which we consider a spectral radius based formulation. We consider arbitrarily specified groups, and interventions that involve both edge and node removal, modeling quarantining and vaccination, respectively. The interventions specify the number x_i of nodes/edges that can be removed within each group C_i ; however, these are chosen randomly within the group. These problems generalize the node level problems and have not been studied before.
- 2) *Effective Algorithms:* We develop efficient theoretical and practical algorithms for the four problem classes we consider, including provable approximation algorithms (SDP and GROUPGREEDYWALK). We find that diverse kinds of techniques are needed for these problems—submodular function maximization on an integer lattice, semidefinite programming, quadratic programming, and the link between closed walks and spectral radius. Our algorithms also leverage prior techniques for analyzing contagion processes, e.g., [6], [12], [13], but require non-trivial extensions.
- 3) *Experimental Evaluation:* We present extensive experiments on multiple real datasets including epidemiological and social networks, and demonstrate that our algorithms outperform other competitors on node and edge deletion at group scale for controlling infection as well as spectral radius minimization.

Outline of the Paper. The rest of the paper is organized as follows. We first discuss the related work in Section 2, and then formulate the Group Immunization Problem in Section 3. Section 4 presents our algorithms for different settings of the problem for both edge and node removal. Experimental results on several datasets are in Section 5. We finally discuss future work, and conclude in Section 6.

2 RELATED WORK

In general, there has been a lot of interest in studying dynamical processes on large graphs like (a) blogs and propagations [14], [15], (b) information cascades [16], [17]; (c) marketing and product penetration [18], [19] and (d) malware prediction [20]. These dynamic processes are all closely related to virus propagation in epidemiology, rumor spread in social media, malware outbreaks in computer networks, etc. In this section, we review related work mainly from four areas: epidemiology, propagation models, immunization and other diffusion based optimization problems. In short, past work concentrates on *individual-based* immunization—in contrast, in this paper we study group-based immunization problems under various models.

Epidemiology. The classical texts on epidemic models and analysis are May and Anderson [8] and Hethcote [21]. Widely studied epidemiological models include *homogeneous models* [8], [9], [22], which assume that every individual has equal contact with others in the population.

Propagation Models. There are broadly two types of propagation models which have been used to describe dynamical processes on graphs: threshold based and cascade style.

Threshold based models are well-motivated in the social science literature [23], [24], [25] to represent ‘threshold’

behaviors, e.g., ideas/spam/rumors on Twitter and Facebook. A classic example is the linear threshold model, which has been extensively studied [10]. In this paper, we study the problem of minimizing propagation for the LT model.

Cascade style models, such as the ‘flu-like’ Susceptible-Infectious-Susceptible (SIS), ‘mumps-like’ Susceptible-Infectious-Recovered (SIR) and its special-case the Independent Cascade (IC) [8], [9], [10], [22], are popular in epidemiology literature to model different epidemiological states of people, and their state-transitions. Much work has gone into in finding the ‘epidemic threshold’ for such models (the minimum virulence of a virus which results in an epidemic over the network). For example, recent studies [12], [26] show that the spectral radius of the underlying network (the largest eigenvalue of the adjacency matrix of the graph) is related to the epidemic threshold for a wide-range of cascade models. Hence, here we investigate how to control an epidemic by minimizing the spectral radius for cascade style models.

Immunization. There has been much work on finding optimal strategies for vaccination and social distancing [1], [3], [4], [5], [6], [7]. Much of the work in the epidemiology literature has been based on differential equation methods [1], [3], [4]. Cohen et al. [5] studied the popular *acquaintance* immunization policy (pick a random person, and immunize one of its neighbors at random). Using game theory, Aspnes et al. [27] developed inoculation strategies for victims of viruses under random starting points. Kuhlman et al. [28] studied two formulations of the problem of blocking a contagion through edge removals under the model of discrete dynamical systems. Tong et al. [7], [29], Van Miegham et al. [30], Prakash et al. [6] proposed various node-based and edge-based immunization algorithms based on minimizing the largest eigenvalue of the graph. Other non-spectral approaches for immunization have been studied by Budak et al. [31], He et al. [32], Khalil et al. [11], Saha et al. [13], and Zhang et al. [33]. All of these papers studied individual-based immunization (where either one targets specific individuals or whole demographics). Here we study group-based problems, where vaccines are distributed randomly inside groups.

Other Diffusion Problems. Other diffusion based optimization problems include the influence maximization problem, which was introduced by Domingos and Richardson [34], and formulated by Kempe et al. [10] as a combinatorial optimization problem. They proved it is NP-Hard and also gave a simple $(1 - 1/e)$ -approximation based on the submodularity of expected spread of a set of starting seeds. Recently the paper by Eftekhar et al. [35] studied this problem at group scale. Other such problems where we wish to select a subset of ‘important’ vertices on graphs, include ‘outbreak detection’ [36] and ‘finding most-likely culprits of epidemics’ [37]. Purohit et al. [38] looked into ‘zooming-out’ of a graph by forming groups based on similar influence.

3 OUR PROBLEM FORMULATIONS

Table 1 lists the main symbols we use throughout the paper. Here we assume our graph $G(V, E)$ is directed and weighted. We refer to both node and edge level interventions as immunization.

Groups in a Graph. For a graph $G(V, E)$, we assume that the edge (node) set is partitioned into groups $C = \{C_1, \dots, C_n\}$

TABLE 1
Terms and Symbols

Symbol	Definition and Description
$G(V, E)$	graph G with the node set V and the edge set E
C	set containing groups
A	set of initial infected nodes
n	the number of groups in the graph
m	budget (the number of vaccines)
p_{uv}	weight on edge $e(u, v)$
$g(v)$	group index of node v , i.e., $g(v) = i$ if $v \in C_i$
$g(u, v)$	group index of edge (u, v) , i.e., $g(u, v) = i$ if $(u, v) \in C_i$
\mathbf{x}	vaccine allocation vector (x_1, \dots, x_n) for edges/nodes
$\sigma_{C,A}(\mathbf{x})$	the expected number of infected nodes at the end when \mathbf{x} is allocated to edges
$\sigma'_{C,A}(\mathbf{x})$	the expected number of infected nodes at the end when \mathbf{x} is allocated to nodes
\mathbf{e}_k	vector with $e_k = 1$ and $e_i = 0$ for $i \neq k$
$\mathbf{M}_{\mathbb{E}}(\mathbf{x})$	$\mathbb{E}[\mathbf{M}(\mathbf{x})]$
$\Delta_{\mathbb{E}}(\mathbf{x})$	maximum expected degree of $G(\mathbf{x})$
$\lambda_{\mathbb{E}}(\mathbf{x})$	expected spectral radius of $\mathbf{M}(\mathbf{x})$
$\lambda(\mathbf{M}_{\mathbb{E}}(\mathbf{x}))$	spectral radius of the expected matrix $\mathbf{M}_{\mathbb{E}}(\mathbf{x})$
$\lambda_{\mathbb{E}}^{\min}$	minimum expected spectral radius over all $\mathbf{M}(\mathbf{x})$, i.e., $\min_{\mathbf{x}} \lambda_{\mathbb{E}}(\mathbf{x})$
\mathbf{x}_{\min}	the allocation vector which minimizes $\lambda(\mathbf{M}_{\mathbb{E}}(\mathbf{x}))$ over all \mathbf{x} , i.e., $\arg \min_{\mathbf{x}} \lambda(\mathbf{M}_{\mathbb{E}}(\mathbf{x}))$
s	number of samples in GROUPGREEDYWALK

for the edge (node) immunization problems. For a node partition, C may correspond to groups of communities, locations, demographics, etc. And edge groups can be induced from node groups. For example, for an edge $e = (u, v)$, if u and v belong to a group C_t , then $e \in C_t$, otherwise it belongs to group $C_{ij} = \{e_{uv} | u \in C_i, v \in C_j\}$. The edge groups we defined ensure that every edge e has a group even if the endpoints of e belong to different node groups. Note that we assume there are no overlaps among groups.

Allocating Vaccines to Groups. We define $\mathbf{x} = (x_1, \dots, x_n)$ as the vaccine allocation vector, i.e., if we give x_i vaccines to group C_i , x_i edges (nodes) will be uniformly randomly removed from C_i , which means those edges/nodes will not be involved in the diffusion process. The objective of our immunization problem is to find an allocation that controls the diffusion process most effectively.

For edge deletion, a good solution tends to give more vaccines to the edge groups where edges inside have high chance to be a part of cuts/walks. Similarly, for node deletion, we prefer the node groups where nodes insides have high impact on the influence/eigenvalue.

Main Idea of Our Problem Definitions. We give two different sets of problems which cover a wide range of contagion-like processes both threshold-based and cascade-style in the next two sections. In addition, all our problems have been carefully formulated to be seamless generalizations of the corresponding individual-level problems.

3.1 Problem Definition under LT Model

Our first set of problems are based on the LT model which is a well-known model for social media and complex propagations [10] suited for representing ‘threshold’ behaviors for activation. As mentioned in the introduction, the vaccination problem here can help to control such processes like spam and rumors on Twitter and Facebook. Under the LT model, our goal is to minimize the expected number of infected nodes at the end of diffusion, in other words, maximize the

expected number of nodes we can save from being infected, by selecting groups for removing edges/nodes.

In the LT model, a node v can be influenced by each neighbor u according to a weight p_{uv} where $\sum_{e(u,v) \in E} p_{uv} \leq 1$. The diffusion process proceeds as follows: at the start, every node u uniformly randomly chooses a threshold θ_u from the range $[0, 1]$, which represents the weighted fraction of u ’s neighbors that must be active to activate u ; an inactive node u becomes active at time $t + 1$ if $\sum_{w \in N_u^t} p_{wu} \geq \theta_u$ where N_u^t is the set of active neighbors of u at time t ; all active nodes will stay active. The process stops when no additional node becomes active. Each group may have some seeds (initial infected nodes). The seeds will spread information/virus by the LT model.

For the edge deletion under the LT model, let $\sigma_{C,A}(\mathbf{x})$ ($\mathbb{Z}^n \rightarrow \mathbb{R}$) denote the expected number of infected nodes in G (the footprint of G), given seed set A and vaccine allocation vector \mathbf{x} for the group set C . Now we are ready to define the edge version of the problem under the LT model.

PROBLEM 1: GROUP IMMUNIZATION under LT model (edge version):

GIVEN: Graph $G(V, E)$, a partition of the edge set $C = \{C_1, \dots, C_n\}$, seed set A and m vaccines (budget). Let \mathbf{x} be the edge vaccine allocation vector.

FIND: The optimum allocation \mathbf{x}_{opt} which maximizes $f(\mathbf{x}) = \sigma_{C,A}(\mathbf{0}) - \sigma_{C,A}(\mathbf{x})$ s.t. $|\mathbf{x}| \leq m$.

Next, we define the node version of this problem. Let $\sigma'_{C,A}(\mathbf{x})$ denote the footprint of G . It is same as $\sigma_{C,A}(\mathbf{x})$ except that the allocation vector \mathbf{x} corresponds to node vaccination.

PROBLEM 2: GROUP IMMUNIZATION under LT model (node version):

GIVEN: Graph $G(V, E)$, a partition of the vertex set $C = \{C_1, \dots, C_n\}$, seed set A and m vaccines (budget). Let \mathbf{x} be the node vaccine allocation vector.

FIND: The optimum allocation \mathbf{x}_{opt} which maximizes $f'(\mathbf{x}) = \sigma'_{C,A}(\mathbf{0}) - \sigma'_{C,A}(\mathbf{x})$ s.t. $|\mathbf{x}| \leq m$.

Hardness of Our Problems. Problems 1 and 2 are NP-hard as their special case, individual-level based immunizations (when each edge/node is a group), are NP-hard themselves [11], [33].

3.2 Problem Definition for Spectral Radius

Our second set of problems are based on the spectral radius formulation [7], [29] for a variety of cascade models including the fundamental SIR (‘mumps-like’ which generalizes the well-known IC model [10]), SIS (‘flu-like’), and SEIS (with incubation period) models. In the SIS/SIR models, every node can be either susceptible (S), infectious (I) or recovered (R). Each infected node u (in state I) can infect each susceptible neighbor v (in state S) with the probability p_{uv} . In the SIS model, each infected node u can switch to the susceptible state with the recovery rate δ . In the SIR model, each infected node u can switch to the recovered state with the recovery rate δ , meaning u cannot be infected again.

Spectral radius, denoted by λ , refers to the largest eigenvalue of the adjacency matrix of a graph G . Recent results [12], [26] have shown that λ is connected to the reproduction number in epidemiology, and determines the phase-transition (‘epidemic threshold’ τ) between epidemic/non-epidemic regimes in a very large range of cascade-style models [12], including SIR, SIS, SEIS and so on. As shown in [12], $\tau \propto \lambda$,

and if $\tau < 1$ the disease will die out quickly irrespective of initial conditions. This gives us the motivation to control the disease spread by minimizing λ in the underlying network.

Tong et al. [7], [29] proposed effective node-based and edge-based individual immunization methods to minimize λ . Following their methodology, in this paper we aim to maximize the drop of the spectral radius of G , $\Delta\lambda$, when vaccines are allocated to groups. Similar to Problems 1 and 2, when x_i vaccines are given to group C_i , we uniformly remove x_i nodes/edge at random. Hence, we want to find the optimal allocation \mathbf{x} such that the expectation of $\Delta\lambda$, $\mathbb{E}[\Delta\lambda](\mathbf{x})$ is maximum. Note that we do not define the problems here based on the ‘footprint’ (as in the previous section for LT) for primarily two reasons: (a) these versions naturally generalize the corresponding individual-level immunization problems studied in past literature [7], [29]; and (b) due to the epidemic threshold results, using the spectral radius allows us to immediately formulate a general problem for multiple cascade-style models (like SIR/SIS/IC) each with differences in their exact spreading process which we can ignore. Formally our problems are:

PROBLEM 3: GROUP IMMUNIZATION for spectral radius (edge version)

GIVEN: Graph $G(V, E)$, a partition of the edge set $C = \{C_1, \dots, C_n\}$, and m vaccines (budget). Let \mathbf{x} be the edge vaccine allocation vector, and let $\mathbb{E}[\Delta\lambda](\mathbf{x})$ denote the expected drop in the spectral radius after the immunization.

FIND: The optimum allocation \mathbf{x}_{opt} which maximizes $\mathbb{E}[\Delta\lambda]$, i.e., $\mathbf{x}_{\text{opt}} = \arg \max_{\mathbf{x}} \mathbb{E}[\Delta\lambda](\mathbf{x})$ s.t. $|\mathbf{x}| \leq m$.

PROBLEM 4: GROUP IMMUNIZATION for spectral radius (node version)

GIVEN: Graph $G(V, E)$, a partition of the node set $C = \{C_1, \dots, C_n\}$, and m vaccines (budget). Let \mathbf{x} be the edge vaccine allocation vector, and let $\mathbb{E}[\Delta\lambda](\mathbf{x})$ denote the expected drop in the spectral radius after the immunization.

FIND: The optimum allocation \mathbf{x}_{opt} which maximizes $\mathbb{E}[\Delta\lambda]$, i.e., $\mathbf{x}_{\text{opt}} = \arg \max_{\mathbf{x}} \mathbb{E}[\Delta\lambda](\mathbf{x})$ s.t. $|\mathbf{x}| \leq m$.

Hardness of Our Problems. Problems 3 and 4 are NP-hard too—their special cases, individual-level immunizations are NP-hard [7], [29].

4 PROPOSED METHODS

We first discuss our algorithms for the GROUP IMMUNIZATION problem under the LT model (Sections 4.1 and 4.2 for Problems 1 and 2), and then the spectral radius versions (Sections 4.3 and 4.4 for Problems 3 and 4).

4.1 Edge Deletion under LT Model

Recall that the function $f(\mathbf{x})$ in Problem 1 is not a simple set function; it is over an *integer lattice*. Hence the submodularity property used in [11] is not applicable to our problem, and we can not simply apply their greedy algorithm. Instead, our approach is to carefully identify a ‘submodularity like’ condition that is satisfied by our function $f(\mathbf{x})$, for which a greedy algorithm gives good performance. Let \mathbf{e}_k be the vector with 1 at the k th index and 0 be the all zeros vector. We consider the following three properties.

$$(P_1) \quad f(\mathbf{x}) \geq 0 \text{ and } f(\mathbf{0}) = 0.$$

$$(P_2) \quad (\text{Non-decreasing}) \quad f(\mathbf{x}) \leq f(\mathbf{x} + \mathbf{e}_k) \text{ for any } k.$$

$$(P_3) \quad (\text{Diminishing returns}) \quad \text{For any } \mathbf{x}' \geq \mathbf{x} \text{ and } k, \text{ we have } f(\mathbf{x} + \mathbf{e}_k) - f(\mathbf{x}) \geq f(\mathbf{x}' + \mathbf{e}_k) - f(\mathbf{x}').$$

The notion of submodularity of set functions has been extended to functions over integer lattices—see, e.g., [39], which shows that a greedy algorithm gives a constant factor approximation to submodular lattice functions with budget constraints. We note that in the context of functions defined on an integer lattice, unlike in the case of set functions, submodularity need not be equivalent to the diminishing return property. Besides, there are multiple non-equivalent definitions of the diminishing return property, as observed in [39]. Next, we show that in Theorem 1 that a greedy algorithm gives an $(1 - 1/e)$ -factor approximation to an integer lattice function satisfying the properties (P_1) , (P_2) and (P_3) above, and our objective function follows all above properties (Lemma 2). Note that it is not clear whether the analysis of [39] implies a similar bound for the kind of functions $f(\mathbf{x})$ we need to consider here.

Lemma 1. Suppose $\mathbf{y} = (y_1, \dots, y_n)^T$ where $y_i \in \mathbb{Z}^*$ and $\sum_j y_j = m$, then $f(\mathbf{x} + \mathbf{y}) - f(\mathbf{x}) \leq \sum_j y_j(f(\mathbf{x} + \mathbf{e}_j) - f(\mathbf{x}))$.

Proof. The proof is in the appendix.¹ \square

Theorem 1. Suppose $f(\mathbf{x})$, $\mathbf{x} \in \mathbb{Z}^n$ satisfies the properties (P_1) , (P_2) and (P_3) above. Then, Algorithm 1 gives a $(1 - 1/e)$ -approximate solution to the problem of maximizing $f(\mathbf{x})$ subject to $\sum_i x_i \leq m$.

Proof. Suppose \mathbf{x} is the solution from the greedy algorithm, and \mathbf{x}^* is the optimal solution. Hence, we have $\sum_j x_j = \sum_j x_j^* = m$. Since $\sigma_C(\mathbf{0})$ is constant, the greedy algorithm is equivalent to

$$C^* = \arg \max_{C_i} f(\mathbf{x} + \mathbf{e}_i) - f(\mathbf{x}).$$

Let us define $\mathbf{x}^{(i)}$ as the solution got from the i th iteration of the greedy algorithm, hence $\mathbf{x} = \mathbf{x}^{(m)}$. And \mathbf{x}^* can be represent as $\sum_j x_j^* \mathbf{e}_j$. We have

$$\begin{aligned} f(\mathbf{x}^*) &\leq f(\mathbf{x}^* + \mathbf{x}^{(i)}) \\ &= f(\mathbf{x}^{(i)}) + (f(\mathbf{x}^* + \mathbf{x}^{(i)}) - f(\mathbf{x}^{(i)})) \\ &\leq f(\mathbf{x}^{(i)}) + \sum_j x_j^*(f(\mathbf{x}^{(i)} + \mathbf{e}_j) - f(\mathbf{x}^{(i)})) \quad (\text{Lemma 1}) \\ &\leq f(\mathbf{x}^{(i)}) + \sum_j x_j^*(f(\mathbf{x}^{(i+1)}) - f(\mathbf{x}^{(i)})) \quad (\text{Greedy Alg.}) \\ &= f(\mathbf{x}^{(i)}) + m(f(\mathbf{x}^{(i+1)}) - f(\mathbf{x}^{(i)})). \end{aligned}$$

Hence, $f(\mathbf{x}^{(i+1)}) \geq (1 - \frac{1}{m})f(\mathbf{x}^{(i)}) + \frac{1}{m}f(\mathbf{x}^*)$. Recursively, we can get $f(\mathbf{x}^{(i)}) \geq (1 - (1 - \frac{1}{m})^i)f(\mathbf{x}^*)$. Therefore, $f(\mathbf{x}) = f(\mathbf{x}^{(m)}) \geq (1 - (1 - \frac{1}{m})^m)f(\mathbf{x}^*) \geq (1 - 1/e)f(\mathbf{x}^*)$. \square

Algorithm 1. Greedy Algorithm

Require: f , budget m

```

1:  $\mathbf{x} = \mathbf{0}$ 
2: for  $j = 1$  to  $m$  do
3:    $i = \arg \max_{k=1,\dots,n} f(\mathbf{x} + \mathbf{e}_k) - f(\mathbf{x})$ 
4:    $\mathbf{x} = \mathbf{x} + \mathbf{e}_i$ 
5: end for
6: return  $\mathbf{x}$ 

```

Now, we will show that the objective function $f(\mathbf{x}) = \sigma_{C,A}(\mathbf{0}) - \sigma_{C,A}(\mathbf{x})$ for the edge deletion problem under

¹ The proof is in the appendix, which can be found at: <http://people.cs.vt.edu/yaozhang/group-immu/>

the LT model satisfies the properties stated in Theorem 1. In the ensuing discussion, we will assume without loss of generality that there is only one seed node. This is because, if there are multiple seed nodes, then, we can merge all of them to a single ‘super’ node (say s) in the following manner: for every vertex $v \in V \setminus A$, set $p_{sv} = \sum_{u \in N(v) \cap A} p_{uv}$, where $N(v)$ is the set of neighbors of v . We note that after this modification the edges between v and its susceptible neighbors are unchanged, and at time 0, $\sum_{w \in N_v} p_{vw} = \sum_{w \in N(v) \cap A} p_{vw} = p_{sv}$. Hence, $\sigma_{C,A}(\mathbf{x}) = \sigma_{C,s}(\mathbf{x})$. Henceforth, we will assume that there is only one seed node, and drop the subscript A from $\sigma_{C,A}(\mathbf{x})$, denoting it by $\sigma_C(\mathbf{x})$.

Lemma 2. *The function $f(\mathbf{x}) = \sigma_C(\mathbf{0}) - \sigma_C(\mathbf{x})$ satisfies the properties (P_1) , (P_2) and (P_3) above.*

Proof. Property 1 is trivially true because, when $\mathbf{x} = \mathbf{0}$, by definition, $f(\mathbf{0}) = 0$, and since vaccination does not increase the number of infections, $\sigma_C(\mathbf{x}) \leq \sigma_C(\mathbf{0})$. For the rest of the proof, since $\sigma_C(\mathbf{0})$ is a constant, we only need to analyze $\sigma_C(\mathbf{x})$. Note that for any $\mathbf{x}' \geq \mathbf{x}$, we can find a sequence of vectors $(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_l)$ for some l such that $\mathbf{x} = \mathbf{z}_1$, $\mathbf{x}' = \mathbf{z}_l$ and $\mathbf{z}_i = \mathbf{z}_{i-1} + \mathbf{e}_{k_{i-1}}$ for some index k_{i-1} . Therefore, it is enough to prove that Properties 1 and 2 hold for $\mathbf{x}' = \mathbf{x} + \mathbf{e}_j$ for some index j . Also, we can assume that $x_j < |C_j|$, for $j = 1, \dots, n$, for if this is not true for some j , then, it implies that all the edges in C_j will be vaccinated, and therefore, we can simply remove all C_j from the analysis and reduce the budget by x_j .

Let $\mathcal{R}(\mathbf{x}) \subseteq 2^V$ be the collection of sets R satisfying $|R \cap C_i| = x_i$. Following the equivalence between influence in the LT model and the directed percolation process [10], we have $\sigma_C(\mathbf{x}) = \sum_{\hat{G}} \Pr[\hat{G}] \sum_{R \in \mathcal{R}(\mathbf{x})} \Pr[R] \gamma_C(\hat{G}, R)$, where the first sum is over all possible live-edge subgraphs \hat{G} of G in the percolation process, $\Pr[R]$ is the probability when the set R is removed, and $\gamma_C(\hat{G}, R)$ is the expected number of infected nodes in \hat{G} at the end of the LT process after the set R is removed. This can be rewritten as $\sigma_C(\mathbf{x}) = \sum_{\hat{G}} \Pr[\hat{G}] \sigma_C(\hat{G}, \mathbf{x})$, where $\Pr[\hat{G}]$ is the probability of sampling \hat{G} , and $\sigma_C(\hat{G}, \mathbf{x}) = \sum_{R \in \mathcal{R}(\mathbf{x})} \Pr[R] \gamma_C(\hat{G}, R)$. Henceforth, we will abbreviate $\gamma_C(\hat{G}, R)$ as $\hat{\gamma}(R)$.

We will show that $\sigma_C(\hat{G}, \mathbf{x})$ is non-increasing, i.e., $\sigma_C(\hat{G}, \mathbf{x}) \geq \sigma_C(\hat{G}, \mathbf{x}')$ where $\mathbf{x}' = \mathbf{x} + \mathbf{e}_j$, thereby showing that $f(\mathbf{x})$ satisfies Property 2. Since the number of nodes reachable from the seed node with R removed is at least as many as those with $R \cup \{e\}$ removed, for any $e \in C_j \setminus R$, we have $\hat{\gamma}(R) \geq \hat{\gamma}(R \cup \{e\})$. Therefore,

$$\begin{aligned} \sigma_C(\hat{G}, \mathbf{x}') &= \sum_{R' \in \mathcal{R}(\mathbf{x}')} \Pr[R'] \hat{\gamma}(R') \\ &= \sum_{R \in \mathcal{R}(\mathbf{x})} \sum_{e \in C_j \setminus R} \frac{1}{|C_j| - x_j} \Pr[R] \hat{\gamma}(R \cup \{e\}) \\ &\leq \sum_{R \in \mathcal{R}(\mathbf{x})} \sum_{e \in C_j \setminus R} \frac{1}{|C_j| - x_j} \Pr[R] \hat{\gamma}(R) \\ &= \sum_{R \in \mathcal{R}(\mathbf{x})} \Pr[R] \hat{\gamma}(R) = \sigma_C(\hat{G}, \mathbf{x}). \end{aligned}$$

Finally, we will show that $\sigma_C(\hat{G}, \mathbf{x} + \mathbf{e}_k) - \sigma_C(\hat{G}, \mathbf{x}) \leq \sigma_C(\hat{G}, \mathbf{x}' + \mathbf{e}_k) - \sigma_C(\hat{G}, \mathbf{x}')$. From the above discussion,

this will imply that $f(\mathbf{x})$ satisfies Property 3. Suppose $\mathbf{x}' = \mathbf{x} + \mathbf{e}_j$, we have two cases to consider: (1). $\mathbf{e}_k = \mathbf{e}_j$; (2). $\mathbf{e}_k \neq \mathbf{e}_j$.

For $1 \leq i \leq n$, let $c_i = |C_i|$ and x_i denote the i th element in \mathbf{x} .

First, we consider case (1) ($\mathbf{e}_k = \mathbf{e}_j$). For $R \in \mathcal{R}(\mathbf{x})$, $\Pr[R] = \prod_i \frac{1}{\binom{c_i}{x_i}} = \rho \frac{1}{\binom{c_k}{x_k}}$, where, $\rho = \prod_{i \neq k} \frac{1}{\binom{c_i}{x_i}}$

$$\begin{aligned} \sigma_C(\hat{G}, \mathbf{x}) - \sigma_C(\hat{G}, \mathbf{x} + \mathbf{e}_k) &= \sum_{R \in \mathcal{R}(\mathbf{x})} \rho \frac{1}{\binom{c_k}{x_k}} \hat{\gamma}(R) - \sum_{R' \in \mathcal{R}(\mathbf{x}') \setminus R} \rho \frac{1}{\binom{c_k}{(x_k+1)}} \hat{\gamma}(R') \\ &= \rho \sum_{R \in \mathcal{R}(\mathbf{x})} \left[\frac{1}{\binom{c_k}{x_k}} \hat{\gamma}(R) - \frac{1}{x_k+1} \sum_{e \in C_k \setminus R} \frac{1}{\binom{c_k}{x_k+1}} \hat{\gamma}(R \cup \{e\}) \right]. \end{aligned}$$

The factor $\frac{1}{x_k+1}$ is due to the fact that $R \cup \{e\}$ comes up in (x_k+1) combinations involving R and e . This simplifies to

$$\begin{aligned} \sigma_C(\hat{G}, \mathbf{x}) - \sigma_C(\hat{G}, \mathbf{x} + \mathbf{e}_k) &= \frac{\rho x_k!(c_k - x_k - 1)!}{c_k!} \sum_{R \in \mathcal{R}(\mathbf{x})} \sum_{e \in C_k \setminus R} \hat{\gamma}(R) - \hat{\gamma}(R \cup \{e\}). \quad (1) \end{aligned}$$

Similarly, we have

$$\begin{aligned} \sigma_C(\hat{G}, \mathbf{x}') - \sigma_C(\hat{G}, \mathbf{x}' + \mathbf{e}_k) &= \frac{\rho(x_k+1)!(c_k - x_k - 2)!}{c_k!} \sum_{R' \in \mathcal{R}(\mathbf{x}') \setminus R} \sum_{e \in C_k \setminus R'} \hat{\gamma}(R') - \hat{\gamma}(R' \cup \{e\}) \\ &= \frac{\rho(x_k+1)!(c_k - x_k - 2)!}{c_k!} \sum_{R \in \mathcal{R}(\mathbf{x})} \frac{1}{(x_k+1)} \sum_{e' \in C_k \setminus R} \hat{\gamma}(R \cup \{e'\}) - \hat{\gamma}(R \cup \{e, e'\}). \end{aligned}$$

From [11, proof of Theorem 6], $\hat{\gamma}(R) - \hat{\gamma}(R \cup \{e\}) \geq \hat{\gamma}(R \cup \{e'\}) - \hat{\gamma}(R \cup \{e, e'\})$ (supermodularity). Therefore, $(c_k - x_k - 1) \sum_e [\hat{\gamma}(R) - \hat{\gamma}(R \cup \{e\})] \geq \sum_{e'} \sum_e [\hat{\gamma}(R \cup \{e'\}) - \hat{\gamma}(R \cup \{e, e'\})]$. Hence proved.

Now, we consider case (2). Let

$$\Pr[R] = \rho' \frac{1}{\binom{c_k}{x_k}} \frac{1}{\binom{c_j}{x_j}},$$

where $\rho' = \prod_{i \neq j, k} \frac{1}{\binom{c_i}{x_i}}$. We can get $\sigma_C(\hat{G}, \mathbf{x}) - \sigma_C(\hat{G}, \mathbf{x} + \mathbf{e}_k)$ from Eqn. (1). And

$$\begin{aligned} \sigma_C(\hat{G}, \mathbf{x}') - \sigma_C(\hat{G}, \mathbf{x}' + \mathbf{e}_k) &= \frac{\rho' x_k!(c_k - x_k - 1)!}{\binom{c_j}{(x_j+1)} c_k!} \sum_{R' \in \mathcal{R}(\mathbf{x}') \setminus R} \sum_{e \in C_k \setminus R'} [\hat{\gamma}(R') - \hat{\gamma}(R' \cup \{e\})] \\ &= \frac{\rho'(x_j+1)!(c_j - x_j - 1)!}{c_j!} \frac{x_k!(c_k - x_k - 1)!}{c_k!} \sum_R \frac{1}{x_j+1} \\ &\quad \sum_{e_j \in C_j \setminus R} \sum_{e \in C_k \setminus (R \cup \{e_j\})} [\hat{\gamma}(R \cup e_j) - \hat{\gamma}(R' \cup \{e, e_j\})]. \end{aligned}$$

Again from [11], $(c_j - x_j - 1) \sum_e [\hat{\gamma}(R) - \hat{\gamma}(R \cup \{e\})] \geq \sum_{e_j} \sum_e [\hat{\gamma}(R \cup \{e_j\}) - \hat{\gamma}(R \cup \{e_j, e\})]$. Hence proved. \square

Algorithm 1 provides a simple greedy algorithm. Here, to estimate $\sigma_C(\mathbf{x})$ when vaccines are uniformly at random allocated within groups, we apply the Sample Average Approximation (SAA) framework. Let $\mathcal{L} \subset \mathcal{R}(x)$, denote a sample set from the set of all possible allocations. $\sigma_C(\mathbf{x}) \approx \hat{\sigma}_C(\mathbf{x}) = \frac{1}{|\mathcal{L}|} \sum_{R \in \mathcal{L}} \gamma_C(R)$, Kempe et al. [10] show that $\gamma_C(R)$ can be estimated by sampling from the set of live-edge graphs. A live-edge graphs T is generated as follows: for each node $v \in V$, independently select at most one of its incoming edges with probability p_{uv} , and with probability $1 - \sum_{w:(u,v) \in E} p_{uw}$ no edge is selected. Let this sample set be denoted by \mathcal{M} . This approach takes $O(|\mathcal{M}||\mathcal{L}|(|E| + |V|))$ time to estimate $\sigma_C(\mathbf{x})$, and $O(mn|\mathcal{M}||\mathcal{L}|(|E| + |V|))$ for the full greedy algorithm, which is not practical for large networks. However, we can speed up this naive greedy algorithm.

Algorithm 2. GREEDY-LT

Require: Graph G , group set C , seed set A , and budget m

- 1: Merge seed set A to I
- 2: Sample live-edge graphs $\mathcal{M} = \{T_X^I, \dots, T_X^{I_{|\mathcal{M}|}}\}$
- 3: For each T_X^I , calculate $r(u, T_X^I)$ for all nodes (in parallel)
- 4: Set $\mathbf{x} = \mathbf{0}$
- 5: **for** $j = 1$ to m **do**
- 6: **for** each T_X^I and C_i **do**
- 7: pick an edge $e_X^{C_i}$ at random for C_i and T_X^I
- 8: **end for**
- 9: $C^* = \arg \max_{C_i} \sum_{e_X^{C_i} \in T_X^I} (r(I, T_X^I) - r(I, T_X^I \setminus e_X^{C_i}))$
- 10: $x_{C^*} = x_{C^*} + 1$
- 11: **for** each T_X^I **do**
- 12: If $e_X^{C^*}(u, v) \in T_X^I$, remove edge $e_X^{C_i}$ and update $r(n, T_X^I)$ for node n (in parallel)
- 13: **end for**
- 14: **end for**
- 15: **return** \mathbf{x}

Speed-Up of the Greedy Algorithm: GREEDY-LT. Since a live-graph sampled from \mathcal{M} is a tree, we can denote it as T_X^s where s is the root, and $r(u, T_X^s) = |\{v|v \in \text{subtree}(u)\}|$, i.e., the number of nodes that are under the subtree of u in T_X^s . GREEDY-LT is summarized in Algorithm 2. It first merges all seeds into a ‘supernode’ s and samples $|\mathcal{M}|$ live-edge graphs, and then compute $r(u, T_X^s)$ in parallel for all nodes in all the live graphs (Lines 1-3). After that we greedily select m vaccines (Lines 4-10): we initially set the allocation vector $\mathbf{x} = \mathbf{0}$, and in each iteration, for each group C_i , we calculate the marginal loss $\Delta_{C_i,s}(\mathbf{x} + \mathbf{e}_i) = \sum_{e \in T_X^s} r(s, T_X^s) - r(s, T_X^s \setminus e)$, i.e., we randomly pick one edge from each group for each live-edge graph, then sum their marginal losses up over T_X^s as C_i ’s marginal loss. Note that $r(s, T_X^s) - r(s, T_X^s \setminus e) = r(v, T_X^s) + 1$, where node v is the endpoint of e [11]. We pick the group C^* with the maximum marginal loss. Finally we removed the edge that has been picked, and update $r(u, T_X^s)$ in parallel (Lines 11-13). There are two cases to update T_X^s if $e(u, v) \in T_X^s$: (1) for v ’s children, we can remove them because it is not reachable from s ; (2) for any ancestor a of v , $r(a, T_X^s \setminus e) = r(a, T_X^s) - r(v, T_X^s) - 1$, which can be done in constant time. Following Theorem 1, GREEDY-LT is a $(1 - 1/e - \epsilon)$ -approximation algorithm where ϵ is the approximation factor for estimating $\sigma_C(\mathbf{x})$.

Running Time of GREEDY-LT. Calculating all $r(u, T_X^I)$ costs $O(|\mathcal{M}||V|)$ time since we can traverse T_X^I once to get all values of $r(u, T_X^I)$. And greedily choosing m vaccine allocation needs $O(mn|\mathcal{M}||V|)$. Hence, the serial version of GREEDY-LT costs $O(mn|\mathcal{M}||V|)$. Note that in practice, we can speed it up by computing and updating $r_i(u, T_X^I)$ in parallel. In addition, since T_X^I is tree, the increasing difference property still holds, hence we can accelerate GREEDY-LT by “lazy evaluation” [36], [40] as well.

4.2 Node Deletion under LT Model

Our algorithm for the node version of the GROUP IMMUNIZATION problem is also the greedy Algorithm 1, as in the edge version in Section 4.1. Without loss of generality, we also assume that all seed nodes in A are merged, and drop the subscript A from $\sigma'_{C,A}(\mathbf{x})$, denoting it by $\sigma'_C(\mathbf{x})$. Our analysis relies on proving that the function $f'(\mathbf{x}) = \sigma'_C(\mathbf{0}) - \sigma'_C(\mathbf{x})$ in Problem 2 satisfies the properties (P_1) , (P_2) and (P_3) from Section 4.1, as discussed below.

Lemma 3. *The function $f'(\mathbf{x}) = \sigma'_C(\mathbf{0}) - \sigma'_C(\mathbf{x})$ satisfies the properties (P_1) , (P_2) and (P_3) .*

Proof. The proof is in the appendix. \square

Lemma 3 suggests that Theorem 1 holds for node version as well: GREEDY algorithm will provide a $(1 - 1/e)$ -approximate solution. We extend GREEDY-LT (Algorithm 2) to the node version: instead of randomly pick edges (Line 7), we randomly pick nodes to calculate the marginal loss (Line 9), and remove the corresponding nodes (Line 12). The observation is that calculating the marginal loss of removing node v in C in constant time holds here as well, i.e., $r(I, T_X^I) - r(I, T_X^I \setminus v) = r(v, T_X^I) + 1$. Hence, the updating process is the same as the edge version of GREEDY-LT.

4.3 Edge Deletion for Spectral Radius

We propose three algorithms for Problem 3 (edge immunization based on spectral radius) with different trade-offs of quality and running time: the first one, SDP, is a constant factor approximation algorithm that minimizes the actual eigendrop; the second algorithm, GROUPGREEDYWALK, is a bicriteria approximation algorithm based on hitting-walks; the third algorithm, LP, is an Linear Programming (LP) based method which uses an estimation of the eigendrop.

SDP is a constant-factor approximation algorithm, which gives us good results, but it is very slow with a $O(|V|^4 \text{polylog}(|V|))$ time complexity. Hence, we develop GROUPGREEDYWALK, a bicriteria approximation algorithm based on hitting closed walks [13]. Though GROUPGREEDYWALK loses a little quality compared to SDP, it is faster with a $O(sm^2|V|^3)$ time complexity (where s corresponds to the number of samples, described later). However, it may still not be scalable to very large networks with millions of nodes. Therefore, we come up with LP, a linear programming based heuristic whose time complexity depends only on the number of groups, not graph size. In reality, the number of groups in a group is typically much smaller than the number of nodes. Hence, LP is much faster than SDP and GROUPGREEDYWALK.

And experimental results demonstrate that it is scalable to networks with millions of nodes, and provides competitive empirical performance (see Section 5).

Note that even though SDP and GROUPGREEDYWALK may not be scalable to very large networks, both have proven performance guarantee. In addition, they are not merely of theoretical interest: they can be used as a baseline to assess the performance of faster heuristics on smaller networks.

Next, we will introduce the SDP algorithm (Section 4.3.1), the GROUPGREEDYWALK algorithm (Section 4.3.2), and the LP heuristic (Section 4.3.3) respectively.

4.3.1 SDP: A Constant Factor Approximation Algorithm

Let $G(V, E)$ be a graph whose edge set is partitioned into n groups C_1, \dots, C_n . Let \mathbf{x} be the edge allocation vector. For an edge (u, v) , let $g(u, v)$ denote the index of the group to which (u, v) belongs. Let $G(\mathbf{x})$ be the random graph obtained by removing each edge in C_i with probability $p_i = x_i/c_i$, where $c_i = |C_i|$. Let $\mathbf{M}(\mathbf{x})$ be its adjacency matrix and $\lambda_{\mathbb{E}}(\mathbf{x}) = \mathbb{E}[\lambda(\mathbf{M}(\mathbf{x}))]$ be the expected spectral radius

$$(\mathbf{M}(\mathbf{x}))_{uv} = \begin{cases} 1, & \text{with prob. } (1 - p_{g(u,v)}) \text{ if } (u, v) \in E(G), \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Let $\mathbf{M}_{\mathbb{E}}(\mathbf{x}) = \mathbb{E}[\mathbf{M}(\mathbf{x})]$ be the expectation of the adjacency matrix of $G(\mathbf{x})$

$$(\mathbf{M}_{\mathbb{E}}(\mathbf{x}))_{uv} = \begin{cases} 1 - p_{g(u,v)}, & \text{if } (u, v) \in E(G), \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

The problem is to find the optimal allocation, i.e., the \mathbf{x} for which $\lambda_{\mathbb{E}}(\mathbf{x})$ is minimized. We will denote this value by $\lambda_{\mathbb{E}}^{\min} := \min_{\mathbf{x}} \lambda_{\mathbb{E}}(\mathbf{x})$.

Remark 4.1. In the SDP formulation, for ease of analysis, we replace the hard budget constraint by an expected budget constraint, i.e., the expected size of the vaccine allocation vector \mathbf{x} is m . This is not a problem since, in reality, the budget is sufficiently high ($\gg \log n$). Hence, with high probability, the number of vaccines in the solution will be very close to the expected budget. Given this small difference, we can force the number of vaccines to be within the budget constraints, with very little effect on the performance.

The SDP Formulation: Finding the Allocation \mathbf{x} with Minimum $\lambda(\mathbf{M}_{\mathbb{E}}(\mathbf{x}))$. Note that, $\mathbf{M}_{\mathbb{E}}(\mathbf{x})_{uv} = (1 - p_{g(u,v)})$, if $(u, v) \in E(G)$. We use a simple SDP to find the allocation which minimizes $\lambda(\mathbf{M}_{\mathbb{E}}(\mathbf{x}))$ and meets the budget constraint m

$$\begin{aligned} & \text{minimize} && t \\ & \text{subject to} && 0 \leq p_i \leq 1, \text{ for } i = 1, \dots, n \\ & && \sum_i p_i |C_i| \leq m, \\ & && tI - \mathbf{M}_{\mathbb{E}}(\mathbf{x}) \succeq 0. \end{aligned} \quad (4)$$

Let \mathbf{x}_{\min} denote the allocation vector corresponding to the solution of the SDP.

Analysis: Relating $\lambda_{\mathbb{E}}^{\min}$ to $\lambda(\mathbf{M}_{\mathbb{E}}(\mathbf{x}_{\min}))$. One can use the following result by Lu and Peng [41] to bound $\lambda_{\mathbb{E}}(\mathbf{x})$ with respect to $\lambda(\mathbf{M}_{\mathbb{E}}(\mathbf{x}))$.

Theorem 2 ([41]). Consider an edge-independent random graph H . Let $\mathbf{M}(H)$ denote its adjacency matrix and $\mathbf{M}_{\mathbb{E}}(H) = \mathbb{E}[\mathbf{M}(H)]$. $\Delta_{\mathbb{E}}(H)$ denotes the maximum expected

degree. If $\Delta_{\mathbb{E}}(H) \gg \log^4 |V|$, then, almost surely $\lambda_i(\mathbf{M}(H)) - \lambda_i(\mathbf{M}_{\mathbb{E}}(H)) \leq (2 + o(1))\sqrt{\Delta_{\mathbb{E}}(H)}$, for $i = 1, \dots, |V|$.

Recall that \mathbf{x}_{\min} is the output of SDP (4), and it corresponds to the allocation vector which minimizes $\lambda(\mathbf{M}_{\mathbb{E}}(\mathbf{x}))$ over all \mathbf{x} . Let $\Delta_{\mathbb{E}}(\mathbf{x}_{\min})$ denote the maximum expected degree of $G(\mathbf{x}_{\min})$. The following lemma proves that the SDP formulation gives us an approximation algorithm with constant factor $O(\sqrt{\Delta_{\mathbb{E}}(\mathbf{x}_{\min})})$.

Lemma 4. If \mathbf{x}_{\min} is such that $\Delta_{\mathbb{E}}(\mathbf{x}_{\min}) \gg \log^4 |V|$, then, $\lambda_{\mathbb{E}}^{\min} \leq \lambda(\mathbf{M}_{\mathbb{E}}(\mathbf{x}_{\min})) + (2 + o(1))\sqrt{\Delta_{\mathbb{E}}(\mathbf{x}_{\min})} + 1$.

Proof. Let $z = \lambda(\mathbf{M}_{\mathbb{E}}(\mathbf{x}_{\min})) + (2 + o(1))\sqrt{\Delta_{\mathbb{E}}(\mathbf{x}_{\min})}$. Applying Theorem 2 to $G(\mathbf{x}_{\min})$, $\lambda(\mathbf{M}(\mathbf{x}_{\min})) \leq z$ almost surely. In fact, for $\Delta_{\mathbb{E}}(\mathbf{x}_{\min}) \gg \log^4 |V|$, it can be shown that $\Pr(\lambda(\mathbf{M}(\mathbf{x}_{\min})) \geq z) \leq 1/|V|$ (see [41, proof of Theorem 6]). Noting that $\lambda(\mathbf{M}(\mathbf{x}_{\min})) \leq \lambda(\mathbf{M})$,

$$\begin{aligned} \lambda_{\mathbb{E}}(\mathbf{x}_{\min}) &= \mathbb{E}[\lambda(\mathbf{M}(\mathbf{x}_{\min}))] \\ &\leq \Pr(\lambda(\mathbf{M}(\mathbf{x}_{\min})) \leq z) \cdot z \\ &\quad + \Pr(\lambda(\mathbf{M}(\mathbf{x}_{\min})) \geq z) \cdot \lambda(\mathbf{M}) \\ &\leq 1 \cdot z + \left(\frac{1}{|V|}\right) \cdot \lambda(\mathbf{M}) < z + 1. \end{aligned}$$

By definition, $\lambda_{\mathbb{E}}^{\min} \leq \lambda_{\mathbb{E}}(\mathbf{x}_{\min})$. Therefore, $\lambda_{\mathbb{E}}^{\min} \leq \lambda_{\mathbb{E}}(\mathbf{x}_{\min}) \leq z + 1$. Hence, proved. \square

Running Time. The SDP step (Eq. (4)) dominates the running time of this algorithm, which is $O(|V|^4 \text{polylog}(|V|))$.

4.3.2 GROUPGREEDYWALK: A Bicriteria Approximation Algorithm

As shown above, SDP with a $(|V|^4 \text{polylog}(|V|))$ time complexity, is too slow for large networks. In this section, we leverage the technique of hitting closed walks [13] for the GROUP IMMUNIZATION problem, and propose a bicriteria approximation algorithm called GROUPGREEDYWALK.

Saha et al. [13] studied the problem of minimizing the spectral radius under a given threshold by removing the smallest number of edges, and developed a greedy based approximation algorithm for it. Different from their work, our goal is to distribute a given budget of vaccines to groups to minimize the spectral radius as small as possible. We can adapt their greedy algorithm to the group immunization, by choosing groups with maximum marginal gain of hitting closed walks. However, it is not clear whether this works, as we need to consider all “possible worlds” for group immunization.

In graph G , a closed walk is a sequence of nodes starting and ending at the same node, with two consecutive nodes adjacent to each other. Closed k -walk is a walk with length k . Let $walks(e, G, k)$ denote the number of closed k -walks in G containing $e = (i, j)$. We say that an edge set E hits a walk w if w contains an edge from E . Recall that $G(\mathbf{x})$ is a random graph obtained by removing a random subset of x_i edges in C_i , where C_1, \dots, C_n is a partition of the edge set E . Let $\mathcal{W}(G, k)$ be the set of all walks of length k in the graph G . Let $n_k(G, e)$ denote the number of walks of length k in G that pass through edge e . Similarly, let $n_k(G, S)$ denote the walks of length k in G that pass through edges in the set S . Let $n_k(G) = n_k(G, E) = |\mathcal{W}(G, k)|$ denote

the number of walks with length k in G . Here we focus on walks of a fixed length $k = \theta(\log |V|)$. Note that for $G(\mathbf{x})$, $n_k(G(\mathbf{x}))$ is a random variable.

Algorithm 3. GROUPGREEDYWALK (G, m)

Require: Graph G , group set C , and budget m

```

1:  $\mathbf{x} = 0$ 
2: for  $j = 1$  to  $m$  do
3:    $i = \arg \max_{k=1,\dots,n} \text{EXPCOUNTWALKS}(\mathbf{x} + \mathbf{e}_k) - \text{EXPCOUNTWALKS}(\mathbf{x})$ 
4:    $\mathbf{x} = \mathbf{x} + \mathbf{e}_i$ 
5: end for
6: return  $\mathbf{x}$ 

```

Algorithm 3 gives the pseudocode of our GROUPGREEDYWALK algorithm. EXPCOUNTWALKS(G, \mathbf{x}) returns the expected number of walks surviving in $G(\mathbf{x})$. Note that EXPCOUNTWALKS is different from COUNTWALKS in [13], as it returns the expected number of hitting walks when a vaccine allocation vector \mathbf{x} is assigned to groups, while COUNTWALKS in [13] is based on removing a set of edges given a budget constraint. The idea of GROUPGREEDYWALK is that, each time we select a group C_i with the maximum marginal gain in EXPCOUNTWALKS(G, \mathbf{x}), when allocating one vaccine to C_i .

Algorithm 3 follows the framework of the individual based GREEDYWALK algorithm [13]. Instead of picking edges, it chooses groups to maximize marginal gain of eigendrop. The main challenge here is to show GROUPGREEDYWALK is a provable approximation algorithm. Let $\mathbf{x}^{opt}(m)$ be the optimum solution corresponding to budget m , and $T = \lambda_1(G(\mathbf{x}^{opt}(m)))$ (the spectral radius after vaccine allocation for the optimum solution). We can prove the following theorem:

Theorem 3. Let $\mathbf{x}^{opt}(m)$ be the optimum solution corresponding to budget m of edges removed. Let \mathbf{x}^g be the allocation returned by GROUPGREEDYWALK ($G, c_1 m \log^2 |V|$), for a constant c_1 . Then we have $\lambda_1(G(\mathbf{x}^g)) \leq c' T$ for a constant c' , where $\lambda_1(G(\mathbf{x}^g))$ is the spectral radius after allocating vaccines based on \mathbf{x}^g .

Remark 4.2. Theorem 3 shows that GROUPGREEDYWALK is a $(c_1 \log^2 |V|, c')$ -bicriteria approximation algorithm. Different from the analysis of traditional approximation algorithms, in order to bound the result of GROUPGREEDYWALK w.r.t to the optimal solution, we need a larger budget $c_1 \log^2 |V| m$. Typically, $\log^2 |V|$ is much smaller than the budget m . And when the budget m is very large, the marginal gain of eigendrop for a larger budget $c_1 \log^2 |V| m$ will tend to be very close to the marginal gain of eigendrop for the budget m . Hence, adding such small factor into the budget m will have little effect on the performance.

We will use Lemmas 5 and 6 to prove this theorem. Intuitively, Lemma 5 shows the expected spectral radius is upperbounded by T if the number of walk $k = O(\log |V|)$; while Lemma 6 shows that the expected number of walks with length k can be upperbounded by T as well.

Lemma 5. If $E[n_k(G(\mathbf{x}))] = O(|V|2^k T^k)$ for $k = O(\log |V|)$, then $E[\lambda_1(G(\mathbf{x}))] \leq c_3 T$ for a constant c_3 .

Proof. The proof is in the appendix. \square

Lemma 6. Let $\mathbf{x}^{opt}(m)$ be the optimum allocation such that $T = E[\lambda_1(G(\mathbf{x}^{opt}(m)))]$. Let \mathbf{y} be defined as

$$y_i = \begin{cases} x_i^{opt}, & \text{if } x_i^{opt} \leq m_i/2, \\ m_i & \text{otherwise,} \end{cases}$$

where m_i is the number of edges in group C_i . Then, we have $E[n_k(G(\mathbf{y}))] \leq |V|2^k T^k$.

Proof. The proof is in the appendix. \square

Now, we prove Theorem 3.

Proof of Theorem 3. Let $g(\mathbf{x})$ denote the expected number of walks in $\mathcal{W}(G, k)$ hit by the edges that are removed in $G(\mathbf{x})$. Then $g(\mathbf{x})$ has the diminishing returns property, i.e., for $\mathbf{x} \leq \mathbf{x}'$, we have $g(\mathbf{x} + \mathbf{e}_i) - g(\mathbf{x}) \geq g(\mathbf{x}' + \mathbf{e}_i) - g(\mathbf{x}')$. The proof of the diminishing returns follows the proof of Lemma 2.

We will compare $g(\mathbf{x}^g)$ to $g(\mathbf{y})$ where \mathbf{y} is defined as

$$y_i = \begin{cases} x_i^{opt}, & \text{if } x_i^{opt} \leq m_i/2, \\ m_i & \text{otherwise,} \end{cases}$$

where m_i is the number of edges in group C_i . Note that $\sum_i y_i \leq 2 \sum_i x_i^{opt} \leq 2m$.

Let $\mathbf{x}^{(i)}$ denote the vector after i th iteration of GROUPGREEDYWALK. Since $g(\mathbf{x})$ has the diminishing returns property, it follows the proof of Theorem 1 that $f(\mathbf{x}^{(i)}) \geq (1 - (1 - \frac{1}{2m})^i) f(\mathbf{y})$. Therefore, for $i = O(m \log^2 |V|)$, we have $1 - (1 - \frac{1}{2m})^{O(m \log^2 |V|)} \geq 1 - (1/e)^{\log^2 |V|} \geq 1 - \frac{1}{|V|^{\log |V|}} \geq 1 - \frac{1}{N}$, where N is the number of total walks in the original graph G .

From Lemma 6, we have $E[n_k(G(\mathbf{y}))] \leq |V|^c T^k$ for a constant c . This implies $f(\mathbf{y}) \geq N - |V|^c T^k$. Therefore, $f(\mathbf{x}^g) \geq (1 - \frac{1}{|V|})(N - |V|^c T^k) \geq N - 1 - |V|^c T^k$. This implies that $n_k(G(\mathbf{x}^g)) \leq O(|V|^c T^k)$. From Lemma 5, it follows that $\lambda_1(G(\mathbf{x}^g)) \leq c' T$. \square

Implementation Notes. Given the adjacency matrix A of G , the number of k -length walks from u to v is given by A_{uv}^{k-1} . It also corresponds to the number of walks hit by the edge (u, v) . We implement the algorithm as follows. In each iteration, we randomly sample a set of edges of the G according to \mathbf{x} . For each sample, we compute the expected decrease in the number of walks for the removal of one edge in group i (for computing the effect of allocation vector $\mathbf{x} + \mathbf{e}_i$) as follows: We construct $G(\mathbf{x})$, compute $A' = A(G(\mathbf{x}))^{k-1}$ and take the average over all $A'(u, v)$ elements where (u, v) belongs to group i . We perform this for each sample (number of samples is s) and take the average over all the samples. Finally, we choose that i which gives the maximum average and update \mathbf{x} by adding \mathbf{e}_i to it.

Running Time. For budget m , A^{m-1} can be computed in time $O(m^2 |V|^3)$. For each sample of \mathbf{x} , we compute $A(G(\mathbf{x}))^{m-1}$. Note that, computing the effect of removing \mathbf{e}_i for each sample takes only $O(|V|^2)$ time. Therefore, for a sample size of s , the algorithm overall takes $O(sm^2 |V|^3)$ time. If $m = O(\log |V|)$, the time complexity is $O(s |V|^3 \log^2 |V|)$.

4.3.3 LP: A Fast Heuristic

GROUPGREEDYWALK is a good approximation algorithm like SDP, however, it may not be scalable to very large networks with millions of nodes. In this section, the propose a much faster heuristic based on estimating eigendrop.

The eigendrop when removing edges in the set E_T can be approximated by $\phi(T) = \sum_{(i,j) \in E_T} \mathbf{M}_{ij} u_i u_j$ where $\mathbf{Mu} = \lambda \mathbf{u}$ and $\mathbf{u} = (u_1, \dots, u_i, \dots)$ [29]. Given the allocation vector \mathbf{x} , the expected drop in spectral radius is then given by

$$\begin{aligned} \mathbb{E}[\Delta\lambda] &\approx \phi(\mathbf{x}) \\ &= \sum_{i,j \in E} \mathbf{M}_{ij} u_i u_j \Pr((i,j) \text{ is removed}) \\ &= \sum_{a \in C} \sum_{(i,j) \in C_k} \mathbf{M}_{ij} u_i u_j x_a. \end{aligned} \quad (5)$$

If we define $\alpha_a = \sum_{(i,j) \in C_a} \mathbf{M}_{ij} u_i u_j$, then, $\phi(\mathbf{x}) = \sum_a \alpha_a x_a$. We want to maximize $\phi(\mathbf{x})$ subject to the budget constraints. This can be formulated as a linear program as given below

$$\begin{aligned} \text{maximize } & \sum_a \alpha_a x_a \\ \text{subject to } & \sum_a x_a |C_a| \leq m \\ & 0 \leq x_a \leq 1. \end{aligned} \quad (6)$$

Running Time. The LP takes $O(n^4)$ time where n is the number of groups. Note that it is not a function of the graph size. Typically, the number of groups is small, hence this algorithm is very fast.

4.4 Node Deletion for Spectral Radius

Here, we propose an algorithm for solving Problem 4: the group node immunization problem with respect to eigendrop. It is based on the approximate eigendrop method which was discussed in Section 4.3. The eigendrop when removing nodes in S can be approximated as follows [7]:

$$\Delta\lambda \approx \phi(S) = \sum_{j \in S} 2\lambda u_j^2 - \sum_{i,j \in S} \mathbf{M}_{ij} u_i u_j, \quad (7)$$

where $\mathbf{Mu} = \lambda \mathbf{u}$ and $\mathbf{u} = (u_1, \dots, u_i, \dots)$. Recall that C is the set of groups and $\mathbf{x} = (x_1, \dots, x_i, \dots)$ is the allocation vector where, x_i is the fraction of nodes vaccinated in group C_i . For the group vaccination problem, the expected eigendrop can be approximated by applying (7) as follows:

$$\begin{aligned} \mathbb{E}[\Delta\lambda] &\approx \phi(\mathbf{x}) = \sum_{j \in V} 2\lambda u_j^2 \Pr(j \text{ is vaccinated}) \\ &\quad - \sum_{i,j \in V} \mathbf{M}_{ij} u_i u_j \Pr(i \& j \text{ are vaccinated}). \end{aligned} \quad (8)$$

Let $g(v)$ denote the index of the group to which v belongs to, i.e., if $v \in C_i$, then, $g(v) = i$. The probability that j is vaccinated is $x_{g(j)}$ and the probability that both i and j are vaccinated is

$$\Pr(i \& j \text{ are vaccinated}) = \begin{cases} x_{g(i)} x_{g(j)}, & \text{if } g(i) \neq g(j), \\ x_{g(i)}^2 \frac{|C_{g(i)}|}{|C_{g(i)}|-1}, & \text{otherwise.} \end{cases} \quad (9)$$

Applying the above to (8),

$$\begin{aligned} \phi(\mathbf{x}) &= \sum_a \sum_{j \in C_a} 2\lambda u_j^2 x_a - \sum_a \sum_{i,j \in C_a} \mathbf{M}_{ij} u_i u_j x_a^2 \frac{|C_a|}{|C_a|-1} \\ &\quad - \sum_{a \neq b} \sum_{i \in C_a, b \in C_b} \mathbf{M}_{ij} u_i u_j x_a x_b. \end{aligned}$$

Observing that \mathbf{M}_{ij} , u_i and x_a are constants, defining $\alpha_a = \sum_{j \in C_a} 2\lambda u_j^2 \frac{|C_a|}{|C_a|-1}$, $\beta_a = \sum_{i,j \in C_a} \mathbf{M}_{ij} u_i u_j$, and $\Gamma_{ab} = \sum_{i \in C_a, j \in C_b} \mathbf{M}_{ij} u_i u_j$, we get,

$$\phi(\mathbf{x}) = \sum_a \alpha_a x_a - \sum_a \beta_a x_a^2 - \sum_{a \neq b} \Gamma_{ab} x_a x_b.$$

Our aim is to find that \mathbf{x} which maximizes $\phi(\mathbf{x})$. This can be formulated as a quadratic program

$$\begin{aligned} \text{minimize } & \sum_a \beta_a x_a^2 + \sum_{a \neq b} \Gamma_{ab} x_a x_b - \sum_a \alpha_a x_a \\ & = \frac{1}{2} \mathbf{c}^T \mathbf{Q} \mathbf{x} + \mathbf{c}^T \mathbf{x} \\ \text{subject to } & \sum_a x_a |C_a| \leq B \\ & 0 \leq x_a \leq 1, \end{aligned} \quad (10)$$

where, $\mathbf{Q}_{aa} = 2\beta_a$ and for $a \neq b$, $\mathbf{Q}_{ab} = 2\Gamma_{ab}$ and $\mathbf{c}_a = -\alpha_a$. If \mathbf{Q} is not semi-definite, the problem is NP-Hard [42]. In that case, we use a low-rank matrix $\hat{\mathbf{Q}}$ formed by all its eigenvectors corresponding to non-negative eigenvalues. The QP on $\hat{\mathbf{Q}}$ can be solved in polynomial time using the ellipsoid method [42]. Let $\hat{\phi}(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \hat{\mathbf{Q}} \mathbf{x} + \mathbf{c}^T \mathbf{x}$, and \mathbf{x}_Q , $\hat{\mathbf{x}}_Q$ correspond to the best allocation vectors corresponding to \mathbf{Q} and $\hat{\mathbf{Q}}$ respectively. The next lemma shows that $\hat{\mathbf{x}}_Q$ is a good approximation to \mathbf{x}_Q .

Lemma 7. $|\hat{\phi}(\hat{\mathbf{x}}_Q) - \phi(\mathbf{x}_Q)| \leq \frac{n}{2} \cdot \|\mathbf{Q} - \hat{\mathbf{Q}}\|_F$, where n is the number of groups in the graph.

Proof. The proof is in the appendix. \square

Running Time. The QP takes $O(n^4)$ time. Again, note that n is the number of groups. Hence, it is fast when the number of groups is small.

5 EMPIRICAL STUDY

We present a detailed experimental evaluation now.

5.1 Experimental Setup

We implemented the algorithms in Python,² and conducted the experiments using a 4 Xeon E7-4850 CPU with 512 GB of 1,066 MHz main memory.

Datasets. Table 2 briefly summarizes the dataset. We run our experiments on multiple datasets, which were chosen for their size as well as different domains where the GROUP IMMUNIZATION problem is especially applicable. Note that all our datasets are networks, not diffusion traces. If diffusion traces are provided as inputs instead of a network, there are state-of-the-art algorithms (such as [43]) which can be applied to learn edge weights first, and then apply our algorithm.

2. Code: <http://people.cs.vt.edu/yaozhang/group-immu/>

TABLE 2
Datasets

Dataset	Num. of nodes	Num. of edges	Num. of groups
SBM	1,500	5,000	20
Protein	2,361	7,182	13
OregonAS	10,670	22,002	100
YouTube	50 K	450 K	5,000
Portland	0.5 million	1.6 million	91
Miami	0.6 million	2.1 million	91

- 1) SBM (Stochastic Block Model) [44] is a well-known model to generate synthetic graphs with groups. We generate small networks from the Stochastic Block Model to test the effectiveness of all our methods.
- 2) Protein³ is a protein-protein interaction network in budding yeast. There are 13 classes of proteins, which are naturally treated as groups. It is a biological network, where our immunization algorithms can be potentially applied to block protein interactions.
- 3) OregonAS⁴ is the Oregon AS router graph collected from the Oregon router views, and groups here are based on router conductivities. We use Louvain [45], a fast community detection algorithm to specify groups. It is a computer network where our algorithms can be used to stop malware outbreaks.
- 4) YouTube⁵ is a friendship network in which users can form groups. We create an induced graph by selecting nodes that are in the top 5,000 communities. It is a social media network where we can apply our algorithms to control rumor spread.
- 5) Portland and Miami are social-contact graphs based on detailed microscopic simulations of large US cities, which has been used in national smallpox and influenza modeling studies using the SIR model [2]. We divided people into groups by ages ranging from 0-90 (hence 91 groups in both networks). They are both contact networks where our algorithms can be adopted to minimize virus propagation.

Settings. For LT model, we uniformly randomly choose 1 percent nodes as the infected nodes (seeds) at the start. And we use the same method in [11] to generate the probabilities on the edges: for a node v , we assign each its incoming edge (u, v) with a probability \hat{p}_{uv} uniformly at random, then we uniformly randomly give a probability w_v to v representing v 's incoming edges fail to activate it. Then we get the normalized weight $p_{uv} = \hat{p}_{uv}/(\sum_{u \in V} \hat{p}_{uv} + w_v)$. We construct 1,000 live-edge graphs in our algorithm for LT model. For robustness, each data point we show is the mean of 1,000 runs of randomly sampling removed edges/nodes from groups. In the edge deletion version, edge communities are induced from node communities, i.e., for an edge $e = (u, v)$, if both u and v belong to a group C_t , then $e \in C_t$, otherwise it belongs to group $C_{ij} = \{e_{uv} | u \in C_i, v \in C_j\}$.

Baselines. As we are not aware of any direct competitor tackling our group immunization problems, we construct

three baselines for both node and edge deletion to better judge their performance. Analogous versions of these baselines have been regularly used in state-of-the-art individual immunization studies [7], [13], [29].

- (1) RANDOM: uniformly randomly assign vaccines to groups for both node deletion and edge deletion.
- (2) DEGREE: for node deletion, we calculate the average degree d_{C_i} of each group C_i , and independently assign vaccines to C_i with probability $d_{C_i} / \sum_{C_k \in C} d_{C_k}$; for edge deletion, we first calculate the product degree d_e [30] of each edge $e = (u, v)$, i.e., $d_e = d_u * d_v$, then similar to node deletion, we calculate the average product degree d_{C_i} of C_i , and assign vaccines to C_i with probability $d_{C_i} / \sum_{C_k \in C} d_{C_k}$.
- (3) EIGEN: Eigenvalue centrality has been widely used in the immunization literature [7], [29], even as a baseline for LT model [11]. Let \mathbf{u} be the eigenvector corresponding to the first eigenvalue of the graph. The eigenscore of node a is u_a , while the eigenscore of edge $e(a, b)$ is $|u_a u_b|$ [29]. For both node and edge deletion, we calculate the average eigenscore u_{C_i} of each group C_i , and independently assign vaccines to C_i with probability $u_{C_i} / \sum_{C_k \in C} u_{C_k}$.

Remark 5.1. Note that we do not compare and run the individual based immunization methods [7], [11] “as-is” on the original graph because these methods directly pick nodes which we do not allow in our problems. Instead, we aim to pick the best groups, and then uniformly at random allocate vaccines within the group. In addition, we did study the effect of our algorithm w.r.t. the size of groups (see Fig. 6). If each node is a group, GROUP IMMUNIZATION reduces to the individual based immunization. Indeed the reason we formulate the group immunization problems in this paper is that it is typically not feasible to force targeted individuals to be vaccinated in practice (as discussed before in the introduction).

5.2 Results

In short, we demonstrate that our methods outperform other baselines on all datasets. We also show how the behaviors of our methods change as groups vary. Finally, we conduct a case study to analyze the vaccine allocations at group scale.

5.2.1 Performance

Fig. 1 shows experimental results under LT model for group edge deletion, while Fig. 2 demonstrates the results for node deletion. In all networks, GREEDY-LT consistently outperform other competitors. Since we have same budgets for both edge and node deletion, clearly node removal should perform better than edge deletion as node deletion removes more edges. Our results demonstrate this fact. As shown in Fig. 1, GREEDY-LT performs pretty well for edge deletion compared with other competitors, e.g., in YouTube, GREEDY-LT can reduce about 25 percent of the infection if 500 edges are removed, while for RANDOM, DEGREE and EIGEN, the infection almost remains the same even removing 500 edges. For node deletion (Fig. 2), GREEDY-LT performs even better: it reduces more than 30 percent of the infection given the maximum budgets.

3. <http://vlado.fmf.uni-lj.si/pub/networks/data/bio/Yeast/Yeast.htm>

4. <http://snap.stanford.edu/data/oregon1.html>

5. <http://snap.stanford.edu/data/com-Youtube.html>

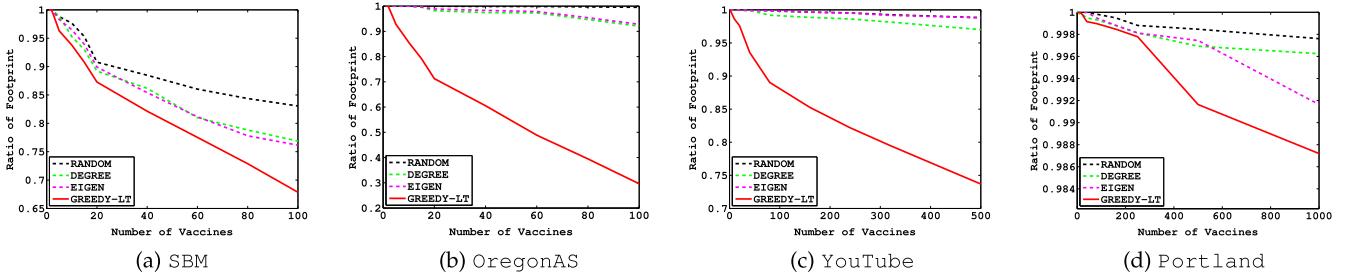


Fig. 1. Effectiveness for LT model various Real Datasets (edge deletion). Footprint ratio ($\frac{\text{footprint when vaccines are given}}{\text{footprint without giving vaccines}}$) versus number of vaccines. Lower is better. GREEDY-LT consistently outperforms other baseline algorithms.

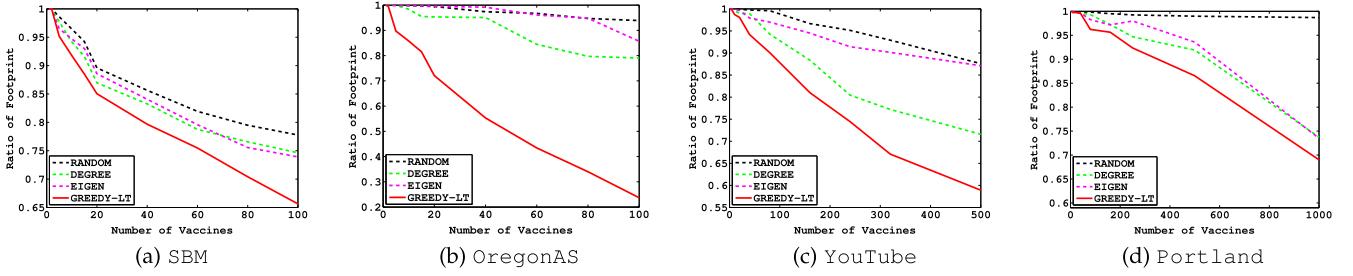


Fig. 2. Effectiveness for LT model various Real Datasets (node deletion). Footprint ratio ($\frac{\text{footprint when vaccines are given}}{\text{footprint without giving vaccines}}$) versus number of vaccines. Lower is better. GREEDY-LT consistently outperforms other baseline algorithms.

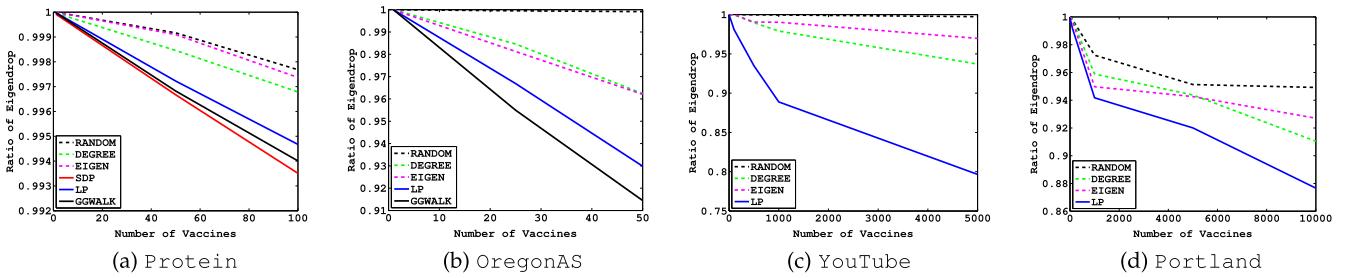


Fig. 3. Effectiveness for the change of the first eigenvalue various Real Datasets (edge deletion). Eigendrop ratio ($\frac{\lambda'_G}{\lambda_G}$) versus number of vaccines (λ'_G is the expected eigenvalue after allocating vaccines). Lower is better. SDP, GROUPGREEDYWALK, and LP consistently outperform other baseline algorithms.

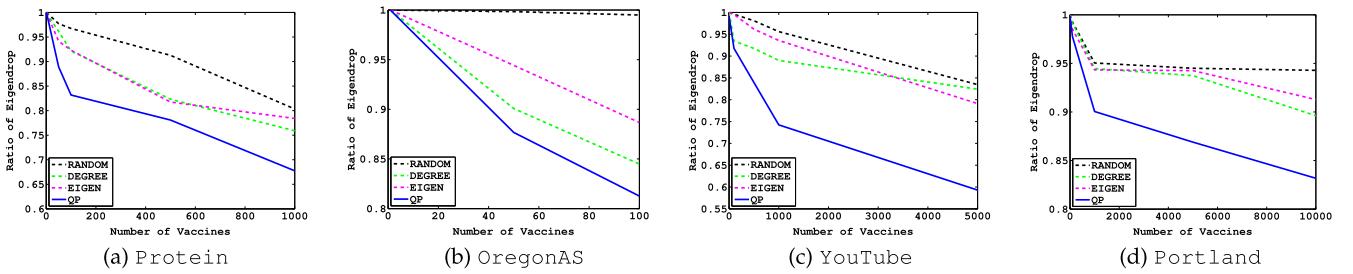


Fig. 4. Effectiveness for the change of the first eigenvalue various Real Datasets (node deletion). Eigendrop ratio ($\frac{\lambda'_G}{\lambda_G}$) versus number of vaccines (λ'_G is the expected eigenvalue after allocating vaccines). Lower is better. QP consistently outperforms other baseline algorithms.

Fig. 3 shows experimental results of edge version of group immunization for spectral radius, while Fig. 4 demonstrates the results for node deletion. In all networks, SDP, GROUPGREEDYWALK, LP and QP consistently outperform other competitors. SDP gives the best results for Protein, however, it is not scalable to large networks with more than thousands of nodes. GROUPGREEDYWALK gives the second best performance, and it works for graphs with about 10 K nodes. For very large networks like YouTube and Portland with millions of nodes, approximate algorithms like SDP and GROUPGREEDYWALK can not finish within an allocated time. LP for edge deletion and QP for node deletion, perform very well for large networks.

For edge deletion (Fig. 3), RANDOM, DEGREE and EIGEN cannot decrease more than 10 percent of the first eigenvalue in YouTube when 5 k vaccines are given to groups, while LP can reduce more than 20 percent of the eigenvalue. For node deletion (Fig. 4), QP can get more than twice reduction of eigenvalue compared to other competitors. When comparing between node and edge deletion, we get the same result as Figs. 1 and 2: given same vaccines to both edge and node, node removal can get a larger decrease of the spectral radius.

As mentioned above, the problems of minimizing the spectral radius are motivated by the epidemic threshold [12]; an epidemic will quickly die out if the spectral radius is very

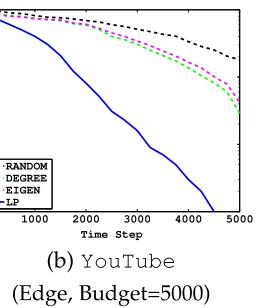
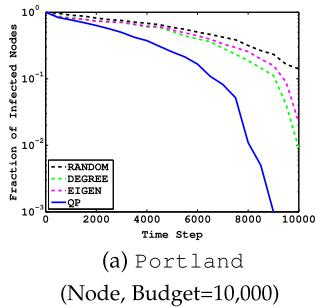


Fig. 5. SIS simulations after vaccine allocation. The fraction of infected nodes (in log-scale) versus the time step. Lower is better. SDP, GROUP-GREEDYWALK, and LP consistently outperform other baseline algorithms.

small. Hence as an example, we also run the SIS model to show how effective our algorithms are to prevent an epidemic from breaking out. We assume all nodes are in the infectious states at the beginning, and the recovery rate is 0.6. Fig. 5 shows the results on Portland and YouTube for node deletion and edge deletion respectively, which is averaged over 1,000 runs (note that we got the similar results on other networks). We observe that LP and QP consistently outperform other competitors: they both have the least number of infected nodes in the network when vaccines are allocated.

5.2.2 Varying Groups

We would like to see the effect of the change of granularity of vaccine allocation. We changed the number of groups on Portland, YouTube and OregonAS. For Portland, age ranges from 0 to 90, hence there are initially 91 groups. We decrease the number of groups by randomly merging two adjacency age groups. For OregonAS, we use community detection algorithm Louvain [45] to find different number of groups. For YouTube, we randomly merge ground true communities to form smaller size of groups.

Figs. 6a and 6b show the performance of QP and LP as the number of groups changes. First, both of them outperform other baselines for Portland and YouTube. Second, as the number of groups increases, the spectral radius decreases more for all algorithms (except for RANDOM) due to the fact that the randomization of allocating vaccines deceases. The extreme case is that when there is only one group, QP, DEGREE and EIGEN are uniformly randomly allocate vaccine to the whole graph, which is exactly the same as RANDOM. On the contrary, when the number of groups is equal to the number of nodes, group immunization

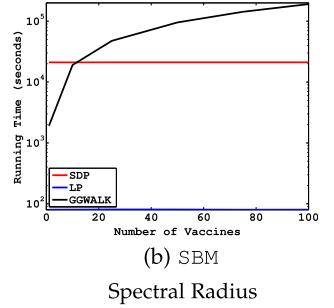
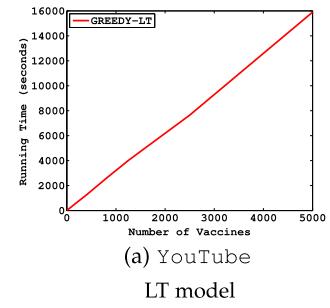


Fig. 7. Running Time (seconds). Running time versus number of vaccines.

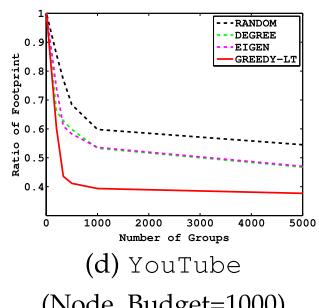
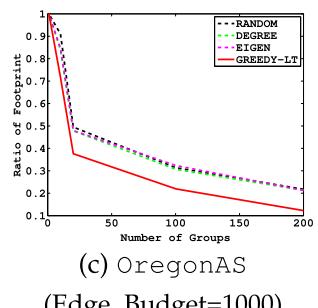
becomes individual immunization which is effective but much more expensive. Figs. 6c and 6d show the performance of GREEDY-LT as the number of groups varies. Similar to QP and LP, it consistently outperforms other baselines. And the performance improvement is even more obvious: when the graph size increases from 1 to 200, GREEDY-LT almost reduces 90 percent of the infection.

5.2.3 Scalability

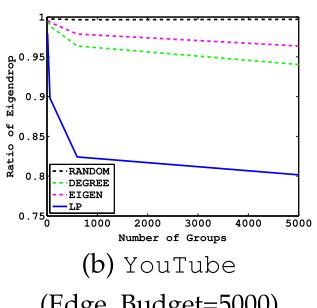
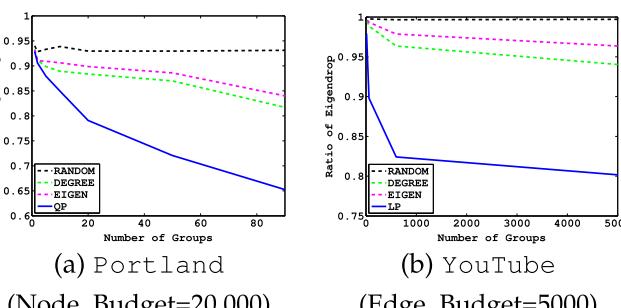
Although our algorithms are polynomial-time, we show some running time results to demonstrate the scalability of our algorithms. Fig. 7 shows the running time of our algorithms w.r.t. the number of vaccines. We did not show the running time of RANDOM, DEGREE and EIGEN, because they are faster heuristics. First, as expected from the time complexity of GREEDY-LT, when the number of vaccines m increases, the running time of GREEDY-LT increases linearly (Fig. 7a). Second, since the time complexities of SDP and LP are irrelevant to m , as shown in Fig. 7b, the running time of them remains almost constant. Furthermore, we observe that when m is small, GROUPGREEDYWALK ran faster than SDP. As the performances of GROUPGREEDYWALK and SDP are very close, in large graphs with a relatively small budget, we could get very good solution from GROUPGREEDYWALK.

5.2.4 Case Study

We now study the group vaccination problem on realistic social contact networks, Portland and Miami, using age based groups; as discussed earlier, age based directives are commonly used by public health agencies. Fig. 8 shows the number of vaccines assigned to different age groups, for a total of 10,000 vaccines, using the QP algorithm. We find the groups with age 70-79 and 60-66 get the maximum allocation,



LT model



Spectral Radius

Fig. 6. (a) and (b): Eigendrop ratio versus number of groups. (c) and (d): Footprint ratio versus number of groups. Lower is better. Our algorithms consistently outperform other baseline algorithms as the number of groups changes as well as the size of groups changes.

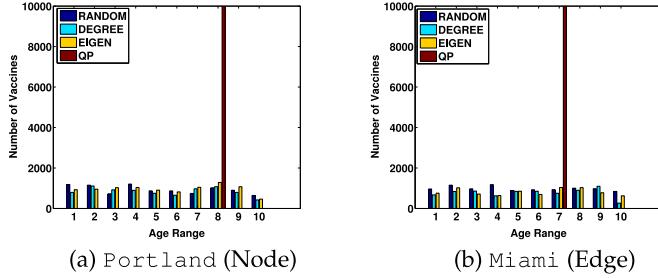


Fig. 8. Vaccine Distributions for Portland and Miami (Budget = 10,000). Number of vaccines versus Age. (Age range '0-9': 1; '10-19': 2; '20-29': 3; '30-39': 4; '40-49': 5; '50-59': 6; '60-69': 7; '70-79': 8; '80-89': 9; '90+'. 10.)

for the Portland and Miami networks, respectively. This contrasts with CDC recommendations, and the strategy proposed by Medlock et al. [1], as CDC recommendations include children through age 18, and Medlock et al. suggests to prioritization of schoolchildren and adults aged 30 to 39 years. This might be because these results do not use the detailed network structure. We believe this is an interesting result which merits further study.

6 DISCUSSION AND CONCLUSION

This paper addresses the problems of controlling epidemics by means of interventions that can be implemented at a group level. We formulate the GROUP IMMUNIZATION problem in the LT model as well as SIS/SIR models (considering the spectral radius minimization) for both edge-level and node-level interventions. We develop algorithms with rigorous performance guarantees and good empirical performance for all these problem classes. Our algorithms require a diverse class of techniques, including submodular function maximization, linear programming, quadratic programming, semidefinite programming, and hitting closed walks. Finally, we evaluate them on real networks of diverse scales. We demonstrate that our algorithms significantly outperform other heuristics, and adapt to the group structure. Some of our algorithms, e.g., SDP is fairly time intensive, though it runs in polynomial time. However, it is important to keep in mind that these algorithms are expected to be run before an epidemic outbreak, where the solution quality is much more critical than the run time.

Currently our SDP and GROUPGREEDYWALK algorithm work for edge deletion. Developing provable approximation algorithms for node deletion by leveraging SDP and GROUPGREEDYWALK, can be another future direction. In addition, our formulations capture the uncertainty, lack of control and compliance at a fine granularity in immunization interventions in public health and social media. Another important practical consideration is the economies of scale that arise in such group level formulations—these could be the result of decreasing per unit cost of production or distribution within a group. Such constraints can be modeled as $\sum_{i=1}^n \phi_i(x_i) \leq B$, where $\phi_i(x_i)$ is a concave function and x_i is the allocation to group C_i , and B is a budget constraint. Extending our algorithms to handle such constraints with our formulation is an interesting future work.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their comments. This work has been partially supported by

the following grants: DTRA Grant HDTRA1-11-1-0016, DTRA CNIMS Contract HDTRA1-11-D-0016-0010, US National Science Foundation Career CNS 0845700, US National Science Foundation ICESCCF-1216000, US National Science Foundation NETSE Grant CNS-1011769, US National Science Foundation DIBBS Grant ACI-1443054, US National Science Foundation Grant IIS-1353346, NEH Grant HG-229283-15, Maryland Procurement Office under contract H98230-14-C-0127, and a Facebook faculty gift. This work is also supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center (DoI/NBC) contract number D12PC000337, the US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Any opinions, findings, and conclusions or recommendations express in this material are those of the author(s) and do not necessarily reflect the views of the respective funding agencies.

REFERENCES

- [1] J. Medlock and A. P. Galvani, "Optimizing influenza vaccine distribution," *Science*, vol. 325, pp. 1705–1708, 2009.
- [2] S. Eubank, et al., "Modelling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, no. 6988, pp. 180–184, May 2004.
- [3] D. Z. Roth and B. Henr, "Social distancing as a pandemic influenza prevention measure: Evidence review," National Collaborating Centre for Infectious Diseases, 2011.
- [4] E. Shim, "Optimal strategies of social distancing and vaccination against seasonal influenza," *Math. Biosciences Eng.*, vol. 10, no. 5/6, pp. 1615–1634, 2013.
- [5] R. Cohen, S. Havlin, and D. ben Avraham, "Efficient immunization strategies for computer networks and populations," *Phys. Rev. Lett.*, vol. 91, no. 24, Dec. 2003, Art. no. 247901.
- [6] B. A. Prakash, L. A. Adamic, T. J. Iwashyna, H. Tong, and C. Faloutsos, "Fractional immunization in networks," in *Proc. SIAM Int. Conf. Data Mining*, 2013, pp. 659–667.
- [7] H. Tong, B. A. Prakash, C. E. Tsourakakis, T. Eliassi-Rad, C. Faloutsos, and D. H. Chau, "On the vulnerability of large graphs," in *Proc. IEEE Int. Conf. Data Mining*, 2010, pp. 1091–1096.
- [8] R. M. Anderson and R. M. May, *Infectious Diseases of Humans*. London, U.K.: Oxford Univ. Press, 1991.
- [9] N. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications*. London, U.K.: Griffin, 1975.
- [10] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2003, pp. 137–146.
- [11] E. B. Khalil, B. Dilkina, and L. Song, "Scalable diffusion-aware optimization of network topology," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 1226–1235.
- [12] B. A. Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, and C. Faloutsos, "Threshold conditions for arbitrary cascade models on arbitrary networks," *Knowl. Inf. Syst.*, vol. 33, pp. 549–575, 2012.
- [13] S. Saha, A. Adiga, B. A. Prakash, and A. K. S. Vullikanti, "Approximation algorithms for reducing the spectral radius to control epidemic spread," in *Proc. SIAM Int. Conf. Data Mining*, 2015, pp. 568–576.
- [14] D. Gruhl, R. V. Guha, D. Liben-Nowell, and A. Tomkins, "Information diffusion through blogspace," in *Proc. 13th Int. Conf. World Wide Web*, May 2004, pp. 491–501.
- [15] R. Kumar, J. Novak, P. Raghavan, and A. Tomkins, "On the bursty evolution of blogspace," in *Proc. 12th Int. Conf. World Wide Web*, 2003, pp. 568–576.
- [16] S. Bikhchandani, D. Hirshleifer, and I. Welch, "A theory of fads, fashion, custom, and cultural change in informational cascades," *J. Political Economy*, vol. 100, no. 5, pp. 992–1026, 1992.
- [17] J. Goldenberg, B. Libai, and E. Muller, "Talk of the network: A complex systems look at the underlying process of word-of-mouth," *Marketing Lett.*, vol. 12, pp. 211–223, 2001.
- [18] E. M. Rogers, *Diffusion of Innovations*, 5th ed. New York, NY, USA: Free Press, Aug. 2003.

- [19] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The dynamics of viral marketing," in *Proc. 7th ACM Conf. Electron. Commerce*, 2006, pp. 228–237.
- [20] E. E. Papalexakis, T. Dumitras, D. H. P. Chau, B. A. Prakash, and C. Faloutsos, "Spatio-temporal mining of software adoption & penetration," in *Proc. IEEE/ACM Int. Conf. Advances Social Netw. Anal. Mining*, 2013, pp. 878–885.
- [21] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM Rev.*, vol. 42, pp. 599–653, 2000.
- [22] A. G. McKendrick, "Applications of mathematics to medical problems," *Proc. Edinburgh Math. Soc.*, vol. 44, pp. 98–130, 1925.
- [23] M. Granovetter, "Threshold models of collective behavior," *Amer. J. Sociology*, vol. 83, pp. 1420–1443, 1978.
- [24] D. J. Watts, "A simple model of global cascades on random networks," *Proc. Nat. Academy Sci. USA*, vol. 99, no. 9, pp. 5766–5771, 2002.
- [25] D. Centola and M. Macy, "Complex contagions and the weakness of long ties," *Amer. J. Sociology*, vol. 113, no. 3, pp. 702–734, 2007.
- [26] A. Ganesh, L. Massoulie, and D. Towsley, "The effect of network topology on the spread of epidemics," in *Proc. IEEE INFOCOM*, 2005, pp. 1455–1466.
- [27] J. Aspnes, K. Chang, and A. Yampolskiy, "Inoculation strategies for victims of viruses and the sum-of-squares partition problem," in *Proc. 16th Annu. ACM-SIAM Symp. Discr. Algorithms*, 2005, pp. 43–52.
- [28] C. J. Kuhlman, G. Tuli, S. Swarup, M. V. Marathe, and S. Ravi, "Blocking simple and complex contagion by edge removal," in *Proc. IEEE 13th Int. Conf. Data Mining*, 2013, pp. 399–408.
- [29] H. Tong, B. A. Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos, "Gelling, and melting, large graphs by edge manipulation," in *Proc. 21st ACM Int. Conf. Inf. Knowl. Manage.*, 2012, pp. 245–254.
- [30] P. V. Mieghem, et al., "Decreasing the spectral radius of a graph by link removals," *Phys. Rev. E*, vol. 84, 2011, Art. no. 016101.
- [31] C. Budak, D. Agrawal, and A. E. Abbadi, "Limiting the spread of misinformation in social networks," in *Proc. 20th Int. Conf. World Wide Web*, 2011, pp. 665–674.
- [32] X. He, G. Song, W. Chen, and Q. Jiang, "Influence blocking maximization in social networks under the competitive linear threshold model," in *Proc. SIAM Int. Conf. Data Mining*, 2012, pp. 463–474.
- [33] Y. Zhang and B. Prakash, "DAVA: Distributing vaccines over large networks under prior information," in *Proc. SIAM Int. Conf. Data Mining*, 2014, pp. 46–54.
- [34] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2002, pp. 61–70.
- [35] M. Eftekhari, Y. Ganjali, and N. Koudas, "Information cascade at group scale," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2013, pp. 401–409.
- [36] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. S. Glance, "Cost-effective outbreak detection in networks," in *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2007, pp. 420–429.
- [37] T. Lappas, E. Terzi, D. Gunopulos, and H. Mannila, "Finding effectors in social networks," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2010, pp. 1059–1068.
- [38] M. Purohit, B. A. Prakash, C. Kang, Y. Zhang, and V. Subrahmanian, "Fast influence-based coarsening for large networks," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2014, pp. 1296–1305.
- [39] T. Soma, N. Kakimura, K. Inaba, and K.-i. Kawarabayashi, "Optimal budget allocation: Theoretical guarantee and efficient algorithm," in *Proc. 31st Int. Conf. Mach. Learn.*, 2014, pp. 351–359.
- [40] M. Minoux, "Accelerated greedy algorithms for maximizing submodular set functions," in *Optimization Techniques*. Berlin, Germany: Springer, 1978, pp. 234–243.
- [41] L. Lu and X. Peng, "Spectra of edge-independent random graphs," *Electron. J. Combinatorics*, vol. 20, no. 4, 2013, Art. no. P27.
- [42] S. Sahni, "Computationally related problems," *SIAM J. Comput.*, vol. 3, no. 4, pp. 262–279, 1974.
- [43] A. Goyal, F. Bonchi, and L. V. Lakshmanan, "Learning influence probabilities in social networks," in *Proc. 3rd ACM Int. Conf. Web Search Data Mining*, 2010, pp. 241–250.
- [44] A. F. McDaid, T. B. Murphy, N. Friel, and N. J. Hurley, "Improved Bayesian inference for the stochastic block model with application to large networks," *Comput. Statist. & Data Anal.*, Elsevier, vol. 60, pp. 12–31, 2013.
- [45] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Statistical Mech.: Theory Experiment*, vol. 2008, no. 10, 2008, Art. no. P10008.



Yao Zhang received the bachelor's and master's degrees in computer science from Nanjing University, China. He is working toward the PhD degree in the Department of Computer Science, Virginia Tech. His current research interests include graph mining and social network analysis with focus on understanding and managing information diffusion in networks. He has published several papers in top conferences and journals such as the *Knowledge Discovery in Databases*, ICDM, SDM, and the *Transactions on Knowledge Discovery from Data*.



Abhijin Adiga received the PhD degree from the Department of Computer Science & Automation, Indian Institute of Science, in 2011. He is a research assistant professor in the Network Dynamics and Simulation Science Laboratory, Biocomplexity Institute of Virginia Tech. His interests include network science, modeling, algorithms, combinatorics, and game theory with current focus on dynamical processes over networks and design & implementation of complex simulation systems.



Sudip Saha received the BSc degree in computer science and engineering from the Bangladesh University of Engineering and Technology, in 2006, and the MS degree in computer science from the University of Memphis, Tennessee, in 2010. He is currently working toward the PhD degree in computer science at Virginia Tech. His research interests include graph mining, game theory, and cybersecurity.



Anil Vullikanti received the undergraduate degree from the Indian Institute of Technology, Kanpur, and the PhD degree from the Indian Institute of Science, Bangalore. He is an associate professor in the Department of Computer Science and the Biocomplexity Institute of Virginia Tech. He was a post-doctoral researcher with the Max-Planck Institute for Informatics, and a technical staff member with the Los Alamos National Laboratory. His research interests include the broad areas of approximation and randomized algorithms and dynamical systems, and their applications to computational epidemiology and the modeling, simulation and analysis of socio-technical systems.



B. Aditya Prakash received the BTech (computer science) degree from the Indian Institute of Technology (IIT)—Bombay, in 2007, and the PhD degree from the Computer Science Department, Carnegie Mellon University, in 2012. He is an assistant professor in the Computer Science Department, Virginia Tech. His research interests include data mining, applied machine learning, and databases, with emphasis on big-data problems in large real-world networks and time-series. He has published more than 40 refereed papers in major venues, holds two US patents and has given two tutorials (VLDB 2012 and ECML/PKDD 2012) at leading conferences. His work has also received a best paper award, two best-of-conference selections (CIKM 2012, ICDM 2012, and ICDM 2011), and multiple travel awards. His work has been funded through grants/gifts from US NSF, DoE, NSA, NEH and from companies like Symantec. He received a Facebook Faculty Gift Award in 2015. He is also an affiliated faculty member in the Discovery Analytics Center, Virginia Tech.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.

To Delay or Not: Temporal Vaccination Games on Networks

Abhijin Adiga*, Srinivasa Venkat* and Anil Vullikanti*†

*Biocomplexity Institute of Virginia Tech

†Department of Computer Science, Virginia Tech

Email: {abhijin,vsriniv,akumar}@vbi.vt.edu

Abstract—Interventions such as vaccinations or installing anti-virus software are common strategies for controlling the spread of epidemics and malware on complex networks. Typically, nodes decide whether to implement such an intervention independently, depending on the costs they incur. A node can be protected by *herd immunity*, if enough other nodes implement such an intervention, making the problem of determining strategic decisions for vaccination a natural game-theoretical problem. There has been a lot of work on vaccination and network security game models, but all these models assume the vaccination decisions are made at the start of the game. However, in practice, a lot of individuals defer their vaccination decision, and the reasons for this behavior are not well understood, especially in network models.

In this paper, we study a novel repeated game formulation, which considers vaccination decisions over time. We characterize Nash equilibria and the social optimum in such games, and find that a significant fraction of vaccinations might be deferred, in general. This depends crucially on the network structure, and the information and the vaccination delay. We show that finding Nash equilibria and the social optimum are NP-hard in general, and we develop an approximation algorithm for the social optimum whose approximation guarantee depends on the delay.

I. INTRODUCTION

The spread of epidemics on social contact networks and malware in computer networks is commonly modeled by diffusion processes on networks, in which the infection spreads from a node to its neighbors. The threat of malware is becoming increasingly critical, as the number of connected devices grows rapidly. Similarly, infectious diseases continue to pose a significant public health problem, despite significant advances in medicines. There exist effective interventions to control the spread of epidemics and malware, e.g., by taking vaccinations in the case of diseases, and installing antivirus software patches in the case of malware. However, these interventions incur a certain cost for the individual (e.g., the economic cost of the vaccine or the patch). Further, an individual can get protected without any intervention if enough nodes in the network are protected—this is referred to as *herd immunity* in mathematical epidemiology, and is a natural setting for a game-theoretical analysis. This has been a very active area of research both in epidemiology and network security, e.g. [3], [2], [13], [4], [17], [14], [7], [10], [9], [19], [20].

An important limitation of all these approaches is that they typically consider vaccination decisions only at the start of the epidemic, in a simultaneous game setting, as discussed later in Section VI. In practice, very few people get vaccinated early in the season, despite active campaigns by public health agencies, such as the CDC. Vaccination rates increase as the epidemic spreads, and understanding this remains a big challenge. There has been some work on temporal vaccination strategies in the mathematical epidemiology literature, e.g., [19], but this is based on differential equation models, and does not consider realistic network structure. A recent work [11] studied the case of developing a vaccine distribution plan to suppress a pandemic on a network, thus considering a temporal graph. However, the paper focuses on centralized policymaking, while our paper is the first study of temporal vaccination decision making by individuals in the complex network.

Our main contributions are summarized below.

- 1) We develop a repeated game formulation, TEMPORAL-VACCINATION, which considers vaccination decisions at multiple time steps. We characterize the structure of Nash equilibria (NE) in such games, and find that they depend very crucially on the network structure and the delay parameters. Further, we show that NE need not always exist, and deciding if there exists a NE is NP-hard, in general.
- 2) Both the NE and optimal strategies exhibit interesting temporal structure. Even if vaccination decisions are allowed to be made at multiple time steps, we show that all decisions are made either at time 0 (before the start of the epidemic), or the first time $T > 0$ when the next round is played, if the source of the infection is known before T . Further, there can be significant variation in the number of nodes that choose to vaccinate at time T , instead of at time 0. Additionally, we find that even if no nodes are vaccinated at time 0 in the NE, the social optimum might choose to have a significant fraction at the start.
- 3) Computing the social optimum turns out to be a challenging stochastic optimization problem. We show that it is NP-complete, and develop an approximation algorithm which gives a strategy with cost at most $2T$ times the optimum, where T is as defined above.
- 4) We study the characteristics of NE in different real and synthetic networks empirically. We use best response strategies, which converge to NE very fast, in general. We find the number of nodes which get vaccinated at the

start is very sensitive to the ratio of the vaccination and infection costs. Further, a significant fraction of nodes defer their vaccination decision. We also find that high degree nodes appear to be more likely to get vaccinated initially.

II. PRELIMINARIES

Setting. We consider the spread of a *highly infectious disease* on a graph $G = (V, E)$, modeled as a simple discrete time **SI** model of an epidemic, where S and I denote Susceptible and Infectious states (see, e.g., [1]). If a node v is infected at time t , all of its uninfected neighbors v' will be infected at time $t + 1$, unless it is vaccinated at or before $t + 1$. Let C_v^t denote the cost for node v to get vaccinated at time t . A node need not decide to get vaccinated at time $t = 0$, but might defer the decision to a future time. If node v gets infected, we assume it incurs a cost L_v , such that $L_v > C_v^t$ for all t , i.e., infection is costlier than vaccination. We assume that the vaccine has 100% efficacy and starts protecting the node immediately. We discuss these assumptions later in Section VII. We refer to Table I for some additional notation and definitions needed for the rest of the paper.

A. Multi-stage game formulation

We formally define TEMPORALVACCINATION as a multi-stage game. We denote a game instance by $(G, \mathcal{T}, \mathbf{C}, \mathbf{L})$, where $G = (V, E)$ is a graph on n vertices, $\mathcal{T} = \{t_0 = 0, \dots, t_k\}$ is a set of times at which vaccination decision can be made, $\mathbf{C} = \{C_v^t \mid v \in V, t \in \mathcal{T}\}$ is the set of vaccination costs, and $\mathbf{L} = \{L_v \mid v \in V\}$ is the set of infection costs. Each node $v \in V$ is a player. The strategy function is denoted by Y . The strategy for node v at time t , given that the source of infection is s , is $Y(v, s, t) \in \{0, 1\}$. Since the source is not revealed at $t = 0$, we denote the strategy at time 0 by $Y(v, \cdot, 0)$. Let \mathbf{Y} be the set of all strategy functions. Y_t corresponds to the strategies at time t , i.e., $\{Y(v, s, t) : v, s \in V\}$. Let $Y_{<t}$ correspond to the strategies until time t , i.e., $\{Y(v, s, t') : v, s \in V, t' < t\}$.

The TEMPORALVACCINATION($G, \mathcal{T}, \mathbf{C}, \mathbf{L}$) game is played in the following manner:

- 1) At time $t = 0$, all the nodes play a simultaneous vaccination game to decide whether to get vaccinated or not. If node v gets vaccinated at this time, we denote this by $Y(v, \cdot, 0) = 1$. Note that the vaccine takes effect immediately.
- 2) A randomly chosen node $s \in V$ is selected to be the source of the epidemic. We assume that if $Y(s, \cdot, 0) = 1$, it remains immune, and the epidemic does not start. If $Y(s, \cdot, 0) = 0$, then s gets infected and the infection spreads to each uninfected neighbor in subsequent times. We also assume perfect information, so all nodes know the identity of source s , and the entire network.
- 3) For each $t = 1, 2, \dots$, we have the following two steps:
 - (a) If $t \in \mathcal{T}$, a simultaneous vaccination game is played at time t , and each node v decides whether to get vaccinated at this time or not— this is denoted by $Y(v, s, t) \in \{0, 1\}$, with 1 denoting vaccination.

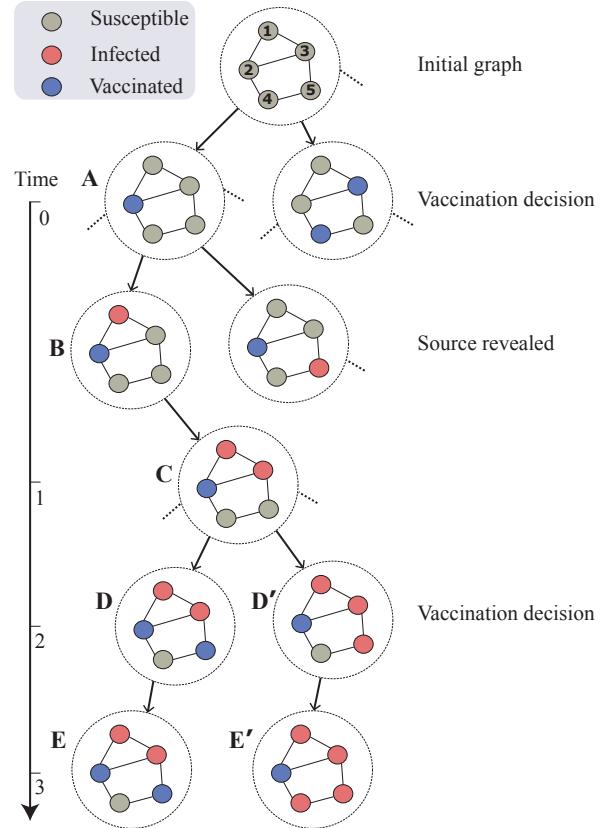


Fig. 1: Example illustrating the different rounds of the TEMPORALVACCINATION game on a graph with 5 nodes, with $\mathcal{T} = \{0, 2\}$. At $t = 0$, some nodes are vaccinated (blue). After this, the infection source is revealed (red). Since $1 \notin \mathcal{T}$, there is no vaccination decision (Step 3(a)) of the game specification) at time 1, and instead only the epidemic spreading step (Step 3(b)) occurs. Next, since $2 \in \mathcal{T}$, vaccination decisions are taken, and thus both 3(a) and 3(b) occur. In configuration D node 5 chooses to vaccinate, while in D' it chooses not to vaccinate.

- (b) Let I_{t-1} denote the set of nodes which are infected at time $t - 1$. For each node $u \in I_{t-1}$, each uninfected neighbor $v \in N(u)$, $v \notin I_{t-1}$ will get infected at time t , unless v is already vaccinated or vaccines at t . Recall that vaccination takes precedence over infection in our model. Define set $I_t = I_{t-1} \cup \{v : v \text{ gets infected at time } t\}$ to be the new set of infected nodes.
- 4) The game stops at time t if there are no more uninfected nodes that can be infected from their neighbors, and there are no more vaccination games to be played, i.e., $t' \notin \mathcal{T}$ for all $t' \geq t$. Each node v incurs cost L_v if it ever got infected, i.e., $v \in I_t$. It incurs cost $C_v^{t'}$ if it got vaccinated at time $t' \leq t$. The overall cost for node v is the expectation over all possible choices of the source.

The configurations representing the states of players during the course of the TEMPORALVACCINATION game can be viewed in the form of a tree— note that this is not the

same as the game tree in the extensive form representation, and we describe it here only for clarifying the structure of the game. The configuration at some time t of the game can be specified by the tuple $(\{Y_{<t}(v, s, t) : v \in V\}, I_t)$, where $\{Y_{<t}(v, s, t) : v \in V\}$ represents all the vaccination decisions in the game rounds played so far and I_t is the current set of infected nodes. The root node of this tree is the tuple (ϕ, ϕ) , since initially there are no infections, and no vaccination game has been played. The root node has 2^n child configurations (the “level 1” configurations) of the form (Y_0, ϕ) , each corresponding to a set $V' \subset V$ of nodes that get vaccinated at time 0. Each level-1 configuration has n children (the “level 2” configurations), $(Y_0, \{s\})$, each corresponding to a random source s . From this point on, the tree expands based on Step 3 of the game specification above. If $t \in \mathcal{T}$, then step 3(a) happens and there are 2^n child configurations, each corresponding to a subset $V' \subset V$ of nodes that get vaccinated at time t . After this, step 3(b) occurs, where the set I_t is updated and each configuration corresponds to $(Y_{<t}, I_t)$. The costs of all the nodes are specified for the leaf configurations of this tree, as discussed in Step 4 of the game specification.

Example. A fragment of the tree of configurations is shown in Figure 1. Configuration **A** can be written down as (Y_0, ϕ) where Y_0 is such that $Y(2, \cdot, 0) = 1$ and $Y(v, \cdot, 0) = 0$ for all $v \neq 2$. There are 2^n such configurations in level 1, which are children of the root (ϕ, ϕ) . The child configuration marked **B** corresponds to the tuple $(Y_0, \{1\})$; there are n such configurations which are children of **A**. **C** represents a level 3 configuration, where the disease spreads at time $t = 1$. Configurations **D** and **D'** depict the vaccination decisions (and subsequent epidemic spread) that occur at time 2, with **E** and **E'** respectively showing the final states.

Simplified uniform cost game. In the rest of the paper, we will frequently consider a simplified game instance where $\mathcal{T} = \{0, T\}$ and every vertex has the same infection cost L and vaccination costs C^0 and C^T . We will use the term $(G, T, \{C^0, C^T\}, L)$ to denote such instances. Further, when $C^0 = C^T = C$, we will use (G, T, C, L) .

Cost model. As discussed above, the cost of a node depends on the final configuration—node v incurs cost L_v if it ever got infected, and it incurs cost C_v^t if it got vaccinated time t . The overall cost for a node is the expected cost over all the random sources (which is the only source of stochasticity in the model). Formally, we define the cost incurred by v under the strategy profile $Y(\cdot)$ as:

$$\begin{aligned} \text{cost}(v, Y(\cdot)) &= C_v^0 Y(v, \cdot, 0) \\ &+ \sum_{s \in V} \frac{1}{n} \left(\sum_t C_v^t Y(v, s, t) + L_v \mathbf{1}_{s \rightarrow v | Y(\cdot)} \right), \end{aligned}$$

where $\mathbf{1}_{s \rightarrow v | Y(\cdot)} = 1$ if v gets infected due to s under strategy profile $Y(\cdot)$. We define $\text{cost}(Y(\cdot)) = \sum_v \text{cost}(v, Y(\cdot))$.

B. Nash equilibria and social optimum

For a strategy profile $Y(\cdot)$, let $Y_{-v}(\cdot)$ be the strategy profile for all the remaining players. We say that a strategy $Y(\cdot)$ is a Nash equilibrium (NE) [16] if for each $v \in V$: $\text{cost}(v, Y'(\cdot)) \geq \text{cost}(v, Y(\cdot))$ where $Y'(\cdot)$ is any strategy profile such that $Y'_{-v}(\cdot) = Y_{-v}(\cdot)$, i.e., $Y'(\cdot)$ has the same

TABLE I: Notation summary

Network	
$G = (V, E)$	simple undirected graph with vertex set V and edge set E
n	number of vertices in G
$G[V']$	subgraph induced by $V' \subseteq V$
$N(v, G)$	neighborhood of v in G
$d_G(s, i)$	shortest path distance between nodes s and i in the graph G
$B(G, v, l)$	$\{u : d_G(u, v) \leq l\}$ i.e., l -ball
$P(G, v, l)$	$\{u : d_G(u, v) = l\}$ i.e., l -perimeter
Strategy	
\mathcal{T}	set of time instants at which the game is played
T	$\min\{t \in \mathcal{T} : t > 0\}$, smallest positive time in \mathcal{T}
\mathbf{Y}	set of all strategy profiles
$Y(\cdot)$	a strategy profile
$Y(v, s, t)$	strategy for node v at time t , given source s ; takes values from $\{0, 1\}$; $Y(v, \cdot, 0)$ for time 0
Y_t	strategies at time t , $\{Y(v, s, t) : v, s \in V\}$
$Y_{<t}$	strategies until time t , $\{Y(v, s, t') : v, s \in V, t' < t\}$
$Y_{-v}(\cdot)$	strategies of all players except v
$V_0(Y)$	$\{v \in V : Y(v, \cdot, 0) = 1\}$
Costs	
C_v^t	Cost for node v to vaccinate at time t
\mathbf{C}	$\{C_v^t \mid v \in V, t \in \mathcal{T}\}$
L_v	Cost of infection for node v
\mathbf{L}	$\{L_v \mid v \in V\}$
$\text{cost}(v, t)$	cost incurred by v at time t
Game	
$(G, \mathcal{T}, \mathbf{C}, \mathbf{L})$	TEMPORALVACCINATION game instance
$(G, T, \{C^0, C^T\}, L)$	two-stage uniform cost game
(G, T, C, L)	further simplified game instance with $C^0 = C^T$

strategies as $Y(\cdot)$ for all other players $v' \neq v$. In other words, no player v can reduce its expected cost by unilaterally changing its strategy, given that the other players’ strategies are fixed.

We define the *social optimum* as a strategy $Y(\cdot)$ that has the minimum cost, over the space of all possible strategies—this is not necessarily (and is not usually) a pure NE. Therefore, the cost of a pure NE relative to the social cost is an important measure, and the maximum such ratio over all possible pure NE is known as the *price of anarchy* [12].

III. CHARACTERIZATION OF NE IN THE TEMPORALVACCINATION GAME

We start with a characterization of pure NE, which will be repeatedly used in our subsequent discussions. Section III-A will cover the hardness of computing a NE. This will be followed by bounds on the Price of Anarchy. In Section III-C and III-D, we study the structure of the NE and social optimum of complete graphs and Erdős-Rényi graphs respectively.

Lemma 3.1: For a game instance $(G, \mathcal{T}, \mathbf{C}, \mathbf{L})$, let $T = \min\{t \in \mathcal{T} : t > 0\}$ be the smallest positive time in \mathcal{T} . Let $V_0(Y) = \{i \mid Y(i, \cdot, 0) = 1\}$, the set of nodes which vaccinated at time 0. For any $i \in V$, let $G_i = G[V - V_0(Y) \cup \{i\}]$. A strategy profile $Y(\cdot)$ is a pure NE iff

- (1) $\forall i \in V$, $Y(i, \cdot, 0) + \sum_t Y(i, s, t) \leq 1$, for each $s \in V$.
- (2) Every node $i \in V$ such that $Y(i, \cdot, 0) = 1$ satisfies $|B(G_i, i, T-1)| \frac{L_i}{n} + |P(G_i, i, T)| \frac{C_i^T}{n} > C_i^0$.

- (3) Every node $i \in V$ such that $Y(i, \cdot, 0) = 0$ satisfies $|B(G_i, i, T-1)|\frac{L_i}{n} + |P(G_i, i, T)|\frac{C_i^T}{n} \leq C_i^0$.
- (4) Every node $i \in V$ such that $Y(i, s, T) = 1$ satisfies $|B(G_i, i, T-1)|\frac{L_i}{n} + |P(G_i, i, T-1)|\frac{C_i^T}{n} \leq C_i^0$ and $d(G_i, s, i) = T$.

Proof: Statement (1) follows from the fact that a node becomes immune the first time it gets vaccinated. Therefore, vaccinating again only increases its cost, and will not happen in a NE. Consider any insecure node i at time 0 (i.e., $Y(i, \cdot, 0) = 0$). If $d_{G_i}(s, i) < T$, node i will be infected even before it gets a chance to vaccinate at T , and so it has $Y(i, s, T) = 0$. If $d_{G_i}(s, i) = T$, then node i is better off getting vaccinated, since $C_i^T < L_i$. Therefore, all nodes at distance T from s in $G[V - V_0(Y)]$ get vaccinated at time T .

From the above discussion, the expected cost of not vaccinating at time 0 is $\Pr(i \text{ will get infected whether it vaccines at time } T \text{ or not}) \times L_i + \Pr(i \text{ will vaccinate at time } T \text{ and become secure}) \times C_i^T = |B(G_i, i, T-1)|\frac{L_i}{n} + |P(G_i, i, T)|\frac{C_i^T}{n}$. A node vaccines at time 0 if the expected cost of not vaccinating is greater than C_i^0 (which implies statement (2)). If this expected cost is less than C_i^0 , then the node does not get vaccinated at time 0, i.e., $Y(i, \cdot, 0) = 0$ (which implies statement (3)). In this case, it vaccinates at time T (i.e., $Y(i, s, T) = 1$) iff it is at distance T from s , which implies statement (4). ■

Lemma 3.1 implies that once the source is revealed, all nodes make their vaccination decisions at the earliest possible time. Therefore, it suffices to study the TEMPORALVACCINATION game with $\mathcal{T} = \{0, T\}$.

Corollary 3.1: Let $Y(\cdot)$ be a pure NE for a TEMPORALVACCINATION instance $(G, \mathcal{T}, \mathbf{C}, \mathbf{L})$. Let $T = \min\{t \in \mathcal{T} : t > 0\}$ be the smallest positive time in \mathcal{T} . Then, we have $Y_i(t) = 0$ for all $t > T$.

It is easy to verify that Lemma 3.1 and Corollary 3.1 hold for mixed NE as well.

A. Existence of pure NE

Consider an instance $(G, \mathcal{T}, \mathbf{C}, \mathbf{L})$ where $\mathcal{T} = \{0, T\}$ and for each $v \in V$, $C_v^T = 0$. This game belongs to the family GNS(δ) studied in [13], where each node makes a decision to vaccinate or not at time 0, and the infection spreads from a random source node to a distance of upto δ . Here, $\delta = T - 1$. In [13], it was shown that for $\delta = 1$, pure NE always exists, while for $1 < \delta < \infty$, there are instances of GNS(δ) for which pure NE does not exist, and that determining if an instance has a pure NE is NP-complete. Therefore, we have the following result.

Lemma 3.2: The problem of determining if an instance $(G, \mathcal{T}, \mathbf{C}, \mathbf{L})$ has a pure NE is NP-complete. For the case $T = 2$, and $C_i^T = 0$ for all nodes $i \in V(G)$, a pure NE always exists.

Since pure NE don't always exist, we consider sufficient conditions in which they exist. Given $(G, \mathcal{T}, \mathbf{C}, \mathbf{L})$, we define an auxiliary graph \widehat{G} as follows: Let $\mathcal{T} = \{0, T\}$. The

vertex set of \widehat{G} , \widehat{V} comprises of all vertices v which satisfy $\text{cost}(v, G) = |B(G, v, T-1)|\frac{L_i}{n} + |P(G, v, T)|\frac{C_v^T}{n} > C_v^0$. These are precisely the nodes which have the incentive to vaccinate at time 0. We draw an edge between two vertices u and v in \widehat{G} if and only if $B(G, u, T) \cap B(G, v, T) \neq \emptyset$, i.e., u is adjacent to v if and only if removing one reduces the incentive to vaccinate for the other. Now we show the following:

Lemma 3.3: For an instance $(G, \mathcal{T}, \mathbf{C}, \mathbf{L})$ where, $\mathcal{T} = \{0, T\}$, a pure NE exists if \widehat{G} is bipartite.

Proof: First, we note that if v is not adjacent to w in \widehat{G} , then, removing w does not change cost, i.e., $\text{cost}(v, G - \{w\}) = \text{cost}(v, G)$. Also, if the cost is above C_v^0 even after removing all the neighbors of v in \widehat{G} , then, it implies that v will vaccinate at time 0 in every pure NE. Such nodes can be ignored in our analysis. Therefore, from now on we will assume that G does not have such nodes and for every v , $\text{cost}(v, G - N(v, \widehat{G})) \leq C_v^0$. Suppose \widehat{G} is a bipartite graph with bipartition $(\widehat{V}_1, \widehat{V}_2)$. We now show that the strategy Y satisfying $Y(v, \cdot, 0) = 1$ if $v \in \widehat{V}_1$, and $Y(v, \cdot, 0) = 0$ if $v \in V \setminus \widehat{V}_1$ corresponds to a pure NE. If a vertex $v \in \widehat{V}_1$ changes its strategy from secure to insecure, then, since $N(v, \widehat{G}) \cap \widehat{V}_1 = \emptyset$, it implies that $B(G(Y_v), v, T) = B(G, v, T)$. Therefore, $\text{cost}(v, G(Y')) = \text{cost}(v, G) > C_v^0 \geq \text{cost}(v, G(Y))$. For a vertex $v \in \widehat{V}_2$, since $N(v, \widehat{G}) \subseteq \widehat{V}_1$, $G(Y)$ does not contain $N(v, \widehat{G})$. Hence by the earlier assumption, $\text{cost}(v, G(Y)) = \text{cost}(v, G - N(v, \widehat{G})) \leq C_v^0$. Therefore, v has no incentive to secure itself. The nodes in $V(G) \setminus V(\widehat{G})$ satisfy $\text{cost}(v, G(Y)) \leq \text{cost}(v, G) \leq C_v^0$. Therefore, they have no incentive to vaccinate either. Further, for every $v \in V$, we set $Y(v, s, T) = 1$ if and only if $v \in P(G[V \setminus \widehat{V}_1], s, T)$ and $Y(v, \cdot, 0) = 0$. From Lemma 3.1, it follows that Y is a pure NE. ■

B. Bounds on the Price of Anarchy

Now, we will show that the price of anarchy for the TEMPORALVACCINATION game can be as high as $\Theta(n)$.

Lemma 3.4: Let C and L be any real numbers satisfying $0 < C < L$ and let $r = \lceil L/C \rceil$ and let $T > 1$. There exists a graph on $n > 2r^2T$ vertices such that the price of anarchy of the simplified TEMPORALVACCINATION instance (G, T, C, L) is at least $\frac{(n-2r^2T)^2}{n(2T+1)r}$. In particular, if C, L and T are constants, then, the price of anarchy is $\Theta(n)$.

Proof: The graph G is constructed as follows. The vertex set V is the disjoint union $\{\ell\} \cup \bigcup_{i=1}^r K_i \cup P_i$ where each $K_i = \{c_{i1}, \dots, c_{i(n-2r^2T)/r}\}$ induces a cycle on the vertices c_{ij} , and each $P_i = \{p_{i1}, \dots, p_{iT}\}$ induces the path $p_{i1}p_{i2} \cdots p_{iT}$. In addition, each p_{i1} is adjacent to all vertices in K_i , and each p_{iT} is adjacent to ℓ . Note that there is a deficit of $r^2T - rT - 1$ nodes. One way to rectify this is to add a path with these many nodes. This will not affect the proof in any way. Therefore, we will ignore this set of nodes for the rest of the proof. Let $\mathcal{P} = \bigcap_{i=1}^r P_i$ and $\mathcal{K} = \bigcap_{i=1}^r K_i$.

First, we will show that for every $v \in V(G)$, $\text{cost}(v, G) = |B(G, v, T-1)|\frac{L_i}{n} + |P(G, v, T)|\frac{C_v^T}{n} \leq C$. If $v \in K_i$, then,

$B(G, v, T-1) \subseteq K_i \cup P_i$ and $P(G, v, T) = \{p_{iT}\}$. Therefore, $\text{cost}(v, G) = \left(\frac{n-2r^2T}{r} + T\right)\frac{L}{n} + \frac{C}{n} \leq \frac{(n-1)L}{n} + \frac{C}{n} \leq C$. Suppose $v \in P_i$. If $v = p_{iT}$ then, $B(G, v, T-1) \subseteq \mathcal{P} \cup \{\ell\}$ and $P(G, v, T) = K_i \cup \{p_{j(T-1)} \mid j \neq i\}$. If $v = p_{ik}$, $k \neq T$, then, $B(G, v, T-1) \subseteq K_i \cup \mathcal{P} \cup \{\ell\}$ and $P(G, v, T) = \{p_{j(T-k)} \mid j \neq i\}$. In either case, $\text{cost}(v, G) \leq \left(\frac{n-2r^2T}{r} + rT + r + 1\right)\frac{L}{n} \leq C$. If $v = \ell$, then, $B(G, v, T-1) \subseteq \mathcal{P} \cup \{\ell\}$ and $P(G, v, T) = \{p_{iT} \mid 1 \leq i \leq r\}$. Therefore, $\text{cost}(v, G) \leq (rT + 1)\frac{L}{n} < C$. This implies that at time $t = 0$, no vertex gets vaccinated and therefore, there always exists a pure NE for this game where the strategy of each vertex only depends on the source. Hence, the cost of the NE is only due to the second stage of vaccination.

Let s be the source. Now, we will compute a lower bound for the expected cost. The sets of vertices which are infected and vaccinated respectively are $B(G, s, T-1)$ and $P(G, s, T)$. It is sufficient to consider the case $s \in \mathcal{K}$. Suppose $s \in K_i$ for some i . Then, $B(G, s, T-1) = K_i \cup P_i \setminus \{p_{iT}\}$ and $P(G, s, T) = \{p_{iT}\}$. The cost of the NE is $\left(\frac{n-2r^2T}{r} + T - 1\right)L + 1 > \frac{n-2r^2T}{r}L$. Hence,

$$\text{cost}(Y) > \Pr(s \in \mathcal{K}) \frac{n-2r^2T}{r}L = \left(\frac{n-2r^2T}{n}\right) \frac{n-2r^2T}{r}L,$$

which is bounded by $\frac{(n-2r^2T)^2 L}{nr}$. Now, we bound the cost of the social optimum from above. Suppose at time $t = 0$, we vaccinate p_{i1} for all $i = 1, \dots, r$ and ℓ . Then, the residual graph is disconnected, and comprises of the following components: (1) cycles induced by K_i and (2) paths induced by $P_i \setminus \{p_{i1}\}$. It is easy to see that at time $T-1$, the number of nodes infected is at most $2(T-1) + 1$, and the number of nodes to be vaccinated at $t = T$ to stop further spread is at most two. Hence, the cost of the social optimum is at most $(2(T-1) + 1)L + 2C \leq (2T + 1)L$. Hence, proved. ■

C. Equilibria in complete graph

For the complete graph, we note that if $T > 1$, then any node which did not vaccinate at time $t = 0$ will get infected if the source is an infected node. The result below gives a lower bound on the number of nodes that vaccinate at $t = 0$ in any NE when G is a complete graph. Its proof is omitted for brevity.

Lemma 3.5: For the instance $(G, \mathcal{T}, \mathbf{C}, \mathbf{L})$ where $G = (V, E)$ is a complete graph on n nodes and $T > 1$ where, $T = \min\{t \in \mathcal{T} \mid t > 0\}$, the number of nodes that vaccinate at $t = 0$, n_0 is at least $\min_{v \in V} (1 - \frac{C_v^0}{L_v})n$.

For the two-stage uniform cost game, we can show the following stronger result.

Lemma 3.6: For the instance (G, T, C, L) where $G = (V, E)$ is a complete graph on n nodes, in every pure NE strategy $n_0 = \lceil(1 - \frac{C}{L})n\rceil$ nodes vaccinate at $t = 0$. Further, if $T > 1$, the PoA is $\approx (1 - \frac{C}{4L})^{-1}$ for $C, L \ll n$.

Proof: By Lemma 3.5, every pure NE strategy should have at least $(1 - \frac{C}{L})n$ nodes vaccinating at $t = 0$. If $n_0 = \lceil(1 - \frac{C}{L})n\rceil$, then, for every remaining node, the expected cost is $\frac{n-n_0}{n}L \leq \frac{1}{n}(n - (1 - \frac{C}{L})n)L \leq C$. Also, making a secure node insecure will increase this cost for every

remaining insecure node. Hence, this is a pure NE and we have proved the first part. Now we will consider the Price of Anarchy. When $T > 1$, the expected cost of any strategy is $n_0C + \frac{n-n_0}{n}(n - n_0)L = n_0C + \frac{(n-n_0)^2}{n}L$, with minima at $n_0 = (1 - \frac{C}{2L})n$. Therefore, the cost of social optimum is $(1 - \frac{C}{2L})nC + (\frac{C^2}{4L^2})nL = nC(1 - \frac{C}{4L})$. By substituting $n_0 = \lceil(1 - \frac{C}{L})n\rceil$ above, the cost of any pure NE is $\approx nC$. Therefore, PoA is $(1 - \frac{C}{4L})^{-1}$ ■

D. Equilibria in Erdős-Rényi graphs

Here, we study the structure of pure NE on instances of the $G(n, p)$ model. We consider two regimes: (R1) $\frac{1}{n} < p < \frac{\log n}{n}$ and $T \leq \frac{\log(\frac{nC}{4L \log n})}{1.5 + \log np}$, and (R2) $p > \frac{c \log n}{n}$ and $T \leq \frac{\log(\frac{nC}{c'(L+nPC)})}{\log np}$ where $c \leq 2$ is a constant and c' is only a function of c . We say that an event $A(n)$ occurs *almost surely* if $\Pr(A(n)) \rightarrow 1$ as $n \rightarrow \infty$.

Lemma 3.7: Consider the two-stage, uniform cost game (G, T, C, L) where, $G \in G(n, p)$. In both regimes (R1) and (R2), almost surely, pure NE exist, and any such strategy Y has $Y(v, \cdot, 0) = 0$ for all $v \in V(G)$. Further, the expected cost of a pure NE Y satisfies $\text{cost}(Y) \leq 2T^3 \log n (np)^T (L + C)$ for (R1) and $\text{cost}(Y) \leq c'(np)^T (L + C)$ for (R2) almost surely.

Proof: Recall that the expected cost incurred by v for not vaccinating at time 0 is $\frac{1}{n}(|B(G, v, T-1)|L + |P(G, v, T)|C)$. First, we will prove the results concerning (R1). We will use the following result from [5]: For $np > 1$, any $v \in V$, $|B(G, v, \ell)| \leq 2\ell^3(np)^\ell \log n$ and $|P(G, v, \ell)| \leq 2\ell^2(np)^\ell \log n$ with probability at least $1 - o(n^{-1.5})$. Henceforth, we will assume that for all v , this condition is satisfied. Note that, by union bound, this happens with probability $1 - o(n^{-0.5})$. Applying this, the expected cost incurred by v for not vaccinating at time 0 is

$$\begin{aligned} &\leq \frac{1}{n}(2(T-1)^3(np)^{T-1}(\log n)L + 2T^2(np)^T(\log n)C) \\ &\leq \frac{2T^3(np)^T \log n(L + C)}{n} < \frac{2(L + C) \log n}{n} e^{3 \log T + T \log(np)}. \end{aligned}$$

In the second inequality, we use the assumption $np > 1$. Since, in (R1) $T \leq \frac{\log(\frac{nC}{2(L+C) \log n})}{1.5 + \log np}$, and $3 \log T + T(\log(np)) \leq T(1.5 + \log(np))$, the cost is at most C . Therefore, from Lemma 3.1, it follows that $Y(v, \cdot, 0) = 0$ with probability $1 - o(n^{-0.5})$ for all $v \in V(G)$. Again, with probability $1 - o(n^{-0.5})$, the expected cost of any pure NE strategy is,

$$\begin{aligned} \text{cost}(Y) &\leq \mathbb{E}_s[|B(G, s, T-1)|L + \mathbb{E}_s[|P(G, s, T)|]C] \quad (1) \\ &\leq 2(T-1)^3(np)^{T-1}(\log n)L + 2T^2(np)^T(\log n)C \\ &\leq 2T^3(np)^T(L + C) \log n. \end{aligned}$$

We proceed similarly for (R2). Here, we use the following result from [5]: For $p > \frac{c \log n}{n}$, where $c \leq 2$ is a constant, for any $v \in V(G)$, $|B(G, v, \ell)| \leq c'(np)^\ell$ and $|P(G, v, \ell)| \leq c'(np)^\ell$ with probability at least $1 - o(n^{-1.5})$, where c' is only a function of c . Again, this implies that (by union bound) with probability at least $1 - o(n^{-0.5})$, these conditions hold for every $v \in V(G)$. The expected cost incurred by v

for not vaccinating at time 0 is at most $\frac{1}{n}(c'(np)^{T-1}L + c'(np)^TC) \leq \frac{c'(np)^T(L+C)}{n} \cdot \log(\frac{nC}{c'(L+C)})$. Since for (R2), $T \leq \frac{\log(\frac{nC}{c'(L+C)})}{\log np}$, it follows that the cost is at most C . Hence, $Y(v, \cdot, 0) = 0$ almost surely. As in the previous case, the cost of the pure NE Y is $\text{cost}(Y) \leq c'(np)^{T-1}L + c'(np)^TC = c'(np)^T(L + C)$. ■

Social optimum: Let n_0 denote the number of nodes which are vaccinated at time 0. We show the following:

Lemma 3.8: In the two-stage, uniform cost game (G, T, C, L) where, $G \in G(n, p)$ and $p > \frac{c \log n}{n}$, the number of nodes vaccinated at time 0 in any optimal strategy is $n_0 \leq n - \frac{1}{p} \left(\frac{nC}{2c'L(T+1)} \right)^{1/T}$, almost surely, where c' is a function of c .

Proof: Let G' denote the residual graph obtained after removing the vaccinated nodes at time 0. The cost of the optimum strategy Y^{OPT} is $\text{cost}(Y^{\text{OPT}}) = n_0C + \mathbb{E}_s[|B(G', s, T-1)|]L + \mathbb{E}_s[|P(G', s, T)|]C$. Note that $G' \in G(n - n_0, p)$. Therefore, $\text{cost}(Y^{\text{OPT}})$

$$\begin{aligned} &= n_0C + \sum_{s \in V(G')} \frac{1}{n} (|B(G', s, T-1)|L + |P(G', s, T)|C) \\ &\leq n_0C + \frac{n - n_0}{n} (c'((n - n_0)p)^{T-1}L + c'((n - n_0)p)^TC) \\ &< n_0C + \frac{c'(n - n_0)^{T+1}p^T(L + C)}{n}. \end{aligned}$$

Note that the last expression is a convex function in n_0 with minima at $n - \frac{1}{p} \left(\frac{nC}{c'(L+C)(T+1)} \right)^{1/T}$. ■

IV. SOCIAL OPTIMUM

We first show that computing the social optimum of the TEMPORALVACCINATION game is NP-complete, and then we develop an approximation algorithm using the approach of two-stage stochastic optimization.

Lemma 4.1: Computing the social optimum of a TEMPORALVACCINATION instance $(G, \mathcal{T}, \mathbf{C}, \mathbf{L})$ is NP-complete.

Proof: For a given strategy $Y(\cdot)$, we can estimate the expected cost in polynomial time, since there are only n random choices, corresponding to the random source. Therefore the decision version of this problem, i.e., deciding if the social optimum has cost at most B for a parameter B is in NP. When $C_j^t = 0$ for all $t \in \mathcal{T}$ and $j \in V$, this problem is equivalent to the GNS(d) game on the graph G , which is NP-hard [13]. ■

Approximating the social optimum. We now discuss an approximation algorithm for computing the social optimum. Our algorithm is based on the two-stage stochastic optimization approach. We first define some quantities which are needed for our algorithm. Let P_{ij}^T denote the set of all simple paths between nodes i and j having length at most T in G . We start with an integer programming formulation for the social optimum. We have the following variables

- 1) y_{0j} for each $j \in V$, which is an indicator for node j being vaccinated at time 0.

2) y_{sj} for each $j, s \in V$, which is an indicator for node j being vaccinated at time T , when the random source is s .

3) z_{ij} for each $i, j \in V$, which is an indicator for the event that there is no path $P \in P_{ij}^T$ consisting entirely of unvaccinated nodes.

Our integer programming \mathcal{P} has the following structure:

$$\min f(\mathbf{y}, \mathbf{z}) = \sum_v C_v^0 y_{0v} + \frac{1}{n} \sum_{k,v} C_v^T y_{kv} + \frac{1}{n} \sum_{k,j} L_j (1 - z_{kj})$$

$$\text{s.t. } \sum_{v \in p} y_{0v} \geq z_{sj}, \forall p \in P_{sj}^T, \text{ with length } < T \quad (2)$$

$$\sum_{v \in p} y_{0v} + y_{sj} \geq z_{sj}, \forall p \in P_{sj}^T, \text{ with length } T \quad (3)$$

$$y_{0v}, y_{sv}, z_{ij} \in \{0, 1\}, \forall s, v, i, j \in V.$$

Lemma 4.2: Let (\mathbf{y}, \mathbf{z}) be the solution to the above integer program \mathcal{P} . Then, the strategy $Y(v, 0) = y_{0v}$ and $Y(v, s, T) = y_{sv}$ is an optimal solution to $(G, \mathcal{T}, \mathbf{C}, \mathbf{L})$.

Proof: Let $Y^{\text{opt}}(\cdot)$ be the socially optimal strategy to the given instance of the TEMPORALVACCINATION instance. If $Y^{\text{opt}}(j, \cdot, 0) = 0$, node j incurs a cost of L_j if the source is s and $d_{G[V - V_0(Y^{\text{opt}})]}(s, j) < T$, and it incurs a cost of C_j^T if $d_{G[V - V_0(Y^{\text{opt}})]}(s, j) = T$, since we assume that $C_j^t < L_j$ for all t, j . If $Y^{\text{opt}}(j, 0) = 1$, node j incurs cost C_j^0 .

Let $Y(\cdot)$ be the strategy constructed from (\mathbf{y}, \mathbf{z}) in the lemma. We observe that $f(\mathbf{y}, \mathbf{z})$ equals $\text{cost}(Y)$, and $Y(\cdot)$ satisfies the same property as Y^{opt} above. For the nodes $j \in V_0(Y)$, we have $y_{0j} = 1$, which accounts for the cost $\sum_{j \in V_0(Y)} C_j^0$ of vaccination at time 0. Next, assume node j was not vaccinated at time t , i.e., $Y(j, \cdot, 0) = 0$. In the second stage, if j has distance less than T from the random source s in the residual graph, then there exists a path $p \in P_{sj}^T$ such that $y_{0v} = 0$ for all $v \in p$, including s and j . Then, constraint (2) in \mathcal{P} causes $z_{sj} = 0$, so that node j incurs an infection cost of $L_j(1 - z_{sj}) = L_j$ in this case. On the other hand, if $d_{G[V - V_0(Y)]}(s, j) = T$, by constraint (3), we have $z_{sj} = 1$ if and only if $y_{sj} = 1$. This leads to a cost of $C_j^T y_{sj} = C_j^T$ for node j . Note that $y_{sj} = 0$ would have caused $z_{sj} = 0$, leading to an infection cost of $L_j(1 - z_{sj})$, which would be higher than C_j^T . Since s is the source with probability $\frac{1}{n}$, this cost component is scaled in $f(\mathbf{y}, \mathbf{z})$ by $\frac{1}{n}$. This implies $f(\mathbf{y}, \mathbf{z}) = \text{cost}(Y)$, and $Y(\cdot)$ satisfies the same property as Y^{opt} . ■

Lemma 4.3: For any T , the linear relaxation \mathcal{P}_L can be solved in polynomial time.

Proof: When T is a constant, the linear program \mathcal{P}_L is of polynomial size, since $|P_{sj}^T| = O(n^T)$. Therefore, in this case, we can directly solve the LP and find the fractional solution $(\mathbf{y}^1, \mathbf{z}^1)$. When T is not a constant, \mathcal{P}_L has super-polynomially many constraints, and cannot be solved explicitly. We use the Ellipsoid method [21], which gives a polynomial time algorithm for finding a feasible solution x from a convex polytope $K \subseteq \mathbb{R}^N$. We refer to [21] for complete details, but note here that the key component needed for this method to work is to design a polynomial time “separation oracle”, which, given a candidate solution

Algorithm 1: APPROXSOCOPT

input : $G, T, \mathbf{C}, \mathbf{L}$
output : Strategy $Y(\cdot)$

- 1 Solve a linear relaxation \mathcal{P}_L of the program \mathcal{P} , in which the constraints $y_{0v}, y_{sv} \in \{0, 1\}$ and $z_{ij} \in \{0, 1\}$ are replaced by
$$y_{0v}, y_{sv} \in [0, 1], \quad \forall s, v \in V$$

$$z_{ij} \in [0, 1], \quad \forall i, j \in V$$
- 2 If T is not a constant, \mathcal{P}_L has exponentially many constraints, and we use the ellipsoid method to solve it, as discussed in Lemma 4.3.
- 3 Let $(\mathbf{y}^1, \mathbf{z}^1)$ denote the optimum fractional solution to \mathcal{P}_L .
- 4 We construct a new fractional solution $(\mathbf{y}^2, \mathbf{z}^2)$ in the following manner:
 - 5 **for each** i, j **do**
 - 6 set $z_{ij}^2 = 0$ if $z_{ij}^1 \leq 1/2$, and $z_{ij}^2 = 1$ otherwise
 - 7 **for** $j \in V$ **do**
 - 8 Set $y_{0j}^2 = \min\{2y_{0j}^1, 1\}$ and
 - 9 $y_{sj}^2 = \min\{2y_{sj}^1, 1\}$, for all $s \in V$.
 - 10 **for** $j \in V$ **do**
 - 11 if $y_{0j}^2 > 1/T$, we set $Y(\cdot, j, 0) = 1$, else we set $Y(\cdot, j, 0) = 0$.
 - 12 if $y_{sj}^2 > 1/T$, we set $Y(s, j, T) = 1$, else we set $Y(s, j, T) = 0$.

(\mathbf{y}, \mathbf{z}) , can decide if it is feasible, or finds a constraint that is infeasible.

We convert the \mathcal{P}_L into a feasibility problem by “guessing” the cost of the objective, and adding a constraint that $f(\mathbf{y}, \mathbf{z}) \leq B$, where B is the estimate of the objective value. Such a separation oracle can be designed for the program \mathcal{P}_L as follows:

- 1) For each pair s, j :
 - a) Define weight $w_v = y_{0v}$ in the graph G
 - b) Compute the shortest path distance $\text{dist}(s, j, T - 1)$ from s to j restricted to paths with length at most $T - 1$, based on the weights \mathbf{w} . If $\text{dist}(s, j, T - 1) < z_{sj}$, we return $\sum_{v \in p} y_{0v} \geq z_{sj}$ as the violated constraint.
 - c) Compute the shortest path distance $\text{dist}(s, j, T)$ from s to j restricted to paths with length at most T , based on the weights \mathbf{w} . If $\text{dist}(s, j, T) < z_{sj} - y_{sj}$, we return $\sum_{v \in p} y_{0v} + y_{sj} \geq z_{sj}$ as the violated constraint.
- 2) Finally, if $f(\mathbf{y}, \mathbf{z}) > B$, it is returned as a violated constraint.

These steps can be implemented in polynomial time. Therefore the separation oracle runs in polynomial time, so that the above algorithm returns a solution to program \mathcal{P}_L in polynomial time, with cost at most the objective value, if it exists. We refer to [21] for complete details of the Ellipsoid method and its proof. ■

Lemma 4.4: Algorithm APPROXSOCOPT gives a $2T$ -approximation to the social optimum.

Proof: Let $Y^{opt}(\cdot)$ be the optimum solution. From Lemma 4.2, it follows that $f(\mathbf{y}^1, \mathbf{z}^1) \leq \text{cost}(Y^{opt})$.

We first argue that $(\mathbf{y}^2, \mathbf{z}^2)$ is feasible and $f(\mathbf{y}^2, \mathbf{z}^2) \leq 2f(\mathbf{y}^1, \mathbf{z}^1)$. By the construction, we have $1 - z_{ij}^2 \leq 2(1 - z_{ij}^1)$, whether $z_{ij}^1 \leq 1/2$ or $z_{ij}^1 > 1/2$. Constraints (2) and (3) corresponding to a pair s, j , continue to hold if $z_{sj}^2 = 0$. The constraints corresponding to pairs s, j , for which $z_{sj}^2 = 1$ also continue to hold, since $z_{sj}^2 \leq 2z_{sj}^1$ in this case, and we have $y_{0j}^2 = \min\{2y_{0j}^1, 1\}$ and $y_{sj}^2 = \min\{2y_{sj}^1, 1\}$, for all $j, s \in V$.

For each pair s, j , if $z_{sj}^2 = 1$, it follows that for every path $p \in P_{sj}^T$: (1) if $\text{len}(p) < T$, there must be some node $v \in p$ such that $y_{0v}^2 \geq 1/T$, which implies $Y(\cdot, v, 0) = 1$; and (2) if $\text{len}(p) = T$, either there exists some node $v \in p$ such that $y_{0v}^2 \geq 1/T$, or $y_{sj}^2 \geq 1/T$. In the former case, we have $Y(\cdot, v, 0) = 1$, whereas in the latter case, we have $Y(s, j, T) = 1$.

Finally, by construction, we have $Y(\cdot, j, 0) \leq Ty_{0j}^2$ and $Y(s, j, T) \leq Ty_{sj}^2$. This implies that $\text{cost}(Y(\cdot)) \leq 2T \cdot \text{cost}(Y^{opt})$. ■

V. EXPERIMENTS

We now study characteristics of Nash equilibria of the TEMPORALVACCINATION game in several social/communication networks and two random graph models, as summarized in Table II. In light of Lemma 3.2, we use a best response strategy to search for NE. We study the number and characteristics of nodes that get vaccinated at times 0 or T , and how this is affected by the relative vaccination costs at these times. Our main observations are summarized below.

TABLE II: Networks used in our experiments and their relevant properties: two real [15] and two synthetic graphs. We study the synthetic graphs with varying edge densities.

Network	Nodes (n)	Edges ($ E $)
Ca-GRQC (co-authorship network)	4158	13422
AS20000102 (autonomous system network)	6474	12572
Barabasi-Albert	1000	varying
Erdős-Rényi	1000	varying

1. Number of vaccinated nodes at $t = 0$ and $t = T$
The number of nodes getting vaccinated initially, i.e., $|V_0(Y)|$, is very sensitive to $\frac{C}{L}$, and drops rapidly, as shown in Figure 2(c) and (d). Surprisingly, the number of nodes vaccinating at time T , i.e., $|V_T(Y)|$, however, is fairly stable across $\frac{C}{L}$ in both the networks. Further, there seems to be a cut-off point for C/L , where $|V_0(Y)|$ falls below $|V_T(Y)|$, which might be useful in policy design.

2. Performance of best response strategies

We find that the best response strategy generally converges to NE in linear number of rounds, as illustrated in Figure

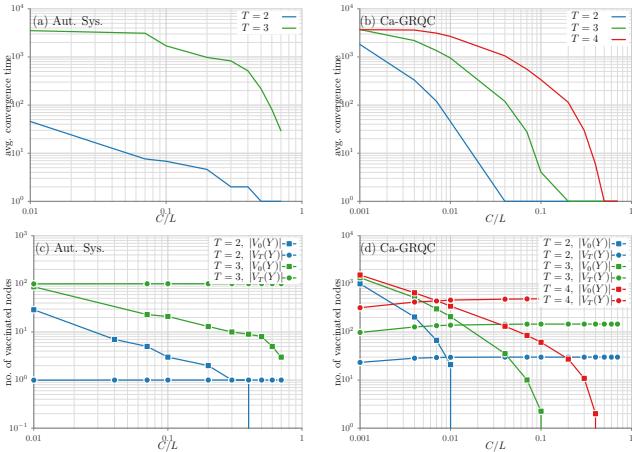


Fig. 2: Convergence time to NE using best response for (a) Autonomous System and (b) Co-authorship network. Variation of $|V_0(Y)|$ and $|V_T(Y)|$ with C/L , for different values of T in (c) Autonomous System and (d) Co-authorship network.

2. Moreover, the convergence time decreases rapidly with the C/L ratio, and increases with T . This suggests a high correlation between nodes the number of nodes vaccinating at time 0 and convergence time, which is an interesting topic for further investigation. The two real networks in Figure 2(a) and (b) exhibit very different behaviors, which might be the result of their structural differences. We also find that performance of best response is sensitive to the initial condition.

3. Effect of network density

Figure 3 (a) and (c) show the effect of network density, by varying the edge probability, p , for the Erdős-Rényi graph $G(n, p)$ and the number of edges per new node, m for the $BA(n, m)$ graph. We see that as the network becomes denser, more nodes vaccinate at time 0. Further, dense networks exhibit low diameter, and thus the game quickly approaches $GNS(\infty)$ of [13] for small increments in T .

4. Effect of heterogeneous costs

Figure 3 (b) considers the scenario when $C^T \neq C^0$. Note that for fixed C^0 , as C^T increases, more nodes vaccinate at time 0.

5. Effect of T

From Figure 3 (d), as T , the minimum waiting time to next vaccination, increases, more nodes vaccinate at time 0. Further, this effect is more pronounced for denser networks. Qualitatively, this behavior is expected; as T increases, its T -ball becomes larger and the likelihood of a node getting infected before time T increases.

6. Correlation between degree and likelihood of vaccination at time 0

Figure 3 (e) shows the degree of nodes vaccinating at time 0 in NE plotted along with the degree of all nodes, for a Barabasi-Albert network with $N = 1000$, $m = 3$, $C/L = 0.1$, $T = 3$. We notice that almost all the nodes that get vaccinated at time 0 are among the top degree nodes. We observe a similar behavior in other networks, and these results are omitted because of space constraints.

VI. RELATED WORK

There is a large literature on the use of non-cooperative game models for controlling the spread of epidemics and malware. We briefly summarize some of the main areas that are directly relevant to our paper.

A common approach in the mathematical epidemiology literature is based on differential equation models, e.g., [4], [10], [9], [6], [18], [19], [22]. These models are based on simplified assumptions about uniform mixing among the players, which allows for rigorous analysis. For instance, Bauch et al. [4] show that the NE can be completely characterized in terms of the reproductive number, which is the expected number of secondary infections caused by an infected individual. These models are deterministic and usually only consider vaccination strategies before the start of the epidemic. Reluga et al. [19] develop an approach that combines population games with Markov decision process, and consider decisions at different times.

Such differential equation models do not capture the complexity of interactions in real social contact networks. The work of Aspnes et al. [3] was among the first to study a network based formulation for vaccination games. They characterize NE in terms of the network structure and develop approximation algorithms for computing the social optimum. The utility function in their formulation requires the estimation of the probability that a node gets infected, which requires a lot of information. Kumar et al. [13] extend this formulation by restricting the amount of information needed by an individual. Our work builds on this formulation. Mean-field approximations have been used for detailed analysis in the SIS model, in which nodes switch from Susceptible to Infectious state, thereby capturing a more realistic epidemic model, e.g., [17], [22]. Saha et al. [20] consider a different formulation based on the spectral radius (the first eigenvalue of the network), in which the utility is based on whether or not the spectral radius is above a threshold or not—this is based on a characterization of the time to die out in the SIS model in terms of the spectral properties. However, all these approaches only consider vaccination decisions at the start of the epidemic, in a one-shot simultaneous game formulation. We also note that game-theoretical methods have been also used in other network security applications, e.g., [7], [8].

VII. DISCUSSION AND CONCLUSIONS

The TEMPORALVACCINATION game shows that vaccination decisions over time exhibit a very rich behavior. In both synthetic and real networks, and across a broad class of parameter regimes, we find that a significant fraction of nodes choose to get vaccinated later. The initially vaccinated fraction drops significantly as the ratio of vaccination and infection costs increases. Further, the timing of vaccination depends crucially on the network structure and information about the source and disease incidence. Therefore, the effect of delays in vaccination decisions has important implications for public health policy planning, and needs to be taken into careful consideration. In particular, vaccine availability needs to reflect the structure of equilibria. As a result, computing properties

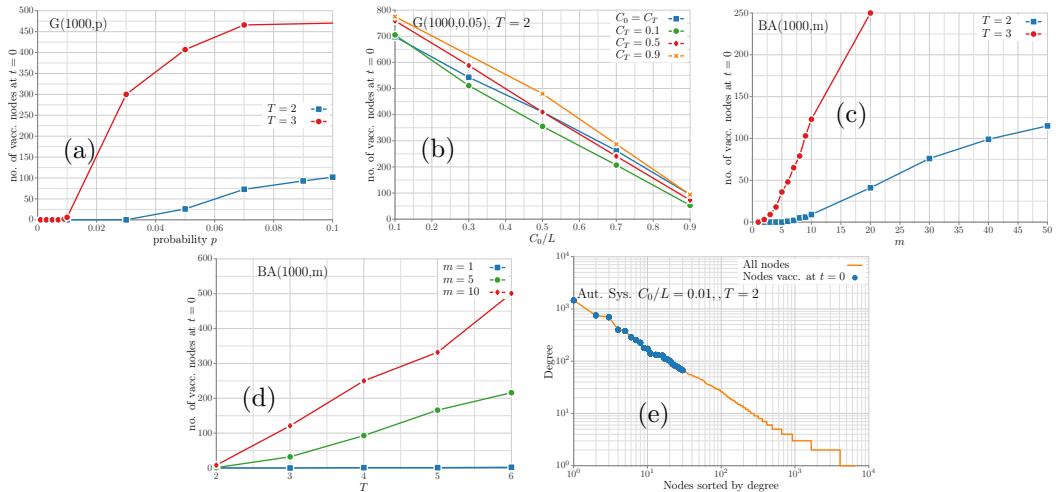


Fig. 3: The first four plots correspond to nodes vaccinating at $t = 0$ as a function of system parameters for the synthetic networks in Table II. The fifth plot shows the correlation between node degree and the chances of the node vaccinating at time 0.

of such games, including NE and social optimum, is an important issue. Admittedly, we have made a number of simplifying assumptions in our formulations. We only focus on a simple version of the **SI** model with infection probability 1, with perfect information about the source and other infections. Further, we assume vaccines have 100% efficacy and no delay. Relaxing all these assumptions are important directions for future work. We note that for many of our results, a vaccine delay of τ can be taken into account by considering an effective decision time of $T + \tau$, instead of T . Finally, we do not assume any resource constraints (e.g., a bound on the number of vaccines available at any time). In [1], we show that resource constraints change the structure of the game significantly.

Acknowledgments. This work has been partially supported by the following grants: DTRA Grant HDTRA1-11-1-0016, DTRA CNIMS Contract HDTRA1-11-D-0016-0010, NSF ICES CCF-1216000 and NSF NETSE Grant CNS-1011769.

REFERENCES

- [1] A. Adiga and A. Vullikanti. Temporal vaccination games under resource constraints. In *Proc. of AAAI*, 2016.
- [2] J. Aspnes, K.L. Chang, and A. Yampolskiy. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. *J. Comput. Syst. Sci.*, 2006.
- [3] J. Aspnes, N. Rustagi, and J. Saia. Worm versus alert: Who wins in a battle for control of a large-scale network? *OPODIS*, 2007.
- [4] C.T. Bauch and D.J.D. Earn. Vaccination and the theory of game. *PNAS*, 2004.
- [5] Fan Chung and Linyuan Lu. The diameter of sparse random graphs. *Advances in Applied Mathematics*, 26(4):257–279, 2001.
- [6] A. Galvani, T. Reluga, and G. Chapman. Long-standing influenza vaccination policy is in accord with individual self-interest but not with the utilitarian optimum. *PNAS*, 104(13):5692–5697, March 2007.
- [7] J. Grossklags, N. Christin, and J. Chuang. Secure or insure? a game-theoretic analysis of information security games. In *World Wide Web Conference (WWW)*, 2008.
- [8] M. Jain, V. Conitzer, and M. Tambe. Security scheduling for real-world networks. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2013.
- [9] M. H. R. Khouzani, Eitan Altman, and Saswati Sarkar. Optimal quarantining of wireless malware through reception gain control. *IEEE Trans. Automat. Contr.*, 57(1):49–61, 2012.
- [10] M. H. R. Khouzani, Saswati Sarkar, and Eitan Altman. Saddle-point strategies in malware attack. *IEEE Journal on Selected Areas in Communications*, 30(1):31–43, 2012.
- [11] Donghyun Kim, Hao Guo, Yuchao Li, Wei Wang, Sung-Sik Kwon, and Alade O Tokuta. Social relation based long-term vaccine distribution planning to suppress pandemic. In *Proc. of CSoNet*, 2015.
- [12] E. Koutsoupias and C. H. Papadimitriou. Worst-case equilibria. In *Proceedings of STACS*, 1999.
- [13] V. S. Anil Kumar, R. Rajaraman, Z. Sun, and R. Sundaram. Existence theorems and approximation algorithms for generalized network security games. In *Proceedings of IEEE ICDCS*, 2010.
- [14] M. Lelarge and J. Bolot. Economic incentives to increase security in the internet: The case for insurance. In *IEEE Infocom*, 2009.
- [15] J. Leskovec. Snap: Stanford network analysis project, 2011.
- [16] Kevin Leyton-Brown and Yoav Shoham. *Essentials of Game Theory: A Concise, Multidisciplinary Introduction*. Morgan and Claypool Publishers, 1st edition, 2008.
- [17] J. Omic, A. Orda, and P. Van Mieghem. Protecting against network infections a game theoretic perspective. In *INFOCOM*, 2009.
- [18] T. Reluga. Game theory of social distancing in response to an epidemic. *PLOS Computational Biology*, 6(5), 2010. e1000793.
- [19] T. Reluga and A. Galvani. A general approach to population games with application to vaccination. *Mathematical Biosciences*, 2011.
- [20] Sudip Saha, Abhijin Adiga, and Anil Kumar S. Vullikanti. Equilibria in epidemic containment games. In *The 28th AAAI Conference on Artificial Intelligence (AAAI)*, 2014.
- [21] A. Schrijver. *Theory of Linear and Integer Programming*. Wiley, 1986.
- [22] S. Trajanovski, Y. Hayel, E. Altman, H. Wang, and P. Van Mieghem. Decentralized protection strategies against sis epidemics in networks. *Control of Network Systems, IEEE Transactions on*, (99), 2015.

Temporal Vaccination Games Under Resource Constraints

Abhijin Adiga[†] and Anil Vullikanti^{*,†}

[†]Network Dynamics and Simulation Science Laboratory,
Biocomplexity Institute of Virginia Tech

^{*}Department of Computer Science, Virginia Tech
Email: {abhijin, akumar}@vbi.vt.edu

Abstract

The decision to take vaccinations and other protective interventions for avoiding an infection is a natural game-theoretic setting. Most of the work on vaccination games has focused on decisions at the start of an epidemic. However, a lot of people defer their vaccination decisions, in practice. For example, in the case of the seasonal flu, vaccination rates gradually increase, as the epidemic rate increases. This motivates the study of temporal vaccination games, in which vaccination decisions can be made more than once. An important issue in the context of temporal decisions is that of resource limitations, which may arise due to production and distribution constraints. While there has been some work on temporal vaccination games, resource constraints have not been considered.

In this paper, we study temporal vaccination games for epidemics in the SI (susceptible-infectious) model, with resource constraints in the form of a repeated game in complex social networks, with budgets on the number of vaccines that can be taken at any time. We find that the resource constraints and the vaccination and infection costs have a significant impact on the structure of Nash equilibria (NE). In general, the budget constraints can cause NE to become very inefficient, and finding efficient NE as well as the social optimum are NP-hard problems. We develop algorithms for finding NE and approximating the social optimum. We evaluate our results using simulations on different kinds of networks.

Introduction

Despite a lot of progress in medical diagnostics and pharmaceutical tools, infectious diseases remain a major challenge for governments all over the world. Even the annual influenza epidemic in the US has a significant social and economic burden, which is estimated to exceed \$87.1 billion (e.g., (Molinari et al. 2007)). For many diseases, especially the annual influenza, there exist vaccines, though their efficacy might be quite variable. However, taking a vaccine has a cost (economic cost, inconvenience and health effects). Further, an individual can get protected without any intervention if enough other people he/she comes in contact with in the population are protected—this is referred to as *herd immunity* in mathematical epidemiology, and is a natural setting for a game-theoretical analysis. This has been a very

active area of research both in epidemiology and network security, e.g. (Bauch and Earn 2004; Grossklags, Christin, and Chuang 2008; Khouzani, Sarkar, and Altman 2011; Reluga and Galvani 2011; Saha, Adiga, and Vullikanti 2014); see the related works section for a more detailed discussion.

In practice, most people do not take a preventive vaccine before the start of the epidemic, and instead wait for some time. Often, the vaccination rate grows with the epidemic outbreak rate. There are many different and complex reasons for vaccination decisions being made at different times, and understanding this remains a big open problem, as well as an important public health issue. Almost all the work on vaccination games only considers vaccination decision made only at the start of the epidemic, in a simultaneous game setting. The only studies on temporal vaccination are by (Reluga and Galvani 2011; Adiga, Venkataraman, and Vullikanti 2016). The work by (Reluga and Galvani 2011) uses a differential equation approach, which assumes simplified and homogeneous connectivity among individuals.

We build on the approach of (Adiga, Venkataraman, and Vullikanti 2016), which studies this as a repeated game on a network, in the SI (Susceptible-Infectious) model of epidemics. They characterize Nash equilibria in such games, and show that these exhibit interesting temporal structure, such as a large fraction of nodes defer their vaccination decisions. An important open question from their work is the effect of resource constraints—they assume there are no bounds on the number of individuals who can get vaccinated at any time, and this is reflected in the solutions they find. In practice, there are important resource constraints of various kinds, e.g., capacity of hospitals and pharmacies in administering vaccines, or the production capacity—see e.g., (CDC ; Orenstein and Schaffner 2008). In this paper, we study the temporal vaccination problem with resource constraints at different times. Our main contributions are:

1. We formalize the temporal vaccination game with resource constraints, BUDGETTEMPVACC , as a multi-stage game on a network, and study the structure of Nash equilibria (NE) in such games. There can be multiple NE even when the network of interactions is a tree, and finding one with the minimum cost is NP-complete.
2. We show that many nodes defer their vaccination deci-

sions, and the budget constraints lead to very significant differences from the solution of (Adiga, Venkataramanan, and Vullikanti 2016). Specifically, there exist families of instances, where small changes in the budgets (while keeping all other components of the problem fixed) lead to very high inefficiencies. Further, unlike their formulation (in which there is no need to consider more than two time steps), we find that there can be nodes that choose to vaccinate at each possible time step.

3. Computing the social optimum turns out to be a challenging optimization problem. We show that it is NP-hard to approximate within a factor of $O(n^\alpha)$ for any $\alpha \in (0, 1)$, without violating any budget constraints. For the special case of BUDGETTEMPVACC, with only two times at which vaccination decisions are made, we design an algorithm that approximately satisfies the budget constraints.
4. We study the characteristics of NE in different kinds of networks experimentally. We use best response strategies, and find that they usually converge to NE quickly. Corroborating our theoretical results, we find very high sensitivity of the solution cost, as well as the number of nodes that defer vaccination decisions, to the budget constraints and vaccination delays.

Some of the details, including proofs and experimental results have been omitted from this abstract, because of space limitations. They are presented the full version (Adiga and Vullikanti 2015).

Preliminaries and Model

We extend the approach of (Adiga, Venkataramanan, and Vullikanti 2016) and formally define BUDGETTEMPVACC as a repeated game on an instance $(G, \mathbf{p}, \mathcal{T}, \mathcal{B}, \mathbf{C}, \mathbf{L})$ in the following manner:

1. V is a set of $n = |V|$ players, who form the nodes of a graph $G = (V, E)$, with an edge $(u, v) \in E$ if the epidemic can spread from u to v ;
2. $p_i = \Pr[\text{source of infection is node } i]$, with $\sum_i p_i = 1$.
3. $\mathcal{T} = \{t_0 = 0, \dots, t_k\}$ is a set of time instants, at which the vaccination decisions will be made;
4. $\mathcal{B} = \{B^t \mid t \in \mathcal{T}\}$ specifies the vaccination budget, where B^t is the number of vaccines allocated for time t ;
5. $\mathbf{C} = \{C_v^t \mid v \in V, t \in \mathcal{T}\}$ is the set of vaccination costs,
6. $\mathbf{L} = \{L_v \mid v \in V\}$ is the set of infection costs.

Epidemic model and strategies. We assume nodes are in Susceptible (S) and Infectious (I) states, and the epidemic spreads according to a simple discrete time SI (Easley and Kleinberg 2010). In this model, if node v gets infected at time t , each uninfected neighbor u will be infected at time $t + 1$, unless it gets vaccinated. The vaccine is assumed to have 100% efficacy, and starts protecting the node immediately after taking it. A node's strategy is to decide if and when to get vaccinated. The strategy of node v at time 0 (before the source of the infection is known) is denoted by $Y(v, \cdot, 0) \in \{0, 1\}$, with $Y(v, \cdot, 0) = 1$ if v gets vaccinated at time 0. The source of the infection is known after time $t = 0$, and

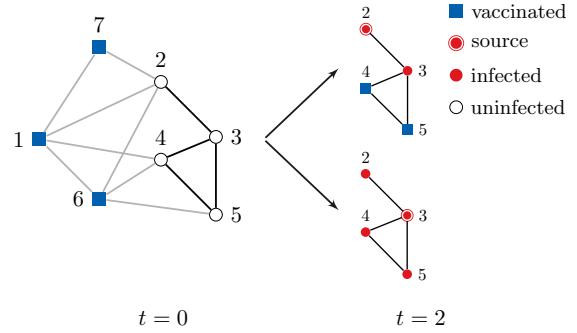


Figure 1: Example NE: Uniform vaccination cost $C = 0.5$ and infection cost $L = 1$ for all nodes, $\mathcal{T} = \{0, 2\}$, $p_2 = 0.5$, $p_3 = 0.5$, $B^0 = 3$, $B^2 = 2$.

$Y(v, s, t)$ denotes the strategy of node v at time t given that the source of infection is s . If node v gets vaccinated at time t , it incurs a cost C_v^t , whereas if it gets infected at any time, it incurs a cost L_v . Throughout, we assume that $L_v > C_v^t$ for all t , so that it is always cheaper to get vaccinated, instead of getting infected. The complete strategy vector is denoted by $Y(\cdot)$, and the expected cost associated with it is denoted by $\text{cost}(Y)$.

Example. Fig. 1 shows an example where $Y(1, \cdot, 0) = Y(6, \cdot, 0) = Y(7, \cdot, 0) = 1$ and $Y(v, \cdot, 0) = 0$ for $v = 2, 3, 4, 5$. For $t = 2$, $Y(4, 2, 2) = Y(5, 2, 2) = 1$ and $Y(v, s, 2) = 0$ for all other v and s combinations. Note that if node 1 decides to become insecure, then, the cost incurred is $p_2 L + p_3 C > 0.5 = C$. The same holds for 6 and 7. At $t = 2$, if 2 is the source, then only 4 and 5 can be saved, while if 3 is the source, none can be saved. By definition no other node can be the source.

Resource constraints. We assume at most B^t people can be vaccinated at time t —this can capture resource constraints, e.g., due to production or distribution limits. Therefore, a strategy vector $Y(\cdot)$ is *feasible* if for any time t : (1) $\sum_v Y(v, \cdot, 0) \leq B^0$, if $t = 0$, and (2) $\sum_v Y(v, s, t) \leq B^t$, for any source s . Let \mathcal{F} denote the set of all feasible strategy vectors.

Stages of BUDGETTEMPVACC. This involves the following rounds:

1. At time $t = 0$, all the nodes play a simultaneous vaccination game to decide whether to get vaccinated or not. As mentioned earlier, the vaccination takes effect immediately. If node v gets vaccinated at this time, we denote this by $Y(v, \cdot, 0) = 1$. As mentioned above, at most B^0 nodes can get vaccinated at this time, so that $\sum_v Y(v, \cdot, 0) \leq B^0$.
2. A randomly chosen node $s \in V$ is selected to be the source of the epidemic. We assume that if $Y(s, \cdot, 0) = 1$, it remains immune, and the infection does not spread. If $Y(s, \cdot, 0) = 0$, then s gets infected and the infection spreads to each uninfected neighbor in subsequent times. We also assume perfect information, so all nodes know about the source s .
3. For each $t = 1, 2, \dots$, we have the following two steps:
 - (a) If $t \in \mathcal{T}$, a simultaneous vaccination game is played at time t , and each node v decides whether to get vaccinated

at this time or not—this is denoted by $Y(v, s, t) \in \{0, 1\}$, with 1 denoting vaccination and s the source. Further, for feasibility, we have $\sum_v Y(v, s, t) \leq B^t$.

- (b) Let I_{t-1} denote the set of nodes which are infected at time $t-1$. For each node $u \in I_{t-1}$, each uninfected neighbor $v \in N(u)$ will get infected at time t , (i.e., $v \notin I_{t-1}$), unless v gets vaccinated at or before time t . Recall that vaccination takes precedence over infection in our model. Define set $I_t = I_{t-1} \cup \{v : v \text{ gets infected at time } t\}$ to be the set of all infected nodes till this time.
4. The game stops at time t if there are no more uninfected nodes that can be infected from their neighbors, and $t' \notin \mathcal{T}$ for all $t' \geq t$ (i.e., there are no more vaccination games to be played). Each node v incurs cost L_v if it ever got infected, i.e., $v \in I_t$. It incurs cost $C_v^{t'}$ if it got vaccinated at time $t' \leq t$. The overall cost for node v is the expectation over all possible choices of the source.

Cost of a feasible strategy. For every $Y \in \mathcal{F}$, we define the cost incurred by $v \in V(G)$ given strategy Y as: $\text{cost}(v, Y) = C_v^0 Y(v, \cdot, 0) + \sum_{s \in V} p_s (\sum_{T \in \mathcal{T}} C_v^T Y(v, s, T) + L_v I(v, s, Y))$, where $I(s, v, Y) = 1$ if v gets infected due to s in the strategy Y . We define $\text{cost}(Y) = \sum_v \text{cost}(v, Y)$.

Nash equilibria and social optimum. For a strategy profile $Y(\cdot)$, let $Y_{-v}(\cdot)$ be the strategy profile for all the remaining players. We say that a strategy $Y(\cdot)$ is a Nash equilibrium (NE) (Leyton-Brown and Shoham 2008) if for each $v \in V$: $\text{cost}(v, Y') \geq \text{cost}(v, Y)$ where Y' is any strategy profile such that $Y'_{-v}(\cdot) = Y_{-v}(\cdot)$, i.e., $Y'(\cdot)$ has the same strategies as $Y(\cdot)$ for all other players $v' \neq v$. In other words, no player v can reduce its expected cost by unilaterally changing its strategy, given that the other players' strategies are fixed. We define the *social optimum* as a strategy $Y(\cdot)$ that has the minimum cost, over the space of all possible strategies—this is not necessarily (and is not usually) a pure NE. Therefore, the cost of a pure NE relative to the social cost is an important measure, and the maximum such ratio over all possible pure NE is known as the *price of anarchy* (Koutsoupias and Papadimitriou 1999).

Source probability. For simplifying our discussion, henceforth, we will assume that the source is chosen uniformly at random, i.e., $p_s = \frac{1}{n}$ for all $s \in V$. Most of our results extend to general source distributions.

Characterization of Nash Equilibria

First we will define valid strategy.

Definition 1. A strategy Y is valid if it satisfies the budget constraints, i.e., $\sum_v Y(v, \cdot, 0) \leq B^0$ and $\forall t \in \mathcal{T} \setminus \{0\}$, $\sum_v Y(v, s, t) \leq B^t$.

From the definition of BUDGETTEMPVACC, $Y(v, s, T) = 1$ implies that a vaccine is “reserved” for node v if s is the source. However, even if s is the source, it is possible that it is secure (i.e., $Y(s, \cdot, 0) = 1$) or the infection never reaches v (because of other nodes that chose to be vaccinated at time T). In both cases, v will not incur the cost of C_v/n . This notion has important implications for the structure of NE, in the

sense that it does not hurt a node v to choose $Y(v, s, T) = 1$ if the budget constraints allow for this.

We now discuss a characterization of the pure NE for BUDGETTEMPVACC for the special case where $\mathcal{T} = \{0, T\}$. We start with some definitions. Let $G[V']$ be the subgraph of G induced by the set V' of nodes. Let $V_0(Y) = \{v : Y(v, \cdot, 0) = 1\}$ and $V_s(Y) = \{v : Y(v, s, T) = 1\}$ be the sets of nodes that are vaccinated at time 0 and time T when the source is s , respectively. Let $C(v, G')$ be the connected component containing node v in the graph G' . For $v, s \in V$, let $I(v, s, G[V - V_0(Y) - V_s(Y)]) = 1$ if $s \in C(v, G')$, $Y(v, \cdot, 0) = 0$ and $Y(s, \cdot, 0) = 0$, i.e., $I(\cdot)$ is the indicator of the event “ v will be infected if s is the source.” Let $d(x, y, G)$ denote the distance of x from y in G . Let $A(v, Y)$ be the set of potential sources s for which it is still possible for node v to get vaccinated at time T , i.e. $A(v, Y) = \{s \mid Y(v, s, T) = 0, |V_s(Y)| < B^T, d(v, s, G[V - V_0(Y) - V_s(Y)]) \geq T, I(v, s, G[V - V_0(Y) - V_s(Y)]) = 1\}$.

Lemma 2. Consider an instance $(G, \mathbf{p}, \mathcal{T} = \{0, T\}, \mathcal{B}, \mathcal{C}, \mathcal{L})$ of BUDGETTEMPVACC. A valid strategy $Y(\cdot)$ is a pure NE iff

1. If $Y(v, \cdot, 0) = 0$ then either (a) $\sum_s I(v, s, G[V - V_0(Y) - V_s(Y)]) \frac{L_v}{n} \leq C_v$ or (b) $|V_0| = B^0$.
2. If $Y(v, \cdot, 0) = 1$ then the following conditions hold: (a) $\sum_s I(v, s, G[V - V_0(Y) - V_s(Y) \cup \{v\}]) \frac{L_v}{n} > C_v$, and (b) $\sum_{s \in V - A(v, Y)} I(v, s, G[V - V_0(Y) - V_s(Y)]) \frac{L_v}{n} > C_v$.
3. If $Y(v, s, T) = 0$ and $Y(v, \cdot, 0) = 0$ then either (a) $I(v, s, G[V - V_0(Y) - V_s(Y)]) = 0$ or (b) $|V_s(Y)| = B^T$.

The proof is in the full version (Adiga and Vullikanti 2015).

Existence and Complexity of finding NE

In this section, we first propose a best-response algorithm to compute NE (if it exists) for the special case $\mathcal{T} = \{0, T\}$. This will be followed by results on minimum cost pure NE and the price of anarchy.

Best response strategy: Consider an instance $(G, \mathbf{p}, \mathcal{T} = \{0, T\}, \mathcal{B}, \mathcal{C}, \mathcal{L})$ of BUDGETTEMPVACC.

1. We start with a feasible strategy $Y(\cdot)$ (one possibility is to start with all nodes insecure).
2. If there exists node v such that $Y(v, \cdot, 0) = 0$ and $\sum_s I(v, s, G[V - V_0(Y) - V_s(Y)]) \frac{L_v}{n} > C_v$, then:
 - (a) If $|V_0| < B^0$, then switch the strategy of node v : set $Y(v, \cdot, 0) = 1$. Set $Y(v, s, T) = 0$ for all s .
 - (b) If $|V_0| = B^0$: For each $s \in A(v, Y)$, we set $Y(v, s, T) = 1$.
3. If there exists node v such that $Y(v, \cdot, 0) = 1$ and $\sum_{s \in V - A(v, Y)} I(v, s, G[V - V_0(Y) - V_s(Y) \cup \{v\}]) \frac{L_v}{n} \leq C_v$, we switch the strategy of node v and set $Y(v, \cdot, 0) = 0$. Then, for $s \in A(v, Y)$, we set $Y(v, s, T) = 1$.

When there are no budget constraints a pure NE need not exist (Adiga, Venkataraman, and Vullikanti 2016). However, for $B^T = 0$ and uniform vaccination and infection costs, it

can be shown that the best response strategy always converges to a pure NE. The proof is based on a potential function argument by (Aspnes, Chang, and Yampolskiy 2006), where they prove the same result for the special case where B^0 is unbounded. The same proof holds for $B^0 < n$.

Lemma 3. *For an instance of BUDGETTEMPVACC, where (1) $\mathcal{T} = \{0, T\}$, and (2) vaccination and infection costs are uniform for all the nodes (i.e., there exist c, L such that $C_v = C$ and $L_v = L$ for all $v \in V$), and (3) $B^T = 0$ for all $t \in \mathcal{T}$, the above algorithm converges to a pure NE.*

In Experimental Results section, we discuss the performance of the best response strategy on several networks.

Minimum cost NE. By a reduction from the Firefighter problem (Finbow et al. 2007), we showed that it is hard to compute the minimum cost NE.

Lemma 4. *Finding a minimum cost pure NE is NP-complete even on instances of BUDGETTEMPVACC where G is a tree.*

The proof is in the full version (Adiga and Vullikanti 2015).

Price of anarchy (PoA). When there are no budget constraints, it was shown in (Adiga, Venkataraman, and Vullikanti 2016) that the price of anarchy can be $\Omega(n)$. Clearly this holds for BUDGETTEMPVACC as well. However, with the budget constraints, the PoA is $\Omega(n)$ even for trees.

Lemma 5. *There exist instances $\mathcal{I} = (G, \mathbf{p}, \mathcal{T}, \mathcal{B}, \mathbf{C}, \mathbf{L})$ of BUDGETTEMPVACC, where G is a tree with root s , $p_s = 1$ and $B^0 = 0$ for which the price of anarchy is $\Omega(n)$.*

Effect of budget constraints

We now show that the budget constraints have a very significant impact on the cost of a NE, as well as on the social optimum.

Lemma 6. *There exist instances $\mathcal{I} = (G, \mathbf{p}, \mathcal{T}, \mathcal{B}, \mathbf{C}, \mathbf{L})$ of BUDGETTEMPVACC, for which there exist strategies $Y(\cdot)$ and $Y'(\cdot)$ such that:*

1. $Y(\cdot)$ is a NE for \mathcal{I} that satisfies the budget constraints,
2. $Y'(\cdot)$ is a NE for the instance \mathcal{I}' obtained by removing all the budget constraints for $T \neq 0$ in \mathcal{T} in instance \mathcal{I} (i.e., setting $B^T = \infty$ for all $T \in \mathcal{T}, T \neq 0$), and
3. $\text{cost}(Y)/\text{cost}(Y') = \Theta(n)$.

The proof is in the full version (Adiga and Vullikanti 2015).

Extending the instance constructed in the above proof, we can see that in the absence of budget constraints, for each random source s , all nodes at distance T from s choose to get vaccinated, where $T = \min\{t \in \mathcal{T} : t > 0\}$ is the first time in \mathcal{T} when the vaccination decisions can be made. As a result, it suffices to consider only one time step in \mathcal{T} at which the vaccination game needs to be played, as was observed in (Adiga, Venkataraman, and Vullikanti 2016). In contrast, because of the budget constraints in an instance of BUDGETTEMPVACC, vaccination decisions might be made in each round. Further, nodes far away from the source might choose to get vaccinated, which might cause a large outbreak. Further, it does not help to increase the budget at time $t = 0$ alone, as discussed in the next lemma.

Lemma 7. *There exist instances $\mathcal{I} = (G, \mathbf{p}, \mathcal{T}, \mathcal{B}, \mathbf{C}, \mathbf{L})$ of BUDGETTEMPVACC, with $B^0 = \infty$, for which removing the budget constraint B^T (i.e., setting $B^T = \infty$) can reduce the cost of NE by a factor of $\Theta(n)$.*

Finding the social optimum

Since BUDGETTEMPVACC generalizes the formulation studied in (Adiga, Venkataraman, and Vullikanti 2016), it is easy to verify that finding the social optimum is NP-complete. We show that even approximating the social optimum of BUDGETTEMPVACC is very hard. This motivates bicriteria approximation algorithms, in which the budget constraints can be violated.

Theorem 8. *The social optimum of an instance $(G, \mathbf{p}, \mathcal{T}, \mathcal{B}, \mathbf{C}, \mathbf{L})$, of BUDGETTEMPVACC cannot be approximated within a factor of n^α in polynomial time for $\alpha \in (0, 1)$, unless $P = NP$.*

The proof is by a reduction from a game-theoretic version of the FireFighter problem studied in (Anshelevich et al. 2009) (see in the full version (Adiga and Vullikanti 2015)).

Bicriteria Approximation Algorithm

Approximating the social optimum. We now discuss an approximation algorithm for computing the social optimum for the special case of $\mathcal{T} = \{0, T\}$. Our algorithm builds on (Hayrapetyan et al. 2005), and involves a linear-programming (LP) rounding approach. We use the following notation below: (1) $x(j, s)$ is an indicator variable, which is 1 if node j gets infected due to source s ; (2) $y(j)$ is an indicator which is 1 if node j is vaccinated at time 0; (3) $y(j, s)$ is an indicator which is 1 if node j is vaccinated at time T , when the source is s ; (4) $P^d(s, j)$ is the set of paths from s to j of length d ; (5) $P^{<d}(s, j)$ is the set of paths from s to j of length less than d .

We now describe an integer programming formulation \mathcal{P} for the social optimum below. Let, $f(\mathbf{x}, \mathbf{y}) = \sum_v C_v^0 y(v) + \frac{1}{n} \sum_{s,v} C_v^T y(v, s) + \frac{1}{n} \sum_{s,j} L_j x(j, s)$.

$$\min_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) \text{ such that}$$

$$\sum_{v \in p} y(v) \geq 1 - x(j, s), \forall s, j, \forall p \in P^{<T}(s, j) \quad (1)$$

$$\sum_{v \in p} y(v) + y(j, s) \geq 1 - x(j, s), \forall s, j, \forall p \in P^T(s, j) \quad (2)$$

$$y(j) + y(j, s) \geq x(u, s) - x(j, s), \forall u \in N(j), \forall s, j \quad (3)$$

$$x(s, s) = 1 - y(s)$$

$$\sum_v y(v) \leq B^0 \quad (4)$$

$$\sum_v y(v, s) \leq B^t, \forall s, t \quad (5)$$

$$y(v), y(v, s), x(i, j) \in \{0, 1\}, \forall s, v, i, j \in V \quad (6)$$

Constraint (1) ensures that if a node j is within T hops of a source s , if node j does not get infected due to source

s , then at least some node on every path from s to j of path length less than T has a vaccinated node. In constraint (2) for paths of length T , node j could be vaccinated at time T to ensure the same effect. Constraint (3) ensures that if j is not vaccinated at either 0 or T , then, it gets infected if any of its neighbors does. Budget constraints are captured by (4) and (5).

Lemma 9. *The optimum solution (\mathbf{x}, \mathbf{y}) to the program \mathcal{P} is the social optimum of the BUDGETTEMPVACC instance.*

We describe an approximation algorithm based on relaxing and rounding the solution of \mathcal{P} for the special case where $B^0 = 0$. For $A \subseteq V$, let $N(A) = \{v \notin A \mid \exists u \in A, u \text{ is adjacent to } v\}$. The algorithm involves the following steps:

1. Solve a linear relaxation \mathcal{P}_R with the constraints $y(i), y(i, j), x(i, j) \in [0, 1]$, for all i, j . Let $\mathbf{x}^f, \mathbf{y}^f$ denote the fractional solution to this relaxed program.
2. We round $\mathbf{x}^f, \mathbf{y}^f$ to an integral solution in the following manner:
 - (a) For each j, s , define $x'(j, s) = \min\{2x^f(j, s), 1\}$, $y'(j) = \min\{2y^f(j), 1\}$, $y'(j, s) = \min\{2y^f(j, s), 1\}$.
 - (b) For each j , if $y'(j) \geq \frac{1}{T-1}$, set $Y(j, \cdot, 0) = 1$.
 - (c) For each j, s , if $y'(j, s) \geq \frac{1}{T}$, set $Y(j, s, T) = 1$.
 - (d) For each $s \in V$, pick “suitable” $r_s \in [0, 1]$ (as discussed in the proof of Theorem 10)
 - i. Let $A(s, r_s) = \{j : x'(j, s) \geq r_s\}$ and $C(s, r_s) = N(A(s, r_s)) - A(s, r_s)$.
 - ii. For each $j \in C(s, r_s)$, define $Y(j, s, T) = 1$.

We show below that this randomized strategy $Y(\cdot)$ satisfies the budget constraints approximately, and $\text{cost}(Y)$ is within a constant factor of the fractional LP value. Ensuring that all the budget constraints are satisfied is a challenging open problem.

Theorem 10. *Suppose $B^0 = B^T = B$. The solution $Y(\cdot)$ constructed by the above algorithm satisfies the following properties: (1) $\sum_v Y(v, \cdot, 0) \leq (2T) \cdot B$, (2) for each s , $\sum_v Y(v, s, T) \leq (8T) \cdot B$, (3) $\text{cost}(Y) \leq (8T) \cdot f(\mathbf{x}^f, \mathbf{y}^f)$.*

Experimental Results

We now explore the structure of the NE obtained through the best response strategy. In particular, we study how it behaves with respect to the budget at time 0, B_0 , time T , the budget at time T , B_T , and the vaccination cost C . The algorithm was applied on three synthetic graphs: Erdős-Rényi with 100 nodes and average degree 7 (ER); random power-law graph generated using the Chung-Lu model with power-law index $\gamma = 2.5$, 93 nodes and average degree 4 (CL), and a random regular graph with 100 nodes and average degree 4 (RR). We chose these networks because they are commonly used in social network analysis (see, e.g., (Newman 2003)), and together model both homogeneous and heterogeneous degree networks. Recall that the best response algorithm requires computing $I(v, s, Y)$ each time a node changes its strategy. For this we need to use the all-pairs shortest path

algorithm. Therefore, we study relatively small networks here.

To keep the framework simple, we assumed uniform vaccination cost $C < 1$ and infection cost $L = 1$ for all nodes. We ran the best response algorithm for various values of B_0 , B_T , T , and C . The results presented are averaged across 20 iterations, each producing a NE. Where the results are consistent across networks, plots are shown for only CL network. The remaining plots are in the full version (Adiga and Vullikanti 2015).

Structure of NE with respect to B_0 and B_T . The results are shown in Fig. 2. We used two criteria to study the effects of B_0 and B_T , viz. average cost of the NE ($\text{cost}(Y)$) and the average number of nodes which vaccinate at time 0 ($\mathbb{E}[|V_0|]$). Comparing column 1 to column 3, we note that across all networks, the effect of B_0 on the $\text{cost}(Y)$ is more pronounced than that of B_T . The plots in column 2 are quite interesting in the sense that they exhibit a threshold phenomenon. We note that $\mathbb{E}[|V_0|]$ increases steadily (possibly linearly) with B_0 and then tapers off. The point at which it hits a plateau is a function of B_T . Qualitatively, these plots indicate that greater the guarantee of vaccine availability at a later stage, the higher the inclination to postpone vaccination. Possible directions from here are to determine when these regime changes occur, exploring the role of network structure in this behavior.

Effect of vaccination cost C . We observe that the vaccination cost has a significant influence on the average cost of the NE—see Fig. 3 first plot, which shows the average cost of solution vs. B_0 for different values of T and B_T . We have only shown results for CL due to lack of space.

Convergence time. We observed that the best response algorithm generally converges quickly and shows little variation with budget constraints (Fig. 3 second plot). However, it would be interesting to see how it performs as the network size increases.

Effect of time T . From Fig. 4, we observe that the effect of T depends on B^T ; higher the B^T , the more sensitive is the cost to T . Note that when B^T is small, even if the second stage of vaccination happens early, it is not enough to secure enough nodes from the source. A possible future direction is to quantify this effect based on network properties such as density, expansion, etc.

Related Work

There is a large literature on vaccination games, which can be broadly partitioned into differential equation based or network based models. We summarize them below.

One of the earliest works on vaccination games is by (Bauch and Earn 2004), who use a differential equation based model to study vaccination decisions made at the start of the epidemic. They show that the vaccinated fraction in NE can be expressed in terms of the reproductive number, R_0 , which is the expected number of secondary infections caused by an infected individual. Vaccination is a special kind of behavioral change in the context of epidemics. There has been a lot of research on different kinds of game formulations arising out of epidemic

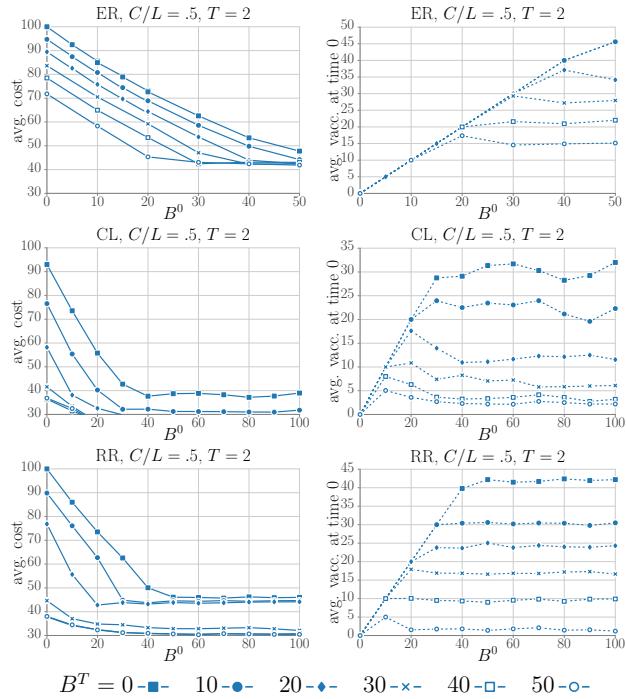


Figure 2: The effect of budget constraints on the structure of NE from the best response algorithm: Each row corresponds to a particular network. The first column is a plot of average cost of the NE vs. B^0 , for various values of B^T . The second column corresponds to similar plots for average number of nodes vaccinating at time 0.

behaviors, e.g., (Bauch and Earn 2004; Khuzani, Sarkar, and Altman 2012; Khuzani, Altman, and Sarkar 2012; Khuzani, Sarkar, and Altman 2011; Galvani, Reluga, and Chapman 2007; Reluga 2010; Reluga and Galvani 2011; Trajanovski et al. 2015; Manfredi and D’Onofrio 2013). Reluga et al. (Reluga and Galvani 2011) develop a general approach that combines population games with Markov decision process, and allows decisions at different times. There has also been work on repeated game formulations that take information and past experience into account. (Conforth et al. 2011) study the effects of vaccination decisions based on past epidemics, and find that individuals with number of contacts above a threshold get vaccinated, whereas individuals with fewer than these many contacts do not, leading to erratic flu behavior. (Bauch and Bhattacharyya 2012) use an evolutionary game theory based model to study the feedback between

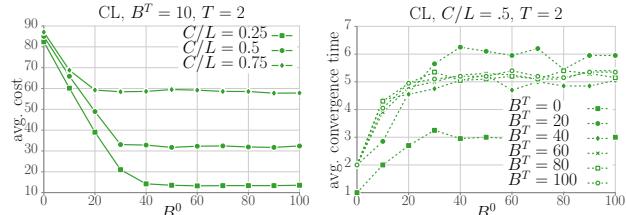


Figure 3: Best response algorithm on Chung-Lu graph CL: Plot (1) corresponds to effect of C on the average cost. Plot (2) shows how average convergence time varies with respect to budget constraints.

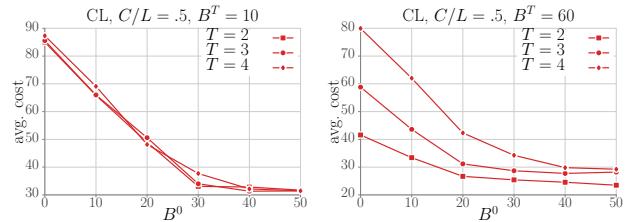


Figure 4: The effect of T on the cost of NE. Plots (1) and (2) show cost vs. B^0 for $B^T = 10$ and 60 respectively.

disease prevalence and vaccinating behavior, especially in the context of vaccine scares.

A different area of research involves the use of network based models. One of the earliest works in this direction is (Aspnes, Rustagi, and Saia 2007). They formalize the network version of the vaccination game problem of (Bauch and Earn 2004), and design algorithms for computing NE and the social optimum. This turns out to be a challenging problem, and they develop an approximation algorithm for the social optimum, that is based on linear programming rounding. This approximation bound was improved by (Chen, David, and Kempe 2010; Kumar et al. 2010). All these approaches have assumed a simplistic model of epidemic spread, which models highly contagious diseases. There has been work on more realistic SIS (Susceptible-Infectious-Susceptible) models, e.g., (Omic, Orda, and Mieghem 2009; Trajanovski et al. 2015). An alternative approach was studied by (Saha, Adiga, and Vullikanti 2014), who develop a formulation based on the spectral radius (the first eigenvalue of the network), in which the utility is based on whether or not the spectral radius is above a threshold or not. One limitation of all of these works is that they only consider vaccination decisions at the start of the epidemic, in a one-shot simultaneous game formulation. The work most closely related to our paper is by (Adiga, Venkataraman, and Vullikanti 2016), who formalize a temporal vaccination game, in which vaccination decisions can be done at multiple times. However, there are no budget constraints in their formulation. As mentioned earlier, this has a very significant impact on both the structure and complexity of the problem.

Conclusions

In this paper, we formalize the temporal vaccination problem with resource constraints on complex networks, as a repeated game **BUDGETTEMPVACC**. This captures many realistic aspects of epidemic spread in real networks. The budget constraints have a very significant impact on the structure of the game solutions, as well as on the complexity of finding equilibria. A significant fraction of nodes that do get vaccinated, choose to defer their vaccination decision—the specific effects depend on the network structure, the budgets, and the vaccination and infection costs. Computing properties of such repeated games becomes more challenging, compared to those of standard vaccination games. Some of the interesting open problems that arise out of our work include: (1) understanding networks in which pure NE exist and can be computed efficiently, and (2) extending our results to other epidemic models, and improving the algo-

rithmic bounds. This work has been partially supported by the following grants: DTRA Grant HDTRA1-11-1-0016, DTRA CNIMS Contract HDTRA1-11-D-0016-0010, NSF ICES CCF-1216000, NSF NETSE Grant CNS-1011769, NIH MIDAS Grant 5U01GM070694, NSF DIBBS Grant ACI-1443054.

References

- Adiga, A., and Vullikanti, A. K. S. 2015. Temporal vaccination games under resource constraints. http://staff.vbi.vt.edu/abhijin/papers/aaai16_budget_temporalvacc.pdf. Full version.
- Adiga, A.; Venkataraman, S.; and Vullikanti, A. 2016. To delay or not: temporal vaccination games on networks. Accepted in INFOCOM.
- Anshelevich, E.; Chakrabarty, D.; Hate, A.; and Swamy, C. 2009. Approximation algorithms for the firefighter problem: Cuts over time and submodularity. In Dong, Y.; Du, D.-Z.; and Ibarra, O., eds., *Algorithms and Computation*, volume 5878 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 974–983.
- Aspnes, J.; Chang, K.; and Yampolskiy, A. 2006. Inoculation strategies for victims of viruses and the sum-of-squares partition problem. *J. Comput. Syst. Sci.*
- Aspnes, J.; Rustagi, N.; and Saia, J. 2007. Worm versus alert: Who wins in a battle for control of a large-scale network? *OPODIS*.
- Bauch, C., and Bhattacharyya, S. 2012. Evolutionary game theory and social learning can determine how vaccine scares unfold. *PLoS Comput Biol.*
- Bauch, C., and Earn, D. 2004. Vaccination and the theory of game. *PNAS*.
- CDC. Vaccine shortages and delays. http://www.cdc.gov/vaccines/vac-gen/shortages/default.htm?s_cid=cs_000.
- Chen, P.; David, M.; and Kempe, D. 2010. Better vaccination strategies for better people. In *Proceedings 11th ACM Conference on Electronic Commerce (EC-2010), Cambridge, Massachusetts, USA, June 7-11, 2010*, 179–188.
- Conforth, D. M.; Reluga, T. C.; Shim, E.; Bauch, C. T.; Galvani, A. P.; and Meyers, L. A. 2011. Erratic flu vaccination emerges from short-sighted behavior in contact networks. *PLoS Computational Biology*.
- Easley, D., and Kleinberg, J. 2010. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press.
- Finbow, S.; King, A.; MacGillivray, G.; and Rizzi, R. 2007. The firefighter problem for graphs of maximum degree three. *Discrete Mathematics* 307(16):2094 – 2105.
- Galvani, A.; Reluga, T.; and Chapman, G. 2007. Long-standing influenza vaccination policy is in accord with individual self-interest but not with the utilitarian optimum. *PNAS* 104(13):5692–5697.
- Grossklags, J.; Christin, N.; and Chuang, J. 2008. Secure or insure? a game-theoretic analysis of information security games. In *World Wide Web Conference (WWW)*.
- Hayrapetyan, A.; Kempe, D.; Pál, M.; and Svitkina, Z. 2005. Unbalanced graph cuts. In *Algorithms - ESA 2005, 13th Annual European Symposium, Palma de Mallorca, Spain, October 3-6, 2005, Proceedings*, 191–202.
- Khouzani, M. H. R.; Altman, E.; and Sarkar, S. 2012. Optimal quarantining of wireless malware through reception gain control. *IEEE Trans. Automat. Contr.* 57(1):49–61.
- Khouzani, M. H. R.; Sarkar, S.; and Altman, E. 2011. A dynamic game solution to malware attack. In *INFOCOM*, 2138–2146.
- Khouzani, M. H. R.; Sarkar, S.; and Altman, E. 2012. Saddle-point strategies in malware attack. *IEEE Journal on Selected Areas in Communications* 30(1):31–43.
- Koutsoupias, E., and Papadimitriou, C. H. 1999. Worst-case equilibria. In *Proceedings of STACS*.
- Kumar, V. S. A.; Rajaraman, R.; Sun, Z.; and Sundaram, R. 2010. Existence theorems and approximation algorithms for generalized network security games. In *Proceedings of IEEE ICDCS*.
- Leyton-Brown, K., and Shoham, Y. 2008. *Essentials of Game Theory: A Concise, Multidisciplinary Introduction*. Morgan and Claypool Publishers, 1st edition.
- Manfredi, P., and D’Onofrio, A. 2013. *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases*. Springer.
- Molinari, N.-A. M.; Ortega-Sánchez, I. R.; Messonnier, M. L.; Thompson, W. W.; Wortley, P. M.; Weintraub, E.; and Bridge, C. B. 2007. The annual impact of seasonal influenza in the US: Measuring disease burden and costs. *Vaccine* 25(27):5086–5096.
- Newman, M. 2003. The structure and function of complex networks. *SIAM Review* 45.
- Omic, J.; Orda, A.; and Mieghem, P. V. 2009. Protecting against network infections a game theoretic perspective. In *INFOCOM*.
- Orenstein, W. A., and Schaffner, W. 2008. Lessons learned: Role of influenza vaccine production, distribution, supply, and demandwhat it means for the provider. *The American Journal of Medicine* S22–S27.
- Reluga, T., and Galvani, A. 2011. A general approach to population games with application to vaccination. *Mathematical Biosciences*.
- Reluga, T. 2010. Game theory of social distancing in response to an epidemic. *PLOS Computational Biology* 6(5). e1000793.
- Saha, S.; Adiga, A.; and Vullikanti, A. K. S. 2014. Equilibria in epidemic containment games. In *The 28th AAAI Conference on Artificial Intelligence (AAAI)*.
- Trajanovski, S.; Hayel, Y.; Altman, E.; Wang, H.; and Van Mieghem, P. 2015. Decentralized protection strategies against SIS epidemics in networks. *Control of Network Systems, IEEE Transactions on PP(99)*:1–1.