

Sune Lehmann  
Yong-Yeol Ahn *Editors*

# Complex Spreading Phenomena in Social Systems

Influence and Contagion in Real-World  
Social Networks

# **Computational Social Sciences**

# Computational Social Sciences

---

A series of authored and edited monographs that utilize quantitative and computational methods to model, analyze and interpret large-scale social phenomena. Titles within the series contain methods and practices that test and develop theories of complex social processes through bottom-up modeling of social interactions. Of particular interest is the study of the coevolution of modern communication technology and social behavior and norms, in connection with emerging issues such as trust, risk, security and privacy in novel socio-technical environments.

Computational Social Sciences is explicitly transdisciplinary: quantitative methods from fields such as dynamical systems, artificial intelligence, network theory, agent based modeling, and statistical mechanics are invoked and combined with state-of-the-art mining and analysis of large data sets to help us understand social agents, their interactions on and offline, and the effect of these interactions at the macro level. Topics include, but are not limited to social networks and media, dynamics of opinions, cultures and conflicts, socio-technical coevolution and social psychology. Computational Social Sciences will also publish monographs and selected edited contributions from specialized conferences and workshops specifically aimed at communicating new findings to a large transdisciplinary audience. A fundamental goal of the series is to provide a single forum within which commonalities and differences in the workings of this field may be discerned, hence leading to deeper insight and understanding.

## Series Editors

Elisa Bertino Purdue University, West Lafayette, IN, USA	Claudio Cioffi-Revilla George Mason University, Fairfax, VA, USA	Jacob Foster University of California, Los Angeles, CA, USA	Nigel Gilbert University of Surrey, Guildford, UK	Jennifer Golbeck University of Maryland, College Park, MD, USA	Bruno Gonçalves New York University, New York, NY, USA	James A. Kitts Columbia University, Amherst, MA, USA	Larry S. Liebovitch Queens College, City University of New York, Flushing, NY, USA	Sorin A. Matei Purdue University, West Lafayette, IN, USA	Anton Nijholt University of Twente, Enschede, The Netherlands	Andrzej Nowak University of Warsaw, Warsaw, Poland	Robert Savit University of Michigan, Ann Arbor, MI, USA	Flaminio Squazzoni University of Brescia, Brescia, Italy	Alessandro Vinciarelli University of Glasgow, Glasgow, Scotland, UK
--	--	---	--	--	--	--	--	---	---	---	---	---	---

More information about this series at <http://www.springer.com/series/11784>

Sune Lehmann • Yong-Yeol Ahn  
Editors

# Complex Spreading Phenomena in Social Systems

Influence and Contagion in Real-World Social  
Networks



Springer

*Editors*

Sune Lehmann  
Technical University of Denmark  
Lyngby, Denmark

Yong-Yeol Ahn  
Indiana University  
Bloomington, IN, USA

ISSN 2509-9574  
Computational Social Sciences  
ISBN 978-3-319-77331-5  
<https://doi.org/10.1007/978-3-319-77332-2>

ISSN 2509-9582 (electronic)  
ISBN 978-3-319-77332-2 (eBook)

Library of Congress Control Number: 2018941566

© Springer International Publishing AG, part of Springer Nature 2018

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature.

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Contents

## Part I Introduction to Spreading in Social Systems

<b>Complex Contagions: A Decade in Review .....</b>	<b>3</b>
Douglas Guilbeault, Joshua Becker, and Damon Centola	
<b>A Simple Person’s Approach to Understanding the Contagion Condition for Spreading Processes on Generalized Random Networks ...</b>	<b>27</b>
Peter Sheridan Dodds	
<b>Challenges to Estimating Contagion Effects from Observational Data ....</b>	<b>47</b>
Elizabeth L. Ogburn	

## Part II Models and Theories

<b>Slightly Generalized Contagion: Unifying Simple Models of Biological and Social Spreading .....</b>	<b>67</b>
Peter Sheridan Dodds	
<b>Message-Passing Methods for Complex Contagions .....</b>	<b>81</b>
James P. Gleeson and Mason A. Porter	
<b>Optimal Modularity in Complex Contagion .....</b>	<b>97</b>
Azadeh Nematzadeh, Nathaniel Rodriguez, Alessandro Flammini, and Yong-Yeol Ahn	
<b>Probing Empirical Contact Networks by Simulation of Spreading Dynamics .....</b>	<b>109</b>
Petter Holme	
<b>Theories for Influencer Identification in Complex Networks .....</b>	<b>125</b>
Sen Pei, Flaviano Morone, and Hernán A. Makse	

**Part III Observational Studies**

<b>Service Adoption Spreading in Online Social Networks .....</b>	151
Gerardo Iñiguez, Zhongyuan Ruan, Kimmo Kaski, János Kertész, and Márton Karsai	
<b>Misinformation Spreading on Facebook.....</b>	177
Fabiana Zollo and Walter Quattrociocchi	
<b>Scalable Detection of Viral Memes from Diffusion Patterns .....</b>	197
Pik-Mai Hui, Lilian Weng, Alireza Sahami Shirazi, Yong-Yeol Ahn, and Filippo Menczer	
<b>Attention on Weak Ties in Social and Communication Networks .....</b>	213
Lilian Weng, Márton Karsai, Nicola Perra, Filippo Menczer, and Alessandro Flammini	
<b>Measuring Social Spam and the Effect of Bots on Information Diffusion in Social Media.....</b>	229
Emilio Ferrara	
<b>Network Happiness: How Online Social Interactions Relate to Our Well Being .....</b>	257
Johan Bollen and Bruno Gonçalves	
<b>Information Spreading During Emergencies and Anomalous Events.....</b>	269
James P. Bagrow	
<b>Part IV Controlled Studies</b>	
<b>Randomized Experiments to Detect and Estimate Social Influence in Networks .....</b>	289
Sean J. Taylor and Dean Eckles	
<b>The Rippling Effect of Social Influence via Phone Communication Network .....</b>	323
Yan Leng, Xiaowen Dong, Esteban Moro, and Alex ‘Sandy’ Pentland	
<b>Network Experiments Through Academic-Industry Collaboration.....</b>	335
Robert M. Bond, Christopher J. Fariss, Jason J. Jones, and Jaime E. Settle	
<b>Spreading in Social Systems: Reflections .....</b>	351
Sune Lehmann and Yong-Yeol Ahn	
<b>Index .....</b>	359

**Part I**

**Introduction to Spreading in Social  
Systems**

# Complex Contagions: A Decade in Review



Douglas Guilbeault, Joshua Becker, and Damon Centola

## 1 Introduction

Most collective behaviors spread through social contact. From the emergence of social norms, to the adoption of technological innovations, to the growth of social movements, social networks are the pathways along which these “social contagions” propagate. Studies of diffusion dynamics have demonstrated that the structure (or topology) of a social network can have significant consequences for the patterns of collective behavior that will emerge.

Over the last 45 years, questions about how the structure of social networks affects the dynamics of diffusion have been of increasing interest to social scientists. Granovetter’s [1] “Strength of Weak Ties” study ushered in an era of unprecedented interest in how network dynamics, and in particular diffusion on networks, affect every aspect of social life, from the organization of social movements to school segregation to immigration. Granovetter’s study showed that “weak ties” between casual acquaintances can be much more effective in promoting diffusion and social integration than “strong ties” between close friends. This is because although casual friendships are relationally weak, they are more likely to be formed between socially distant actors with few network “neighbors” in common. These “long ties” between otherwise distant nodes provide access to new information and greatly increase the rate at which information propagates, despite the relational weakness of the tie as a conduit.

In the last two decades, the explosion of network science across disciplines such as physics, biology, and computer science has produced many important advances for understanding how the structure of social networks affect the dynamics

---

D. Guilbeault · J. Becker · D. Centola (✉)

Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA, USA

e-mail: [dcentola@asc.upenn.edu](mailto:dcentola@asc.upenn.edu)

of diffusion. The full impact of Granovetter's original insight was not realized until Watts and Strogatz's [2] "small world" model demonstrated that bridge ties connecting otherwise distant nodes can dramatically increase the rate of propagation across a network by creating "shortcuts" between remote clusters. Introducing "long ties" into a network can give even highly clustered networks the "degrees of separation" characteristic of a small world. This model of network dynamics has had a tremendous impact on fields as diverse as computer science, physics, epidemiology, sociology, and political science.

Building on the idea of pathogenic contagions, this research combines the diverse domains of ideas, information, behaviors, and diseases into the generic concept of a universal contagion. The attractive implication is that the mathematical tools developed by epidemiologists for studying the spread of disease can be generically used to study the dynamics of social, cultural, and political change. In particular, the properties of social networks that have been shown to accelerate the spreading dynamics of disease diffusion—such as small world topologies, weak ties, and scale-free degree distributions—can also be used to make inferences about the role of networks in the domains of social and political behavior. Regardless of whether a given contagion is a prophylactic measure to prevent HIV infection or the HIV infection itself, Granovetter's groundbreaking claim was that "whatever is to be diffused can reach a larger number of people, and traverse a greater social distance, when passed through weak ties rather than strong" [1], p. 1366).

However, while this theory is useful for understanding the rapid spread of HIV infections through networks of weak ties, it has not shed light on the remarkable failure of these same networks to spread prophylactic measures for preventing HIV [3]. The reason for this disturbing asymmetry between the spread of infectious diseases and the diffusion of preventative measures is that infectious diseases are typically *simple contagions*—that is, contagions for which a single activated source can be sufficient for transmission—while preventive measures are typically *complex contagions*, that is, behaviors, beliefs, or attitudes for which transmission requires contact with multiple sources of activation. While repeated contact with the same person can increase the likelihood of transmitting a simple contagion, the transmission of complex contagions requires reinforcement from several contacts. Any social contagion that is costly, difficult, or unfamiliar is likely to be a complex contagion, requiring social reinforcement to spread.

The primary consequence of the distinction between simple and complex contagions for diffusion through social networks is that as "worlds" become very small, the speed of simple contagions increases, while complex contagions become harder to spread. As Centola and Macy write,

For simple contagions, too much clustering means too few long ties, which slows down cascades. For complex contagions, too little clustering means too few wide bridges, which not only slows down cascades but can prevent them entirely (2007, p. 723).

Centola and Macy [4] identify several reasons why contagions may be complex, including the need for social legitimization, the need for credibility, or the complementarity of a behavior. For instance, a contagion might be complex due

to externalities, in which the value of the contagion increases with the number of adopters. The value of a communication technology such as a fax machine rests heavily on the number of people who use it. When only one person has a fax machine, it holds no value. A single contact with someone who has a fax machine provides little reason for someone else to adopt it. Even if the adopter provides repeated signals, a single person alone cannot do much to increase the complementary value of the fax machine. However, if a potential adopter comes into contact with several independent sources who have all adopted fax machines, the complementary value of the technology increases. After exposure to a sufficient number of reinforcing contacts, a person with no inherent interest in fax machines can be convinced that it is a necessary investment.

A different kind of reason why a contagion might be complex is due to uncertainty. For instance, physicians are often resistant to adopting new medical technologies for fear of placing themselves at risk of acting outside accepted protocols. Early studies on adoption patterns among physicians found that physicians were unlikely to adopt a new medical technology, even though it had been formally approved and was expected to be very effective, until they observed several of their colleagues using it [5]. For similar reasons, complexity in diffusion can also be a result of normative pressures. This is often the case with the diffusion of managerial practices among elite firms. Because the choice of corporate governance strategy can impact the reputation of a firm, the adoption of new practices is often dependent upon social reinforcement from competing firms within the same industry. Corporate boards concerned about the risk of social sanction are often unwilling to adopt new managerial practices until they have already seen them adopted by several peer institutions [6].

In the last decade, the literature on complex contagions has rapidly evolved both empirically and theoretically. In this review, we discuss recent developments across four empirical domains: health, innovation, social media, and politics. Each domain adds new complexities to our understanding of how contagions emerge, spread, and evolve, as well as how they undergird fundamental social processes. We also discuss how these empirical studies have spurred complementary advancements in the theoretical modeling of contagions, which concern the effects of network topology on diffusion, as well as the effects of variation in threshold dynamics. Importantly, while the empirical studies reviewed in this paper complement existing theoretical work, they also present many opportunities for new theoretical extensions. We suggest three main directions for future development of research on complex contagions. The first concerns the study of how multiple contagions interact within the same network and across networks, in what may be called an ecology of complex contagions. The second concerns the study of how the structure of thresholds and their behavioral consequences can vary by individual and social context. The third area concerns the recent discovery of diversity as a causal variable in diffusion, where diversity can refer either to the diversity of demographic profiles among local peers, or to the broader structural diversity that local peers are situated within. Throughout, we take effort to anticipate the theoretical and empirical challenges that may lie ahead.

## 2 Empirical Advances

### 2.1 Applications to Health

For the past few decades, the study of public health has concerned not only biological contagions, but also social contagions concerning health behaviors, for example, medication, vaccines, exercise, and the ideologies related to each (Christakis and Fowler 2012). It has been found that simple contagions do not adequately capture the network dynamics that govern the diffusion of health behaviors [4, 7–9]. Social health behaviors often require reinforcement from peers, and they are strongly influenced by cultural practices and group norms.

The Framingham Heart Study suggested that obesity spread socially through a densely interconnected network of 12,067 people, assessed from 1971 to 2003 [10]. However, this study posited that either biological or normative mechanisms might play a role in the diffusion process, where each mechanism would be expected to yield very different diffusion dynamics.

A clearer hypothesis came from a follow-up study examining the spread of smoking behavior [11]. This study found evidence that the likelihood a smoker will quit depends on their exposure to multiple contacts, in part because smoking is often explicitly social and thus shaped by the dynamics of social norms. The role of complexity in smoking behavior (and cessation) has been supported by a more recent study using data from the National Longitudinal Study of Adolescent Health, which simulated the complex contagion dynamics of smoking under conditions where smokers can revert to smoking after quitting [12, 13]. By examining peer interactions over QuitNet—a social media platform for smokers attempting to quit—it was found that smokers were more likely to abstain if exposed to reinforcing contact from several abstinent users [14]. Kuhlman et al. [13] discuss how the diffusion of smoking behavior is filtered by both pro- and anti-smoking norms. This insight into the complexity of the quitting process helps to refine earlier models of smoking diffusion, in which threshold outcomes are represented by the binary decision to adopt without consideration of countervailing influences from non-adopters. Norms empower people to exert different kinds of influence—that is, for and against behavior—which amplifies the role of complexity in situations where non-adopters exhibit countervailing influences.

Exercise has similarly been found to exhibit the dynamics of complexity when peers influence each other to adopt new exercise behaviors. The characteristics of peers play an important role in influence dynamics, as both homophily and diversity have been shown to amplify the impact of reinforcing signals on the likelihood of behavior change. Centola [8] demonstrated a direct causal relationship between homophily and the diffusion of complex contagions, indicating that the effects of social reinforcement were much stronger when individuals shared a few key health characteristics in common. Further, Centola and van de Rijt [15] showed that social selection among “health buddy” peers in a fitness program led to connections among peers who were homophilous on the same key health characteristics: gender, age, and BMI. Aral and Nicolaides [16] elaborate in showing

that social reinforcement from similar peers is strengthened when those peers come from different social groups, highlighting the value of structural diversity in the dynamics of complexity. Another recent study of exercise behavior used an online intervention to demonstrate that exposure to social influence from a reinforcing group of anonymous online “health buddies” could directly increase participants’ levels of offline exercise activity [17].

An interesting twist in the relationship between complexity and health came from a series of studies which showed how clustered networks that facilitate the spread of social norms (e.g., anti-vaccination behavior) can thereby make populations susceptible to epidemic outbreaks of simple contagions (e.g., such as the measles) [18, 19]. These studies model the diffusion of anti-vaccine attitudes as a complex contagion that pulls people into echo chambers that amplify the likelihood of disease outbreak in the overall population. This work points to a vital direction for future research into how health behaviors and attitudes toward health interact in a broader, multilayered network of both complex contagions and disease diffusion.

Moreover, there are even some surprising instances where biological pathogens may also be complex. Infectious diseases are complex in situations where patients suffer simultaneous “co-infections” from multiple pathogens. In these cases, each disease increases a patient’s susceptibility to the other one, making it more likely that both infections will take hold in a patient. For instance, infection with the influenza virus can increase the likelihood of coinfection with other respiratory diseases, such as the *Streptococcus pneumoniae* bacterium (a leading cause of pneumonia). Each one creates susceptibility to the other, increasing the likelihood that joint exposure will lead a patient to become infected with both.

While a single virus can efficiently use weak ties to spread across a network, several viruses from different sources cannot be so easily transmitted the same way. For these kinds of illnesses, clustered social networks significantly increase the likelihood that individuals who are exposed to complementary infections, such as pneumonia and flu, or syphilis and HIV, will spread reinforcing coinfections. Contrary to most epidemiological intuitions, in random networks incidence rates of “complex synergistic co-infections” typically drop to zero, while clustered social networks are surprisingly vulnerable to epidemic outbreaks [20].

## 2.2 Diffusion of Innovations

Economists, marketers, and organizational theorists have long been interested in how technological innovations diffuse through a population. Bass [21] developed one of the first influential models of innovation diffusion, where technological adoption was understood as a simple contagion. As the uptake of innovations came to be viewed as inseparable from social networks, Schelling [22] started to formulate a threshold-based model of innovation adoption based on the influence of multiple peers. It has since been found that complex contagions characterize the diffusion of technologies in multiple areas of social life.

A number of controlled experiments illustrate that innovations diffuse through populations as complex contagions. Bandiera and Rasul [23] showed how farmers in Mozambique were more likely to adopt a new kind of crop if they had a higher number of network neighbors who had adopted. Oster and Thornton [24] show that the adoption of menstrual cups in women depends on influence from multiple peers, because of the transference of technology-relevant knowledge. Based on these findings, Beaman et al. [25] used complex contagion models to design seeding strategies for the distribution of pit planting in Malawi. Pit planting is a traditional West African technology which is largely unknown in Malawi, and it has the potential to significantly improve maize yields in arid areas of rural Africa. Beaman et al. compared the seeding strategies recommended by complex contagion models to a benchmark treatment where village leaders used local knowledge to select seeds. Seeding, in this experiment, involved training specific people in each village on how to use pit planting, given evidence that trained adopters of a technology are most effective in distributing new technologies [26]. 200 different villages were randomly treated with seeding strategies from either complex contagion models or traditional approaches based on local expertise. They found that seeding strategies informed by complex contagion models increased adoption more than relying on extension workers to choose seeds. Further, Beaman et al. observe no diffusion of pit planting in 45% of the benchmark villages after 3 years. In villages where seeds were selected using the complex contagion model, there was a 56% greater likelihood of uptake in that village.

Complex contagions have also been shown to characterize the diffusion of software innovations. Karsai et al. [27] examined the uptake of Skype—the world’s largest Voice over Internet protocol service—from September 2003 to March 2011. They find that the probability of adoption via social influence is proportional to the fraction of adopting neighbors. Interestingly, they find that while adoption behaves like a complex contagion process, termination of the service occurs spontaneously, without any observable cascade effects. These results suggest that there may be an asymmetry in the dynamics of adoption (which are socially driven) versus the dynamics of termination (which may depend on nonsocial factors).

Ugander et al. [28] also observe complex contagion dynamics in the initial growth of Facebook, which now has over a billion users worldwide. Facebook initially grew through peer recruitment over e-mail. The results showed a complex diffusion process, in which people were more likely to adopt Facebook if they received requests from multiple friends, especially if these friends belonged to separate network components. This finding on the value of structural diversity for amplifying reinforcing signals for adoption suggests interesting new theoretical directions for research on the connections between homophily, diversity, and complexity (see Sect. 4.3).

A parallel stream of research has focused on the role of mass-media marketing in spreading the complex diffusion of innovations. Toole et al. [29] show that while mass media served to measurably increase the adoption of Twitter, peer-to-peer social influence mechanisms still account for the lion’s share of the adoption

patterns that were observed, where local reinforcement played a major role in individuals' decisions to adopt Twitter. So much so, that the online microblogging platform exhibited strong spatial diffusion patterns in its initial growth, as it spread through densely clustered networks of peer reinforcement. Similar findings are echoed by Banerjee et al. [26]. These studies suggest that the local peer influence dynamics of complexity can initiate global cascades in the adoption of innovations. For marketing to propel the diffusion of new technologies, mass-media strategies need to account for how messages are dynamically filtered by social networks [30]. Evidence suggests that advertising campaigns initially diffuse like simple contagions with the first media broadcast, but diffuse more like complex contagions once they begin spreading through social networks [31]. The interaction between mass-media diffusion and social influence in the adoption of technology (particularly complementary technologies) suggests that the complexity of a diffusion process is determined in part by interactions across several scales of a population.

The study of innovation diffusion is expanding in response to a novel kind of complexity introduced by technologies themselves. A new direction for future research concerns the role that social media technologies play in shaping the evolution of other contagions, once the social media technologies themselves are adopted. Due to their explicitly complementary design, social media technologies, including Facebook, Twitter, and Skype, all exhibited the dynamics of complexity in their diffusion, spreading most effectively through networks of peer reinforcement. Once these technologies diffuse, they allow individuals to grow larger networks that communicate at much faster rates than were previously possible in word-of-mouth exchanges. Thus, in addition to the spread of social media technologies, the domain of social media itself has become its own space for studying the complex dynamics of the diffusion of collective behavior.

### 2.3 Social Media

Social media has significantly shaped and, in some cases, augmented the diversity of complex contagions that can spread, the speed at which they can spread, and the overall size of the populations they are able to reach via global cascades [32]. Kooti et al. [33] show that one of the first methods for retweeting was established as the successor of various competing complex contagions, in an ecology of possible conventions. Barash [34] and Weng et al. [35] find that most tweets spread via complex contagions in retweet networks. This finding reappears with Harrigan et al. [36] who show that tweets are more likely to diffuse through retweeting within clustered communities, where twitter users are able to observe their friends retweeting the same message. Complex contagions are observed across other platforms as well. Photo-tagging in Flickr exhibits the hallmarks of diffusion via influence from multiple peers [34]. A recent massive-scale randomized experiment over Facebook showed that user-generated stories diffused like complex

contagions [37]. Meanwhile, social media websites gather an unprecedented amount of data on communication flows, permitting novel insights into how complex contagions emerge and operate.

One of the most interesting findings of social media research is that the content of a contagion matters for whether it behaves in a complex manner. Wu et al. [38] show that the modality of information that structures a contagion influences its life span: viral videos long outlive their textual counterparts. Romero et al. [39] find that there are distinct contagion dynamics for different kinds of hashtags. Political hashtags are found to behave like complex contagions, where exposure to multiple people using the hashtag is strongly correlated with adoption. But hashtags based on idioms or memes, by contrast, behave like simple contagions. Barash and Kelly [40] and Fink et al. [41] replicate this finding by showing that political hashtags behave like complex contagions, whereas news-based hashtags, broadcast by mass media, spread like simple contagions.

Using the massively multiplayer virtual world of Second Life, Bakshy et al. [42] uncover complex contagions in the exchange of user-created content. Specifically, they focus on the spread of conventionalized avatar gestures constructed by players, which can only spread through peer-to-peer sharing mechanisms. Bakshy et al. unveil subtle interactions between user degree and diffusion: Users who are most effective at initiating cascades of gestures do not have the highest degree; rather, they collect rare gestures that other users are more likely to adopt. This result points to uncharted territory in complex contagions research, relating to how the quality or style of a contagion influences its likelihood of spreading via social influence.

Undoubtedly, the source of complexity in these online dynamics of spreading behavior lies partly in the sociological significance that the content of an online contagion holds. For instance, Romero et al. suggest that political hashtags, such as #TCOT (which stands for “Top Conservatives on Twitter”) and #HCR (which stands for “Health Care Reform”), were “riskier to use than conversational idioms . . . since they involve publicly aligning yourself with a position that might alienate you from others in your social circle” (2011, p. 3). The implication is that users have to be motivated enough to use the hashtag despite social costs, as a result of either personal political engagement or peer influence. In this case, the authors found “that hashtags on politically controversial topics are particularly persistent, with repeated exposures continuing to have unusually large marginal effects on adoption” (p. 3).

It is also likely that the level of complexity in diffusion depends, in part, on the design of interfaces and the kinds of sociological processes that platforms facilitate. Readymade communication buttons—such as the “share” button on Facebook or the “retweet” button on Twitter—automatically enable the spread of information as a simple contagion. However, State and Adamic [43] show how simple contagions do not account for the spread of digital artifacts that require more effort to construct. Using a dataset of over three million users, they show that the adoption of new conventions for profile pictures are best described as complex contagions. They argue that the difference pertains to the amount of effort it takes to adopt the

behavior: Certain informational contagions behave in a simple manner because it takes no time to click and share after one exposure. But when a contagion requires more effort, such as manually changing a profile picture, users require evidence that several of their peers have expended the energy for the contagion, thereby justifying its weight in terms of social capital.

Conversely, platform design can also prevent complex contagions from emerging and spreading by constraining the ability for people to perceive and share potential contagions [31, 37, 44, 45]. Doerr et al. [46] find that, over the social news aggregator *Digg*, users do not seem to preferentially share the content of their peers. This result is likely to be specific to the *Digg* environment, because the culture of the platform is based on sharing news that your friends do not already know. Studies of social media thus reveal how environmental design alters the capacity for diffusion by shaping the salience of peer behaviors and the culture of interaction altogether.

Going forward, social media environments are likely to serve as a powerful tool for studying complex contagions experimentally. Centola [7, 8, 47] developed a method for designing social media platforms that embed participants into engineered social networks, which allow researchers to test the effects of network topology and other variables on the dynamics of social diffusion. In a less controlled study, Kramer et al. [48] modified the newsfeeds of Facebook users to examine emotional contagion. For some users, they reduced the amount of positive content, whereas for other users, they reduced the amount of negative content. As a result, they were able to systematically alter the emotional content of users' posts. While this study could not eliminate endogeneity within user networks, the randomization of messages allowed for suggestive experimental results on the ways that social exposure to messages influences user behavior.<sup>1</sup>

Another related approach to experimentation on social media comes from the advent of experimental methods that use algorithmically controlled social media accounts called bots to manipulate users' experiences [51]. Mønsted et al. [52] released a network of bots into twitter and tested whether they could prompt the uptake of specific hashtags. They show that bots can initiate the uptake of new hashtags and that these hashtags spread as complex contagions, whereby the probability of using the new hashtag drastically increased if multiple bots and users were seen using it.

---

<sup>1</sup>Kramer et al.'s study also raised the important point about the ethics of experimentation on social media. While previous social media studies using experimentally designed social platforms [7, 8, 49, 50] enrolled subjects into their online platform with an explicit process of informed consent, Kramer et al.'s study on Facebook used existing networks of peers without their explicit consent. It is an important topic of ongoing discussion how to properly use existing peer networks, such as Facebook and Twitter, to conduct experiments that manipulate user behavior.

## 2.4 Politics

Political processes have been a long-standing topic of interest for threshold-based contagion models. Granovetter's [1, 53] original threshold model of collective action gave special attention to the start-up problem for political protests and riots. He observed that individuals have different degrees of willingness (i.e., thresholds) to participate in a riot, where their willingness is dependent on how many of their neighbors they observe participating in the riot. Granovetter observed that riots can emerge as a result of cascades, where a subset of instigator individuals with low thresholds trigger the spread of rioting. The first efforts to describe the emergence of social movements with agent-based modeling maintained that population diversity was essential for getting a movement off the ground. Without long ties connecting communities, it was thought that social movements would not be able to diffuse through a population and reach critical mass.

More recent models extend the study of diversity in political processes by emphasizing the supporting role of homophily during the growth phase of social movements. Centola [54] argues that because social movements involve risky and costly forms of deviant behavior, people require reinforcement from multiple peers to participate, where homophily is useful for establishing a critical mass of like-minded peers.

Again, this raises an interesting connection between diversity and homophily. For organizing a critical mass, dense, homophilous communities are necessary for getting social movements off the ground because like-mindedness facilitates group solidarity, which may be necessary to withstand the normative backlash that comes from deviant behavior. On these grounds, Centola designed an agent-based model to show that weak ties hinder the spread of social movements by increasing exposure to counter-norm pressures, while also reducing the group transitivity needed to reinforce group interests. Homophily and clustering thus reinforce one another. However, once homophilous networks gain enough local reinforcement, they can create a critical mass that allows the movement to achieve sufficient salience in the whole population and to expand to diverse communities through the aid of mass media.

In an empirical study of the effects of communication networks on mobilization, Hassanpour [55] explored the spread of armed conflict as a complex contagion in Damascus, Syria. On November 29, 2012, Internet and cellular communications were shut down all across Syria for over a day. The shutdown, according to Hassanpour, resulted in the loss of communication with long ties to individuals across the city. At the same time, the shutdown immediately preceded an unprecedented increase in the diffusion of armed conflict throughout the city. Using a geolocated dataset of daily conflict locations in Damascus, Hassanpour uncovers signs that the likelihood of conflict in a region was influenced by whether there had been conflict in multiple neighboring regions. Hassanpour suggests that this indicates the spread of conflict as a complex contagion, which was allowed to emerge when long ties were broken and interaction within local clusters became the strongest determiners of armed conflict.

In other results, González-Bailón et al. [56] shows that protest recruitment in Spain, 2011, diffused over Twitter as a complex contagion via peer influence. González-Bailón et al. [56] used k-core decomposition to show that the users who are in the core of the network were most effective at initiating cascades of recruitment. In a complementary study, Steinhert-Threlkeld [57] offered evidence that users in the periphery of social media networks can also trigger global cascades. These studies suggest that social media can influence the rise and spread of political complex contagions that inspire on-the-ground political action.

Other recent empirical work has uncovered complex contagions within a wide range of political processes, including campaign donations [58], grassroots mobilization [59, 60], petition signing [61], social control [62, 63], institutional change [64], and administrative management in both rural [65] and urban settings [66].

Barash [34] developed a unique set of measures for characterizing the life span of political contagions over social media. A complex contagion begins by saturating a locally clustered community. Once saturation is reached, the rate of propagation for the contagion decelerates, as the number of potential adopters decreases. If the saturated community has sufficiently wide bridges to other communities, Barash [34] argues that it is possible for a contagion to travel from one community to the next. Diffusion between communities can create a detectable temporal signature, because as a contagion enters a new community, its rate of propagation rapidly increases with the availability of new adopters. Barash explains how changes in the rate of complex propagation can provide a measure for whether a contagion is ramping up for a global cascade, hinting toward the possibility of detecting global cascades, prior to their emergence.

Based on the work of Barash et al. [67], Fink et al. [41] developed a number of measures for characterizing the spread of political hashtags as complex contagions. These measures include *peakedness*, *commitment*, *concentration*, and *cohesion*. Peakedness concerns the duration of global activity associated with a contagion, where a peak refers to a day-long period of usage when the average mentions per day are more than two standard deviations away from the average mentions in the preceding days. Peakedness is closely related to burstiness, which has been shown to play an important role in threshold-based cascade dynamics [68]. Commitment refers to the number of people who sustain the life of a complex contagion, even though they endure social costs by not conforming to surrounding norms. Concentration simply refers to the proportion of people using a hashtag during a given time period. And cohesion refers to the network density over the subgraph of all users engaged in a particular contagious phenomena. The authors make use of the idea that complex contagions are incubated in locally dense communities before they colonize other communities via sparse connections.

Using these measures, researchers have made a number of valuable observations. Fink et al. [41] apply these measures to the study of political hashtags in Nigeria. In their sample, they find political hashtags consistently arise with a small proportion of instigators (roughly 20%) who are densely connected, and that almost 60% of late adopters for political hashtags had two or more previous adopter friends. News hashtags, by contrast, are first propagated by largely unconnected instigators who

constitute between 50% and 90% of the network, where less than 10% of adopters had two or more previous adopter friends. Consistent with Romero et al. [39], the authors suggest that political hashtags require influence from multiple peers because they have higher social costs, especially in countries like Nigeria where surveillance by governments and extremists groups looms over users. Compared to other hashtags, the researchers also find that hashtags related to social movements have a higher density of ties among early adopters, consistent with the argument that political movements require a coalition of homophilous, densely connected users [54]. Fink et al. further illustrate that it is possible to map the virality of political hashtags using Barash's measures for the temporal signatures of diffusion. They show how the #bringbackourgirls hashtag went viral shortly after a period of decreased usage among early adopters, which indicated saturation in a local community prior to the spread of the contagion to other communities.

In a related paper, Barash and Kelly [40] use the same measures to model the spread of complex contagions over Russian Twitter, a significantly different cultural setting. Yet again, these researchers find that politically salient hashtags diffuse like complex contagions where news hashtags from mass media do not. Importantly, this analysis shows how the heterogeneous distribution of adoption thresholds is critical for understanding political contagions. They find that engagement in political issues is nonuniform across the population, and different communities have distinct patterns of engagement and adoption, based on how the community relates to the content captured by the hashtag. Users belonging to groups that oppose the political regime engage with controversial topics over a long period, as a committed minority. Contrary to expectations, they find that when hot button issues relevant to the opposition make it into the mainstream, they are much more likely to sustain global saturation, even amongst pro-government users. These results suggest that reinforcement dynamics can drive the spread of politically salient content over social media.

The diffusion of political contagions online interacts with both the structure of the subcommunities that they reach and the group identities that they activate. In March 2013, three million Facebook users changed their profile picture to an “equals sign” to express support for same-sex marriage. Consistent with earlier work, State and Adamic [43] found that the equals sign a profile picture spread as a complex contagion. Their data suggests that mass media created only about 58,000 spontaneous adopters, while roughly 106 million users adopted based on peer exposures. They find that it took, on average, exposure to eight different peer adopters for a person to adopt. When examining this threshold, the authors uncover intricate dependencies between the identity of a user and their willingness to adopt. Users were more likely to be exposed and thereby more likely to adopt if they were female, liberal, nonheterosexual, and between the ages of 25–34. These findings suggest that thresholds for adopting contagions are modulated by online identity signaling regarding political values and beliefs. Similar, smaller-scale studies of behavior on Facebook find that a user’s demographic characteristics do not determine their influence in generating cascades, but instead most cascades rely on multiple users to trigger spreading [69].

The study of political contagions—offline and online—reveals a number of subtleties in how thresholds operate in sociological contexts. Political identity is a driving motivation for behavior change, suggesting that homophily and clustering in social networks can be essential for incubating the early growth of a political behavior over social media. Furthermore, gender, race, and religion are also strong predictors of whether someone will be exposed and receptive to a political contagion. A recent study by Traag [58] shows that campaign donations diffuse as complex contagions, but the findings here emphasize the value of diversity. The growth of support for a candidate increases when people are exposed to donors from separate communities, particularly if those donors supported the opposite party. Diversity can thus complement homophily when it signals wider support for a candidate, and thereby increases the likelihood that the candidate will be more effective in achieving bi-partisan goals. The details are subtle, however, since there are also situations where diverse support for a candidate might signal mixed allegiances and compromise the candidate’s party loyalty. The complementary roles of homophily and diversity in supporting complexity depend upon the content of the political messages that are used and the identities that they activate.

### 3 Theoretical Advances

Recent research into the formal model of complex contagions has explored two general directions. The first direction investigates how complex contagions spread within large networks of varying topologies. To date, researchers have examined threshold-based contagion models within power-law [67], locally tree-like [70], degree-correlated [71, 72], directed [73], weighted [74], small-world [9], modular [75], clustered [76, 77], temporal [78, 79], multiplex [80–82], and interdependent lattice networks [83]. Researchers have used different topologies to simulate how external factors like mass media influence cascade dynamics [84], and how topologies influence percolation processes [85]. A pivotal theoretical finding is that complex contagions require a critical mass of infected nodes to initiate global cascades, and it has been shown that critical mass dynamics depend in sensitive ways on network topology and the distribution of node degree and adoption thresholds [34]. There have also been efforts to provide analytic proofs for the global dynamics of complex contagions [34, 86]. At the cutting edge is research into how complex contagions spread in coevolving, coupled, time-varying, and multilayered networks [68, 87].

The second major direction in theoretical complex contagion research concerns mechanisms of diffusion at the node level, concerning individual attributes and thresholds. Wang et al. [88, 89] propose a contagion model which shows that the final adoption size of the network is constrained by the memory capacities of agents and the distribution of adoption thresholds. Perez-Reche et al. [90] simulate complex contagion dynamics with synergistic effects among neighbors, and McCullen et al. [91] structure the motivation for an agent to adopt a behavior as a combination of

personal preference, the average of the states of one's neighbors, and the global average. Dodds et al. [92] attempt to explicitly encode sociological processes into their models by building agents with a preference for imitation but an aversion for complete conformity. Melnik et al. [93] model *multistage* complex contagions, in which agents can assume different levels of personal involvement in propagating the contagion, at different times in their life cycle. They find that multistage contagions can create multiple parallel cascades that drive each other, and that both high-stage and low-stage influencers can trigger global cascades. Huang et al. [94] build agents with a persuasiveness threshold which determines their ability to initiate adoption. This new parameter can cause networks to become more vulnerable to global cascades, especially heterogeneous networks. Further incorporating sociological considerations, Ruan et al. [95] simulated how conservativeness among nodes—that is, the reluctance to adopt new norms—interacts with cascades caused by spontaneous adopters.

The latest theoretical developments have informed research on how to design network interventions and seeding strategies to stop the spread of harmful complex contagions [96]. Such interventions are based on the use of oppositional nodes that are permanently unwilling to adopt a behavior, regardless of peer influence. Kuhlman et al. [96] offers two heuristics for using seeding methods to determine critical nodes for inhibiting the spread of complex contagions. The first heuristic is to select the nodes with the highest degree, and the second heuristic is to select nodes from the 20 core, determined by k-core decomposition. They show how the second heuristic is more effective at initiating and preventing global cascades, because selecting from the 20 core increases the likelihood that nodes are adjacent and thereby capable of reinforcing each other's influence. Centola [97] shows that similar ideas can be used to evaluate the tolerance of networks against error and attack. Albert et al. [98] showed that scale-free networks are robust against network failures, defined in terms of the inability to diffuse simple contagions. When it comes to diffusing complex contagions, Centola shows that scale-free networks are much less robust than exponential networks. Thus, moving from simple to complex contagions changes the robustness properties of scale-free networks. Building on this work, Blume et al. [99] investigate which topologies are more susceptible to what they call *cascading failures*, which refers to the outbreak of negative complex contagions that are harmful for social networks. Siegel [62, 63] shows how these developments can inform models for repressing social movements and performing crowd control on behalf of governments.

While early models of diffusion consider individual contagions as independent and spreading in isolation, a number of studies have begun to investigate evolutionary dynamics among multiple complex contagions. Myers and Leskovec [100] develop a statistical model wherein competing contagions decrease one another's probability of spreading, while cooperating contagions help each other in being adopted throughout the network. They evaluate their model with 18,000 contagions simultaneously spreading through the Twittersphere, and they find that interactions between contagions can shape spreading probability by 71% on average. Jie et al. [101] construct a similar model to simulate competing rumor

contagions in a homogenous network. Empirical evidence is accumulating that multiple contagions frequently interact in real-world social systems. For instance, the study of social contagions in the health domain has shown competitive dynamics among positive and negative health practices, for example, smoking vs. jogging [12, 13]. Most interestingly, health research has uncovered ecological interactions among contagions at different scales, such as the interaction between complex contagions (e.g., health-related attitudes and lifestyle choices), and the spread of simple contagions (e.g., biological pathogens) [18].

## 4 New Directions

Recent work on complex contagions points to three main directions for future development. The first concerns the study of how multiple contagions interact within the same network and across networks, in what may be called an ecology of complex contagions. The second concerns the study of how the structure of thresholds and their behavioral consequences can vary by social context. The third area concerns the interaction of diversity and homophily in the spread of complex contagions, where diversity can refer to either the diversity of demographic profiles among one's local peers, or to the broader structural diversity that local peers may be situated within.

### 4.1 Ecologies of Complex Contagions

Past theoretical research has made significant progress in mapping the behavior of complex contagions within a range of network topologies. Newer work has begun to explore the complexities that arise when multiple kinds of contagions interact in the same network [102]. Moreover, while the content of a contagion undoubtedly influences the spread and interaction of competing behaviors, it may have an impact on network structure as well. An important area of future research concerns how complex contagions shape network structure and how network structure shapes complex contagions, as part of a coevolutionary process of network formation [103].

The process of modeling ecologies of contagions goes hand in hand with a growing effort to model complex contagions in several new domains of collective behavior. Among the most recent applications is the examination of complex contagions in swarm behaviors. One study showed that complex contagions provided the most robust model of escape reflexes in schools of golden shiner fish, where frightened individuals trigger cascades of escape responses on the basis of a fractional threshold among multiple peers [104]. Another direction for application concerns the role of complex contagions in cognitive science. Simulation results suggest that complex contagions may be able to account for the emergence and spread of new categories, at the level of both perception and language, consistent

with the long-standing view that cultural artifacts depend on principles of emergence and diffusion [49, 105, 106]. Related extensions concern the role of contagion in the structuring of collective memory [107]. Situating complex contagions at this level will extend existing perspectives on how processes of social diffusion are woven into the foundations of culture and cognition.

## 4.2 *Mapping Heterogeneous Thresholds in Context*

Extant models represent threshold heterogeneity in terms of distributions of values along a numerically defined scale, from 0 to 1 [108]. Applied studies of contagion dynamics show how thresholds vary by individual differences and contextual dependencies relating to the content of the contagion and its sociological significance. For instance, there appear to be a different set of thresholds that govern the adoption of a contagion (e.g., a technology) and the termination of the contagion [27]. Similarly, the study of health contagions suggests that people are susceptible to influence by those supporting a positive health behavior and to those resisting it, where individuals may vary in their responses to processes of support and resistance [12, 14].

In the context of social media, readymade sharing buttons alter the cost structure for certain contagions, allowing memes to be adopted simply with a click. Interface design can also make certain contagions costlier, thereby impacting the thresholds of individuals and their willingness to adopt. Certain complex contagions, such as political hashtags, appear to require exposure from 2 to 5 peers [40], whereas changes in profile pictures appear to require exposure of up to eight or more peers [43]. One conjecture is that thresholds are fractional, and therefore depend sensitively on the number of connections that a person has. The more connections there are, the higher the thresholds are likely to be.

Finally, identity appears to play a structural role in defining thresholds. Identity has been used in two ways: group identity and personal identity. A few recent studies have excluded group identity and focused narrowly on personal identity, such as demographic characteristics [16, 109]. However, the role of demographic characteristics such as gender and race on adoption thresholds is hard to understand independently of social context. Depending on the social context and the identities that are activated, people will react differently to a political contagion than to a health contagion. By contrast, other work has suggested that demographic traits play an important role in defining group identity, which in turn interacts with people's thresholds for adoption [8].

Political studies further show how identity-based responses to contagions can take a variety of forms, where thresholds do not simply represent the binary outcome of adoption—they also represent whether an individual will join a committed minority, or whether they will actively attempt to punish deviant behavior [54]. Parkinson [60] uses ethnographic methods to suggest that part of the reason why

identity influences contagion thresholds is because identities correspond to different functional roles in a social system, which entail different kinds of behavioral responses that mediate diffusion. These studies help to expose how group membership and normative pressures give rise to individual variation in threshold dynamics. It is likely that individuals differ in the kinds of thresholds they adopt toward a potential contagion based on how they categorize the contagion, relative to their political identity [43, 58]. It may therefore be useful to consider different types of thresholds that vary along sociological and psychological dimensions, where key differences are marked by how contagions interface with the identity-based responses of individuals and groups.

### 4.3 The Roles of Homophily and Diversity in Diffusion

There are two forms of diversity in the literature on diffusion. Researchers use the term to refer to cases where one's local neighborhood in the network consists of people with different demographic profiles and personality traits. We may call this *identity-based* diversity. At other times, researchers use the term to refer to *structural diversity* where one's local neighborhood consists of people who belong to separate components of the network, identified by removing the ego node from the ego network. The first kind of diversity tends to limit diffusion of complex contagions, while the second kind tends to amplify it.

Looking at identity-based diversity, Centola [8] compared complex contagion dynamics on homophilous networks to the dynamics on non-homophilous networks, keeping network topology constant. The results showed that homophily (i.e., reduced identity-diversity) significantly improved the spread of complex contagions. The reason for this is that greater similarity among contacts in a health context made peers more relevant. Women were more likely to adopt from women, and obese people were likely to adopt from obese people. Reinforcing signals from irrelevant (i.e., diverse) peers were largely ignored, while reinforcing signals from relevant (i.e., similar) peers were influential in getting individuals to adopt a new health behavior. This result was most striking for obese individuals. Exposure among obese individuals was the same across conditions, yet there was not a single obese adopter in any of the diverse networks, while the number of obese adopters in homophilous networks was equivalent to the total number of overall adopters in the diverse networks—resulting in a 200% increase in overall adoption as a result of similarity among peers.

The effects of homophily can be complemented by structural diversity. In studying the complexity of campaign donations, Traag [58] suggests that structural diversity can increase the credibility of a complex contagion. If one belongs to an echo chamber, where one's peers are highly similar and densely connected, then peer agreement may undermine credibility, since their agreement may be the result of induced homophily and pressures for conformity. By contrast, if one's peers

are from different components of the network, their opinions may reasonably be viewed as independent and mutually confirming. What unites these arguments is the supposition that people use the identity composition of their local network neighbors to infer the broader structural diversity of their network. However, structural diversity does not imply reduced homophily. Individuals may be similar to their friends in different ways. They may be the same gender as some, have the same professional role as others, and participate in the same volunteer organizations as yet others. While identity diversity can correlate with structural diversity, it does not always provide a reliable way for inferring it. Receiving reinforcing encouragement from individuals who belong to different parts of a person's social network strengthens the independence of their signals, and may therefore be more likely to trigger adoption.

Similarly, Ugander et al. [28] identify how the mechanism of structural diversity can boost the influence of social reinforcement. Their study of Facebook shows that people are more likely to adopt a social media technology when they receive invites from people belonging to separate components of their ego network. Structural diversity does not, however, entail identity-based diversity. Ugander et al.'s study leaves open the possibility that structural diversity alone—without identity-based diversity—can modulate adoption thresholds.

This observation is especially interesting in light of State and Adamic's [43] finding that while the number of friends a user had scaled linearly with their chances of adoption, adoption probabilities plummeted as soon as a user possessed 400 friends or more. The authors propose that having too many friends on social media can stifle the spread of complex contagions by exposing users to a variety of content so vast that they fail to receive repeated exposure by different peers to any given phenomenon. Consistent with earlier results on political hashtags and social movement mobilization, these findings suggest that more contentious complex contagions tend to benefit from clustered, homophilous networks that can foster social change without being overwhelmed by countervailing influences.

## 5 Conclusion

Complex contagions are found in every domain of social behavior, online and off. Early theoretical developments in complex contagions showed that topology and the distribution of adoption thresholds can be decisive for determining whether global saturation is possible. More recent theoretical modeling concerns the interaction of multiple different contagions in the same network, where individuals are attributed different motivations and behavioral responses to each contagion. One of the critical challenges ahead involves mapping heterogeneous thresholds in context, where political identity, group membership, and even the content of contagions can affect individual thresholds and, by consequence, diffusion. Another valuable area for future research concerns the ways in which individuals use information about

global network structure to inform their adoption patterns, as is demonstrated by the effects of structural diversity on diffusion. Investigations in this direction will benefit from studying how individuals infer global structure from local interactions, and how new social media environments are augmenting these inferences by supplying information about one's broader ego network. As shown by the literature accumulated over the last decade, examining complex contagions in various applied domains has been enormously fruitful. Each new domain has revealed new elements of diffusion dynamics that require new theoretical explanations and elaborated modeling techniques, revealing new areas of cumulative progress in understanding the collective dynamics of social diffusion.

## References

1. Granovetter M (1973) The strength of weak ties. *Am J Sociol* 78:1360–1380
2. Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393(6684):440–442
3. Coates TJ, Richter L, Caceres C (2008) Behavioural strategies to reduce HIV transmission: how to make them work better. *Lancet* 372(9639):669–684
4. Centola D, Macy M (2007) Complex contagions and the weakness of long ties. *Am J Sociol* 113(3):702–734
5. Coleman JS, Katz E, Menzel H (1966) Medical innovation; a diffusion study. Bobbs-Merrill Co, Indianapolis
6. Davis G, Greve H (1997) Corporate elites and governance changes in the 1980s. *Am J Sociol* 103(1):1–37
7. Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194–1197
8. Centola D (2011) An experimental study of homophily in the adoption of health behavior. *Science* 334(6060):1269–1272
9. Centola D, Eguiluz V, Macy M (2007) Cascade dynamics of complex propagation. *Phys A* 374:449–456
10. Christakis NA, Fowler JH (2007) The spread of obesity in a large social network over 32 years. *N Engl J Med* 357(4):370–379
11. Christakis NA, Fowler JH (2008) The collective dynamics of smoking in a large social network. *N Engl J Med* 358(21):2249–2258
12. Kuhlman C, Kumar A, Tuli G (2011) A bi-threshold model of complex contagion and its application to the spread of smoking behavior. AAAI fall symposium: complex adaptive systems
13. Kuhlman CJ, Kumar VSA, Marathe MV, Ravi SS, Rosenkrantz DJ (2011) Effects of opposition on the diffusion of complex contagions in social networks: an empirical study. In: Salerno J, Yang SJ, Nau D, Chai SK (eds) Social computing, behavioral-cultural modeling and prediction. SBP 2011, Lecture notes in computer science, vol 6589. Springer, Berlin
14. Myneni S, Fujimoto K, Cobb N, Cohen T (2015) Content-driven analysis of an online community for smoking cessation: integration of qualitative techniques, automated text analysis, and affiliation networks. *Am J Public Health* 105(6):1206–1212
15. Centola D, van de Rijt A (2014) Choosing your network: social preferences in an online health community. *Soc Sci Med* 125:19–31. <https://doi.org/10.1016/j.socscimed.2014.05.019>
16. Aral S, Nicolaides C (2017) Exercise contagion in a global social network. *Nat Commun* 8:14753

17. Zhang J, Brackbill D, Yang S, Becker J, Herbert N, Centola D (2016) Support or competition? How online social networks increase physical activity: a randomized controlled trial. *Prev Med Rep* 4:453–458
18. Campbell E, Salathe M (2013) Complex social contagion makes networks more vulnerable to disease outbreaks. *Sci Rep* 3:1905
19. Salathe M, Bonhoeffer S (2008) The effect of opinion clustering on disease outbreaks. *J R Soc Interface* 5(29):1505–1508
20. Hébert-Dufresne L, Althouse BM (2015) Complex dynamics of synergistic Coinfections on realistically clustered networks. *Proc Natl Acad Sci* 112(33):10551–10556
21. Bass F (1969) A new product growth for model consumer durables. *Manag Sci* 15(5):215–227
22. Schelling T (1973) Hockey helmets, concealed weapons, and daylight saving: a study of binary choices with externalities. *J Confl Resolut* 17(3):381–428
23. Bandiera O, Rasul I (2006) Social networks and technology adoption in northern Mozambique. *Econ J* 116(514):869–902
24. Oster E, Thornton R (2012) Determinants of technology adoption: peer effects in menstrual cup take-up. *J Eur Econ Assoc* 10(6):1263–1293
25. Beaman L, Yishay A, Magruder J, Mobarak M (2015) Can network theory-based targeting increase technology adoption? Working paper
26. Banerjee A, Chandrasekhar AG, Duflo E, Jackson MO (2013) The diffusion of microfinance. *Science* 341:6144
27. Karsai M, Iniguez G, Kaski K, Kertesz J (2014) Complex contagion process in spreading of online innovation. *J R Soc Interface* 11(101):20140694
28. Ugander J, Backstrom L, Marlow C, Kleinberg J (2012) Structural diversity in social contagion. *PNAS* 109(16):5962–5966
29. Toole J, Meeyoung C, Gonzalez M (2012) Modeling the adoption of innovations in the presence of geographic and media influences. *PLoS One* 7(1):1–9
30. Berger J (2013) Contagious: why things catch on. Simon and Schuster, London
31. Bakshy E, Eckles D, Yan R, Rosenn I (2012) Social influence in social advertising: evidence from field experiments. arXiv
32. Borge-Holthoefer J, Baños RA, González-Bailón S, Moreno Y (2013) Cascading behaviour in complex soci-technical networks. *J Complex Netw* 1:3–24. <https://doi.org/10.1093/comnet/cnt006>
33. Kooti F, Yang H, Cha M, Gummadi PK, Mason WA (2012) The emergence of conventions in online social networks. *ICWSM*
34. Barash V (2011) The dynamics of social contagion. PhD Dissertation, Cornell University
35. Weng L, Menczer F, Ahn YY (2013) Virality prediction and community structure in social networks. *Sci Rep* 3:2522
36. Harrigan N, Achananuparp P, Lim E (2012) Influentials, novelty, and social contagion: the viral power of average friends, close communities, and old news. *Soc Netw* 34(4):470–480
37. Bakshy E, Rosenn I, Marlow C, Adamic L (2012) The role of social networks in information diffusion. In: Proceedings of ACM WWW 2012
38. Wu S, Tan C, Kleinberg J, Macy M (2011) Does bad news go away faster? In: Fifth international AAAI conference on weblogs and social media, ICWSM. AAAI
39. Romero DM, Meeder B, 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, ACM, pp 695–704
40. Barash V, Kelly J (2012) Salience vs commitment: dynamics of political hashtags in Russian Twitter. Berkman Center for Internet and Society, Research Publication No. 2012–9
41. Fink C, Schmidt A, Barash V, Cameron C, Macy M (2016) Complex contagions and the diffusion of popular Twitter hashtags in Nigeria. *Soc Networks* 6(1):1
42. Bakshy E, Karrer B, Adamic L (2009) Social influence and the diffusion of user-created content. In Proceedings of the 10th ACM conference on electronic commerce, pp 325–334

43. State B, Adamic L (2015) The diffusion of support in an online social movement: evidence from the adoption of equal-sign profile pictures. In: CSCW'15 proceedings of the 18th ACM conference on computer supported cooperative work and social computing, pp 1741–1750
44. Gomez-Rodriguez M, Gummadi K, Schölkopf B (2014) Quantifying information overload in social media and its impact on social contagions. In: Proceedings of the eighth international conference on weblogs and social media, pp. 170–179
45. Hodas NO, Lerman K (2012) How visibility and divided attention constrain social contagion. Privacy, security, risk and trust (PASSAT). In: 2012 international conference on social computing (SocialCom)
46. Doerr C, Blenn N, Tang S, Miegham P (2012) Are friends overrated? A study for the social news aggregator Digg.com. *Comput Commun* 35(7):796–809
47. Centola D (2013) Social media and the science of health behavior. *Circulation* 127(21):2135–2144
48. Kramer A, Guillory J, Hancock J (2014) Experimental evidence of massive-scale emotional contagion through social networks. *PNAS* 111(24):8788–8790
49. Centola D, Baronchelli A (2015) The spontaneous emergence of conventions: an experimental study of cultural evolution. *PNAS* 112(7):1989–1994
50. Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856
51. Krafft P, Macy M, Pentland A (2016) Bots as virtual confederates: design and ethics. The 20th ACM conference on computer-supported cooperative work and social computing (CSCW)
52. Mønsted B, Sapieżyński P, Ferrara E, Lehmann S (2017) Evidence of complex contagion of information in social media: an experiment using Twitter bots. *PLoS One* 12(9):e0184148. <https://doi.org/10.1371/journal.pone.0184148>
53. Granovetter M (1978) Threshold models of collective behavior. *Am J Sociol* 83:1420–1443
54. Centola D (2013) Homophily, networks, and critical mass: solving the start-up problem in large group collective action. *Ration Soc* 25(1):3–40
55. Hassanpour N (2017) A quasi-experimental study of contagion and coordination in urban conflict: evidence from the Syrian civil war in damascus. Unpublished manuscript
56. González-Bailón S, Borge-Holthoefer J, Rivero A, Moreno Y (2011) The dynamics of protest recruitment through an online network. *Sci Rep* 1:197. <https://www.nature.com/articles/srep00197>
57. Steinhert-Threlkeld Z (2017) Spontaneous collective action: peripheral mobilization during the Arab spring. *Am Polit Sci Rev* 111(2):379–403
58. Traag V (2016) Complex contagion of campaign donations. *PLoS One* 11(4):e0153539
59. Parigi P, Gong R (2014) From grassroots to digital ties: a case study of a political consumerism movement. *J Consum Cult* 14(2):236–253
60. Parkinson S (2013) Organizing rebellion: rethinking high-risk mobilization and social networks in war. *Am Polit Sci Rev* 107(3):418–432
61. Yasseri T, Hale S, Margetts H (2014) Modeling the rise in Internet-based petitions. arXiv. <https://arxiv.org/abs/1308.0239>
62. Siegel D (2009) Social networks and collective action. *Am J Polit Sci* 53(1):122–138
63. Siegel D (2011) When does repression work? Collective action in social networks. *J Politics* 73(4):993–1010
64. DellaPosta D, Nee V, Opper S (2017) Endogenous dynamics of institutional change. *Ration Soc* 29(1):5–48
65. Catlaw, T., and Stout, M. (2016). Governing small-town America today: the promise and dilemma of dense networks. The American Association for Public Administration, Washington, 76(2), 225–229
66. Pan W, Ghoshal G, Krumme C, Cebrian M, Pentland A (2013) Urban characteristics attributable to density-driven tie formation. *Nat Commun* 4:1961
67. Barash V, Cameron C, Macy M (2012) Critical phenomena in complex contagions. *Soc Networks* 34:451–461

68. Takaguchi T, Masuda N, Holme P (2013) Bursty communication patterns facilitate spreading in a threshold-based epidemic dynamics. *PLoS One* 8(7):e68629
69. Sun E, Rosenn I, Marlow C, Lento T (2009) Gesundheit! modeling contagion through Facebook news feed. In: Third international AAAI conference on weblogs and social media
70. Gleeson JP, Cahalane DJ (2007) Seed size strongly affects cascades on random networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 75:056103
71. Dodds PS, Payne JL (2009) Analysis of a threshold model of social contagion on degree-correlated networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 79(6 Pt 2):066115
72. Gleeson JP (2008) Cascades on correlated and modular random networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 77(4 Pt 2):046117
73. Gai P, Kapadia S (2010) Contagion in financial networks. *Proc R Soc A* 466:2401–2423
74. Hurd TR, Gleeson JP (2013) On watts cascade model with random link weights. *J Complex Netw* 1:25–43
75. Galstyan A, Cohen P (2007) Cascading dynamics in modular networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 75(3 Pt 2):036109
76. Hackett A, Melnik S, Gleeson JP (2011) Cascades on a class of clustered random networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 83(5 Pt 2):056107
77. Ikeda Y, Hasegawa T, Nemoto K (2010) Cascade dynamics on clustered network. *J Phys* 221:012013
78. Backlund VP, Saramaki J, Pan RK (2014) Effects of temporal correlations on cascades: threshold models on temporal networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 89(6):062815
79. Karimi K, Holme P (2013) Threshold model of cascades in empirical temporal networks. *Phys A* 392:3476–3483
80. Brummitt CD, Lee KM, Goh KI (2012) Multiplexity-facilitated cascades in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 85(4 Pt 2):045102
81. Lee KM, Brummitt CD, Goh KI (2014) Threshold cascades with response heterogeneity in multiplex networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 90(6):062816
82. Yağan O, Gligor V (2012) Analysis of complex contagions in random multiplex networks. *Phys Rev E* 86:036103
83. Shu P, Gao L, Zhao P, Wang W, Stanley H (2017) Social contagions on interdependent lattice networks. *Sci Rep* 7:44669
84. Bassett D, Alderson D, Carlson J (2012) Collective decision dynamics in the presence of external drivers. arXiv. <https://arxiv.org/abs/1206.1120>
85. Zhao JH, Zhou HJ, Liu YY (2013) Inducing effect on the percolation transition in complex networks. *Nat Commun* 4:2412
86. O'Sullivan D, Keefe G, Fennell P, Gleeson J (2015) Mathematical modeling of complex contagion on clustered networks. *Front Phys* 8:71
87. Pastor-Satorras R, Castellano C, Mieghem PV, Vespignani A (2015) Epidemic processes in complex networks. arXiv. <https://arxiv.org/abs/1408.2701>
88. Wang W, Shu P, Zhu YX, Tang M, Zhang YC (2015) Dynamics of social contagions with limited contact capacity. *Chaos* 25(10):103102
89. Wang W, Tang M, Zhang HF, Lai YC (2015) Dynamics of social contagions with memory of nonredundant information. *Phys Rev E Stat Nonlin Soft Matter Phys* 92(1):012820
90. Perez-Reche FJ, Ludlam JJ, Taraskin SN, Gilligan CA (2011) Synergy in spreading processes: from exploitative to explorative foraging strategies. *Phys Rev Lett* 106(21):218701
91. McCullen N, Rucklidge A, Bale C, Foxon T, Gale W (2013) Multiparameter models of innovation diffusion on complex networks. *J Appl Dyn Syst* 12:515–532
92. Dodds PS, Harris KD, Danforth CM (2013) Limited imitation contagion on random networks: chaos, universality, and unpredictability. *Phys Rev Lett* 110(15):158701
93. Melnik S, Ward JA, Gleeson JP, Porter MA (2013) Multi-stage complex contagions. *Chaos* 23:013124
94. Huang W, Zhang I, Xu X, Fu X (2016) Contagion on complex networks with persuasion. *Sci Rep* 6:23766. <https://doi.org/10.1038/srep23766>

95. Ruan Z, Iniguez G, Karsai M, Kertesz J (2015) Kinetics of social contagion. *Phys Rev Lett* 115(21):218702
96. Kuhlman CJ, Kumar VSA, Marathe MV et al (2015) Inhibiting diffusion of complex contagions in social networks: theoretical and experimental results. *Data Min Knowl Disc* 29:423. <https://doi.org/10.1007/s10618-014-0351-4>
97. Centola D (2009) Failure in complex networks. *J Math Sociol* 33:64–68
98. Albert R, Jeong H, Barabási A (2000) Error and attack tolerance of complex networks. *Nature* 406:378–382
99. Blume L, Easley D, Kleinberg J, Kleinberg R, Tardos E (2011) Which networks are least susceptible to cascading failures? In: IEEE 52nd annual symposium on foundations of computer science (FOCS)
100. Myers S, Leskovec J (2013) Clash of the contagions: cooperation and competition in information diffusion. In: IEEE 12th international conference on data mining (ICDM)
101. Jie R, Qiaoa J, Xub G, Menga Y (2016) A study on the interaction between two rumors in homogeneous complex networks under symmetric conditions. *Phys A* 454:129–142
102. Su Y, Zhang X, Liu L et al (2016) Understanding information interactions in diffusion: an evolutionary game-theoretic perspective. *Front Comput Sci* 10:518. <https://doi.org/10.1007/s11704-015-5008-y>
103. Teng C, Gong L, Eecs A, Brunetti C, Adamic L (2012) Coevolution of network structure and content. In: Proceedings of WebSci '12 proceedings of the 4th annual ACM web science conference, pp 288–297
104. Rosenthal SB, Twomey CR, Hartnett AT, Wu HS, Couzin ID (2015) Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *PNAS* 112(15):4690–4695
105. Dimaggio P (1997) Culture and cognition. *Annu Rev Sociol* 23:263–287
106. Puglisi A, Baronchelli A, Loreto V (2008) Cultural route to the emergence of linguistic categories. *PNAS* 105(23):7936–7940
107. Coman A, Momennejad I, Drach RD, Geana A (2016) Mnemonic convergence in social networks: the emergent properties of cognition at a collective level. *PNAS* 113(29):8171–8176
108. Morris S (2000) Contagion. *Rev Econ Stud* 67:57–78
109. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *PNAS* 106(51):21544–21549
110. Alvarez-Galvarez J (2015) Network models of minority opinion spreading: using agent-based Modeling to study possible scenarios of social contagion. *Soc Sci Comput Rev* 34(5):567–581

# A Simple Person’s Approach to Understanding the Contagion Condition for Spreading Processes on Generalized Random Networks



Peter Sheridan Dodds

## 1 Introduction

Given a local contagion mechanism acting on a random network, and a seed set of nodes  $\mathcal{N}_0$ , we would like to know the answers to a series of increasingly specific questions:

- Q1: Is a global spreading event possible? We’ll define a “global spreading event” as one that reaches a non-zero fraction of a network in the infinite limit.
- Q2: If a global spreading event is possible, what’s the probability of one occurring?
- Q3: What’s the distribution of final sizes for all spreading events?
- Q4: Global or not, how does the spreading from the seed set  $\mathcal{N}_0$  unfold in time?

Now, if we know the full time course of a spreading event (Q4) (see [11]), we evidently will be able to answer questions 1, 2, and 3. We might be tempted to take on only the more challenging analytical work and call it day (or appropriate time frame of suffering required). But it turns out to be useful to address each question separately.

While we will take on these questions for simple model distillations only, their real-world counterparts are some of the most important ones we face. What’s the probability that a certain fraction of a population will contract influenza? Could an ecosystem collapse? Indeed, the biggest question for many systems is:

- Q5: If we have limited knowledge of a network and limited control, how do we optimally facilitate or prevent spreading [21, 37]?

---

P. S. Dodds (✉)

Vermont Complex Systems Center, Computational Story Lab, the Vermont Advanced Computing Core, Department of Mathematics & Statistics, The University of Vermont, Burlington, VT, USA  
e-mail: [peter.dodds@uvm.edu](mailto:peter.dodds@uvm.edu)

In this chapter, we'll focus on Q1, determining the *contagion condition* for a range of contagion processes on random networks including bipartite ones. We will do so by plainly encoding the course of the spreading process itself into the contagion condition.

We will take the basic contagion mechanism to be one for which there are node states: Susceptible (S) and Infected (I). We will also prevent nodes from recovering or becoming susceptible; once nodes are infected, they remain so. In mathematical epidemiology, such models are referred to as SI, where S stands for Susceptible and I for Infected. Two other commonly studied models are SIR and SIRS, where a recovered immune state R is allowed for both and the possibility of cycling in the latter.

For the most part, we will be considering infinite random networks. If needed, we will define such networks as the limit of a one parameter family of networks (e.g., Erdős-Rényi networks with increasing  $N$  and mean degree held constant). As a rough guide for simulations, using around  $N = 10^4$  nodes is typically sufficient for yield results that visually conform well to theoretical ones (e.g., fractional size of the largest component in Erdős-Rényi networks).

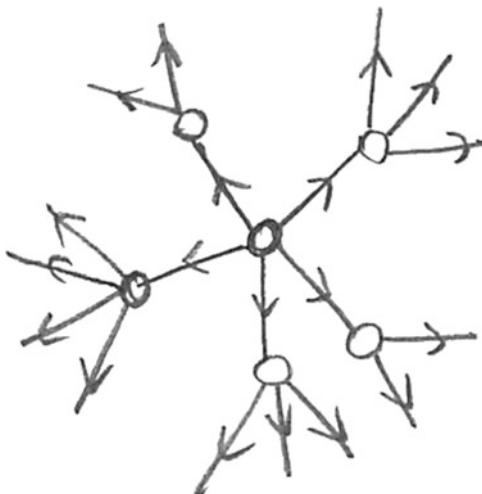
## 2 Elements of Simple Contagion on Random Networks

The key feature of random networks for spreading is that they are locally pure branching structures. This remains true for a large number of variations on random networks such as correlated random networks and bipartite affiliation graphs. Successful spreading away from a single seed (which could be one of many seeds) can only occur if nodes are susceptible when just one of their neighbors is infected (see Fig. 1). We will refer to these easily susceptible nodes as critical nodes (called vulnerable nodes in [36]). Denoting a network's entire node set as  $\Omega$ , global spreading will only be possible if there is a connected subnetwork of critical nodes that forms a giant component, the critical mass network  $\Omega_{\text{crit}}$ .

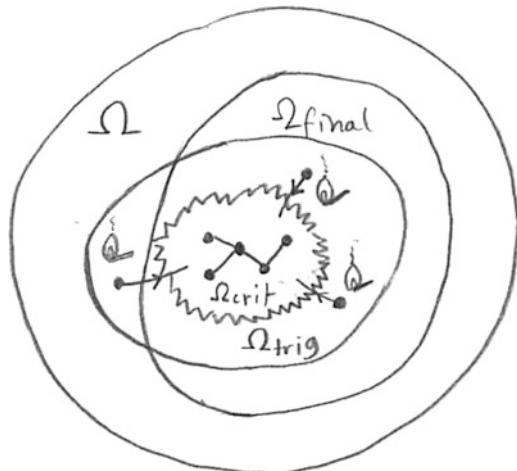
This set of critical nodes behaves in the same way as a critical mass one does for collective action [13, 27–29] but there is now an internal dynamic. If one node is infected within the critical mass network  $\Omega_{\text{crit}}$ , then spreading to some fraction of the critical mass network and beyond is possible, depending on the probabilistic nature of the contagion process.

There are two other subnetworks that need to be characterized to understand spreading on random networks. First, containing the critical mass network and all non-critical nodes connected to the critical mass network is the triggering component,  $\Omega_{\text{trig}}$ . Knowledge of this structure is required to determine the probability of a global spreading event [17]. Second, we have  $\Omega_{\text{final}}$  which is the extent of infection realized for any spreading event. For random networks, the distribution of the fractional size of  $\Omega_{\text{final}}$  will be either unimodal (the contagion process always succeeds) or bimodal (initial failure is possible).

**Fig. 1** Random networks are locally pure branching structures. For the initial stages of the spread shown, nodes can only experience the infection from a single neighbor. For spreading to take off from a simple seed, the network must contain a connected macroscopic critical mass network  $\Omega_{\text{crit}}$  of nodes susceptible to a single neighbor becoming infected



**Fig. 2** One possible arrangement of the three essential subnetworks for a contagion process on a random network: the critical mass network  $\Omega_{\text{crit}}$ , the triggering component  $\Omega_{\text{trig}}$ , and the final extent of a global spreading process,  $\Omega_{\text{final}}$ . In general,  $\Omega_{\text{crit}} \subset \Omega_{\text{trig}}$ ,  $\Omega_{\text{crit}} \subset \Omega_{\text{final}}$ , and  $\Omega_{\text{trig}}, \Omega_{\text{final}} \subset \Omega$



In Fig. 2, we show how the three subnetworks  $\Omega_{\text{crit}}$ ,  $\Omega_{\text{trig}}$ , and  $\Omega_{\text{final}}$  potentially overlap. A global spreading event is only possible if  $\Omega_{\text{crit}}$  takes up a non-zero fraction of the network. Some limiting cases allow for surprising kinds of robust-yet-fragile contagion, such as  $\Omega_{\text{crit}}$  being vanishingly small while any successful infection spreads to the full network [36].

### 3 The Contagion Condition

We would like to devise some kind of general, quick test algorithm into which we would be able to feed any contagion mechanism and any network, whether constructed or real. Such an algorithm would generate what we'll call a Contagion

Condition, and would only be worthwhile if it avoided simulating all possible spreading events and instead computed a composite test statistic. Upon running a system through our algorithm we would simply receive a “Yes” or “No.” Scaling up, we could then test an array of systems in parallel and for the “Yes” responses, we would proceed to explore those systems in detail (e.g., those cities which are susceptible to Zombie outbreaks [24]).

### 3.1 Contagion Condition for One-Shot Spreading Processes

For random network models, our test algorithm can be formulated in a physically-minded way. We will step through the building of the contagion condition for one-shot, permanent infection spreading on generalized, uncorrelated random networks and then expand from there.

By one-shot spreading, we mean that each newly infected node has one chance in the next time step to infect its uninfected neighbors. That is, if node  $i$  fails to infect a specific neighbor  $i'$ , then  $i$  cannot attempt to infect  $i'$  again in any following time step. Permanent infection means that nodes do not recover.

For a node  $i$  with degree  $k$ , we will write  $i$ ’s probability of infection given  $j$  of its neighbors are infected as  $B_{kj}$ . While our focus on the initial spread on random networks means we need only consider the probability nodes are infected by one of their neighbors,  $B_{k1}$ , we must consider the response to multiple simultaneous infections for later stages of global spreading on random networks [10, 11], more complicated contagion mechanisms, and, more importantly if we care about the real world, networks with non-zero clustering [25, 38].

As is often the case with networks, we open up better ways to understand and explain phenomena if we focus on edges rather than nodes. This is not entirely natural as for many problems we are ultimately concerned with how nodes behave and, for contagion especially, we can readily map ourselves directly onto individual nodes (will my next movie fail?). But once we lose this anchoring and shift to thinking first about edges with nodes in the background, clearer paths emerge.

So, instead of framing spreading as rooted in node infection rates, we consider the dynamics of infected edges. For our purposes, an infected edge will be one emanating from an infected node, and we will have to consider direction even for undirected networks.

We need to determine one number for our system, what we’ll call the gain ratio,  $\mathbf{R}$  [6]. We define  $\mathbf{R}$  as the expected number of newly infected edges that will be generated by a single infected edge leading to an uninfected node. (In epidemiology, the gain ratio would be equivalent to the reproduction number,  $R_0$ .)

For the moment, let’s assume we have computed  $\mathbf{R}$  for a system. Because sparse random networks are locally pure branching structures (see Fig. 1), the spread emanating from a single seed will also be a simple branching one. Early on, there will be no interactions between any two newly infected edges leading to the same uninfected node.

The fraction of newly infected edges at time  $t$ ,  $f_{(\cdot)}^{\text{inf}}(t)$ , must then follow an elementary evolution:

$$f_{(\cdot)}^{\text{inf}}(t) = \mathbf{R} f_{(\cdot)}^{\text{inf}}(t-1). \quad (1)$$

The subscript for the count  $f^{\text{inf}}$  will indicate the edge's type which for our initial system is irrelevant, hence  $(\cdot)$ .

The early growth will therefore be exponential with

$$f_{(\cdot)}^{\text{inf}}(t) = \mathbf{R}^t f_{(\cdot)}^{\text{inf}}(0), \quad (2)$$

where  $f_{(\cdot)}^{\text{inf}}(0)$  equals the degree of the seed node. We might guess that we can write down the exact evolution as  $f_{(\cdot)}^{\text{inf}}(t) = \mathbf{R}^t f_{(\cdot)}^{\text{inf}}(0)$ , but the initial step is sneakily different. Well get to this issue later on.

Global spreading will evidently be possible only if

$$\mathbf{R} > 1, \quad (3)$$

and this very simple criterion will be our Contagion Condition.

The above equations maintain the same form if we consider not one seed but a random seed set taking up a non-zero fraction of the random network. Writing  $\rho_t$  as the fraction of edges emanating from newly infected nodes at time  $t$ , we have, again for the initial phase of spreading:

$$\rho_t = \mathbf{R} \rho_{t-1}, \quad (4)$$

which leads to

$$\rho_t = \mathbf{R}^t \rho_0. \quad (5)$$

We now determine the gain ratio  $\mathbf{R}$  for the simple class of one-shot contagion on random network systems. In doing so, we show that the Contagion Condition is worthwhile beyond being a simple diagnostic as, with the right treatment, it can be also seen to carry physical intuition.

In determining  $\mathbf{R}$ , there are three (3) pieces to consider: two are structural and a function of the network, and the third couples the contagion mechanism to the network.

1. We start on an edge that has just become infected and look toward the uninfected node that has now become exposed. The properly normalized probability that this node has degree  $k$  is

$$Q_k = \frac{k P_k}{\langle k \rangle} \quad (6)$$

because each degree  $k$  node can be reached along its  $k$  edges. This skewing of the degree distribution is a result of some renown as it drives the Simon-like rich-get-richer models of network growth of Price [4, 5] and Barabási and Albert [2],

and also underlies the friendship paradox and its generalizations [7, 23]: Your friends are quite likely to be different from you, and often in disappointing ways such as by having more friends or wealth on average.

2. Second, we have the action of contagion mechanism. As have already defined, with probability  $B_{k1}$  the node of degree  $k$  is infected by the single incoming infected edge. With probability  $1 - B_{k1}$ , the infection fails.
3. Depending on whether or not the infection is successful, we know that in the next time step the contagion mechanism will generate either 0 or  $k - 1$  new infected edges.

Putting these pieces together, we have

$$\begin{aligned} \mathbf{R} = & \sum_{k=0}^{\infty} \underbrace{\frac{k P_k}{\langle k \rangle}}_{\text{prob. of connecting to a degree } k \text{ node}} \bullet \underbrace{B_{k1}}_{\text{Prob. of infection}} \bullet \underbrace{(k-1)}_{\# \text{ outgoing infected edges}} \\ & + \sum_{k=0}^{\infty} \underbrace{\frac{k P_k}{\langle k \rangle}}_{\text{prob. of connecting to a degree } k \text{ node}} \bullet \underbrace{(1 - B_{k1})}_{\text{Prob. of no infection}} \bullet \underbrace{(0)}_{\# \text{ outgoing infected edges}} \end{aligned} \quad (7)$$

The second piece evaporates and we have our contagion condition:

$$\mathbf{R} = \sum_{k=0}^{\infty} \frac{k P_k}{\langle k \rangle} \bullet B_{k1} \bullet (k-1) > 1. \quad (8)$$

Again, the value here is that this structure of  $\mathbf{R}$  encodes the contagion mechanism in a clear way. As such, we resist any urge to rearrange the form of Eq. (8) for a more elegant form. As we move to more general systems, the three part form of two pieces for the network and one for the contagion mechanism will be maintained, and the criterion of a single number exceeding unity,  $\mathbf{R} > 1$ , will elevate to being the largest eigenvalue of a gain ratio matrix exceeding unity.

We now move through a few examples of other kinds of systems involving contagion mechanisms acting on network structures.

### 3.2 Contagion Condition for Multiple-Shot Spreading Processes

We have presumed a one-shot contagion process in our derivation of Eq. (8). In loosening this restriction to spreading processes that may involve repeated attempts

to infect a node with the possible recovery of the infected node allowed as well, we can compute  $B_{k1}$  as the long-term probability of infection. The form of gain ratio remains the same and therefore so does the contagion condition given in Eq. (8).

### 3.3 Remorseless Spreading and the Giant Component Condition

We step back from contagion momentarily to show that we can also determine whether or not a random network has a giant component. This is now a structural test absent any processes. A network will have a giant component if it is, on average, locally expanding. That is, if we travel along a randomly chosen edge, we will reach a node which has, on average, more than one other edge emanating from it. But this is just a remorseless version of our one-shot contagion mechanism, one where infection always succeeds, i.e.,  $B_{k1} = 1$ .

Setting  $B_{k1} = 1$  in Eq. (8), we have the giant component condition:

$$\mathbf{R} = \sum_{k=0}^{\infty} \frac{k P_k}{\langle k \rangle} \bullet (k - 1) > 1, \quad (9)$$

where we have again used the physical sense of a gain ratio.

### 3.4 Simple Contagion on Generalized Random Networks

If  $B_{k1} = B < 1$ , a fraction  $(1-B)$  of all edges will not transmit infection, and the contagion condition becomes

$$\mathbf{R} = \sum_{k=0}^{\infty} \frac{k P_k}{\langle k \rangle} \bullet B \bullet (k - 1) > 1. \quad (10)$$

This is a bond percolation model [33], and Eq. (10) can be seen as a giant component condition for a network with  $(1-B)$  of its edges removed. The resultant network has a degree distribution  $\tilde{P}_k = B^k \sum_{i=k}^{\infty} \binom{i}{k} (1-B)^{i-k} P_i$ , and evidently, as  $B$  decreases, only increasingly more connected networks will be able to facilitate spreading.

### 3.5 Other Routes to Determining the Contagion and Giant Component Conditions

There are many other ways to arrive at the contagion condition in Eq. (8) and the giant component condition in Eq. (9). The path taken affects the form of the

condition and may limit understandability [6]. For example, the giant component condition was determined by Molloy and Reed [22] in 1995 and presented as

$$\sum_{k=0}^{\infty} k(k-2)P_k > 0. \quad (11)$$

While equivalent to Eq. (9), the framing of local expansion is obscured.

For a simple spreading mechanism with  $B_{k1} = B$ , Newman [25], for example, used generating functionology methods [40] to first determine the average size of finite components and then find when this quantity diverged. For Granovetter's social contagion threshold model on random networks [13], Watts took the same approach [36]. This size divergence is a hallmark of phase transitions in statistical mechanical systems in general, and while it can be used to find the critical point, doing so would ideally be at the level of a consistency check.

For the giant component condition, a somewhat more direct approach using generating functions [26] is based on the probability distribution that the node at the randomly chosen end of a randomly chosen edge has  $k$  other edges is

$$R_k = Q_{k+1} = \frac{1}{\langle k \rangle} (k+1) P_{k+1}. \quad (12)$$

Writing the generating function for the degree distribution as  $F_P(x) = \sum_{k=0}^{\infty} P_k x^k$ , we have  $F_R(x) = F'_P(x)/F'_P(1)$ , where we have used  $\langle k \rangle = F'_P(1)$ , an elementary result for determining averages with generating functions [40]. The average number of other edges found at a randomly-arrived-at node is  $F'_R(1) = F''_P(1)/F'_P(1) = \frac{\langle k(k-1) \rangle}{\langle k \rangle}$ . This is exactly our gain ratio and we now have

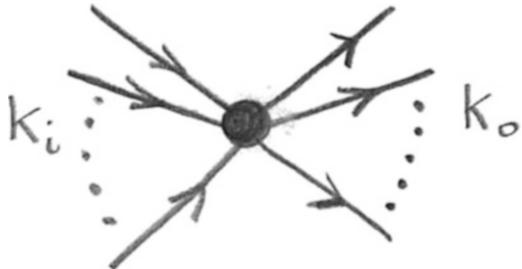
$$\frac{\langle k(k-1) \rangle}{\langle k \rangle} > 1 \quad (13)$$

for the giant component condition. Again, while Eqs. (9) and (13) are equivalent, the latter does not have an immediate physical interpretation—it's just a condition.

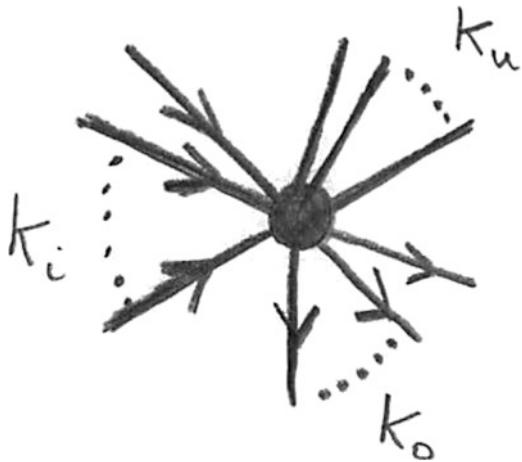
### 3.6 Simple Contagion on Generalized Directed Random Networks

For purely directed networks, we allow each node to have an in-degree  $k_i$  and an out-degree  $k_o$  with probability  $P_{k_i, k_o}$  (see Fig. 3). The same arguments that gave us Eq. (8) now end with:

**Fig. 3** For general directed networks, a node has  $k_i$  incident edges and  $k_o$  emanating edges governed by a joint distribution  $P_{k_i, k_o}$



**Fig. 4** Nodes in mixed random networks have  $k_u$  undirected edges,  $k_i$  incident edges, and  $k_o$  emanating edges. Node degree is represented by the vector  $\mathbf{k} = [k_u \ k_i \ k_o]^T$  and degrees are sampled from a joint distribution  $P_{\mathbf{k}}$



$$\mathbf{R} = \sum_{k_i=0}^{\infty} \sum_{k_o=0}^{\infty} \frac{k_i P_{k_i, k_o}}{\langle k_i \rangle} \bullet B_{k_i, 1} \bullet k_o > 1. \quad (14)$$

The three components of the contagion condition have the same interpretation as before (Fig. 4).

### 3.7 Simple Contagion on Mixed, Correlated Random Networks

We jump to a more complex possibility of mixed random networks with a combination of directed and undirected (or bidirectional) edges as well as arbitrary degree-degree correlations between nodes, as introduced in [3].

Nodes may have three types of edges:  $k_u$  undirected edges,  $k_i$  incoming directed edges, and  $k_o$  outgoing directed edges. The degree distribution is now a function of a three-vector:

$$P_{\mathbf{k}} \text{ where } \mathbf{k} = [k_u \ k_i \ k_o]^T. \quad (15)$$

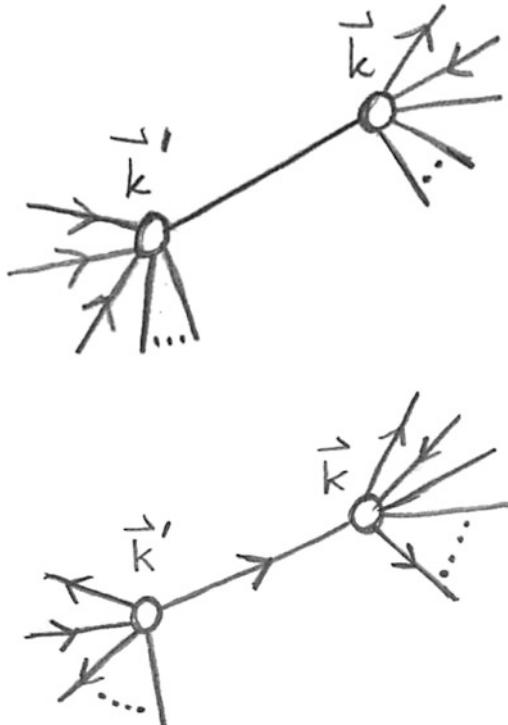
As for directed networks, we require in- and out-degree averages to match up:  $\langle k_i \rangle = \langle k_o \rangle$ . We add two point correlations per [3, 6] through three conditional probabilities:

- $P^{(u)}(\mathbf{k} | \mathbf{k}')$  = probability that an undirected edge leaving a degree  $\mathbf{k}'$  nodes arrives at a degree  $\mathbf{k}$  node.
- $P^{(i)}(\mathbf{k} | \mathbf{k}')$  = probability that an edge leaving a degree  $\mathbf{k}'$  nodes arrives at a degree  $\mathbf{k}$  node is an in-directed edge relative to the destination node.
- $P^{(o)}(\mathbf{k} | \mathbf{k}')$  = probability that an edge leaving a degree  $\mathbf{k}'$  nodes arrives at a degree  $\mathbf{k}$  node is an out-directed edge relative to the destination node.

We now require more refined (detailed) balance along both undirected and directed edges (see Fig. 5). Specifically, we must have [3, 6]:  $P^{(u)}(\mathbf{k} | \mathbf{k}') \frac{k'_u P(\mathbf{k}')}{\langle k'_u \rangle} = P^{(u)}(\mathbf{k}' | \mathbf{k}) \frac{k_u P(\mathbf{k})}{\langle k_u \rangle}$ , and  $P^{(i)}(\mathbf{k} | \mathbf{k}') \frac{k'_o P(\mathbf{k}')}{\langle k'_o \rangle} = P^{(o)}(\mathbf{k}' | \mathbf{k}) \frac{k_i P(\mathbf{k})}{\langle k_i \rangle}$ .

For all example systems so far, the gain ratio has been a single number. For mixed random networks, infections along directed edges may cause infections along undirected edges and so on. We will need to count undirected and directed edge infections separately, the growth of infections for a one-shot contagion process will obey the following dynamic:

**Fig. 5** For mixed random networks, node degree correlations may be measured along undirected and/or directed edges



$$\begin{bmatrix} f_{\mathbf{k}}^{(u)}(t+1) \\ f_{\mathbf{k}}^{(o)}(t+1) \end{bmatrix} = \sum_{\mathbf{k}'} \mathbf{R}_{\mathbf{k}\mathbf{k}'} \begin{bmatrix} f_{\mathbf{k}'}^{(u)}(t) \\ f_{\mathbf{k}'}^{(o)}(t) \end{bmatrix}, \quad (16)$$

where we now identify a gain ratio tensor:

$$\mathbf{R}_{\mathbf{k}\mathbf{k}'} = \begin{bmatrix} P^{(u)}(\mathbf{k} | \mathbf{k}') \bullet B_{\mathbf{k}\mathbf{k}'} \bullet (k_u - 1) & P^{(i)}(\mathbf{k} | \mathbf{k}') \bullet B_{\mathbf{k}\mathbf{k}'} \bullet k_u \\ P^{(u)}(\mathbf{k} | \mathbf{k}') \bullet B_{\mathbf{k}\mathbf{k}'} \bullet k_o & P^{(i)}(\mathbf{k} | \mathbf{k}') \bullet B_{\mathbf{k}\mathbf{k}'} \bullet k_o \end{bmatrix}. \quad (17)$$

For a gain ratio matrix or tensor, our contagion condition is now a test of whether or not the largest eigenvalue exceeds 1.

### 3.8 Contagion on Correlated Random Networks with Arbitrary Node and Edge Types

We make one last step of generalization for correlated random networks [6]. As per Fig. 6, we allow arbitrary types of nodes and edges along with arbitrary correlations between node-edge pairs. For multi-shot contagion, we have

$$f_{\alpha}(d+1) = \sum_{\alpha'} R_{\alpha\alpha'} f_{\alpha'}(d) \quad (18)$$

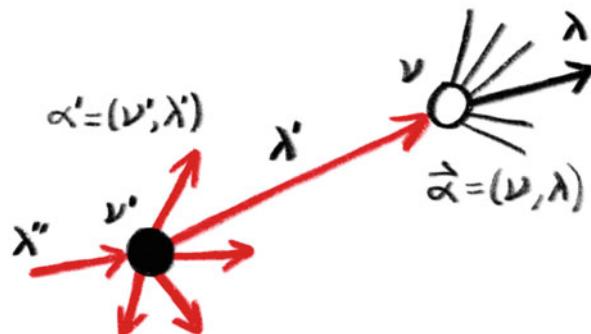
where  $R_{\alpha\alpha'}$  is the gain ratio matrix and has the form:

$$R_{\alpha\alpha'} = P_{\alpha\alpha'} \bullet k_{\alpha\alpha'} \bullet B_{\alpha\alpha'}. \quad (19)$$

Here,

- $P_{\alpha\alpha'}$  = conditional probability that a type  $\lambda'$  edge emanating from a type  $\nu'$  node leads to a type  $\nu$  node.

**Fig. 6** Element of a general correlated random network where edges and nodes may take on arbitrary characteristics. Node and edge type are specified as  $\alpha = (\nu, \lambda)$



- $k_{\alpha\alpha'} =$  potential number of newly infected edges of type  $\lambda$  emanating from nodes of type  $\nu$ .
- $B_{\alpha\alpha'} =$  probability that a type  $\nu$  node is eventually infected by a single infected type  $\lambda'$  link arriving from a neighboring node of type  $\nu'$ .

Finally, we can write down our generalized contagion condition as:

$$\max |\mu| : \mu \in \sigma(\mathbf{R}) > 1, \quad (20)$$

where  $\sigma(\mathbf{R})$  denotes the eigenvalue spectrum of  $\mathbf{R}$ .

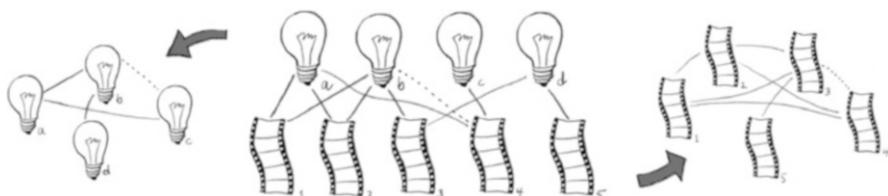
### 3.9 Simple Contagion on Bipartite Random Networks

Bipartite networks (or affiliation graphs) connect two populations through some association, and induce networks within each population [1, 9, 12, 18, 26, 34]. Bipartite structures and variants are natural representations of many real networked systems with a classic example being boards and directors. The induced distributions are formed by connecting all pairs of boards that share at least one director and all pairs of directors that belong to the same board.

Base models for real bipartite systems are random bipartite networks which are formed by randomly connecting two populations with specified degree distributions. Random bipartite networks are able to reproduce induced degree distributions, which may be non-trivial in form [26].

To help with our analysis, we'll consider a random bipartite network between stories and tropes [35]. Each story contains one or more trope, and each trope is part of one or more stories. Stories sharing tropes are then linked as are tropes found in the same story. In Fig. 7, we show a small example (center) along with the induced trope-trope and story-story networks.

For spreading between stories we may wish to imagine we're in the BookWorld of the Thursday Next series [8].



**Fig. 7** Example of a bipartite affiliation network and the induced networks. Center: A small story-trope bipartite graph. The induced trope network and the induced story network are on the left and right. The dashed edge in the bipartite affiliation network indicates an edge added to the system, resulting in the dashed edges being added to the two induced networks

We'll use this notation for our two inter-affiliated types:  $\blacksquare$  for stories and  $\heartsuit$  for tropes.

Consider a story-trope system with  $N_{\blacksquare}$  denoting the number of stories,  $N_{\heartsuit}$  the number of tropes, and  $m_{\blacksquare, \heartsuit}$  the number of edges connecting stories and tropes.

Let's have some underlying distributions for numbers of affiliations:  $P_k^{(\blacksquare)}$  (a story has  $k$  tropes) and  $P_k^{(\heartsuit)}$  (a trope is in  $k$  stories).

Some bookkeeping arises with balance requirements. Writing  $\langle k \rangle_{\blacksquare}$  as the average number of tropes per story, and  $\langle k \rangle_{\heartsuit}$  as the average number of stories containing a given trope, we must have:  $N_{\blacksquare} \cdot \langle k \rangle_{\blacksquare} = m_{\blacksquare, \heartsuit} = N_{\heartsuit} \cdot \langle k \rangle_{\heartsuit}$ .

Let's first get to the giant component condition before talking about contagion.

Just as for random networks, we focus on edges begetting edges, and we will need the distributions analogous to  $Q_k$ , Eq. (6). We randomly select an edge connecting a story  $\blacksquare$  to a trope  $\heartsuit$ . Traveling from the trope to the story, we have that the probability the story  $\blacksquare$  contains  $k$  total tropes is:

$$Q_k^{(\blacksquare)} = \frac{k P_k^{(\blacksquare)}}{\sum_{j=0}^{N_{\blacksquare}} j P_j^{(\blacksquare)}} = \frac{k P_k^{(\blacksquare)}}{\langle k \rangle_{\blacksquare}}. \quad (21)$$

Heading instead towards the trope  $\heartsuit$ , we find the probability that the trope  $\heartsuit$  is in  $k$  total stories is

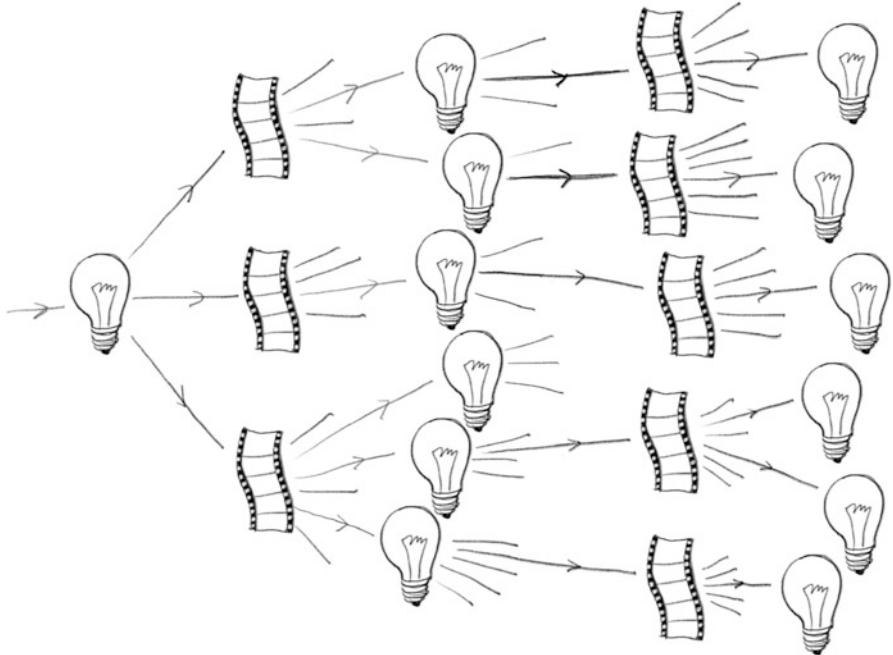
$$Q_k^{(\heartsuit)} = \frac{k P_k^{(\heartsuit)}}{\sum_{j=0}^{N_{\heartsuit}} j P_j^{(\heartsuit)}} = \frac{k P_k^{(\heartsuit)}}{\langle k \rangle_{\heartsuit}}. \quad (22)$$

To determine the giant component condition for the induced network of stories (to choose a side), let's start with a randomly chosen edge and travel from the story to the trope. As shown starting on the left of Fig. 8, we hit the trope and then travel to the other stories containing that trope. This bouncing back and forth between tropes and stories continues and because the connections are random and if the system is large enough, no story or trope is returned to early on. Just as for random networks, there are no short loops (technically, finitely many in the infinite limit).

We are thus able to depict the expanding branching in Fig. 8 and we can see that the giant component condition will involve the product of the gain ratio for each distribution.

$$\mathbf{R} = \mathbf{R}_{\blacksquare} \cdot \mathbf{R}_{\heartsuit} = \left[ \sum_{k=0}^{\infty} \frac{k P_k^{(\blacksquare)}}{\langle k \rangle_{\blacksquare}} \bullet (k-1) \right] \left[ \sum_{k=0}^{\infty} \frac{k P_k^{(\heartsuit)}}{\langle k \rangle_{\heartsuit}} \bullet (k-1) \right] > 1 \quad (23)$$

As for gain ratios for random networks we can arrive at this result through the use of generating functions and other approaches. Regardless of the path, more mathematically pleasing variants are always available such as [26]:



**Fig. 8** Spreading on a random bipartite network can be seen as bouncing back and forth between the two connected populations. The gain ratio for simple contagion on a bipartite random network is the product of two gain ratios as shown in Eq. (23)

$$\sum_{k=0}^{\infty} \sum_{k'=0}^{\infty} kk'(kk' - k - k') P_k^{(\boxplus)} P_{k'}^{(\boxtimes)} = 0, \quad (24)$$

but, again, we have stripped the physics away.

Introducing a simple contagion can be done as before by allowing tropes to infect other tropes in the same story (with probability  $B_{k1}^{(\boxtimes)}$ ) and stories to affect other stories if they share a trope (with probability  $B_{k1}^{(\boxplus)}$ ). We adjust Eq. (23) to obtain:

$$\begin{aligned} \mathbf{R} = \mathbf{R}_{\boxplus} \cdot \mathbf{R}_{\boxtimes} &= \left[ \sum_{k=0}^{\infty} \frac{k P_k^{(\boxplus)}}{\langle k \rangle_{\boxplus}} \bullet B_{k1}^{(\boxplus)} \bullet (k-1) \right] \\ &\times \left[ \sum_{k=0}^{\infty} \frac{k P_k^{(\boxtimes)}}{\langle k \rangle_{\boxtimes}} \bullet B_{k1}^{(\boxtimes)} \bullet (k-1) \right] > 1 \end{aligned} \quad (25)$$

### 3.10 Threshold Contagion on Generalized Random Networks

We turn to our last example: threshold contagion, an important simple model of social contagion [13–16, 30–32, 36]. In basic threshold contagion models, all

individuals observe the infection status of their neighbors at each time step, and become infected if their internal threshold is exceeded. In the present and following section, we will explore the contagion condition for threshold models on all-to-all networks and random networks, and examine the early course of a global spreading event reflecting on the nature of early adopters.

In Granovetter's mean-field or all-to-all network version [13], individuals are always aware of the overall fraction of the population that is infected. We write the fraction of the population that is infected at time  $t$  as  $a_t$ . If we have a general threshold distribution  $f(\phi)$ , then the fraction of the population whose threshold will be exceeded at time  $t$  and hence be infected at time  $t + 1$  is:

$$\phi_{t+1} = \int_0^{\phi_t} f(u)du = F(u)|_0^{\phi_t} = F(\phi_t) - F(0) \quad (26)$$

where  $F$  is the cumulative distribution of  $f$  (if  $F(0) > 0$ , then the system has nodes that will always be on regardless of the state of others). Thus, we have system whose dynamics are described by a map of the unit interval. We are interested in small seeds for the mean-field version, i.e.,  $\phi_0 \rightarrow 0$ . In this limit, global spreading occurs if (1)  $F(0) > 0$  meaning the population will always activate spontaneously, or (2)  $\phi = 0$  is a fixed point but is unstable (meaning  $F(0) = 0$  and  $F'(0) > 1$ ). If  $\phi = 0$  is a stable fixed point (meaning  $F(0) = 0$  and  $F'(0) < 1$ ), then spreading may still occur but not for vanishingly small seeds. Perhaps surprisingly, the same process on a network may give rise to spreading from a single seed, as we explain this in the next section.

For the random network version due to Watts [36], and again taking a general threshold distribution  $f(\phi)$  a degree  $k$  node will be part of the critical mass network with probability:

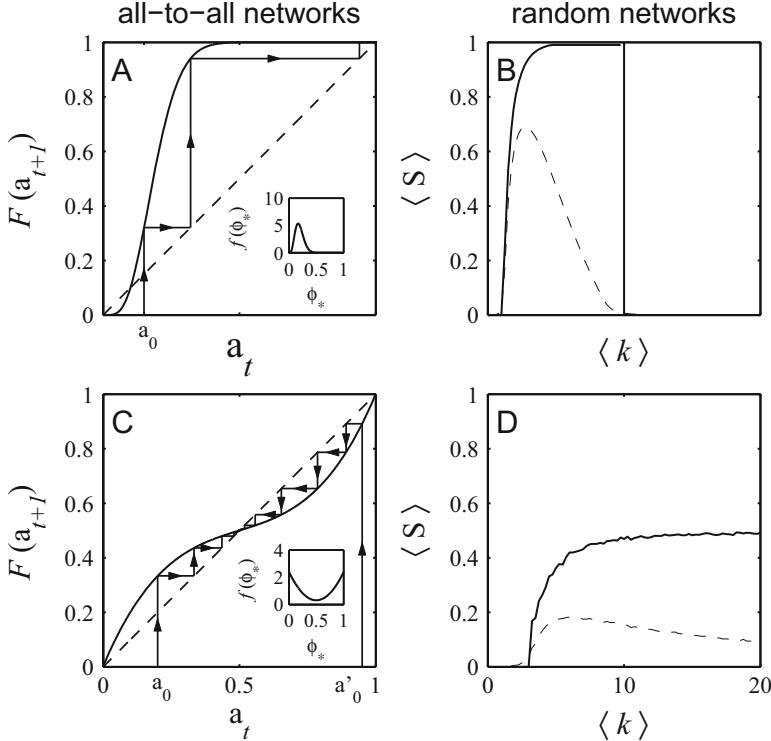
$$B_{k1} = \int_0^{1/k} f(\phi)d\phi. \quad (27)$$

The gain ratio remains the same as the one given in Eq. (8).

We now link the contagion conditions for the all-to-all network and random network versions of social contagion.

### 3.11 Connecting the Contagion Condition for All-To-All and Random Networks for Threshold Contagion

We make the simple observation that if we examine the threshold model's behavior on a random network and allow the average degree  $\langle k \rangle$  to increase, then the results will tend towards what we would observe on an all-to-all network. Since the limiting behavior of the contagion model on all-to-all networks is governed by the presence or absence of fixed points of the cumulative threshold distribution  $F$ , we are therefore able to state what the model's behavior on random networks must tend towards as  $\langle k \rangle$  increases based solely on the form of  $F$ .



**Fig. 9** Plots comparing the behavior of the model on all-to-all networks (plots (a) and (c)) and random networks ((b) and (d)) for two different example threshold distributions. The insets to plots (a) and (c) show the two underlying threshold distributions, which are unimodal and bimodal, respectively, and the corresponding cumulative distributions are presented in the main plots of (a) and (c). Plots (b) and (d) show global spreading event intervals for random networks with the same threshold distributions as (a) and (c), respectively. The black lines in (b) and (d) indicate the average size of global spreading events that exceed  $0.05N$ , and the dashed lines the average size of the largest critical mass network (sizes are normalized by  $N$ ). The threshold distribution in plot A leads to a bounded global spreading event interval on random networks while the distribution in plot (c) leads to an unbounded one. In plot (d), the average size of the largest critical mass network decays to 0 as  $\langle k \rangle \rightarrow \infty$ . The results in plots (b) and (d) are derived from  $10^3$  networks with  $N = 10^4$  and one seed per network

We consider two examples of threshold distribution  $f$  to facilitate our discussion. First, for a general threshold distribution  $f$ , it is useful for us to define a *global spreading event interval* as the range of  $\langle k \rangle$  for which global spreading events are possible on a random network. A simple example involving a bounded global spreading event interval and a non-trivial threshold distribution  $f$  is represented in Fig. 9a, b. The main plot of Fig. 9a shows the cumulative distribution  $F$ , and the inset shows the threshold distribution  $f$ . The all-to-all network model, Fig. 9a, exhibits a simple kind of critical mass behavior: the infection level approaches unity if the initial activated fraction  $\phi_0$  is above the sole unstable fixed point, or else it dies away.

Thus for all-to-all networks, a small initial infection level will always fail to yield global infection. For global spreading events to occur on all-to-all networks, some alternative seeding mechanism (an advertising campaign, perhaps) must precede the word-of-mouth dynamics so as to create a sufficiently large  $\phi_0$ .

By contrast, global spreading events can arise from a *single* infected individual in a sparse random network with exactly the same distribution of thresholds, as shown in Fig. 9b. The reason is that when individuals are connected to a limited number of alters within a population, the fraction of their neighbors that are infected may now be nonzero and thus may exceed their threshold (in infinite all-to-all networks, this fraction is always 0 for finite seeds). By effectively reducing the knowledge individuals have of the overall population—by increasing their ignorance—global spreading events become possible. Related observations invoke pluralistic ignorance [19, 20] and the importance of small groups in facilitating collective action [29] by circumventing the free rider problem.

Thus, when the threshold distribution  $f$  is fixed, we observe a connection between the results for spreading on all-to-all networks and random networks. Bounded global spreading event intervals can only occur when the mean-field version exhibits a critical mass property, i.e., when there exists a stable fixed point at the origin  $\phi = 0$  (i.e.,  $F(0) = 0$  and  $F'(0) < 1$ ). We know this because no small seed will ever be able to generate a global spreading event in the all-to-all case and that as the average degree of a random network increases, so too must its similarity in behavior to that of all-to-all networks. Furthermore, if there is a stable fixed point at the origin, whether or not global spreading events are possible at all in any random network depends on the global spreading event condition being satisfied. In other words, ignorance does not always help the spread of influence—some threshold distributions never lead to the contagion condition being satisfied for any value of  $\langle k \rangle$ .

*Unbounded global spreading event intervals* arise when there are sufficient individuals who will be vulnerable even if their degree is very high, i.e., when the threshold distribution has enough weight at or near  $\phi = 0$ . An example of an unbounded global spreading event interval is given in Fig. 9d with the underlying threshold distribution and its cumulative shown in Fig. 9c. Since small seeds always take off in the all-to-all network version, as network connectivity is increased, global spreading events continue to occur and the global spreading event interval is unbounded. The size of the largest critical mass network is nonzero for all finite  $\langle k \rangle$ , though it tends to 0 in the limit  $\langle k \rangle \rightarrow \infty$ . For highly connected random networks, the final size of the global spreading event again depends on the fixed points of  $F$ . For example, in Fig. 9b, global spreading events typically reach the full size of the giant component which corresponds to an upper stable fixed point of  $F$  at  $\phi = 1$ . In Fig. 9d, we see global spreading events only reach half the size of the population, corresponding to the stable fixed point of  $F$  at  $\phi = 1/2$ .

We thus see that in moving from all-to-all networks to random networks, the behavior of the threshold model changes qualitatively in the sense that there exist threshold distributions for which global spreading events started by a small seed cannot occur on an all-to-all network, yet may occur on sparse, random networks.

## 4 Concluding Remarks

For any parameterized system that may afford global spreading, the contagion condition is a fundamental criterion to determine. We have outlined the contagion condition for a range of contagion mechanisms acting on generalized random networks, showing that the condition can be derived so as to bear a clear imprint of the mechanism at work. A similar approach can be used to lay out the triggering probability of a global spreading event in a readable form [17].

While generating function approaches provided many of the first breakthroughs giving the possibility and probability of spreading [26, 39] and have yielded powerful access to many other results, they have tended to obscure the forms of the simplest ones such as the contagion condition. These techniques are also inherently indirect as they work by avoiding the giant component and characterizing only finite ones. Later work focusing on fractional seeds was able to go directly into the giant component and determine not just the final size but full time dynamics of global spreading events [10, 11], and we recommend continued pursuit of this line of attack going forward.

## References

1. Ahn YY, Ahnert SE, Bagrow JP, Barabási AL (2011) Flavor network and the principles of food pairing. *Nat Sci Rep* 1:196
2. Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286:509–511
3. Boguñá M, Ángeles Serrano M (2005) Generalized percolation in random directed networks. *Phys Rev E* 72:016106
4. de Solla Price DJ (1965) Networks of scientific papers. *Science* 149:510–515
5. de Solla Price DJ (1976) A general theory of bibliometric and other cumulative advantage processes. *J Am Soc Inform Sci* 27:292–306
6. Dodds PS, Harris KD, Payne JL (2011) Direct, physically motivated derivation of the contagion condition for spreading processes on generalized random networks. *Phys Rev E* 83:056122
7. Eom YH, Jo HH (2014) Generalized friendship paradox in complex networks: the case of scientific collaboration. *Nat Sci Rep* 4:4603
8. Fforde J (2001) *The Eyre affair: a thursday next novel*. New English Library, London
9. García-Pérez LP, Serrano MA, Boguñá M (2014) The complex architecture of primes and natural numbers. <http://arxiv.org/abs/1402.3612>
10. Gleeson JP (2008) Cascades on correlated and modular random networks. *Phys Rev E* 77:046117
11. Gleeson JP, Cahalane DJ (2007) Seed size strongly affects cascades on random networks. *Phys Rev E* 75:056103
12. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabási AL (2007) The human disease network. *Proc Natl Acad Sci* 104:8685–8690
13. Granovetter M (1978) Threshold models of collective behavior. *Am J Sociol* 83(6):1420–1443
14. Granovetter MS, Soong R (1983) Threshold models of diffusion and collective behavior. *J Math Sociol* 9:165–179
15. Granovetter MS, Soong R (1986) Threshold models of interpersonal effects in consumer demand. *J Econ Behav Organ* 7:83–99

16. Granovetter M, Soong R (1988) Threshold models of diversity: Chinese restaurants, residential segregation, and the spiral of silence. *Sociol Methodol* 18:69–104
17. Harris KD, Payne JL, Dodds PS (2014) Direct, physically-motivated derivation of triggering probabilities for contagion processes acting on correlated random networks. <http://arxiv.org/abs/1108.5398>
18. Hidalgo CA, Klinger B, Barabási AL, Hausman R (2007) The product space conditions the development of nations. *Science* 317:482–487. <https://doi.org/10.1126/science.1144581>
19. Kuran T (1991) Now out of never: the element of surprise in the east European revolution of 1989. *World Polit* 44:7–48
20. Kuran T (1997) Private truths, public lies: the social consequences of preference falsification, reprint edn. Harvard University Press, Cambridge
21. Lazarsfeld P, Merton R (1954) Friendship as social process: a substantive and methodological analysis. In: Berger M, Abel T, Page C (eds) *Freedom and control in modern society*. Van Nostrand, New York, pp 18–66
22. Molloy M, Reed B (1995) A critical point for random graphs on a fixed degree sequence. *Random Struct Algorithms* 6:161–180
23. Momeni N, Rabbat M (2016) Qualities and inequalities in online social networks through the lens of the generalized friendship paradox. *PLoS One* 11:e0143633
24. Munz P, Hudea I, Imad J, Smith? RJ (2009) When zombies attack!: mathematical modelling of an outbreak of zombie infection. In: Tchuenche JM, Chiyaka C (eds) *Infectious disease modelling research progress*. Nova Science, New York, pp 133–150
25. Newman MEJ (2003) The structure and function of complex networks. *SIAM Rev* 45(2):167–256
26. Newman MEJ, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. *Phys Rev E* 64:026118
27. Oliver PE (1993) Formal models of collective action. *Ann Rev Sociol* 19:271–300
28. Oliver PE, Marwell G, Teixeira R (1985) A theory of the critical mass. I. Interdependence, group heterogeneity, and the production of collective action. *Am J Sociol* 91(3):522–556
29. Olson M (1971) The logic of collective action: public goods and the theory of groups, revised edn. Harvard Economic Studies, Harvard University Press, Cambridge
30. Schelling TC (1971) Dynamic models of segregation. *J Math Sociol* 1:143–186
31. Schelling TC (1973) Hockey helmets, concealed weapons, and daylight saving: a study of binary choices with externalities. *J Confl Resolut* 17:381–428
32. Schelling TC (1978) *Micromotives and macrobehavior*. Norton, New York
33. Stauffer D, Aharony A (1992) *Introduction to percolation theory*, 2nd edn. Taylor & Francis, Washington
34. Teng CY, Lin YR, Adamic LA (2012) Recipe recommendation using ingredient networks. In: *Proceedings of the 3rd annual ACM web science conference, WebSci ’12*. ACM, New York, pp 298–307
35. TV Tropes (2017). <http://tvtropes.org>
36. Watts DJ (2002) A simple model of global cascades on random networks. *Proc Natl Acad Sci* 99(9):5766–5771
37. Watts DJ, Dodds PS (2007) Influentials, networks, and public opinion formation. *J Consum Res* 34:441–458
38. Watts DJ, Strogatz SJ (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393:440–442
39. Watts DJ, Dodds PS, Newman MEJ (2002) Identity and search in social networks. *Science* 296:1302–1305
40. Wilf HS (2006) *Generating functionology*, 3rd edn. A K Peters, Natick

# Challenges to Estimating Contagion Effects from Observational Data



Elizabeth L. Ogburn

## 1 Background

A network is a collection of units, or *nodes*, and the ties, or *edges*, between them. The presence of a tie between two nodes indicates that the nodes share some kind of a relationship; what types of relationships are encoded by network ties depends on the context. Some types of relationships are mutual (undirected), for example familial relatedness and shared place of work; others may go in only one direction. A node whose characteristics we wish to explain or model is called an *ego*; nodes that share ties with the ego are its *alters*. If an ego's outcome may be affected by his contacts' outcomes, then the outcome is said to exhibit *induction*, *contagion*, *peer effects*, or *peer influence*. For consistency we will use the term *contagion* throughout.

A growing body of literature attempts to learn about contagion using observational (i.e., non-experimental) data collected from a single social network. While the conclusions of these studies may be correct, the methods rely on assumptions that are likely—and sometimes guaranteed to be—false, and therefore the evidence for the conclusions is often weaker than it is portrayed to be. Developing methods that do not need to rely on implausible assumptions is an incredibly challenging and important open problem in statistics. Appropriate methods don't (yet!) exist, so researchers hoping to learn about contagion from observational social network data are sometimes faced with a dilemma: they can abandon their research program, or they can use inappropriate methods. This chapter will focus on the challenges and the open problems and will not weigh in on that dilemma, except to mention here that the most responsible way to use any statistical method, especially when it is

---

E. L. Ogburn (✉)

Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

e-mail: [eogburn@jhsph.edu](mailto:eogburn@jhsph.edu)

well-known that the assumptions on which it rests do not hold, is with a healthy dose of skepticism, with honest acknowledgment and deep understanding of the limitations, and with copious caveats about how to interpret the results.

A number of high profile papers have used standard methods like generalized linear models (GLMs) and generalized estimating equations (GEEs) to attempt to infer causal relationships from network data (e.g., [1, 8–11, 24, 40]). There has been backlash from the statistical community [12, 28, 45] because these statistical models are not equipped to deal with network dependence and are rarely appropriate for estimating effects using network data. In some settings it may be possible to use them to test for the presence of network dependence, but it is unclear whether these tests have power to detect contagion and therefore whether rejecting the null hypothesis can safely be interpreted as evidence for the alternative hypothesis [44, 53]. In general, methods that assume independence when in fact network dependence is present result in p-values that are artificially small, confidence intervals that are artificially narrow, and inference that is anticonservative.

Spatial autoregressive (SAR) models have been applied to the study of contagion in network settings (e.g., [15, 25, 26, 35]). The shortcoming of these models stems from the fact that, because the endogenous and exogenous variables are measured at the same time, they parameterize an equilibrium state rather than causal relationships. Causal relationships require the exposure to temporally precede the outcome. Few data generating processes give rise to true equilibrium states [5, 23, 50]; therefore, SAR models may often be misspecified or uninformative about causal relationships.

A hallmark of most of the work to date on outcomes sampled from a network is that it uses models, like GEE, GLM, and SAR models, that were developed for very different settings. Very recently, researchers have begun to develop methods designed specifically for the network setting. Work by van der Laan [51] and Ogburn et al. [34] harnesses independence assumptions that require observing the evolution of the network and outcomes over time. In many settings, however, we will only get a snap shot of the network, or we may observe it at multiple time points but not enough to capture the full evolution of the network. Many methods for interference, which is when one subject’s exposure may affect another subject’s outcome, are highly relevant to the analysis of network data. However, the inferential methods developed in this context generally require observing multiple independent groups of units, which corresponds to observing multiple independent networks, or else they require that the exposure be randomized.

In the rest of the chapter, we will go through the specific challenges to learning about contagion from observational social network data one by one, followed by a quick discussion of successful methods for overcoming these challenges in some settings. First, in Sect. 2 we describe a motivating example that will anchor the discussion throughout. In Sect. 3 we describe causal effects that are of interest in social network settings. In Sects. 4 and 5 we discuss two overwhelming challenges for estimating causal effects in social network settings: confounding and statistical dependence. In Sect. 6 we briefly describe some existing and future directions for solutions to these challenges.

## 2 Motivating Example

Suppose that students attending the residential Faber College are measured and weighed at the start and close of each school year, and a complete social network census is taken, cataloguing all social ties among members of the student body. In addition, researchers have access to basic demographic covariates measured on each student. Researchers are interested in testing whether there is a contagion effect for body mass index (BMI): if one individual—the ego—gains (or loses) weight, does that make his or her social contacts—the alters—more likely to do the same? They are also interested in estimating the contagion effect if one exists: if an ego gains (or loses) weight, what is the expected increase (or decrease) in the alters’ body mass indices?

There are many different procedures one could use to test for or estimate a contagion effect, using different models, different assumptions, different sets of covariates, different ways of calculating intervals or uncertainty, and the list goes on. In order for a procedure to be useful, it has to satisfy two requirements. First, it has to isolate the causal effect of the ego’s change in BMI on the alters’ changes in BMI from potential other sources of similarity between the ego’s and the alters’ outcomes. This has to do with confounding, which is the subject of Sect. 4.

The second requirement for a useful analysis is that it must be generalizable to populations beyond the precise student body used in the analysis. We would like to be able to extrapolate what we learn about contagion from the Faber student body to contagion of BMI in similar college populations across different colleges or even across different years at Faber College. Assume that the student body we observe at Faber College is representative of these other student populations, that is, that the true underlying contagion effect for the observed sample of Faber students is the same as the true underlying contagion effect in the other college populations to which we want to extrapolate. This ensures that whatever quantities we are able to estimate using Faber College data will be unbiased. Then one way to determine what we can learn by extrapolating from Faber students to the other similar groups of students is to calculate a confidence interval for the true contagion effect, based on a model of asymptotic growth of the sample. For example, if the sample is large enough that a central limit theorem approximately holds for the contagion effect estimate, then a Gaussian confidence interval around the sample mean is approximately valid. Under the assumption of the same true underlying contagion effect, our confidence that this interval covers the true contagion effect for Faber College students is the same as our confidence that it covers the true contagion effect for students at a different college or in a different year. As in many settings for statistical inference, asymptotics are appropriate not because we care about an infinite population but because they shed light on finite samples. This requires valid statistical inference, and specifically appropriate methods for calculating the variance of an estimator, which is the subject of Sect. 5.

### 3 Defining Causal Effects

Questions about the influence one subject has on the outcome of another subject are inherently questions about causal effects: contagion is a causal effect on an ego's outcome at time  $t$  of his alter's outcome at time  $s$  for some  $s < t$ . Causal effects are defined in terms of potential or counterfactual outcomes (see, e.g., [19, 42]). In general, a unit-level potential outcome,  $Y_i(z)$ , is defined as the outcome that we would have observed for subject  $i$  if we could have intervened to set that subject's treatment or exposure  $Z_i$  to value  $z$ . A contagion effect of interest for dyadic data might be a contrast of counterfactuals of the form  $Y_{ego}^t(y_{alter}^{t-1})$ , for example  $E \left[ Y_{ego}^t(y) - Y_{ego}^t(y-1) \right]$  would be the expected difference in the ego's counterfactual outcome at time  $t$  had the alter's outcome at time  $t-1$  been set to  $y$  compared to  $y-1$ . In data comprised of independent dyads this contagion effect is well-defined, but social networks represent a paradigmatic opportunity for *interference*, whereby one subject's exposure may affect not only his own outcome but also the outcomes of his social contacts and possibly other subjects. This is a violation of the stable unit treatment value assumption (SUTVA) usually made in causal inference settings, which entails that each subject's potential outcome is a function of his or her own treatment but no other treatments. Under interference, the traditional unit-level potential outcomes are not well-defined. Instead,  $Y_i(\mathbf{z})$  is the outcome that we would have observed if we could have set the vector of exposures for the entire population,  $\mathbf{Z}$ , to  $\mathbf{z} = (z_1, \dots, z_n)$  where for each  $i$ ,  $z_i$  is in the support of  $Z$ . The causal inference literature distinguishes between interference, which is present when one subject's treatment or exposure may affect others' outcomes, and contagion, which is present when one subject's outcome may influence or transmit to other subjects (e.g., [31]), but in fact they are usually intertwined. Consider three Faber students: Alex, Andy, and Ari, all friends with each other. Alex's outcome at time  $t$  depends on both Andy's and Ari's outcomes at time  $t-1$ , Andy's outcome at time  $t$  depends on Alex's and Ari's at time  $t-1$ , and Ari's outcome at time  $t$  depends on Alex's and Andy's at time  $t-1$ . This results in a situation that is hardly distinguishable from the hallmarks of interference:  $Y_{Alex}^t(y_{Andy}^{t-1}, y_{Ari}^{t-1})$ ,  $Y_{Andy}^t(y_{Alex}^{t-1}, y_{Ari}^{t-1})$ , and  $Y_{Ari}^t(y_{Alex}^{t-1}, y_{Andy}^{t-1})$  are potential outcomes that depend on multiple "treatments" and those treatments are overlapping across subjects. Furthermore, just as in settings with interference, a counterfactual outcome for node  $i$  that omits some of the treatments to which node  $i$  is exposed (i.e., the outcomes at time  $t-1$  for some of  $i$ 's alters) is not well-defined. This has been overlooked in most of the literature on contagion in observational social network data, which generally focuses on alter-ego pairs, thereby inherently considering ill-defined counterfactuals like  $Y_{Alex}^t(y_{Andy}^{t-1})$ .

This points to an under-appreciated challenge for the study of contagion in a social network: simply defining the causal effect of interest. If researchers sample non-overlapping alter-ego dyads from the network, then  $Y_{ego}^t(y_{alter}^{t-1})$  may be well-defined, but if they wish to use all of the available data, comprised of overlapping dyads, causal effects must be defined in terms of all of the alters for a particular

ego. In the latter case, we could define a contagion effect that compares the mean counterfactual outcome for an ego had the mean outcome among the alters been set to one value as opposed to a different value. For simplicity, in the remaining sections we will talk about alter-ego pairs rather than clusters of an ego with all of its alters. This is in keeping with the existing applied literature, but it is important to note that close attention should be paid in future work to the definition of causal contagion effects for non-dyadic data. Numerous papers and researchers have addressed the definition of counterfactuals and causal effects in settings with interference (e.g., [4, 17, 18, 20, 21, 31, 39, 41, 48, 49]); similar attention should be paid to contagion effects.

It is also worth noting here that measuring alters' and egos' outcomes at different times is crucial. When all outcomes are measured at the same time, it is impossible to determine the direction of causality. Treatments or exposures must temporally precede outcomes in order for causal effects to be well-defined.

## 4 Confounding

Confounding, is, loosely, the presence of a non-causal association that may be misinterpreted as a causal effect of one variable on another. Most commonly, confounding is due to the presence of a confounder that has a causal effect on both the hypothesized cause and the hypothesized effect. Such a confounder generates an association between the hypothesized cause and effect which, without careful analysis, could be taken as evidence of a causal effect. There are two types of confounding that are nearly ubiquitous and especially intransigent in the context of contagion effects in social networks: homophily is the tendency of people who are similar to begin with to share network ties, and environmental confounding is the tendency of people who share network ties to also share environmental exposures that could jointly affect their outcomes. We elucidate these two types of confounding below.

### 4.1 Homophily

Consider the Faber College student body. Suppose that two students, Pat and Lee, meet in September and bond over the fact that they both used to be competitive runners but recently developed injuries that prevent them from running and from participating in other active hobbies they used to enjoy. Soon Pat and Lee are close friends. Over the course of a few months, the sedentary lifestyle catches up with Pat, who gains a considerable amount of weight. It takes longer for Lee, but by the close of the school year Lee has also gained a lot of weight. If you did not have access to the back story and only observed that Pat gained weight and then Pat's close friend Lee did too, this looks like potential evidence of a causal effect of

Pat's change in BMI on Lee's change in BMI. In fact, this is a case of homophily: unobserved covariates related to the propensity to gain weight (in this case, recent injury) caused Pat and Lee to become friends and also caused them to both undergo changes in BMI.

Some carefully considered studies attempt to control for all sources of homophily (see [45] for details and references), but this is generally not possible unless researchers have a high degree of control over data collection and can collect extremely rich (and therefore expensive!) data on the covariates that affect ties. Any traits that are related to the formation, duration, or strength of ties and to the outcome of interest must be measured. For some outcomes, such as infectious diseases, it may be possible to enumerate and observe all such traits, but for other outcomes, such as BMI, endless permutations of the Pat-and-Lee story are possible (e.g., friendship based on shared body norms, shared love of sugary snacks, shared appreciation for a particular celebrity whose BMI changes could affect both Pat and Lee's, etc.), making it nearly impossible to control for all potentially confounding traits. In addition to the challenge of enumerating the potentially confounding traits, there are huge costs to collecting such rich data, and available social network data are highly unlikely to include adequate covariates.

For these reasons, researchers have developed clever tricks to try to control for homophily using only data the network and the outcome of interest. One such trick is to include both the alter and the ego's outcomes at a time  $t - 2$  as covariates in a regression of the ego's outcome at time  $t$  on the alter's outcome at time  $t - 1$ . The argument used to justify this method is that any traits related to tie formation and to the outcome are fully captured by the similarity in the alter and ego's outcomes at time  $t - 2$ ; any association between the alter's outcome at time  $t - 1$  and the ego's at time  $t$  after controlling for this baseline similarity must be due to contagion. But the story of Pat and Lee demonstrates one flaw in this argument: baseline traits can affect outcome trajectories over time and so conditioning on the outcome at a single time point does not render all future outcome measures independent of the baseline covariates. Another flaw in the argument is that homophily operates not only through the propensity to form ties, but also through the propensity to maintain ties and through the strength of the ties; neither strength nor duration can be captured by past outcomes [30]. Furthermore, Shalizi and Thomas [45] demonstrated that, even if a baseline trait only affects friendship formation (not strength or duration), merely conditioning on the presence of a tie, which is inherent in all analyses focused on alter-ego pairs, creates a spurious association between the alter's outcome at time  $t - 1$  and the ego's outcome at time  $t$ . This is because the presence of a tie is a *collider*: a common effect of two variables, conditioning on which creates a spurious association between the two causes. (For an accessible review of colliders, see [13].)

Another clever trick is to compare the strength of the association between an alter's and an ego's outcomes across different types of ties: undirected, or mutual; directed, with the ego naming the alter as a friend but not vice versa; and directed, with the alter naming the ego as a friend but not vice versa. Suppose Pat claims Lee as a friend but Lee does not claim Pat as a friend. Any similarity in baseline traits that Pat and Lee share is a symmetric relationship, the argument goes, and therefore

if the regression of Pat’s BMI at time  $t$  on Lee’s BMI at time  $t - 1$  results in a larger coefficient than does the regression of Lee’s BMI at time  $t$  on Pat’s BMI at time  $t - 1$ , this is evidence of contagion. Unfortunately, this argument is also flawed [28, 45]. This is because, somewhat counterintuitively, similarity in baseline traits does not have to be symmetric. Suppose Pat claims Lee as a friend because Lee is the only person Pat knows who is going through a painful separation with running and other active hobbies, while Lee participates in a support group for recently injured former runners and considers only one participant, Lou, who has the exact same injury and prognosis, as a friend. By construction, even though Lee is the node with the most baseline similarity to Pat from among all of Pat’s potential friends, the reverse is not true: Lou, not Pat, is the node with the most similarity to Lee from among all of Lee’s potential friends. Therefore, if Lou’s outcome at time  $t - 1$  has a stronger association with Lee’s outcome at time  $t - 1$  than Pat’s does, this could be evidence of greater similarity on baseline characteristics rather than contagion. Furthermore, it can be shown that a similar story results in reciprocated ties having the strongest association of all [28]. Shalizi and Thomas [45] used a slightly different data-generating process to show that purported evidence for contagion due to asymmetry in the association of an alter’s outcome with an ego’s outcome for different types of ties is consistent with homophily rather than contagion.

## 4.2 Shared Environment

Let’s turn to a different pair of Faber students, Cam and Sam, who both decided to move off campus to a neighborhood across town from the college. Over the course of the school year, both the grocery store and the gym in their neighborhood closed down and were replaced with fast food restaurants. Cam immediately starts taking every meal at the fast food joint and gains weight fairly quickly, while Sam holds out for several months, taking the bus to a distant grocery store, but when time winter weather and final exams pile on Sam, too, falls prey to the fast food marketing. By the end of the year both students have gained weight. This is confounding due to shared environment, another source of confounding that plagues attempts to learn about contagion from observational data. People who share network ties tend to live near each other, work together, pay attention to the same information, or work in the same industry, all of which can generate confounding due to shared environment (which need not be restricted to physical environment). Note that confounding due to shared environment is present whether Cam and Sam are friends because they live in the same neighborhood or they moved to the same neighborhood because they were friends. The distinction between homophily and shared environment is not always clear-cut; if Cam and Sam became friends because they lived in the same neighborhood that would simultaneously be an example of homophily and of shared environment. The same strategies described above for dealing with homophily have been used in an attempt to control for confounding due to shared environment, but similar reasoning contradicts their effectiveness.

Cohen-Cole and Fletcher [12] proposed controlling for confounding by shared environment by including fixed effects for “community” in regressions of an ego’s outcome at time  $t$  on an alter’s outcome at time  $t - 1$ . If all such confounding occurs due to clearly delineated and known communities, like well-defined neighborhoods in the example above, this is potentially a good solution, though in many cases the operative communities, or their membership, will likely be unknown.

## 5 Dependence

Suppose confounding is not an issue, because researchers at Faber were well-funded and prescient enough to collect data on every possible confounder of the contagion effect, and further suppose that the researchers have a model—maybe a regression, maybe a propensity-score based method [3], maybe some other model—that they believe gives an estimate of the causal contagion effect. We now turn to the question of how to perform valid statistical inference using a model fit to data from a social network. The issue of valid statistical inference is entirely separate from the issue of confounding or even contagion; it applies whether we want to estimate a simple mean or a complicated causal effect. The key points made in this section apply to *anything* that we want to estimate using social network data. Most estimators of causal effects, including The coefficient on the alter’s outcome at time  $t - 1$  in a regression of the ego’s outcome at time  $t$ , are closely related to sample means (to be technical, they are M-estimators), so *all* of the points made below apply.

Going back to Faber College, administrators are now interested in the simpler problem of estimating the mean BMI for the student body at the end of the school year. There are  $n$  students, or nodes in the social network comprised of students, and each one furnishes an observed BMI measurement  $Y_i$ . Our goal is to perform valid (frequentist) statistical inference about the true mean  $\mu$  of  $Y$  using a sample mean  $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$  of dependent observations  $\mathbf{Y} = (Y_1, \dots, Y_n)$ , where the dependence among observations is determined or informed by network structure. But for the dependence, this is a familiar problem. In general, when we want to use a sample mean to perform inference about a true mean, we take the sample mean as our point estimate, calculate a standard error for the sample mean, and tack on a confidence interval based on that standard error. The unique challenge for the social network setting is the effect of dependence on the standard error. To keep things as simple as possible, let’s assume that  $Y_1, \dots, Y_n$  are identically, though not independently, distributed, so the mean of  $Y_i$  is  $\mu$  and the variance of  $Y_i$  is  $\sigma^2$ , which we assume is finite, for all  $i$ . (In fact, it is easier to deal with observations that are not identically distributed than it is to deal with observations that are dependent, so relaxing this assumption is not too difficult.)

Recall that the standard error of  $\bar{Y}$  is the square-root of its variance, where

$$\begin{aligned} Var(\bar{Y}) &= \frac{1}{n^2} Var\left(\sum_{i=1}^n Y_i\right) \\ &= \frac{1}{n^2} \left\{ \sum_{i=1}^n \sigma^2 + \sum_{i \neq j} cov(Y_i, Y_j) \right\} \\ &= \frac{\sigma^2}{n} + \frac{1}{n^2} \sum_{i \neq j} cov(Y_i, Y_j). \end{aligned}$$

When  $Y_1, \dots, Y_n$  are independent, the covariance term  $cov(Y_i, Y_j)$  is equal to 0 for all  $i \neq j$  pairs, so the variance of  $\bar{Y}$  is  $\frac{\sigma^2}{n}$ , which should be familiar from any introductory statistics or data analysis class. But when  $Y_1, \dots, Y_n$  are *dependent*, in particular when they are positively correlated (which is the type of dependence that we would expect to see in just about every social network setting), the variance of  $\bar{Y}$  is bigger than  $\frac{\sigma^2}{n}$  because it includes the term  $\frac{1}{n^2} \sum_{i \neq j} cov(Y_i, Y_j)$ . This is an average of the pairwise covariances for all of the  $\binom{n}{2}$  pairs of observations; the more dependence the data exhibit the larger this term will be. Define  $b_n = \frac{1}{n} \sum_{i \neq j} cov(Y_i, Y_j)$ . Then

$$var(\bar{Y}) = \frac{\sigma^2}{n / \left(1 + \frac{b_n}{\sigma^2}\right)}$$

and we can see that the factor by which the variance of  $\bar{Y}$  is bigger than what it would be if  $Y_1, \dots, Y_n$  were independent is  $\left(1 + \frac{b_n}{\sigma^2}\right)$ . We call  $n / \left(1 + \frac{b_n}{\sigma^2}\right)$  the *effective sample size* of our sample of  $n$  dependent observations  $Y_1, \dots, Y_n$ . The effective sample size  $n / \left(1 + \frac{b_n}{\sigma^2}\right)$  is smaller than the true sample size  $n$ ; heuristically, this is because each observation  $Y_i$  contains some new information about the target of inference  $\mu$  and some information that is rendered redundant by dependence. Under independence each observation furnishes 1 “bit” of information about  $\mu$ , whereas under dependence each observation furnishes only  $1 / \left(1 + \frac{b_n}{\sigma^2}\right)$  bit of information about  $\mu$ .

In order to explain the impact of this dependence on statistical inference, we first review the standard inferential procedure for independent data. When  $Y_1, \dots, Y_n$  are independent, a typical procedure would be to calculate an approximate 95% confidence interval for  $\mu$  as  $\bar{Y} \pm 1.96 \times \frac{\hat{\sigma}}{\sqrt{n}}$ , where  $\hat{\sigma}$  is the square root of an estimate of the variance of  $Y$ . The factor 1.96 is the 97.5th quantile of the standard Normal distribution; t-distribution quantiles could be used instead to account for the fact that  $\sigma$  is estimated rather than known. This procedure relies on several preliminaries: (1)  $\bar{Y}$  is unbiased for  $\mu$ , (2)  $\bar{Y}$  is approximately Normally distributed, and (3)  $\frac{\hat{\sigma}}{\sqrt{n}}$  is a

good estimate of the variance of  $\bar{Y}$ . These preliminaries hold, at least approximately, in most settings with independent data and moderate to large  $n$ . Dependence doesn't affect (1), but it does affect (2) and (3).

When  $Y_1, \dots, Y_n$  are independent, the Central Limit Theorem (CLT) tells us that  $\sqrt{n}(\bar{Y} - \mu)$  converges in distribution to a Normal distribution as  $n \rightarrow \infty$ . The factor  $\sqrt{n}$  is called the *rate of convergence* and it is needed to make sure that the variance of  $\sqrt{n}(\bar{Y} - \mu)$  is not 0, in which case  $\sqrt{n}(\bar{Y} - \mu)$  would converge to a constant rather than a distribution, and is not infinite, in which case  $\sqrt{n}(\bar{Y} - \mu)$  would not converge at all. The variance of  $\bar{Y}$  (equivalently, the variance of  $\bar{Y} - \mu$ ) is  $\sigma^2/n$ , so the variance of  $\sqrt{n}(\bar{Y} - \mu)$  is  $n \times (\sigma^2/n) = \sigma^2$ , which is a positive, finite constant. When  $Y_1, \dots, Y_n$  are dependent, the rate of convergence may be different (slower) than  $\sqrt{n}$ . (In fact, if the dependence is strong and widespread enough, the CLT may not hold at all; determining what types of social network dependence are consistent with the CLT is an important area for future study.) This is because the rate of convergence is determined by the effective sample size instead of by  $n$ : the variance of  $\bar{Y}$  is  $\sigma^2 / \left\{ n / \left( 1 + \frac{b_n}{\sigma^2} \right) \right\}$ , so (as long as a CLT holds),  $\sqrt{n / \left( 1 + \frac{b_n}{\sigma^2} \right)} (\bar{Y} - \mu)$  will converge to a Normal distribution as  $n \rightarrow \infty$  and the rate of convergence is given by  $\sqrt{n / \left( 1 + \frac{b_n}{\sigma^2} \right)}$  rather than  $\sqrt{n}$ . Sometimes, in particular when  $b_n$  is fixed as  $n \rightarrow \infty$ , this distinction will be meaningless. But sometimes, when  $b_n$  grows with  $n$ , it is a meaningfully slower rate of convergence. (Note that  $b_n/n$  must converge to 0 as  $n \rightarrow \infty$  in order for a CLT to hold, so  $b_n$  must grow slower than  $n$ .) This matters because it informs when the approximate Normality of the CLT kicks in, i.e. at what sample size it is safe to assume that  $\bar{Y}$  is approximately Normally distributed. Many different rules of thumb exist for determining when approximate Normality holds; one popular rule of thumb is that  $n = 30$  suffices. With dependent data, this number is larger, and sometimes considerably so. The effective sample size, rather than  $n$ , should be used to assess whether the sample size is large enough to approximate the distribution of  $\bar{Y}$  with a Normal distribution. When researchers ignore dependence and rely on the Normal approximation in samples that have large enough  $n$  but not large enough effective sample size, there is no reason to think that their 95% confidence intervals will have good coverage properties.

Ignoring dependence is most dangerous when estimating the standard error of  $\bar{Y}$ . Any estimate of  $\text{var}(\bar{Y})$  that is based only on the marginal variances  $\sigma^2$  of  $Y_i$  and ignore the covariances  $\text{cov}(Y_i, Y_j)$  will underestimate the standard error of  $\bar{Y}$ , often severely. Inference that is based on an underestimated standard error is *anticonservative*: confidence intervals are narrower than they should be and p-values are lower than they should be, leading researchers to draw conclusions that are not in fact substantiated by the data. Even if each observation is dependent only on a fixed and finite number of other observations, so that dependence is asymptotically negligible and does not affect the rate of convergence of the CLT, in

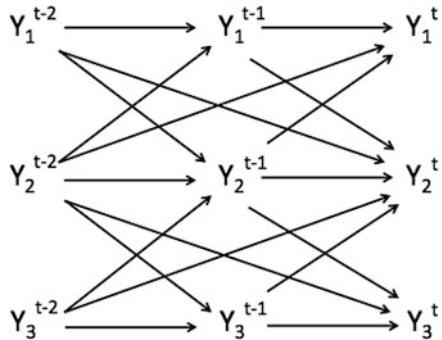
finite samples ignoring the covariance terms in  $\text{var}(\bar{Y})$  could still have substantial implications on inference. This is particularly a problem because no good solutions exist. Statisticians are good at dealing with dependence that arises due to space or time, or even other more complicated processes that can be expressed using Euclidean geometry. But dependence that is informed by a network is very different from these well-understood types of dependence, and, unfortunately, statisticians are only just beginning to develop methods for taking it into account. Most published research about social contagion uses regression models or generalized estimating equations (GEEs) to estimate contagion effects; though some of these models account for the dependence due to observing the same nodes over multiple time points, none of them account for dependence among nodes.

## 5.1 Sources of Network Dependence

In the literature on spatial and temporal dependence, dependence is often implicitly assumed to be the result of latent traits that are more similar for observations that are close in Euclidean distance than for distant observations. This type of dependence is likely to be present in many network contexts as well. In networks, edges present opportunities to transmit traits or information, and contagion or influence is an important additional source of dependence that depends on the underlying network structure.

Latent trait dependence will be present in data sampled from a network whenever observations from nodes that are close to one another are more likely to share unmeasured traits than are observations from distant nodes. Homophily is a paradigmatic example of latent trait dependence. If the outcome under study in a social network has a genetic component, then we would expect latent variable dependence due to the fact that family members, who share latent genetic traits, are more likely to be close in social distance than people who are unrelated. If the outcome were affected by geography or physical environment, latent variable dependence could arise because people who live close to one another are more likely to be friends than those who are geographically distant. Of course, whether these traits are latent or observed they can create dependence, but if they are observed then conditioning on them renders observations independent, so only when they are latent do they result in dependence that requires new tools for statistical inference. Just like in the spatial dependence context, there is often little reason to think that we could identify, let alone measure, all of these sources of dependence. The notions of latent sources of homophily or latent correlates of shared environment are familiar from the discussion of confounding, above, but there is an important distinction to be made between latent sources of confounding and latent sources of dependence: in order to be a source of unmeasured confounding, a latent trait must affect both the exposure (e.g., the alter's outcome at time  $t - 1$ ) and the outcome (ego's outcome at time  $t$ ) of interest. In order to be a source of dependence, a latent trait must

**Fig. 1** Dependence by contagion



affect two or more outcomes of interest. Latent trait dependence is the most general form of dependence, in that it provides no structure that can be harnessed to propel inference. In order to make any progress towards valid inference in the presence of latent trait dependence, some structure must be assumed, namely that the range of influence of the latent traits is primarily local in the network and that any long-range effects are negligible.

Direct transmission of an alter’s treatment or outcome to an ego also results in statistical dependence. Contagion or influence arises when the outcome under study is transmitted from node to node along edges in the network. The diagram in Fig. 1 depicts contagion in a network with three nodes in which node 2 is connected to nodes 1 and 3 but there is no edge between 1 and 3.  $Y_i^t$  represents the outcome for node  $i$  at time  $t$ , and the unit of time is small enough that at most one transmission event can occur between consecutive time points. Dependence due to direct transmission has known, though possibly unobserved, structures that can sometimes be harnessed to facilitate inference; we touch on this briefly in Sect. 6. Crucially, whenever contagion is present so is dependence due to direct transmission, and therefore statistical analysis must take dependence into account in order to result in valid inference.

## 6 Solutions

Researchers have known for decades that learning about contagion from observational data is fraught with difficulty, perhaps most famously expressed by Manski [29]. Recent years have seen incremental methodological progress, but huge hurdles remain. Most of the constructive ideas in [45] involve bounding contagion effects rather than attempting to point identify them; looking for bounds rather than point estimates is a general approach that could prove fruitful in the future. Indeed, Ver Steeg and Galstyan [54] built upon the ideas in [45] and were able to derive bounds on the association due to homophily on traits that do not change over time (“static homophily”). Another general approach is to make use of sensitivity

analyses whenever an estimation procedure relies on assumptions that may not be realistic (e.g., [52]). Some of the problems discussed above have solutions in some settings; below we discuss solutions that exploit features of specific settings rather than providing general approaches to the problem of estimating contagion effects. (Some of the material below was first published in [33].)

## 6.1 Randomization

An in-depth discussion of randomized experiments is given in the chapter “Randomized Experiments to Detect and Estimate Social Influence” by Sean J. Taylor and Dean Eckles in this volume; here we give a very brief overview. If it is possible to randomize some members of a social network to receive an intervention, and if it is known that an alter’s receiving an intervention can only affect the ego’s outcome through contagion (as opposed to directly; see [31] for discussion), then problems of confounding and dependence can be entirely obviated. *Randomization-based inference*, pioneered by Fisher [14] and applied to network-like settings by Rosenbaum [39] and Bowers et al. [6], is founded on the very intuitive notion that, under the null hypothesis of no effect of treatment on any subject (sometimes called the *sharp null hypothesis* to distinguish it from other null hypotheses that may be of interest), the treated and control groups are random samples from the same underlying distribution. Randomization-based inference treats outcomes as fixed and treatment assignments as random variables: quantities that depend on the vector of treatment assignments are the only random variables in this paradigm. Therefore, dependence among outcomes is a non-issue. Typically this type of inference is reserved for hypothesis testing, though researchers have extended it to estimation. We leave the details, including several subtleties and challenges that are specific to the social network context, to a later chapter (see also [33] for a review).

Randomizing the formation of network ties themselves obviates confounding due to the effects of homophily on tie formation. A number of studies have taken advantage of naturally occurring randomizations of this kind, such as the assignment of students to dorm rooms [43] or of children to classrooms [22]. However, this does not suffice to control for the effects of homophily on tie strength or duration, or to control for confounding due to shared environment. If students at Faber College are randomly assigned to dorm rooms at the beginning of their freshman year, then each student’s exposure to his or her roommate’s BMI is unconfounded by design. However, if, for example, randomly assigned roommates who happen to both be athletes form stronger bonds than other types of roommates, then the contagious effect of BMI could still be confounded by homophily: homophily acting on tie strength rather than tie presence.

## 6.2 Parametric Models

If researchers are willing to commit to certain types of parametric models, it may be possible to isolate contagion from confounding [46]. It is a reliance on strong parametric models, for example, that underpins mathematical modeling or agent based modeling approaches to contagion [7, 38, 47].

This might seem benign—after all, most statistical analyses rely on parametric models of one kind or another—but there is a fundamental difference between, for example, using a linear regression when the true underlying relationships is not linear, and relying on parametric models to identify a causal effect that is otherwise hopelessly confounded. In the first case, a misspecified model may bias the estimate we are interested in, often in ways that are well-understood, and often in proportion to the fit of the model to the data (i.e., the worse the misspecification, the greater the bias). In the latter case, at least in the absence of a model-specific proof otherwise, any hint of misspecification undermines the causal interpretation we would like to be able to justify and what looks like evidence of a causal effect could just be evidence of confounding. George Box’s oft-cited aphorism, “all models are wrong but some are useful,” justifies the use of misspecified parametric models in many settings, but when the parametric form of the model is the only bulwark against confounding, the model must (in the absence of a proof to the contrary) in fact be correct in order to be useful.

## 6.3 Instrumental Variable Methods

O’Malley et al. [36] proposed an instrumental variable (IV) solution to the problem of disentangling contagion from homophily. An instrument is a random variable,  $V$ , that affects exposure but has no effect on the outcome conditional on exposure. When the exposure–outcome relation suffers from unmeasured confounding but an instrument can be found that is not confounded with the outcome, IV methods can be used to recover valid estimates of the causal effect of the exposure on the outcome. In this case there is unmeasured confounding of the relation between an alter’s outcome at time  $t - 1$  and an ego’s outcome at time  $t$  whenever there is homophily on unmeasured traits. Angrist and Pischke [2], Greenland [16], and Pearl [37] provide accessible reviews of IV methods.

O’Malley et al. [36] proposed using a gene that is known to be associated with the outcome of interest as an instrument. In their paper they focus on perhaps the most highly publicized claim of peer effects, namely that there are significant peer effects of body mass index (BMI) and obesity [19]. If there is a gene that affects BMI but that does not affect other homophilous traits, then that gene is a valid instrument for the effect of an alter’s BMI on his ego’s BMI. The gene affects the ego’s BMI only through the alter’s manifest BMI (and it is independent of the ego’s BMI conditional on the alter’s BMI), and there is unlikely to be any confounding, measured or unmeasured, of the relation between an alter’s gene and the ego’s BMI.

There are two important challenges to this approach. First, the power to detect peer effects is dependent in part upon the strength of the instrument–exposure relation which, for genetic instruments, is often weak. Indeed, O’Malley et al. [36] reported low power for their data analyses. Second, in order to assess contagion at more than a single time point (i.e., the average effect of the alter’s outcomes on the ego’s outcomes up to that time point), multiple instruments are required. O’Malley et al. [36] suggests using a single gene interacted with age to capture time-varying gene expression, but this could further attenuate the instrument–exposure relation and this method is not valid unless the effect of the gene on the outcome really does vary with time; if the gene-by-age interactions are highly collinear, then they will fail to act as differentiated instruments for different time points.

## 6.4 Data from Multiple Independent Networks

When multiple independent networks are observed, the problems of confounding due to shared environment and of dependence may be considerably easier to deal with. A large literature on interference in causal inference is dedicated to inference in the setting where independent groups of individuals interact and affect one another within, but not between, groups; this is analogous to multiple independent social networks (see, e.g., [20, 21, 27, 48, 49]). If environmental factors can be shared within but not across networks, it may be possible to control for confounding by shared environment via a fixed effect for each network, as in [12]. For our running example, this would entail that the administrators of Faber College join forces with  $n - 1$  other (randomly selected and similar) colleges to generate a sample of  $n$  iid social networks. This magnitude of data is often unavailable in practice.

## 6.5 Highly Structured Dependence

If dependence and the structure is lost, researchers have reason to believe that there is no unmeasured homophily or features of shared environments that contribute to confounding or to dependence, i.e. if direct transmission (see Sect. 5) is the only mechanism giving rise to either dependence or to associations among the outcomes of interest, then there are a few recent methodological advances that can be used to estimate contagion effects [32, 34, 51]. Dependence due to direct transmission has known, though possibly unobserved, structures that can sometimes be harnessed to facilitate inference. Time and distance act as information barriers for dependence due to contagion, giving rise to many conditional independencies that can sometimes be used to make network dependence tractable. Two examples of the many conditional independencies that hold in Fig. 1 are  $[Y_1^t \perp Y_2^t \mid Y_1^{t-2}, Y_2^{t-2}, Y_1^{t-1}, Y_2^{t-1}]$  and  $[Y_1^{t-1} \perp Y_3^t \mid Y_2^{t-2}]$ . The first

conditional independence statement illustrates the principle that outcomes measured at a particular time point are mutually independent conditional on all past outcomes. The second conditional independence statement illustrates the fact that outcomes sampled from two nonadjacent nodes are independent if the amount of time that passed between the two measurements was not sufficiently long for information to travel along the shortest path from one node to the other, conditional any information that could have simultaneously influenced the sampled nodes (in this case  $Y_2^{t-2}$ ). Observing outcomes in a network on a fine enough time scale to observe all transmissions requires a richness of data that will not usually be available, and if the network under a contagious process is observed at a single time point, dependence due to contagion is indistinguishable from latent variable

Ogburn et al. [34] allows for a very limited kind of latent variable dependence in addition to dependence due to direct transmission: if any latent variable dependence only affects friends and friends-of-friends, that is pairs of network nodes separated by no more than two ties, then the methods presented in [34] are valid. This is a first step towards making the methods described above more appropriate for real data, but it is generally unrealistic for latent variable dependence to vanish at a distance of two. Future work is needed to accommodate more realistic kinds of latent variable dependence, which would decay as distance grows rather than suddenly dropping off.

## 7 Conclusion

Interest in and availability of social network data are both on the rise, and statisticians and other methodologists have a lot of work to do to catch up. It is crucial for applied researchers to acknowledge the limitations of many current methods, most notably their inability to control for confounding due to homophily or to account for network dependence; for the curators of social network data to recognize the importance of temporal data and of studies carefully designed to control for homophily; and for statisticians to develop new methods better suited to this new kind of dependent data.

**Acknowledgement** This work was funded by the Office of Naval Research grant N00014-15-1-2343.

## References

- Ali MM, Dwyer DS (2009) Estimating peer effects in adolescent smoking behavior: a longitudinal analysis. *J Adolesc Health* 45(4):402–408
- Angrist JD, Pischke JS (2008) *Mostly harmless econometrics: an empiricist's companion*. Princeton University Press, Princeton
- Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci* 106(51):21544–21549

4. Aronow PM, Samii C (2012) Estimating average causal effects under general interference. Technical report
5. Besag J (1974) On spatial-temporal models and Markov fields. In: Transactions of the seventh Prague conference on information theory, statistical decision functions, and random processes. Springer, Dordrecht, pp 47–55
6. Bowers J, Fredrickson MM, Panagopoulos C (2013) Reasoning about interference between units: a general framework. *Polit Anal* 21(1):97–124
7. Burk WJ, Steglich CE, Snijders TA (2007) Beyond dyadic interdependence: actor-oriented models for co-evolving social networks and individual behaviors. *Int J Behav Dev* 31(4):397–404
8. Cacioppo JT, Fowler JH, Christakis NA (2009) Alone in the crowd: the structure and spread of loneliness in a large social network. *J Pers Soc Psychol* 97(6):977
9. Christakis N, Fowler J (2007) The spread of obesity in a large social network over 32 years. *N Engl J Med* 357(4):370–379
10. Christakis N, Fowler J (2008) The collective dynamics of smoking in a large social network. *N Engl J Med* 358(21):2249–2258
11. Christakis N, Fowler J (2010) Social network sensors for early detection of contagious outbreaks. *PLoS One* 5(9):e12948
12. Cohen-Cole E, Fletcher JM (2008) Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. *J Health Econ* 27(5):1382–1387
13. Elwert F, Winship C (2014) Endogenous selection bias: the problem of conditioning on a collider variable. *Ann Rev Sociol* 40:31–53
14. Fisher RA (1922) On the mathematical foundations of theoretical statistics. *Philos Trans R Soc Lond Ser A* 222:309–368. Containing papers of a mathematical or physical character
15. Goetzke F (2008) Network effects in public transit use: evidence from a spatially autoregressive mode choice model for New York. *Urban Stud* 45(2):407–417
16. Greenland S (2000) An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 29(4):722–729
17. Halloran M, Hudgens M (2011) Causal inference for vaccine effects on infectiousness. The University of North Carolina at Chapel Hill Department of Biostatistics. Technical report series, p 20
18. Halloran ME, Struchiner CJ (1995) Causal inference in infectious diseases. *Epidemiology* 6(2):142–151
19. Hernán MA (2004) A definition of causal effect for epidemiological research. *J Epidemiol Community Health* 58(4):265–271
20. Hong G, Raudenbush S (2006) Evaluating kindergarten retention policy. *J Am Stat Assoc* 101(475):901–910
21. Hudgens M, Halloran M (2008) Toward causal inference with interference. *J Am Stat Assoc* 103(482):832–842
22. Kang C (2007) Classroom peer effects and academic achievement: quasi-randomization evidence from South Korea. *J Urban Econ* 61(3):458–495
23. Lauritzen SL, Richardson TS (2002) Chain graph models and their causal interpretations. *J R Stat Soc Ser B Stat Methodol* 64(3):321–348
24. Lazer D, Rubineau B, Chetkovich C, Katz N, Neblo M (2010) The coevolution of networks and political attitudes. *Polit Commun* 27(3):248–274
25. Lee LF (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72(6):1899–1925
26. Lin X (2005) Peer effects and student academic achievement: an application of spatial autoregressive model with group unobservables. Unpublished manuscript, Ohio State University
27. Liu L, Hudgens MG (2014) Large sample randomization inference of causal effects in the presence of interference. *J Am Stat Assoc* 109(505):288–301
28. Lyons R (2011) The spread of evidence-poor medicine via flawed social-network analysis. *Stat Polit Policy* 2(1):Article 2

29. Manski CF (1993) Identification of endogenous social effects: the reflection problem. *Rev Econ Stud* 60(3):531–542
30. Noel H, Nyhan B (2011) The unfriending problem: the consequences of homophily in friendship retention for causal estimates of social influence. *Soc Netw* 33(3):211–218
31. Ogburn EL, VanderWeele TJ (2014) Causal diagrams for interference. *Stat Sci* 29(4):559–578
32. Ogburn EL, VanderWeele TJ (2014) Vaccines, contagion, and social networks (preprint). arXiv:14031241
33. Ogburn EL, Volfovsky A (2016) Networks. In: Bühlmann P, Drineas P, Kane M, van der Laan MJ (eds) *Handbook of big data*. Chapman and Hall/CRC, Boca Raton
34. Ogburn EL, Sofrygin O, Diaz I, van der Laan MJ (2017, Preprint) Causal inference for social network data. arXiv:170508527
35. O’Malley JA, Marsden PV (2008) The analysis of social networks. *Health Serv Outcomes Res Methodol* 8(4):222–269
36. O’Malley AJ, Elwert F, Rosenquist JN, Zaslavsky AM, Christakis NA (2014) Estimating peer effects in longitudinal dyadic data using instrumental variables. *Biometrics* 70:506–515
37. Pearl J (2000) *Causality: models, reasoning and inference*. Cambridge University Press, Cambridge
38. Railsback SF, Grimm V (2011) *Agent-based and individual-based modeling: a practical introduction*. Princeton University Press, Princeton
39. Rosenbaum P (2007) Interference between units in randomized experiments. *J Am Stat Assoc* 102(477):191–200
40. Rosenquist JN, Murabito J, Fowler JH, Christakis NA (2010) The spread of alcohol consumption behavior in a large social network. *Ann Intern Med* 152(7):426–433
41. Rubin D (1990) On the application of probability theory to agricultural experiments. Essay on principles. section 9. Comment: Neyman (1923) and causal inference in experiments and observational studies. *Stat Sci* 5(4):472–480
42. Rubin DB (2005) Causal inference using potential outcomes: design, modeling, decisions. *J Am Stat Assoc* 100(469):322–331
43. Sacerdote B (2000) Peer effects with random assignment: Results for Dartmouth roommates. Technical report, National Bureau of Economic Research
44. Shalizi CR (2012) Comment on “why and when ‘flawed’ social network analyses still yield valid tests of no contagion”. *Stat Polit Policy* 3(1). <https://doi.org/10.1515/2151-7509.1053>
45. Shalizi CR, Thomas AC (2011) Homophily and contagion are generically confounded in observational social network studies. *Sociol Methods Res* 40(2):211–239
46. Snijders T, Steglich C, Schweinberger M (2007) Modeling the coevolution of networks and behavior. NA
47. Snijders TA, Van de Bunt GG, Steglich CE (2010) Introduction to stochastic actor-based models for network dynamics. *Soc Netw* 32(1):44–60
48. Sobel M (2006) What do randomized studies of housing mobility demonstrate? *J Am Stat Assoc* 101(476):1398–1407
49. Tchetgen Tchetgen EJ, VanderWeele T (2012) On causal inference in the presence of interference. *Stat Methods Med Res* 21(1):55–75
50. Thomas A (2013) The social contagion hypothesis: comment on ‘social contagion theory: examining dynamic social networks and human behavior’. *Stat Med* 32(4):581–590
51. van der Laan MJ (2012) Causal inference for networks. UC Berkeley Division of Biostatistics, Working Paper Series, Working Paper 300
52. VanderWeele TJ (2011) Sensitivity analysis for contagion effects in social networks. *Sociol Methods Res* 40(2):240–255
53. VanderWeele TJ, Ogburn EL, Tchetgen EJT (2012) Statistics, politics, and policy. *Polit Policy* 3(1):4
54. Ver Steeg G, Galstyan A (2010) Ruling out latent homophily in social networks. In: NIPS workshop on social computing

## **Part II**

# **Models and Theories**

# Slightly Generalized Contagion: Unifying Simple Models of Biological and Social Spreading



Peter Sheridan Dodds

## 1 Introduction

Spreading, construed fully, is everywhere: the entropically aspirant diffusive relaxation of all systems; wave motion, for which ubiquitous is assigned with no overstatement; in the propagation of earthquakes; the expansion of species range, so often involving people; power blackouts, now able to affect large fractions of the world population through system growth; the repeated bane of global pandemics; economic prosperity and misery; and the talk of the famously talked about. And understanding how myriad entities spread between people—from diseases to stories, both true and false—is central to our scientific understanding of large populations.

Used for good, as the trope goes, a deep knowledge of contagion mechanisms—contagion science—is necessary to help in our collective efforts to produce a world where individuals can flourish. Used for bad, a path scientific knowledge always offers, malefactors will be empowered in the persuasion and manipulation of populations or the breaking of financial systems. To prevent negative and catastrophic outcomes, contagion science should be able to provide us with algorithms for system defense.

There remain many open questions on contagion. How many types of spreading and contagion mechanisms are there? How can we identify and categorize real-world contagions? But we have only recently moved from the data-scarce period of studying social phenomena to the start of the data-rich stage, and contagion science is still very much developing

---

P. S. Dodds (✉)

Vermont Complex Systems Center, Computational Story Lab, the Vermont Advanced Computing Core, Department of Mathematics and Statistics, The University of Vermont, Burlington, VT, USA

e-mail: [peter.dodds@uvm.edu](mailto:peter.dodds@uvm.edu)

Our goal in this piece is constrained to revisiting our 2004 revisiting of basic mathematical models of contagion surrounding one question [5, 6]: Can we connect models of disease-like and social contagion?

We call the process we constructed for this objective “generalized contagion.” We will give a straightforward explanation of the model here and discuss its most important features.

An incidental contribution with generalized contagion was to make memory a primary ingredient. For contagion, memory comes in many forms, for example, in the development of protection against an infectious disease through an immune response, or through recalling past exposures to some kind of social influence. The core models of biological and social contagion incorporate only the simplest kind of memory, that of the present state.

In proceeding, we first outline the independent and interdependent interaction models of biological and social contagion. Apart from standing as the footing of our generalized model, we will also preserve certain framings and notations. We then describe our model of generalized contagion and discuss the three universality classes of systems identified in the context of small seeds leading to global spreading.

## 2 Independent Interaction Models of Biological Contagion

In mathematical epidemiology, the standard model [17] was first put forward in the 1920s by Reed and Frost and formalized by Kermack and McKendrick [14–16]. These models came to be generally referred to as SIR models in reference to the three epidemiological states:

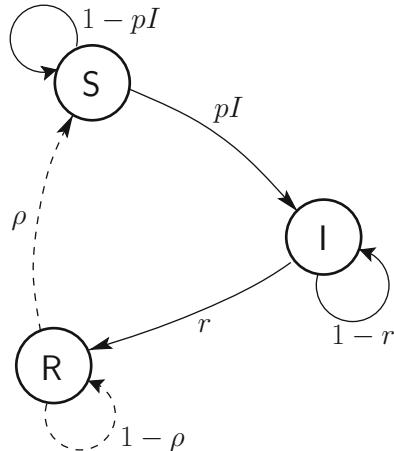
- Susceptible;
- Infective (or Infectious);
- Recovered (or Removed or Refractory).

Individuals cycle through the states **S** to **I** to **R** (and then back to **S** for an SIRS model). The behavior of these initial models was described by differential equations but can be easily realized as a discrete time system, and we will use the latter framework for our generalized model. SIR models are also mass action type models, meaning individuals are represented as normalized fractions of a population which randomly interact with each other.

To connect notation across different models, we will write the fractions in the three states as  $S_t$ ,  $\phi_t$  (normally  $I_t$ ), and  $R_t$ . We must have the constraint  $S_t + \phi_t + R_t = 1$ . There is no memory in these systems other than the current balance of Susceptibles, Infectives, and Recovereds.

Figure 1 shows an example automata for the independent interaction model when time is discrete. From the point of view of an individual agent in a discrete time SIR system, they interact independently, at each time step connecting with

**Fig. 1** Update mechanism for an example discrete version of the basic SIR model. Individuals may be susceptible (state **S**), infective (state **I**), and recovered (state **R**). The three transition probabilities are  $p$  for being infected given contact with infected ( $S \rightarrow I$ ),  $r$  for recovery ( $I \rightarrow R$ ), and  $\rho$  for loss of immunity ( $R \rightarrow S$ ). The model's complication lies in the nonlinear term involved in the transition of susceptibles to infectives



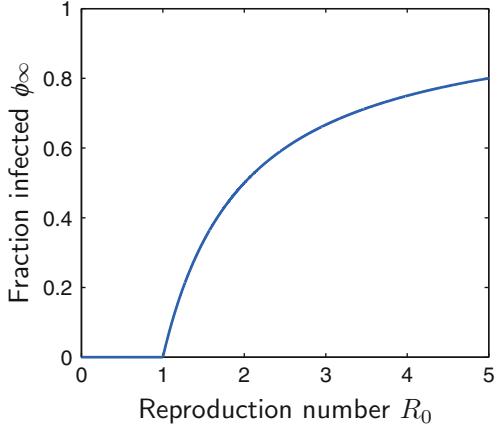
a Susceptible, Infective, or Recovered. The probabilities of each interaction are equal to the normalized fractions  $S_t$ ,  $\phi_t$ , and  $R_t$ . When Susceptibles interact with Infectives (occurring with probability  $\phi_t$ ), they themselves become Infective with probability  $p$ . Regardless of their interactions, Infectives recover with a probability  $r$  and Recovereds become Susceptibles with probability  $\rho$  (for SIR models,  $\rho = 0$ , while for SIRS models,  $\rho > 0$ ).

A traditionally key quantity in mathematical epidemiology is the Reproduction Number  $R_0$  [which is terrible notation given we already have state **R** and  $R_t$ ]. The Reproduction Number is the expected number of infected individuals resulting from the introduction of a single initial infective. The Reproduction Number is easily interpreted and leads to an Epidemic threshold: If  $R_0 > 1$ , an “epidemic” occurs. As with many complex systems, the focus on a single number as a diagnostic is always fraught, and the Reproduction Number ultimately combines too many aspects of the disease itself and population interaction patterns, rendering it a deceptive measure [24]. Nevertheless, for simple models  $R_0$  is important and the notion of an Epidemic Threshold is more generally essential.

For our simple discrete model, we can compute  $R_0$  easily. We introduce one Infective into a randomly mixing population of Susceptibles. At time  $t = 0$ , this single Infective randomly bumps into a Susceptible who is infected with probability  $p$ . The single Infective remains infected with probability  $(1 - r)^t$  at time  $t$ , having attempted to infect  $t$  Susceptibles by this point. The expected number infected by original Infective is therefore:

$$\begin{aligned}
 R_0 &= p + (1 - r)p + (1 - r)^2 p + (1 - r)^3 p + \dots \\
 &= p \frac{1}{1 - (1 - r)} = p/r,
 \end{aligned} \tag{1}$$

**Fig. 2** Stylized example plot of the final fractional size of a spreading event for SIR type models. The reproduction number  $R_0 = p/r$  (Eq. (1)) acts as a phase parameter with a continuous phase transition occurring at  $R_0 = 1$ , the epidemic threshold



and the disease spreads in this system if

$$R_0 = p/r > 1. \quad (2)$$

Figure 2 shows an example of epidemic threshold from our elementary SIR model where the tunable parameter is the Reproduction Number  $R_0 = p/r$ . The final fraction infected exhibits a continuous phase transition (technically a transcritical bifurcation [20]). The epidemic threshold is a powerful story arising from a simple model.

### 3 Interdependent Interaction Models of Social Contagion

In spite of the basic SIR model's failings to represent biological contagion accurately in all cases and particularly at large scales, it has enjoyed a long tenure. There have also been overly courageous attempts to use SIR and its sibling models beyond disease spreading including the adoption of ideas and beliefs [9], the spread of rumors [3, 4], the diffusion of innovations [1], and the spread of fanatical behavior [2].

And while some kinds of social contagion may be disease-like, it is clearly of a different nature for the most part. One of the major departures is due to the fact that people take in information from potentially many sources and weigh their inputs relatively. This observation gives rise to the notion of thresholds, first used in modeling in the early 1970s by Schelling in his efforts to understand segregation [18, 19] (the so-called tipping of neighborhoods, and the origin of “Tipping Point”). Schelling's model played out (literally) on a chessboard and was manifestly spatial.

Later in the same decade and inspired in part by Schelling's work, Granovetter produced a distilled mass action threshold model which would become famous in its own right. While social contagion is arguably more multifaceted than biological contagion, Granovetter's model will serve as our elemental model here.

An individual in Granovetter's model may be framed as having a choice of adopting a behavior or not based on their perception of that behavior's popularity. Each individual  $i$  has a threshold  $d_i^* \in [0, 1]$  drawn from a population-level threshold distribution  $P^{(\text{thr})}$  at  $t = 0$ . We can preserve the SIR model framing of two states: **S** and **I**, with infectives being those who have adopted the behavior. We will continue to use  $\phi_t$  as the fraction individuals who are infected.

At each time step, if individual  $i$  observes the fraction **I** of the total population expressing the behavior as meeting or exceeding their threshold  $d_i^*$ , then they adopt the behavior. The system iterates forward, potentially reaching an asymptotic state.

Without any spatial structure, all of the interesting dynamics of Granovetter's model is generated purely by the threshold distribution  $P^{(\text{thr})}$ . We are in fact in the realm of maps of the interval, the territory where so many extraordinary findings have been made for dynamical systems and chaos [20]. The time evolution of Granovetter's model can be written down as:

$$\phi_{t+1} = \int_0^{\phi_t} P_u^{(\text{thr})} du. \quad (3)$$

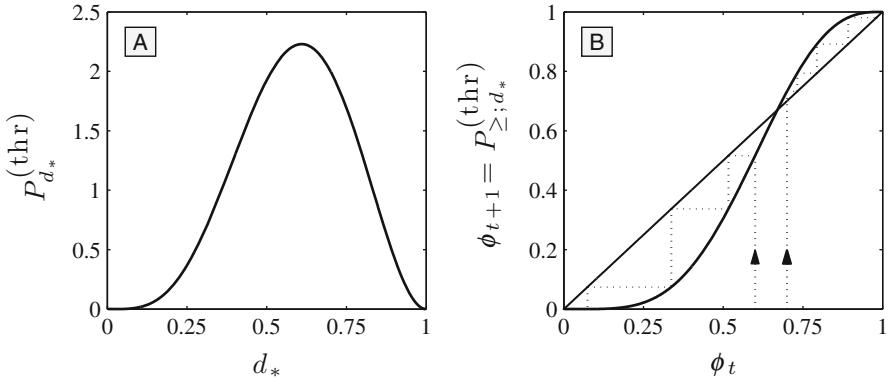
The fraction infected in the next time step  $\phi_{t+1}$  will be exactly the fraction whose threshold is exceeded by the current fraction infected  $\phi_t$ .

Writing  $P_{\geq}^{(\text{thr})}$  as the cumulative function of  $P^{(\text{thr})}$ , we have, compactly, that

$$\phi_{t+1} = P_{\geq; \phi_t}^{(\text{thr})}. \quad (4)$$

The dynamics of Granovetter's model are thus inscribed in  $P_{\geq}^{(\text{thr})}$  particularly in  $P_{\geq}^{(\text{thr})}$ 's fixed points and relative slopes. As an example, Fig. 3 shows how Granovetter's model may represent a critical mass phenomenon. Figure 3a gives the distribution  $P_{\geq}^{(\text{thr})}$  of individual thresholds showing a middle tendency. There are very few extremely gullible people ( $d^* \simeq 0$ ) and very difficult to influence ones ( $d^* \simeq 1$ ). In Fig. 3b, the cumulative function with some example cobweb iterates [20] show that if the initial fraction infected is above the internal fixed point, the fraction adopting the behavior rapidly approaches 1, while any initial fraction starting below the fixed point will see the behavior die out. The initial adoption level  $\phi_0$  must be generated by an exogenous mechanism (e.g., education, marketing) and then the purely imitative dynamics of the system take off.

Granovetter's model and its variants are rich in dynamics and avenues of analysis [10–12, 21, 22]. In reintroducing spatial interactions, Watts transported Granovetter's model to random networks [21] showing that limiting an individual's awareness to a small set of neighbors on a network could lead to large-scale,



**Fig. 3** Example of Granovetter’s model reflecting a Critical Mass system. **a** Distribution of individual thresholds:  $P_{\phi}^{(\text{thr})} = \frac{7}{2}\phi^4(1-\phi)^2(1 - \frac{1}{2}\phi)^2$ . **b** Map of the interval showing the evolution of the model per Eq. (3). The two cobweb iterates indicate how a critical mass is required initially for the contagion to be self-sustained and grow

potentially catastrophic and unexpected spreading [21]. And in moving to more structured, socially realistic networks, even more surprising dynamics open up as possibilities [7, 13, 23].

## 4 Generalized Contagion Model

The SIR and threshold models are of course intended to be simple, extracting the most amount of story from the least amount of stage setting. But let’s list some standard “I have two comments”-type complaints anyway. As we have trumpeted, both models involve no memory other than of the current state traditional disease models assume independence of infectious events. Threshold models only involve proportions:  $17/73 \equiv 170/730$ . Threshold models also ignore the exact sequence of influences and assume immediate and repeated polling. Other issues applying to both models, and ones that we will not attend to here, include the choice between continuous and discrete time, synchronous updating for discrete time models, and the dominant assertion of random mixing populations (even so, network effects are only part of the story as media provides population-scale and sub-population scale signals). (Standard random scientist issue: “You did not cite my work [which you will find out is not related].”)

We would like to bring these basic models of biological and social contagion together, and, if this is possible, see if we can gain some new knowledge about contagion processes in general. Adding memory will be the way forward. Memory has been successfully incorporated into other kinds of social contagion models with a view to modeling real-world behavior online [8, 25].

We explain generalized contagion in the context of a random-mixing model acting on a population of  $N$  individuals. We will again have the three states **S**, **I**, and **R**, for susceptibles, infectives, and recovereds.

The major variation on the previous models is that each individual has a fixed memory length  $T$  drawn from a distribution  $P^{(\text{mem})}$  with  $1 \leq T \leq T_{\max}$ . In [5] and [6],  $T$  was the same for all individuals. At all times, individual  $i$  possesses a record of their last  $T_i$  interactions, a kind of ticker tape memory. Each entry in individual  $i$ 's memory will be either zero or a dose received from a successful interaction with an infective (details below).

As for Granovetter's model, we allow for a general threshold distribution,  $P^{(\text{thr})}$ . All nodes randomly select a threshold  $d^*$  using  $P^{(\text{thr})}$ , and thresholds remain fixed. Both memory and thresholds could be made to vary with time though we do not do this here.

Here's the game play for each step.

At each time step, regardless of their current state, each individual  $i$  will interact with a randomly chosen individual  $i'$  from the population. Next:

1. Individual  $i'$  will be an infective with probability  $\phi_t$ , the current fraction of infectives.
  - a. With probability  $p$ , a dose is successfully transmitted to  $i$ —an exposure. The dose size  $d$  will be drawn from a distribution  $P^{(\text{dose})}$ .
  - b. With probability  $1 - p$ ,  $i$  will not be exposed and they will record a dose size  $d = 0$ .
2. Individual  $i'$  will not be an infective with probability  $1 - \phi_t$  and  $i$  will record  $d = 0$  in its memory.

For the SIR model,  $p$  was the probability of a successful infection whereas now it is the probability of a successful transmission of a dose which is in turn probabilistic.

Node  $i$ 's updates its current dosage level  $D_{t,i}$  as the sum of its last  $T_i$  doses:

$$D_{t,i} = \sum_{t'=t-T_i+1}^t d_{t,i}. \quad (5)$$

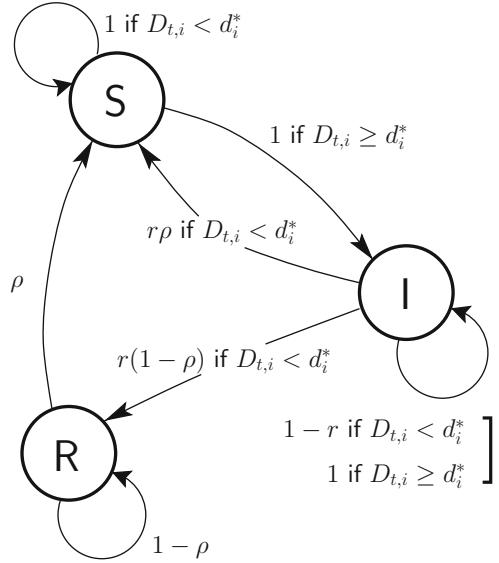
We can now define transition probabilities for individuals in each of the three states. As shown in Fig. 4:

- **S  $\Rightarrow$  I:** Infection occurs if individual  $i$ 's “threshold” is exceeded:

$$D_{t,i} \geq d_i^*. \quad (6)$$

- **I  $\Rightarrow$  R:** Only if  $D_{t,i} < d_i^*$ , individual  $i$  may recover to state **R** with probability  $r$ .
- **R  $\Rightarrow$  S:** An individual  $i$  may become susceptible again with probability  $\rho$ . A detail here is that we allow nodes that arrive in state **R** an immediate chance of returning to **S** in the same time step. Nodes in state **R** are immune and will remain in state **R** even if their dosage level  $D_{t,i}$  exceeds their threshold.

**Fig. 4** Mechanism of the generalized contagion model, developing from the same template used for the SIR model in Fig. 1



## 5 Analysis

We now perform some basic analyses of the generalized contagion model with a focus on determining the potential for a small seed to lead to a global spreading event, and characterizing the abruptness of that spreading if it is possible. In doing so, we will show how the dynamics of the SIR and threshold models are contained within that of generalized contagion.

Expanding on the results of [5, 6], the key quantity for our analysis is the probability that a randomly selected threshold  $d^*$  will be exceeded by  $k$  randomly selected doses drawn from  $P^{(\text{dose})}$ . Using the notation  $P_k^{(\text{inf})}$  we have

$$P_k^{(\text{inf})} = \int_0^\infty dd^* P_{d^*}^{(\text{thr})} \Pr \left( \sum_{j=1}^k d_j \geq d^* \right). \quad (7)$$

The integral is over all thresholds  $d^*$ , and the probability in the integrand is the cumulative distribution of the convolution of  $k$  copies of the dose distribution  $P^{(\text{dose})}$ .

The probabilities  $P_1^{(\text{inf})}$  and  $P_2^{(\text{inf})}$  will prove to be essential. In particular,  $P_1^{(\text{inf})}$ , the probability that one randomly chosen dose will exceed one randomly chosen threshold will determine if SIR-like dynamics are possible. The quantity  $P_1^{(\text{inf})}$  can be interpreted as the population fraction of the most “vulnerable” individuals [21]. Whatever the length of memory  $T$  of these individuals, they typically require only

one dose to become infected, and their high susceptibility enables the contagion to spread. This is a harder story to see and many are readily taken by the simpler, naive ones of “super-spreaders” and “influentials.”

We will consider the SIS version,  $\rho = 1$ , and the case of immediate recovery once an individual dosage drops below its threshold,  $r = 1$ . Although more difficult, some analytic work can be carried out if these probabilities are reduced below 1 (many variations are explored in [6]), and, of course, simulations can always be readily performed.

As with many dynamical systems problems, we are able to determine the main features of the  $\rho = r = 1$  generalized contagion system by examining the system’s fixed points which follow from the system’s update equation:

$$\phi_{t+1} = \sum_{T=1}^{T_{\max}} P_T^{(\text{mem})} \sum_{k=1}^T \binom{T}{k} (p\phi_t)^k (1-p\phi_t)^{T-k} P_k^{(\text{inf})}. \quad (8)$$

Reading through the right-hand side of this fixed point equation, we first have the probability that a randomly chosen individual has a memory of length  $T$ ,  $P_T^{(\text{mem})}$ . The inner sum then computes the probability that an individual with memory of length  $T$ ’s threshold is exceeded after receiving all possible numbers of positive doses,  $k = 1$  to  $k = T$ .

To find a closed form expression for the fixed points of the system, we set  $\phi_{t+1} = \phi_t$ :

$$\phi^* = \sum_{T=1}^{T_{\max}} P_T^{(\text{mem})} \sum_{k=1}^T \binom{T}{k} (p\phi^*)^k (1-p\phi^*)^{T-k} P_k^{(\text{inf})}. \quad (9)$$

In general, curves for  $\phi^*$  as a function of the exposure probability  $p$  will need to be determined numerically. However, for the question of whether a small seed may lead to a global spreading event or not, we can use Eq. (9) to find universal results.

Expanding Eq. (9) for  $\phi^*$  near 0 we obtain:

$$\phi^* = \sum_{T=1}^{T_{\max}} P_T^{(\text{mem})} T p \phi^* P_1^{(\text{inf})} + O(\phi^{*2}). \quad (10)$$

Taking  $\phi^* \rightarrow 0$ , we find the critical exposure probability for the system is therefore given by

$$p_c = \frac{1}{\langle T \rangle P_1^{(\text{inf})}}, \quad (11)$$

where  $\langle T \rangle = \sum_{T=1}^{T_{\max}} T P_T^{(\text{mem})}$  is the average memory length (if all individuals have a memory of uniform length  $T_*$ , as assumed in [5] and [6], Eq. (11) reduces to

$p_c = 1/[T_* P_1^{(\text{inf})}]$ .) We interpret  $p_c$  in the same way as the epidemic threshold of the SIR model. Global spreading from small seeds will occur if  $p > p_c$ , and this will only be feasible if the condition for an epidemic threshold is satisfied:

$$p_c < p < 1. \quad (12)$$

If instead  $p_c > 1$ , then our system will be more social-like. As we will show below, an initial critical mass will be needed for spreading to take off, if any spreading is possible at all.

To make the epidemic threshold criterion for generalized contagion intuitive, we can combine Eqs. (11) and (12) to form the condition:

$$(p\langle T \rangle) \cdot P_1^{(\text{inf})} > 1. \quad (13)$$

For a small seed to take off, the interpretation of Eq. (13) tracks as follows. Consider one infected individual at  $t = 0$  with a one off dose in their memory exceeding their threshold. They will randomly interact with  $T$  different uninfected individuals before they themselves recover. The expected number of exposures they will produce in this time is  $p\langle T \rangle$ , the first term in Eq. (13). Because the seed set of infectives is infinitesimally small, each susceptible individual interacted with by an infective will receive at most one dose. And this dose will infect them with probability  $P_1^{(\text{inf})}$ , the second term in Eq. (13). Thus,  $p\langle T \rangle \cdot P_1^{(\text{inf})}$  is the expected number of new infectives due to one infective, equivalent to the reproduction number  $R_0$  of the SIR model. In short, Eq. (13) is the statement that one infective begets at least one new infective, leading to an initial exponential growth of the contagion.

We now need to take some more care as the epidemic threshold for generalized contagion is not as simple as that of SIR contagion. If  $p_c < 1$ , we must consider whether the transition is continuous or discontinuous. As we saw with the example in Fig. 2, it is always the former for the SIR model.

If the transition is continuous, then when  $p = p_c$  a small seed will not grow, whereas when the transition is discontinuous, spreading will take off rapidly.

To test the phase transition's continuity, we expand Eq. (8) to second order:

$$\begin{aligned} \phi_{t+1} &\simeq (p\phi_t) \sum_{T=1}^{T_{\max}} P_T^{(\text{mem})} T P_1^{(\text{inf})} + (p\phi_t)^2 \sum_{T=1}^{T_{\max}} P_T^{(\text{mem})} T(T-1) \left[ \frac{1}{2} P_2^{(\text{inf})} - P_1^{(\text{inf})} \right] \\ &= (p\phi_t)\langle T \rangle P_1^{(\text{inf})} + (p\phi_t)^2 \langle T(T-1) \rangle \left[ \frac{1}{2} P_2^{(\text{inf})} - P_1^{(\text{inf})} \right]. \end{aligned} \quad (14)$$

Setting  $p = p_c$ , Eq. (11), we have:

$$\phi_{t+1} \simeq \phi_t + (\phi_t)^2 \frac{\langle T(T-1) \rangle}{\langle T \rangle^2 [P_1^{(\text{inf})}]^2} \left[ \frac{1}{2} P_2^{(\text{inf})} - P_1^{(\text{inf})} \right]. \quad (15)$$

A discontinuous phase transition is apparent if the fraction infected  $\phi_t$  grows and this evidently occurs if right-hand side of Eq. (15) is positive, meaning:

$$\begin{aligned} P_2^{(\text{inf})} < 2P_1^{(\text{inf})} &: \text{continuous,} \\ P_2^{(\text{inf})} > 2P_1^{(\text{inf})} &: \text{discontinuous.} \end{aligned} \quad (16)$$

We see that the kind of contagion behavior we observe with social phenomena, that repeated doses combine superlinearly  $P_2^{(\text{inf})} > 2P_1^{(\text{inf})}$ , corresponds with explosive spreading of a small seed at the critical point. Discontinuous phase transitions are phase transitions of surprise—as we increase the exposure probability  $p$  starting well below  $p_c$ , we see no spreading until we reach  $p_c$  (or just below depending on  $\phi_0$ ) when the growth will both be sudden and potentially leading to a large final fraction of infection. If repeated doses combine sublinearly,  $P_2^{(\text{inf})} < 2P_1^{(\text{inf})}$ , then the final fraction of infections will grow continuously from 0 as we move past  $p_c$ . Now, this is for the special case of a pure SIS model and as we later note, the criterion for a vanishing critical mass model,  $P_2^{(\text{inf})} > 2P_1^{(\text{inf})}$ , does not remain so simple as we move to more complicated models. So, while we can observe that a sufficiently nonlinear interaction in doses leads to non-epidemic threshold model, we arguably should not have been able to intuit the simple inequality  $P_2^{(\text{inf})} > 2P_1^{(\text{inf})}$  as being the salient test.

We can now assert that the generalized contagion model produces three distinct universality classes with respect to spreading behavior from a small seed. These are:

- **Epidemic Threshold Class:**

Criteria:

1.  $p_c = 1/(\langle T \rangle P_1^{(\text{inf})}) < 1$ .
2.  $P_1^{(\text{inf})} > P_2^{(\text{inf})}/2$ .

- **Vanishing Critical Mass:**

Criteria:

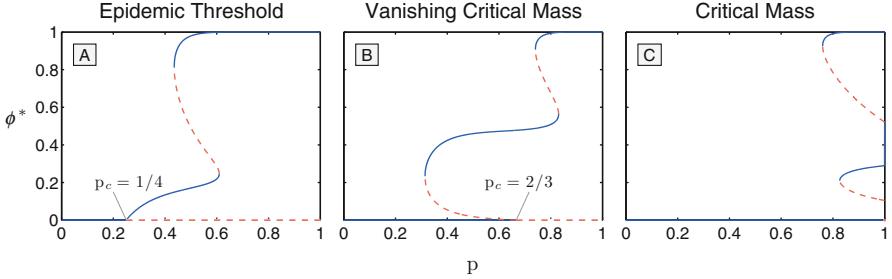
1.  $p_c = 1/(\langle T \rangle P_1^{(\text{inf})}) < 1$ .
2.  $P_1^{(\text{inf})} < P_2^{(\text{inf})}/2$ .

- **Pure Critical Mass:**

Criteria:

1.  $p_c = 1/(\langle T \rangle P_1^{(\text{inf})})$ .
2. Equation (9) is solvable with solutions  $\phi^*(p) \in [0, 1]$ .

In Fig. 5, we show results from numerically solving Eq. (9) for three example dose distributions and  $T = 20$  set uniformly, (see caption for details; adapted from Fig. 9 in [6]).



**Fig. 5** Examples of the three main universality classes with added bifurcative embellishments, arising solely from variable threshold distributions. In the epidemic threshold and vanishing critical mass cases of **a** and **b**, phase transitions for  $p < 1$  are both apparent but are strikingly different. The continuous phase transition in **a** means the system's behavior does not change abruptly as  $p$  moves above  $p_c$  for a small seed  $\phi_0$ . The discontinuous phase transition of **b** however means that the growth will be sudden and large. Vanishing critical mass models with  $r = 1$  have  $P_2^{(\text{inf})} > 2P_1^{(\text{inf})}$  which can be interpreted as meaning that the mutual effect of two doses is greater than their direct sum would suggest. In **c**, we see a critical mass system for which only a non-zero fraction must be initially infected for the contagion to maintain and spread. Mathematically,  $p_c > 1$ , so no small seed can take off. In all three cases, initial seeds will grow if the fixed point curve directly below them is unstable (and necessarily the one above will be stable). Simulation details (adapted from [6]):  $r = \rho = 1$ ,  $P_d^{(\text{mem})} = \delta_{T,1}$ , and  $P_d^{(\text{dose})} = \delta_{d,1}$ . (**a**)  $P_{d^*}^{(\text{dose})} = 0.2\delta(d-1) + 0.8\delta(d-6)$ ; (**b**)  $P_{d^*}^{(\text{dose})} = 0.075\delta(d-1) + 0.4\delta(d-2) + 0.525\delta(d-12)$ ; and (**c**)  $P_{d^*}^{(\text{dose})} = 0.3\delta(d-3) + 0.78\delta(d-12)$ . All curves were obtained from numerically solving Eq. (9)

The three panels correspond in order to the three universality classes. We emphasize that the universality classes we find here relate to the kind of critical point present in the system for  $\phi^* = 0$ , if such a critical point exists. The details of these systems are unimportant as many threshold and dose distributions give same  $P_k^{(\text{inf})}$ . All solid blue curves indicate stable fixed points and dashed red curves unstable fixed points.

For the epidemic threshold in Fig 5a, we see a continuous phase transition occurring at  $p_c = 1/4$ . Small seeds for  $p$  above  $p_c$  will grow but be constrained.

In Fig. 5b, the Vanishing Critical Mass class also shows a epidemic threshold but now the phase transition is discontinuous. Tuning the system from below to above  $p_c = 2/3$ , a small seed moves from ineffectual to suddenly producing successful global spreading to, roughly, half of the population.

The fixed point curves for the Critical Mass model in Fig. 5c show the resilience of this third class to small seeds initiating spreading events. Only if the initial seed is above the dashed red curves of unstable fixed points, will the final extent of spreading be non-zero (this statement is true for all three classes).

For uniform memory length  $T_*$ , the full linearization near  $p$  has the form [6]:

$$\phi^* \simeq \frac{C_1}{C_2 p^2} (p - p_c) = \frac{T_*^2 P_1^3}{(T_* - 1)(P_1 - P_2/2)} (p - p_c), \quad (17)$$

where from the denominator we can again see that  $P_1 - P_2/2 = 0$  locates the transition between Epidemic Threshold models and Vanishing Critical Mass models.

Moving away from systems behavior for small seeds, in all three examples, we see that the threshold distributions are of enough variability to produce non-trivial fixed point curves. Further, both the Epidemic Threshold and Vanishing Critical Mass examples also show that hysteresis dynamics (with respect to  $p$ ) are available for Generalized Contagion systems.

If we relax the recovery probability  $r$  below 1 and/or elevate the immune state transition probability  $\rho$  above 0, then we see the same three universality classes will still emerge. The conditions for the three classes will become more complex [6]. The appealing form of the test separating Epidemic Threshold and Vanishing Critical Mass models,  $P_2^{(\text{inf})} < 2P_1^{(\text{inf})}$ , will no longer be quite so simple. Analytic results are possible for  $r < 1$  and  $\rho = 0$  [6] while systems with  $\rho > 0$  have not yielded, at least to our knowledge, to exact treatments.

## 6 Concluding Remarks

We developed generalized contagion to demonstrate that a single mechanism could be shown to produce both disease-like and social-like spreading behavior. The observation that memory is a natural aspect of real-world spreading phenomena proved to be the binding agent.

The three universal classes of contagion processes pertain to the spectrum of random-mixing models and their dynamics in the fundamental initial condition of an infinitesimally small seed. We see that dramatic changes in behavior are possible, particularly in the Vanishing Critical Mass class.

Generalized contagion is also another example of a model where the vulnerable or gullible population may be more important than a small group of super-spreaders or influentials [22].

Two avenues for changing dynamics are clear. One would be to change the model itself through adjusting its parameters: memory, recovery rates, and the fraction of individuals vulnerable to 1 or 2 doses. ( $T$ ,  $r$ ,  $\rho$ ,  $P_1^{(\text{inf})}$ , and  $P_2^{(\text{inf})}$ ). Given a model with fixed parameters, changing the system's behavior would be possible by changing the probability of exposure ( $p$ ) and/or the initial fraction infected ( $\phi_0$ ).

We hope that this overview of generalized contagion serves as both an introduction to the model itself and an inspiration for the many possible adjacent areas in contagion dynamics available for development. Generalized contagion on social-like networks more complicated than random networks would be one such path. While perhaps this work would be resilient to simple analysis, simulations could prove illuminating.

## References

1. Bass F (1969) A new product growth model for consumer durables. *Manage Sci* 15:215–227
2. Castillo-Chavez C, Song B (2003) Models for the transmission dynamics of fanatic behaviors, vol 28. SIAM, Philadelphia, pp 155–172
3. Daley DJ, Kendall DG (1964) Epidemics and rumours. *Nature* 204:1118
4. Daley DJ, Kendall DG (1965) Stochastic rumours. *J Inst Math Appl* 1:42–55
5. Dodds PS, Watts DJ (2004) Universal behavior in a generalized model of contagion. *Phys Rev Lett* 92:218701
6. Dodds PS, Watts DJ (2005) A generalized model of social and biological contagion. *J Theor Biol* 232:587–604. <https://doi.org/10.1016/j.jtbi.2004.09.006>
7. Dodds PS, Harris KD, Danforth CM (2013) Limited imitation contagion on random networks: chaos, universality, and unpredictability. *Phys Rev Lett* 110:158701
8. Gleeson JP, O’Sullivan KP, Baños RA, Moreno Y (2016) Effects of network structure, competition and memory time on social spreading phenomena. *Phys Rev X* 6(2):021019
9. Goffman W, Newill VA (1964) Generalization of epidemic theory: an application to the transmission of ideas. *Nature* 204:225–228
10. Granovetter MS, Soong R (1983) Threshold models of diffusion and collective behavior. *J Math Sociol* 9:165–179
11. Granovetter MS, Soong R (1986) Threshold models of interpersonal effects in consumer demand. *J Econ Behav Organ* 7:83–99
12. Granovetter M, Soong R (1988) Threshold models of diversity: Chinese restaurants, residential segregation, and the spiral of silence. *Sociol Methodol* 18:69–104
13. Harris KD, Payne JL, Dodds PS (2014) Direct, physically-motivated derivation of triggering probabilities for contagion processes acting on correlated random networks. <http://arxiv.org/abs/1108.5398>
14. Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. *Proc R Soc Lond A* 115:700–721
15. Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. III. Further studies of the problem of endemicity. *Proc R Soc Lond A* 141(843):94–122
16. Kermack WO, McKendrick AG (1927) Contributions to the mathematical theory of epidemics. II. The problem of endemicity. *Proc R Soc Lond A* 138(834):55–83
17. Murray JD (2002) Mathematical biology, 3rd edn. Springer, New York
18. Schelling TC (1971) Dynamic models of segregation. *J Math Sociol* 1:143–186
19. Schelling TC (1978) Micromotives and macrobehavior. Norton, New York
20. Strogatz SH (1994) Nonlinear dynamics and chaos. Addison Wesley, Reading
21. Watts DJ (2002) A simple model of global cascades on random networks. *Proc Natl Acad Sci* 99(9):5766–5771
22. Watts DJ, Dodds PS (2007) Influentials, networks, and public opinion formation. *J Consum Res* 34:441–458
23. Watts DJ, Dodds PS (2009) Threshold models of social influence. In: Hedström P, Bearman P (eds) The Oxford Handbook of analytical sociology. Oxford University Press, Oxford, chap 20, pp 475–497
24. Watts DJ, Muhamad R, Medina D, Dodds PS (2005) Multiscale, resurgent epidemics in a hierarchcial metapopulation model. *Proc Natl Acad Sci* 102(32):11157–11162
25. Weng L, Flammini A, Vespignani A, Menczer F (2012) Competition among memes in a world with limited attention. *Nat Sci Rep* 2:335

# Message-Passing Methods for Complex Contagions



James P. Gleeson and Mason A. Porter

## 1 Introduction

In this chapter, we consider analytical approaches for calculating the expected sizes of cascades in complex contagions.<sup>1</sup> As a concrete example of a complex contagion, we use the Watts threshold model (WTM) [44] (see also [15, 43]) on undirected, unweighted networks. In this model, each node  $i$  of a network has a positive threshold  $r_i$ ; usually, the thresholds are chosen at random from a given probability distribution, but (with some difficulty and arguably circular reasoning) they can also be estimated from empirical data. We focus in particular on the case in which a contagion is initiated by multiple seed nodes, so we assume that a finite (but small) fraction of the network nodes are active at the beginning of contagion dynamics.

Each node can be in one of two states; we will call the states “inactive” and “active.” All nodes, except for the seed nodes, are initially inactive. In each discrete time step, each inactive node  $i$  of a network considers its neighboring nodes, and it becomes active if the fraction of its neighbors that are active exceeds or equals the threshold  $r_i$  of node  $i$ . One can interpret the threshold as a node’s stubbornness and the fraction of active nodes as a “peer pressure” function [26]. Once a node becomes active, it cannot later return to the inactive state, so the cascade grows in a monotonic

---

<sup>1</sup>See [35] and references therein for discussions of cascades on networks and for a “definition” of a complex contagion. See [29] for a friendly introduction to cascades on networks.

J. P. Gleeson (✉)

MACSI, Department of Mathematics and Statistics, University of Limerick, Limerick, Ireland  
e-mail: [james.gleeson@ul.ie](mailto:james.gleeson@ul.ie)

M. A. Porter

Department of Mathematics, University of California, Los Angeles, CA, USA

fashion. An important macroscopic quantity is the fraction  $\rho_n$  of active nodes at time step  $n$ . Because of the monotonic nature of the dynamics,  $\rho_n$  is a nondecreasing function of  $n$ , so (because  $\rho_n \leq 1$ , by definition) the limit  $\rho_\infty = \lim_{n \rightarrow \infty} \rho_n$  exists. We call  $\rho_\infty$  the “steady-state fraction of active nodes,” and we focus our attention on methods for analytically approximating its value.<sup>2</sup> Assuming that a fraction  $\rho_0$  of the nodes are selected uniformly at random as the seed nodes for a contagion, we want to predict the steady-state value  $\rho_\infty$  and to determine the conditions under which  $\rho_\infty$  substantially exceeds  $\rho_0$ . In other words, we want to answer the question “When does a global cascade occur?”<sup>3</sup>

The rest of this chapter is organized as follows. In Sects. 2 and 3, we focus on ensembles of infinite-size random networks (i.e., on asymptotic behavior as the number  $N$  of nodes becomes infinite), both without (see Sect. 2) and with (see Sect. 3) degree-degree correlations. In Sect. 4, we discuss recent progress on calculating  $\rho_\infty$  for finite-size networks. We conclude in Sect. 5.

## 2 Configuration-Model Networks

Let’s begin by assuming that our networks are realizations drawn from a configuration-model ensemble [9]; they are characterized by a given degree distribution  $p_k$ , where  $p_k$  is the probability that a node chosen uniformly at random has  $k$  neighbors, but they are otherwise maximally random. Moreover, our theoretical approach is for the limit of infinitely large networks (sometimes called the “thermodynamic limit”). Because configuration-model networks are locally tree-like [25], one might expect that we can apply mean-field approaches, such as those used for models of biological contagions [33], to approximate the fraction of active nodes. We’ll first briefly summarize what we’ll call a “naive mean-field (MF)” approach, and we’ll then explain why—and how—it can be improved.

### 2.1 Naive Mean-Field Approximation

We define  $\rho_n^{(k)}$  as the probability that a node of degree  $k$  is active at time step  $n$ ; the total fraction of active nodes is then given by

$$\rho_n = \sum_k p_k \rho_n^{(k)}. \quad (1)$$

---

<sup>2</sup>In [12], Gleeson and Cahalane showed that if nodes are updated one at a time in a random order, rather than all simultaneously as described here (i.e., if we use “asynchronous” updating instead of “synchronous” updating [35]), one obtains the same steady-state limit  $\rho_\infty$ , although the temporal evolution of the active fraction does depend on the updating scheme that is used [8]. See Sect. 5.1 of [35] for a description of an algorithm for a stochastic simulation of the WTM.

<sup>3</sup>See the discussion in [35] of ways of measuring cascade sizes.

A node of degree  $k$  is active at time  $n$  either because (i) it was a seed node (with probability  $\rho_0$ ) or (ii) it was not a seed node (with probability  $1 - \rho_0$ ), but it has become active by time step  $n$ . In the latter case, the number  $m$  of its active neighbors at time  $n - 1$  must be large enough so that the fraction  $m/k$  is at least as large as the node's threshold. Treating the  $k$  neighbors as independent of each other, the probability that  $m$  of the  $k$  are active at time  $n$  is given by the binomial distribution

$$\binom{k}{m} (\bar{\rho}_{n-1})^m (1 - \bar{\rho}_{n-1})^{k-m}, \quad (2)$$

where  $\bar{\rho}_{n-1}$  is the probability that the node at the end of a uniformly randomly chosen edge is active at time step  $n - 1$ . Under the usual mean-field assumptions (see, for example, [25]), we write  $\bar{\rho}_{n-1}$  as the weighted mean over the possible degrees of neighbors<sup>4</sup>:

$$\bar{\rho}_{n-1} = \sum_k \frac{k}{z} p_k \rho_{n-1}^{(k)}, \quad (3)$$

where  $z = \sum_k k p_k$  is the mean degree of the network.

If  $m$  neighbors of a node are active, the probability that the node is active is equal to the probability that its threshold is less than  $m/k$ . We write this probability as  $C(m/k)$ , where  $C$  is the cumulative distribution function (CDF) of the thresholds. Putting together these arguments and summing over all possible values of  $m$ , we write the MF approximation for  $\rho_n^{(k)}$  as

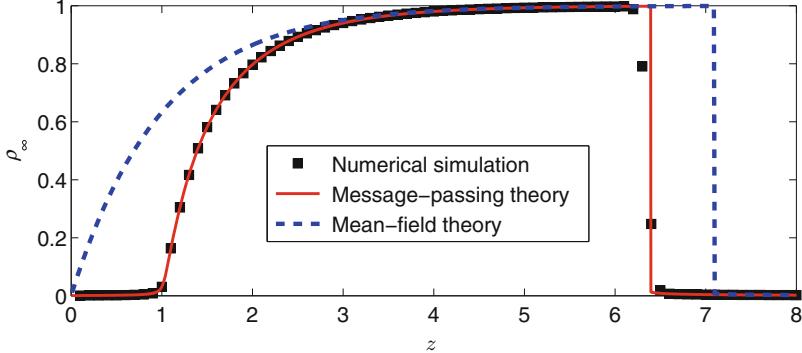
$$\rho_n^{(k)} = \rho_0 + (1 - \rho_0) \sum_{m=0}^k \binom{k}{m} (\bar{\rho}_{n-1})^m (1 - \bar{\rho}_{n-1})^{k-m} C\left(\frac{m}{k}\right). \quad (4)$$

Multiplying Eq. (4) by  $\frac{k}{z} p_k$  and summing over  $k$  gives

$$\bar{\rho}_n = \rho_0 + (1 - \rho_0) \sum_k \frac{k}{z} p_k \sum_{m=0}^k \binom{k}{m} (\bar{\rho}_{n-1})^m (1 - \bar{\rho}_{n-1})^{k-m} C\left(\frac{m}{k}\right), \quad (5)$$

so we now have an expression for  $\bar{\rho}_n$  in terms of  $\bar{\rho}_{n-1}$ . Starting from an initial condition with a fraction  $\bar{\rho}_0 = \rho_0$  of seed nodes (chosen uniformly at random), one can iterate Eq. (5) to determine  $\bar{\rho}_n$  for any later time step, and it converges to  $\bar{\rho}_\infty$

<sup>4</sup>The weighting  $(k/z)p_k$  arises because we are considering the mean over nodes of degree  $k$ , where those nodes are reached by traveling along an edge from the node of interest. It is well-known (see, e.g., [32]) that a node at the end of a uniformly randomly chosen edge of a configuration-model network has degree  $k$  with probability  $(k/z)p_k$ , reflecting the fact that large- $k$  nodes are more likely than small- $k$  nodes to be reached in this way.



**Fig. 1** The expected steady-state fraction  $\rho_\infty$  of active nodes for cascades in the Watts threshold model (WTM) when every node has the same threshold  $r = 0.18$  (so a node becomes active when its fraction of active neighbors is at least as large as 0.18). The networks are Erdős–Rényi random graphs ( $G(N, m)$ , where  $m$  is the total number of edges) with mean degree  $z$  (so they have approximately a Poisson degree distribution  $p_k = z^k e^{-z} / k!$ ), and the initial seed fraction is  $\rho_0 = 10^{-3}$ . The simulation results, shown by the black squares, are a mean over 100 realizations on networks with  $N = 10^5$  nodes. The blue dashed curve shows the result of the naive mean-field approximation given by Eqs. (5) and (6), and the red solid curve comes from the message-passing approach of Eqs. (10) and (12)

as  $n \rightarrow \infty$ . One then calculates the naive MF approximation to the steady-state fraction  $\rho_\infty$  of active nodes from Eqs. (1) and (4) with the formula

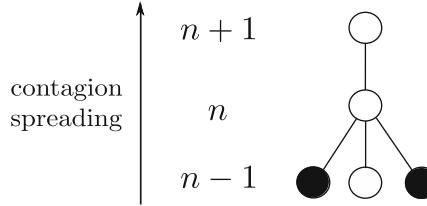
$$\rho_\infty = \rho_0 + (1 - \rho_0) \sum_k p_k \sum_{m=0}^k \binom{k}{m} (\bar{\rho}_\infty)^m (1 - \bar{\rho}_\infty)^{k-m} C\left(\frac{m}{k}\right). \quad (6)$$

However, as we illustrate in Fig. 1, the naive MF approximation calculated using Eqs. (5) and (6) does not accurately match the values of  $\rho_\infty$  from numerical simulations on large networks. In Sect. 2.2, we consider why this mismatch occurs, and we introduce an improved approximation technique, which is of “message-passing” type.

## 2.2 Message-Passing for Configuration-Model Networks

In this section, we present the approach that was first used in [10, 13], who adapted the method used by Dhar et al. [5] for the zero-temperature random-field Ising model on Bethe lattices. Nowadays, the approach is called “message-passing for configuration-model networks.” See, for example, Sect. IV of [40].

The fundamental problem with the naive MF approach of Sect. 2.1 is that it neglects the directionality in the spreading of a contagion. The contagion spreads outwards from the seed nodes, and it can reach inactive nodes only after it has first



**Fig. 2** Schematic for the method described in Sect. 2.2. We suppose that the contagion spreads upward from level  $n - 1$  to level  $n$  and beyond. The assumption of infinite network size allows us to consider the limit of an infinite number of levels, terminating with the “top” (i.e., “root”) node of the tree approximation

infected some of their neighbors. In the schematic in Fig. 2, we assume that the contagion spreads upward from “level”  $n - 1$  to level  $n$  and then to level  $n + 1$ . We number the levels according to their distance from the seed nodes, which we place at level 0. This is a highly stylized approximation, as we are almost always considering networks that are not actually trees (and, e.g., social networks typically have significant clustering), but we see nevertheless that it gives good results (see, e.g., the discussion in [25]). For the synchronous updating that we employ in this chapter, level  $n$  of the tree approximation corresponds to time step  $n$  of the contagion process on the original network. See [10] for details and an extension to asynchronous updating.

We now focus again on the steady-state limit  $n \rightarrow \infty$ . We introduce the variable  $q_n^{(k)}$ , the probability that a node of degree  $k$  on level  $n$  is active, conditional on its parent (at level  $n + 1$ ) being inactive. When we calculate  $q_n^{(k)}$ , we account for the directionality of the contagion spreading, because we assume that the node at level  $n + 1$  in Fig. 2 is inactive at the time when the node at level  $n$  is updating from the inactive to the (possibly) active state. As before, there are two ways in which the node at level  $n$  can be active: either it was a seed node (with probability  $\rho_0$ ) or it was not a seed node (with probability  $1 - \rho_0$ ) but has been activated by its children (i.e., the nodes at level  $n - 1$  in Fig. 2). Because the level- $n$  node has degree  $k$  and one of its edges is adjacent to its (inactive) parent, there are  $k - 1$  children node at level  $n - 1$ . Each of these children is active with probability  $q_{n-1}$ , where (similar to Eq. (3))  $q_n$  is the weighted mean over the  $q_n^{(k)}$  values. That is,

$$q_{n-1} = \sum_k \frac{k}{z} p_k q_n^{(k)}. \quad (7)$$

Therefore, the probability that  $m$  children are active is given by the binomial distribution on  $k - 1$  nodes, where each is active with independent probability  $q_{n-1}$ . That is,

$$\binom{k-1}{m} (q_{n-1})^m (1 - q_{n-1})^{k-1-m}. \quad (8)$$

As with the naive MF case, the activation of a degree- $k$  node with  $m$  active children depends on its threshold being less than the fraction  $m/k$ ; this occurs with probability  $C(m/k)$ . Putting together the preceding arguments, we write

$$q_n^{(k)} = \rho_0 + (1 - \rho_0) \sum_{m=0}^{k-1} \binom{k-1}{m} (q_{n-1})^m (1 - q_{n-1})^{k-1-m} C\left(\frac{m}{k}\right), \quad (9)$$

and we obtain a discrete scalar map for  $q_n$  by multiplying Eq. (9) by  $\frac{k}{z} p_k$  and summing over  $k$ . Using Eq. (7) then yields

$$q_n = \rho_0 + (1 - \rho_0) \sum_k \frac{k}{z} p_k \sum_{m=0}^{k-1} \binom{k-1}{m} (q_{n-1})^m (1 - q_{n-1})^{k-1-m} C\left(\frac{m}{k}\right). \quad (10)$$

Iterating Eq. (10) starting from initial condition  $q_0 = \rho_0$  leads to the steady-state value

$$q_\infty = \lim_{n \rightarrow \infty} q_n. \quad (11)$$

Finally, we use the fact that a node at the “top” (i.e., “root”) of the tree—formally at level  $\infty$ —has  $k$  children with probability  $p_k$  and (assuming that the root node is not a seed node) that each child is active with probability  $q_\infty$ . We then determine the steady-state active fraction of nodes from  $q_\infty$  by calculating

$$\rho_\infty = \rho_0 + (1 - \rho_0) \sum_k p_k \sum_{m=0}^k \binom{k}{m} (q_\infty)^m (1 - q_\infty)^{k-m} C\left(\frac{m}{k}\right). \quad (12)$$

The solid red curve in Fig. 1 shows the result of using Eqs. (10) and (12) to determine the steady-state fraction of active nodes. This approximation method is very accurate, and it is far superior to the naive MF approach of Sect. 2.1. Note that simulation results at the discontinuous transition near  $z = 6.5$  depend strongly on the size of a network, and agreement with the theory improves as one considers larger networks (see Fig. 3 of [12]).

### 2.3 The Criticality Condition (i.e., “Cascade Condition”)

An additional benefit of the analytical approach that we outlined in Sect. 2.2 is that it enables one to determine conditions on the model parameters that control whether or not global cascades occur. This question was first addressed by Watts [44] using a percolation argument, but one can derive the same condition using the approach of Sect. 2.2. For this analysis, we assume that the seed fraction is vanishingly small,

so we take the  $\rho_0 \rightarrow 0$  limit of our general equations. (See [13] for extensions to nonzero  $\rho_0$ .) In this case, Eq. (10) always has the solution  $q_n \equiv 0$  for all  $n$ , corresponding to the case of no contagion. However, for certain parameter regimes, this contagionless solution can be unstable, and then any infinitesimal seed fraction  $\rho_0 > 0$  leads to a global cascade of nonzero fractional size. (The “fractional size” of a contagion is the number of active nodes divided by the total number of nodes.) Therefore, we linearize Eq. (10) about the solution  $q_n \equiv 0$  to determine its (linear) stability. For scalar maps of the form  $q_n = g(q_{n-1})$ , the criterion for instability of the 0 solution is that [41]

$$|g'(0)| > 1. \quad (13)$$

Differentiating the right-hand side of Eq. (10) and setting  $q_{n-1} = 0$  yields the following condition for global cascades to occur (from an infinitesimal seed fraction)<sup>5</sup>:

$$\sum_k \frac{k}{z} p_k (k-1) C\left(\frac{1}{k}\right) > 1. \quad (14)$$

Given a network’s degree distribution  $p_k$  and the CDF  $C$  of thresholds, it is easy to evaluate the condition (14), so Eq. (14) is a very useful criterion for determining whether global cascades can exist (the “supercritical regime”) or not (the “subcritical regime”).

### 3 Networks with Degree–Degree Correlations

We now follow [6, 10, 34] and extend the message-passing approach to networks with nontrivial degree–degree correlations. Let  $p_{kk'}$  be the joint probability distribution function (PDF) for the degrees  $k$  and  $k'$  of the end nodes of a uniformly randomly chosen edge of a network.<sup>6</sup> As in Sect. 2, and referring again to Fig. 2, we define  $q_n^{(k)}$  as the probability that a degree- $k$  node on level  $n$  is active, conditional on its parent (on level  $n+1$ ) being inactive. Similarly, writing  $\bar{q}_n^{(k)}$  for the probability that a child of an inactive level- $(n+1)$  node of degree  $k$  is active, it follows that

$$\bar{q}_n^{(k)} = \frac{\sum_{k'} p_{kk'} q_n^{(k')}}{\sum_{k'} p_{kk'}}, \quad (15)$$

---

<sup>5</sup>Note that  $C(0) = 0$ , because we have assumed that all thresholds are positive.

<sup>6</sup>In configuration-model networks, in which no correlations are imposed in the generative model,  $p_{kk'} = kp_k k' p_{k'}/z^2$ , because the degrees of the nodes at the two ends of an edge are independent.

because a neighbor of the degree- $k$  node has degree  $k'$  with probability  $p_{kk'}/\sum_{k''} p_{kk''}$ . Similar to Eq. (9), we then determine the conditional probabilities for each degree at level  $n$  from the children at level  $n - 1$  using the relation

$$q_n^{(k)} = \rho_0 + (1 - \rho_0) \sum_{m=0}^{k-1} \binom{k-1}{m} \left(\bar{q}_{n-1}^{(k)}\right)^m \left(1 - \bar{q}_{n-1}^{(k)}\right)^{k-1-m} C\left(\frac{m}{k}\right), \quad (16)$$

where  $q_0^{(k)} = \rho_0$  for all  $k$ . The unconditional density of active degree- $k$  nodes at steady-state is

$$\rho_\infty^{(k)} = \rho_0 + (1 - \rho_0) \sum_{m=0}^k \binom{k}{m} \left(q_\infty^{(k)}\right)^m \left(1 - q_\infty^{(k)}\right)^{k-m} C\left(\frac{m}{k}\right), \quad (17)$$

and the total network density is equal to

$$\rho_\infty = \sum_k p_k \rho_\infty^{(k)}. \quad (18)$$

### 3.1 Matrix Criticality Condition

As in Sect. 2.3, one can derive the condition that determines whether global cascades arise from infinitesimal (i.e.,  $\rho_0 \rightarrow 0$ ) seeds by linearizing the system of equations (16) about the zero-contagion solution  $q_n^{(k)} \equiv 0$  for all  $n$  and  $k$ . Note that Eq. (16) includes one equation for each distinct degree class in a network, so the condition for instability of the contagionless solution is an eigenvalue condition on the Jacobian matrix of the system. From Eqs. (16) and (15), we find (see [10]) that the condition for instability (i.e., for the existence of global cascades) is that the largest eigenvalue<sup>7</sup> of the matrix  $\mathbf{M}$  exceeds 1, where  $\mathbf{M}$  is the matrix with entries

$$M_{kk'} = \frac{(k'-1)}{\sum_{k''} p_{kk''}} p_{kk'} C\left(\frac{1}{k'}\right). \quad (19)$$

As noted in [10], a similar condition occurs for bond percolation on networks with degree-degree correlations [31], and such conditions are also relevant for epidemic models on networks [19].

The message-passing method that we have described has also been generalized for networks with community structure [10] and different degree-degree correlations in different communities [27] (where the latter case also has a notable interpretation in the language of multilayer networks [20]), multiplex networks

---

<sup>7</sup>The matrix  $\mathbf{M}$  is not symmetric, but there exists a similarity transformation to a symmetric matrix, so all of its eigenvalues are real.

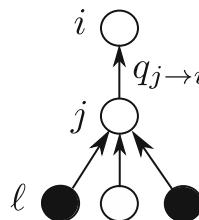
[46], other contagion models [18], dynamics in which nodes can be in more than two states [26], and more. Reference [11] presented an alternative derivation (starting from a so-called “approximate master equation” (AME) framework) of the configuration-model-approximation equations (10) and (12).

## 4 Message-Passing for Finite-Size Networks

In this section, we discuss message-passing approaches [23, 40] that are applicable to finite-size networks, rather than to the ensembles of (infinite-size) networks that we discussed above. Recent papers [23, 40] have shown how a message-passing approach can be applied successfully to networks with a finite number of nodes. In this section, we explain this idea by applying it to the WTM. The resulting equations are computationally very expensive to solve. We close the chapter by deriving the analog of the criticality conditions of Eqs. (14) and (19) for the existence of global cascades in finite-size networks. This criticality condition is relatively tractable to compute.

Suppose that we are given a finite-size network that is unweighted and undirected (and unipartite). The total number of edges in the  $N$ -node network is  $E$ , where  $E = Nz/2$  and  $z$  is the mean degree. To use a message-passing approach, we consider quantities like  $q_{j \rightarrow i}$ , which are specified for a *directed* edge  $j \rightarrow i$ . We consider each undirected edge of a network (such as the one between nodes  $i$  and  $j$ ) as consisting of a reciprocal pair of directed edges ( $i \rightarrow j$  and  $j \rightarrow i$ ), giving a total of  $2E$  directed edges. The direction of the edges gives the local directionality of a contagion, analogous to the ascending levels in Fig. 2.

The edge-based quantity  $q_{j \rightarrow i}$  is the probability that node  $j$  is active, conditional on node  $i$  being inactive. See Fig. 3, and compare it to Fig. 2. To write an equation for  $q_{j \rightarrow i}$ , we consider the effect on  $j$  of all of its neighbors aside from  $i$ . Specifically, if node  $j$  is not a seed node (which is the case with probability  $1 - \rho_0$ ), it is active only if sufficiently many of its neighbors are active. To calculate  $q_{j \rightarrow i}$ , we assume that node  $i$  is inactive,<sup>8</sup> so we must consider whether the number of active nodes among



**Fig. 3** Schematic for the message-passing approach of Sect. 4

---

<sup>8</sup>This assumption has various names: it is called the “cavity approach” in statistical physics [28, 39, 47], and it is closely related to the WOR (“without regarding”) property that was used for financial-contagion cascades in [16].

the remaining neighbors is sufficient to activate node  $j$ . It is convenient to introduce the notation  $\sigma_\ell$  to represent the state of node  $\ell$  in a given realization:  $\sigma_\ell = 1$  if node  $\ell$  is active, and  $\sigma_\ell = 0$  if node  $\ell$  is inactive. One can then write the equation for  $q_{j \rightarrow i}$  as

$$q_{j \rightarrow i} = \rho_0 + (1 - \rho_0) \sum_{\{\sigma_\ell\}: \ell \in \mathcal{N}_j \setminus i} C \left( \frac{\sum_\ell \sigma_\ell}{k_j} \right) \prod_{\sigma_\ell=1} q_{\ell \rightarrow j} \prod_{\sigma_\ell=0} (1 - q_{\ell \rightarrow j}). \quad (20)$$

The summation in Eq. (20) is over all combinations of  $\sigma_\ell$  values. In other words, one sums over the possible states of the neighbors of  $j$  (where  $\mathcal{N}_j$  denotes the set of such neighbors), except for node  $i$ . Given the set  $\{\sigma_\ell\}$  of neighbor states, the fraction of active neighbors of node  $j$  is  $\sum_\ell \sigma_\ell / k_j$ , where  $k_j$  is the degree of node  $j$ . The probability that this fraction is at least as large as the threshold of node  $j$  is given by  $C\left(\frac{\sum_\ell \sigma_\ell}{k_j}\right)$ . Let's consider each of inactive node  $j$ 's neighbors, except for  $i$ . Because each of these nodes  $\ell$  is active with an independent probability of  $q_{\ell \rightarrow j}$ , the first product term of Eq. (20) gives the probability that a specified subset of nodes is active, and the second product term of Eq. (20) gives the probability that the remaining neighbors of  $j$  are inactive. Consequently, multiplying the two product terms gives the probability (assuming that  $j$  is inactive) to have a given combination  $\{\sigma_\ell\}_{\ell \in \mathcal{N}_j \setminus i}$  of neighbors' states, and the sum over all possible combinations plays the same role as the sum over  $m$  in Eqs. (9) and (16).

In principle, one can solve Eq. (20) by iteration to determine  $q_{j \rightarrow i}$  for every directed edge. The probability that node  $i$  is active (similar to Eq. (17)) is then given by

$$\rho^{(i)} = \rho_0 + (1 - \rho_0) \sum_{\{\sigma_j\}: j \in \mathcal{N}_i} C \left( \frac{\sum_j \sigma_j}{k_i} \right) \prod_{\sigma_j=1} q_{j \rightarrow i} \prod_{\sigma_j=0} (1 - q_{j \rightarrow i}), \quad (21)$$

where the sum in Eq. (21) is over the possible states of all neighbors of  $i$  (compare to Eq. (17)). Unfortunately, the summations in both Eqs. (20) and (21) require calculating a combinatorially large numbers of terms. For example, the sum over the sets  $\{\sigma_\ell\}_{\ell \in \mathcal{N}_j \setminus i}$  of the possible states of the neighbors of node  $j$  has  $2^{k_j-1}$  terms, each of which has its own probability measure that needs to be evaluated with the two product terms in Eq. (20). The large number of possible combinations makes the implementation of this message-passing approach extremely computationally expensive, except for very small networks.

On the bright side, one can derive the steady-state equations for the configuration-model ensemble that we discussed in Sect. 2 from the message-passing Eqs. (20) and (21), as is described in detail in [40]. Essentially, in a configuration-model ensemble, each edge-based conditional probability  $q_{\ell \rightarrow j}$  is replaced by the single quantity  $q$  (which we called  $q_\infty$  in Sect. 2). Because all neighbors are treated as identical, the sum in Eq. (20) over  $\{\sigma_\ell\}$  becomes the sum over the number  $m$  of active neighbors, weighted by the binomial coefficient  $\binom{k-1}{m}$ , which gives the number of arrangements of precisely  $m$  active neighbors among the

$k - 1$  neighbors who can be active. Consequently, the sum over  $\{\sigma_\ell\}$  in Eq. (20) reduces to a sum over  $m$  in Eq. (8), yielding the steady-state limit ( $n \rightarrow \infty$ ) of the configuration-model equations (10) and (12).

## 4.1 Criticality Condition for Finite-Size Networks

Although calculating the full message-passing equations (21) is prohibitively expensive for large networks, one can nevertheless apply the same approach as in earlier sections to derive a condition for the existence of global cascades. As before, we take the  $\rho_0 \rightarrow 0$  limit and linearize the governing equation (20) about the zero-contagion equilibrium. Specifically, we linearize Eq. (20) about  $q_{j \rightarrow i} = 0$  for each edge. For very small values of the edge probabilities, the sum in Eq. (20) gives a linear contribution only when a single neighbor is active. The resulting linearization is then given by

$$q_{j \rightarrow i} = \sum_{\ell \in \mathcal{N}_j \setminus i} C\left(\frac{1}{k_j}\right) B_{i \rightarrow j, j \rightarrow \ell} q_{\ell \rightarrow j}, \quad (22)$$

where  $\mathbf{B}$  is the nonbacktracking (Hashimoto) matrix, which has recently been studied in network-science questions such as percolation [17], community detection [21], and centrality [24, 38]. The nonbacktracking matrix is a sparse matrix of dimension  $2E \times 2E$ , where each row (or column) corresponds to a directed edge between two nodes. The elements of  $\mathbf{B}$  are nonzero when the directed edge that corresponds to the row (e.g., the edge  $i \rightarrow j$ ) leads to the directed edge that corresponds to the column (e.g.,  $j \rightarrow \ell$ ) via a common node (which, in this case, is node  $j$ ), provided that the second directed edge does not return to the source node of the original edge (i.e., node  $\ell$  cannot be the same as node  $i$ ).

Rewriting Eq. (22) in a matrix form that is suitable for iteration (analogous to Eqs. (10) and (16)) yields

$$\mathbf{q}_n = \mathbf{DB}\mathbf{q}_{n-1}, \quad (23)$$

where  $\mathbf{q}$  is the  $2E$ -vector of values  $q_{j \rightarrow i}$ . We then immediately see that the linear stability of the  $\mathbf{q} = \mathbf{0}$  solution depends on the largest eigenvalue of the product matrix  $\mathbf{DB}$ , where  $\mathbf{D}$  is a  $2E \times 2E$  diagonal matrix with nonzero elements given by

$$D_{i \rightarrow j, i \rightarrow j} = C\left(\frac{1}{k_j}\right). \quad (24)$$

The criterion that we have derived from the message-passing approach is therefore that the existence of global cascades requires the spectral radius of the  $2E \times 2E$  matrix  $\mathbf{DB}$  to exceed 1. Because the matrix is sparse, one can check this cascade criterion even for large networks.

**Table 1** The critical value of  $\theta$ , the upper limit of the uniform distribution of thresholds, for the WTM on various networks, as calculated using the configuration-model result Eq. (14) for  $\theta_{\text{config}}$  and using the maximum eigenvalue of the **DB** matrix in Eq. (23) to determine  $\theta_{\text{crit}}$

Network	$N$	$z$	$\theta_{\text{config}}$	$\theta_{\text{crit}}$
3-Regular random graph	$10^5$	3	$\frac{2}{3}$	$\frac{2}{3}$
Facebook (Caltech) [42]	762	43.7	0.98	0.98
Facebook (Oklahoma) [42]	17,420	102	0.99	0.99
Gowalla [2, 4]	$1.97 \times 10^5$	9.67	0.90	0.94
PGP network [1, 3]	10,680	4.55	0.78	0.94
Power grid [30, 45]	4941	2.67	0.63	0.78

The network size (i.e., number of nodes) is  $N$  and the mean degree is  $z$ , so the number of undirected edges is  $E = Nz/2$ . Note, as expected, that  $\theta_{\text{config}}$  is identical to  $\theta_{\text{crit}}$  for the 3-regular random graph. The corresponding values for the Facebook networks are also very close, indicating that the configuration-model theory is very accurate for these networks (as also found in [14, 25]). For the other networks, there is a considerable difference between  $\theta_{\text{config}}$  and  $\theta_{\text{crit}}$ , indicating that the configuration-model result is inaccurate for these networks (although it is also known that the message-passing approach, which is based on a tree-like assumption of independence of messages [7], tends to be inaccurate for spatially-embedded networks [36, 38], such as the power-grid example in this table)

In Table 1, we give examples in which we consider the WTM with thresholds uniformly distributed over the interval  $(0, \theta)$ , so the mean threshold value is  $\theta/2$ . If the parameter  $\theta$  is small, all thresholds are small, and a seed node is likely to cause many neighbors to become active, leading quickly to a global cascade. However, a very large value of  $\theta$  implies that many nodes' thresholds are too large to allow them to activate, so no global cascades occur. In Table 1, we report the critical value of the parameter  $\theta$  that separates the global-cascade (i.e., supercritical) regime from the no-global-cascade (i.e., subcritical) regime for several real-world networks using the configuration-model condition given by Eq. (14) and the spectral condition on the matrix **DB** that we described above. In previous work on calculating percolation thresholds for real-world networks [36, 37], using the nonbacktracking matrix has led to more accurate predictions than those found by applying configuration-model theory (which uses only the degree distribution of a network). We therefore anticipate that the cascade threshold identified by the largest eigenvalue of the matrix **DB** will be more accurate than configuration-model predictions and will provide important insights into the structural features of certain networks that enable configuration-model theories to give accurate results [14].

## 5 Conclusions

In this chapter, we reviewed several analytical approaches for complex-contagion dynamics. For concreteness, we focused on the example of the Watts threshold model, but the methods that we discussed can also be applied to other monotonic binary-state dynamics [11]. To provide context, we first introduced a naive mean-

field approach, which has limited accuracy. We then showed that using the methods of [5, 13] gives very accurate results on configuration-model networks. We demonstrated how the methodology can yield a criterion for determining whether global cascades occur, and we briefly reviewed an extension of the method to networks with imposed degree-degree correlations. In Sect. 4, we briefly discussed the approaches of [23, 40] to derive message-passing equations for cascades on finite-size networks. Although the resulting equations are computationally expensive to solve, we showed that they give a condition for global cascades in terms of the spectral radius of a matrix that is related to the nonbacktracking matrix. The nonbacktracking matrix has arisen in prior work from linearizations of belief-propagation algorithms [21], but the product matrix **DB** that determines the cascade condition has not been studied in detail (to our knowledge), and we believe that further investigations of it will yield fascinating insights into the propagation of monotonic complex contagions and other monotonic dynamics [22, 23].

**Acknowledgements** This work was supported by Science Foundation Ireland on grants that were awarded to JPG (grant numbers 16/IA/4470 and 11/PI/1026). We acknowledge the SFI/HEA Irish Centre for High-End Computing (ICHEC) for the provision of computational facilities.

## References

1. Boguñá M, Pastor-Satorras R, Díaz-Guilera A, Arenas A (2004) Models of social networks based on social distance attachment. *Phys Rev E* 70(5):056122
2. Cho E, Myers SA, Leskovec J (2011) Friendship and mobility: User movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, New York, pp 1082–1090
3. Largest connected component of the network of users of the Pretty-Good-Privacy algorithm for secure information interchange. <http://deim.urv.cat/~alexandre.arenas/data/xarxes/PGP.zip>
4. SNAP: Network datasets: Gowalla <http://snap.stanford.edu/data/loc-gowalla.html>
5. Dhar D, Shukla P, Sethna JP (1997) Zero-temperature hysteresis in the random-field Ising model on a Bethe lattice. *J Phys A Math Gen* 30(15):5259
6. Dodds PS, Payne JL (2009) Analysis of a threshold model of social contagion on degree-correlated networks. *Phys Rev E* 79(6):066115
7. Faqeeh A, Melnik S, Gleeson JP (2015) Network cloning unfolds the effect of clustering on dynamical processes. *Phys Rev E* 91(5):052807
8. Fennell PG, Melnik S, Gleeson JP (2016) Limitations of discrete-time approaches to continuous-time contagion dynamics. *Phys Rev E* 94(5):052125
9. Fosdick BK, Larremore DB, Nishimura J, Ugander J (2018) Configuring random graph models with fixed degree sequences. *SIAM Rev* 60(2):315–355
10. Gleeson JP (2008) Cascades on correlated and modular random networks. *Phys Rev E* 77(4):046117
11. Gleeson JP (2013) Binary-state dynamics on complex networks: Pair approximation and beyond. *Phys Rev X* 3(2):021004
12. Gleeson JP, Cahalane DJ (2007a) An analytical approach to cascades on random networks. In: SPIE Fourth International Symposium on Fluctuations and Noise. International Society for Optics and Photonics, Bellingham, 66010W

13. Gleeson JP, Cahalane DJ (2007b) Seed size strongly affects cascades on random networks. *Phys Rev E* 75(5):056103
14. Gleeson JP, Melnik S, Ward JA, Porter MA, Mucha PJ (2012) Accuracy of mean-field theory for dynamics on real-world networks. *Phys Rev E* 85(2):026106
15. Granovetter M (1978) Threshold models of collective behavior. *Am J Sociol* 83(6):1420–1443
16. Hurd TR (2016) *Contagion! Systemic Risk in Financial Networks*. Springer, Cham
17. Karrer B, Newman MEJ, Zdeborová L (2014) Percolation on sparse networks. *Phys Rev Lett* 113(20):208702
18. Karsai M, Ihiguchi G, Kaski K, Kertész J (2014) Complex contagion process in spreading of online innovation. *J R Soc Interface* 11(101):20140694
19. Kiss IZ, Miller JC, Simon PL (2017) *Mathematics of Epidemics on Networks: From Exact to Approximate Models*. Interdisciplinary Applied Mathematics. Springer, Cham
20. Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA (2014) Multilayer networks. *J Complex Netw* 2(3):203–271
21. Krzakala F, Moore C, Mossel E, Neeman J, Sly A, Zdeborová L, Zhang P (2013) Spectral redemption in clustering sparse networks. *Proc Natl Acad Sci U S A* 110(52):20935–20940
22. Lokhov AY, Saad D (2017) Optimal deployment of resources for maximizing impact in spreading processes. *Proc Natl Acad Sci U S A* 114(39):E8138–E8146
23. Lokhov AY, Mézard M, Zdeborová L (2015) Dynamic message-passing equations for models with unidirectional dynamics. *Phys Rev E* 91(1):012811
24. Martin T, Zhang X, Newman MEJ (2014) Localization and centrality in networks. *Phys Rev E* 90(5):052808
25. Melnik S, Hackett A, Porter MA, Mucha PJ, Gleeson JP (2011) The unreasonable effectiveness of tree-based theory for networks with clustering. *Phys Rev E* 83(3):036112
26. Melnik S, Ward JA, Gleeson JP, Porter MA (2013) Multi-stage complex contagions. *Chaos* 23(1):013124
27. Melnik S, Porter MA, Mucha PJ, Gleeson JP (2014) Dynamics on modular networks with heterogeneous correlations. *Chaos* 24(2):023106
28. Mezard M, Montanari A (2009) *Information, Physics, and Computation*. Oxford University Press, Oxford
29. Motter AE, Yang Y (2017) The unfolding and control of network cascades. *Phys Today* 70(1):32–39
30. An undirected, unweighted network representing the topology of the Western States Power Grid of the United States. <http://www-personal.umich.edu/~mejn/netdata/power.zip>. Originally released with Ref. [45].
31. Newman MEJ (2002) Assortative mixing in networks. *Phys Rev Lett* 89(20):208701
32. Newman MEJ (2010) *Networks: An Introduction*. Oxford University Press, Oxford
33. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A (2015) Epidemic processes in complex networks. *Rev Mod Phys* 87(3):925
34. Payne JL, Dodds PS, Eppstein MJ (2009) Information cascades on degree-correlated random networks. *Phys Rev E* 80(2):026125
35. Porter MA, Gleeson JP (2016) *Dynamical Systems on Networks: A Tutorial*. Frontiers in Applied Dynamical Systems: Reviews and Tutorials, Vol. 4. Springer, Cham
36. Radicchi F (2015) Predicting percolation thresholds in networks. *Phys Rev E* 91(1):010801
37. Radicchi F, Castellano C (2015) Breaking of the site-bond percolation universality in networks. *Nat Commun* 6:10196
38. Radicchi F, Castellano C (2016) Leveraging percolation theory to single out influential spreaders in networks. *Phys Rev E* 93(6):062314
39. Rogers T (2015) Assessing node risk and vulnerability in epidemics on networks. *Europhys Lett* 109(2):28005
40. Shrestha M, Moore C (2014) Message-passing approach for threshold models of behavior in networks. *Phys Rev E* 89(2):022805
41. Strogatz SH (2015) *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Second Edition. CRC Press, Boca Raton

42. Traud AL, Kelsic ED, Mucha PJ, Porter MA (2011) Comparing community structure to characteristics in online collegiate social networks. *SIAM Rev* 53(3):526–543
43. Valente TW (1995) Network Models of the Diffusion of Innovations. Hampton Press, Cresskill
44. Watts DJ (2002) A simple model of global cascades on random networks. *Proc Natl Acad Sci U S A* 99(9):5766–5771
45. Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393(6684):440–442
46. Yağan O, Gligor V (2012) Analysis of complex contagions in random multiplex networks. *Phys Rev E* 86(3):036103
47. Zdeborová L, Krzakala F (2016) Statistical physics of inference: Thresholds and algorithms. *Adv Phys* 65(5):453–552

# Optimal Modularity in Complex Contagion



Azadeh Nematzadeh, Nathaniel Rodriguez, Alessandro Flammini,  
and Yong-Yeol Ahn

## 1 Introduction

The previous chapter reviewed the message-passing (MP) framework that can accurately describe the dynamics of spreading processes, and in particular that exhibited by the Watts threshold model [1–3]. In this chapter, we leverage the framework to study how complex contagions are affected by the modular structure of the underlying social network. In particular, we focus on the notion of *optimal modularity* that predicts the occurrence of global cascades when the network exhibits just the right amount of modularity [4].

Modular organization, or community structure, is one of the most ubiquitous properties of real-world networks [5, 6] and therefore it is crucial to understand how information diffusion is affected by a modular structure. Addressing this problem is particularly urgent when one considers spreading phenomena characterized by complex contagion. Unlike the case of simple contagion, where modules simply slow down the spreading, complex contagion may be either enhanced or hampered by modular structure [7, 8]. In contrast to simple, complex contagion requires multiple exposures and those are favored within densely connected communities. At the same time, complex contagion can be strongly hampered at the boundaries of communities due to the lack of the sufficient connectivity needed to provide the required multiple exposures from the activated community to the yet-to-be-activated one. The counter-intuitive phenomenon of optimal modularity arises from the clash and compromise between these opposite tendencies.

---

A. Nematzadeh · N. Rodriguez · A. Flammini · Y.-Y. Ahn (✉)  
Center for Complex Networks and Systems Research, School of Informatics and Computing,  
Indiana University, Bloomington, IN, USA  
e-mail: [azadnema@indiana.edu](mailto:azadnema@indiana.edu); [njrodrig@indiana.edu](mailto:njrodrig@indiana.edu); [aflammin@indiana.edu](mailto:aflammin@indiana.edu);  
[yyahn@indiana.edu](mailto:yyahn@indiana.edu)

The basic setting for our study is as follows. We assume a network of individuals where an individual can be in either an “active” or “inactive” state. At each time step, an inactive node may become active if the node is surrounded by *enough* active nodes. The activation condition is captured in a threshold function  $C(m, k)$  that typically depends on the degree  $k$  of a node, and the number  $m$  of its active neighbors. Here we consider  $C(m, k) = H(\frac{m}{k} - \theta)$ , where  $H(x)$  is a Heaviside step function and  $\theta$  is a threshold value. Throughout this chapter we assume that  $\theta$  is constant across the network. Our analysis leverages the framework introduced in the previous chapter. We focus our analysis only on the ensembles of random networks with arbitrary degree distribution [9], and “message-passing” (MP) and “Tree-Like” (TL) are used interchangeably throughout our chapter.

## 2 Analytical Framework

### 2.1 Mean-Field and Message-Passing Approaches for Configuration Model

As explained in the previous chapter, the steady-state fraction of active nodes  $\rho_\infty$  can be estimated using Mean-Field (MF) or the Message-Passing (MP) approaches. Assuming an underlying infinite networks with a given degree distribution  $p_k$  but otherwise random,  $\rho_\infty$  can be obtained by solving the following self-consistent equations. Using the MF approach,

$$\rho_\infty = \rho_0 + (1 - \rho_0) \sum_k p_k \sum_{m=0}^k \binom{k}{m} (\rho_\infty)^m (1 - \rho_\infty)^{k-m} C\left(\frac{m}{k}\right), \quad (1)$$

where  $\rho_0$  is the initial fraction of seeds. This approach does not aim at describing the evolution from one time step to the other, rather it states that at stationarity, the density of active nodes is the sum of two contributions: the fraction of seed nodes and expected number of nodes that have an above-the-threshold fraction of active neighbors. This last contribution, in turn, is expressed in terms of the degree distribution and of the density of active nodes itself.

The MP (TL) approach assumes that the underlying network is well approximated by a tree structure. To the extent to which such approximation is valid, where each node is affected only by its children. The density of active nodes depends only from the level of the tree where the node is, which is described by the following formula:

$$q_n = \rho_0 + (1 - \rho_0) \sum_k \frac{k}{\langle k \rangle} p_k \sum_{m=0}^{k-1} \binom{k-1}{m} (q_{n-1})^m (1 - q_{n-1})^{k-1-m} C\left(\frac{m}{k}\right), \quad (2)$$

where  $q_n$  is the density of active nodes at the  $n$ -th level of the tree ( $q_0 = \rho_0$ ). Note that *excess degree distribution* is used in the place of degree distribution because each node uses one of its links to connect to its parent and only children nodes affect the status of the node. The final density can be calculated by focusing on the root node:

$$\rho_\infty = \rho_0 + (1 - \rho_0) \sum_k p_k \sum_{m=0}^k \binom{k}{m} (q_\infty)^m (1 - q_\infty)^{k-m} C\left(\frac{m}{k}\right), \quad (3)$$

where  $q_\infty = \lim_{n \rightarrow \infty} q_n$ . See the previous chapter for more details.

## 2.2 Generalization to Modular Networks

The MP framework can be readily generalized to modular networks by introducing density-of-active variables for each community [3]. Consider a network with  $d$  communities, where the connection probabilities between communities are stored in a  $d \times d$  mixing matrix  $\mathbf{e}$ . Here  $e_{ij}$  is the probability that a random edge connects community  $i$  and  $j$ . Consider a node in community  $i$  and at the  $n+1$  level of the spreading tree. The probability to pick one of its active children ( $n$ -th level) can be written as:

$$\bar{q}_n^{(i)} = \frac{\sum_j e_{ij} q_n^{(j)}}{\sum_j e_{ij}}. \quad (4)$$

Equation (2) can be extended to describe the relation between the densities in different communities [3].

$$q_{n+1}^{(i)} = \rho_0^{(i)} + (1 - \rho_0^{(i)}) \sum_k \frac{k}{z^{(i)}} \sum_{m=0}^{k-1} \binom{k-1}{m} (\bar{q}_{n-1}^{(i)})^m (1 - \bar{q}_{n-1}^{(i)})^{k-1-m} C^{(i)}\left(\frac{m}{k}\right), \quad (5)$$

where  $z^{(i)} = \sum_k k p_k^{(i)}$  is the mean degree of community  $i$ . This set of equations can be solved by iteration analogously to Eq.(3). The density of active nodes at stationarity is:

$$\rho_\infty = \sum_i \frac{N^{(i)}}{N} \rho_\infty^{(i)}. \quad (6)$$

See [3] for more details.

### 3 Networks with Two Communities

Having set up the necessary tool, we now turn into investigating how the strength of the modular structure can affect the spreading of complex contagion. The simplest setting one may consider is a network with two equally sized communities. Given a fixed and predefined number  $L$  of links in the network, we first randomly connect  $\mu L$  couple of nodes, where each member of the couple sits in a different community. The remaining links are then used to randomly connect couple of nodes in the same community [5]. If  $\mu = 0$ , no edge is placed between the two communities (the network has two components and is therefore maximally modular); if  $\mu = 0.5$ , the network is an Erdős-Rényi random graph in the infinite size limit. A fraction  $\rho_0$  of active nodes are set in one of the two communities, which we call “seed community.”

As shown in [4] and illustrated in Fig. 1, the density of active nodes at stationarity  $\rho_\infty$  depends non-trivially on the degree of inter-community connectivity, showing a maximum at intermediate values of  $\mu$ .

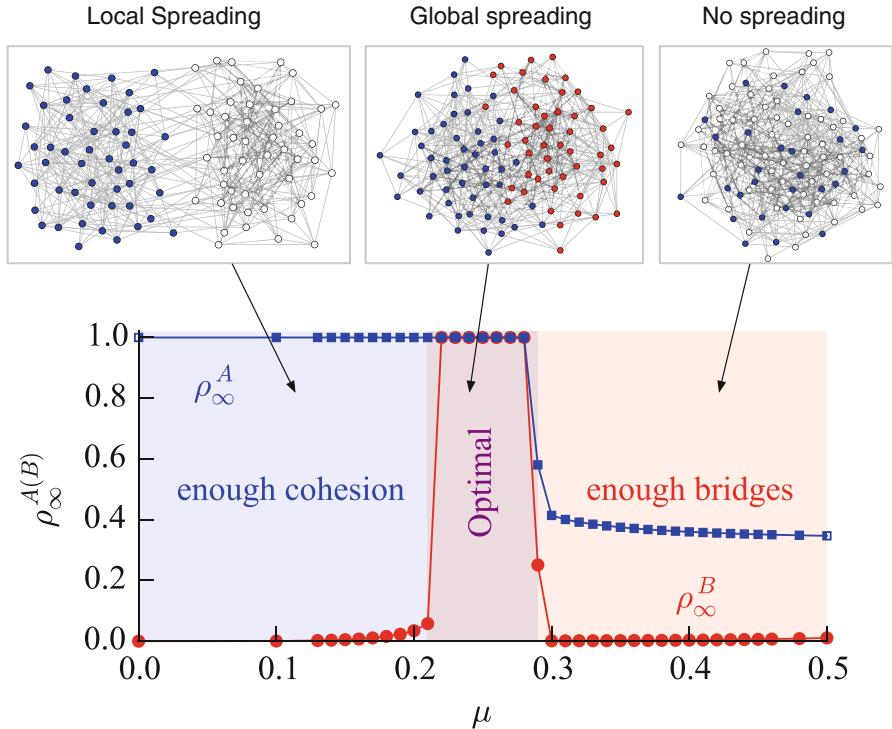
Small values of  $\mu$  allow initial spreading in the seed community, but it is essential to have enough mixing (bridges) between communities to have a cascade that significantly interests the global community. At the same time, when too many links across community are present, since these occur at the expense of the intra-community links, there is insufficient connectivity in the seed community to trigger the initial diffusion of the activation. We name *optimal modularity* the range of  $\mu$  values for which the two mechanisms above find their trade-off to maximize the size of the cascade.

### 4 Optimal Modularity in Networks with Many Communities

A network with just two equally sized communities with all the seed users concentrated in one of those is an obvious starting point for this study, but, in general, a non-realistic assumption. We generalize our finding by first considering multiple communities of the same size and with the same degree of intra and inter-connectivity. We then consider a more general process to generate the network and its modular structure. We consider the family of graphs known as LFR (Lancichinetti-Fortunato-Radicchi) benchmark graphs [10], which allow us to independently modulate both the size and the degree distribution of the individual communities. We finally remove the constraint of having all seed nodes in a single community.

#### 4.1 Spreading from a Seed Community

Figures 2 and 3 show the qualitative behavior of  $\rho_\infty$  as a function of  $\mu$  and  $\rho_0$ , when there are two or more communities and the activation is initiated from a single seed community. In general, a larger number of communities require a smaller

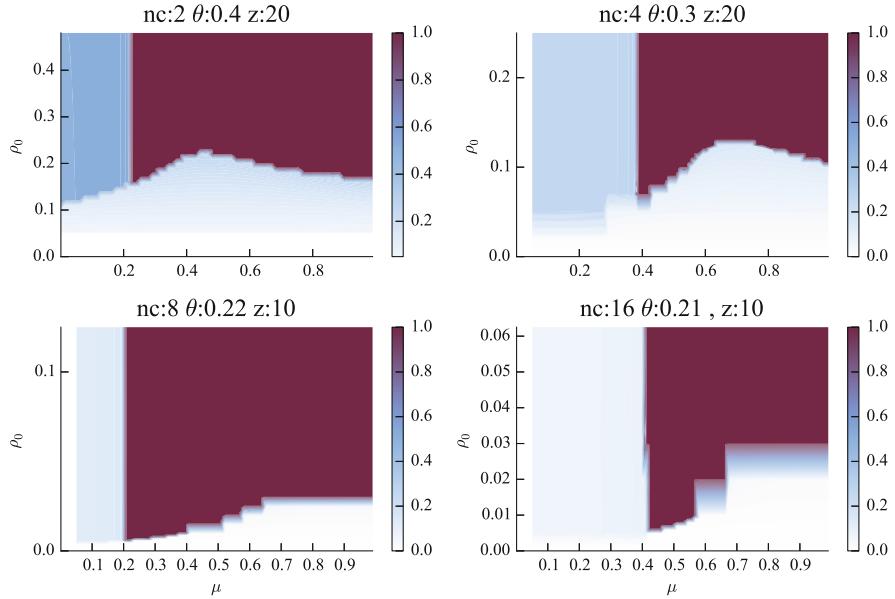


**Fig. 1** The trade-off between intra- and inter-community spreading. Stronger communities (small  $\mu$ ) facilitate spreading within the originating community while weak communities (large  $\mu$ ) provide bridges that allow spreading between communities. Blue and Red imply activation, while white implies inactivity. There is a range of  $\mu$  values that allow both (optimal). The blue squares represent  $\rho_\infty^A$ , the final density of active nodes in the community  $A$ , and the red circles represent  $\rho_\infty^B$ . The parameters for the simulation are:  $\rho_0 = 0.17$ ,  $\theta = 0.4$ ,  $N = 131,056$ , and  $\langle k \rangle = 20$

adoption threshold to allow the cascade to spread over all the network; increasing the number of communities makes the signal outgoing from the seed community less focused. Such signal, therefore, spreads less easily from community to community. Nevertheless, the same trade-off, and thus optimal modularity, exists between local spreading due to clustering and inter-community spreading due to bridges.

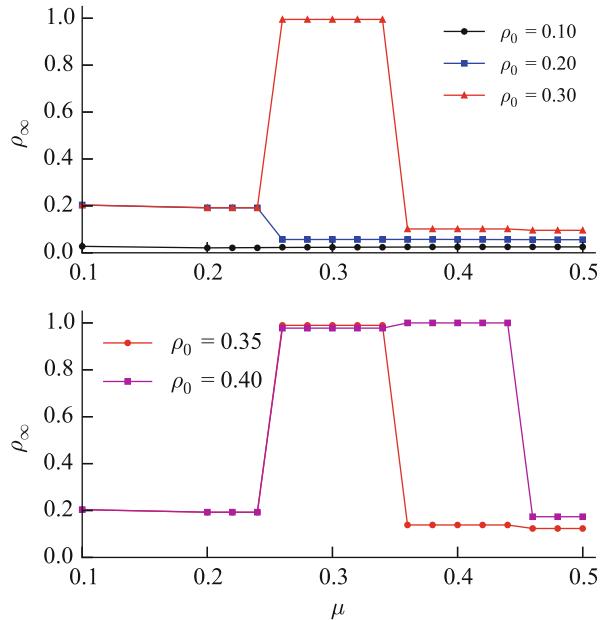
## 4.2 Spreading from Randomly Distributed Seeds

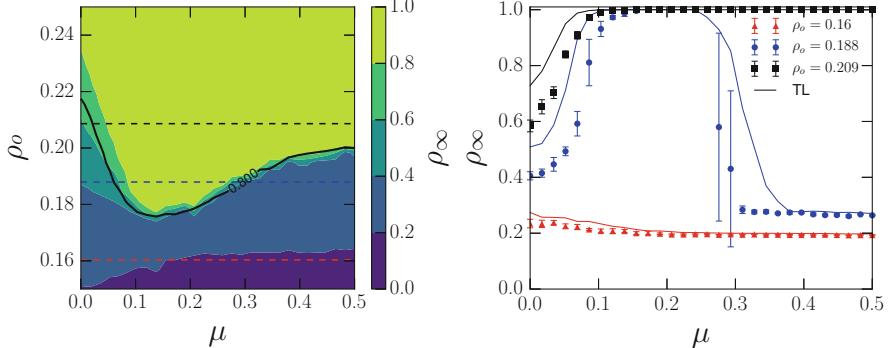
Next we consider the more general scenario in which the initial signal is distributed across the whole network. The MP approach as described by Eqs. (4) and (5) is sufficiently general to handle the scenario at hand. In particular we will have:



**Fig. 2** Optimal modularity arises even when there are many communities. The phase diagrams are calculated using the MP framework with different number of communities.“nc” refers the number of communities.  $\theta$  and  $z$  values are varied to demonstrate the existence of optimal modularity clearly

**Fig. 3** The behavior of threshold model in the presence of community structures generated by LFR benchmark, with  $N = 25,000$ ,  $z = 10$ ,  $t_1 = 2.5$  (degree exponent),  $t_2 = 1.5$  (community size exponent),  $k_{max} = 30$  and  $\theta = 0.3$ . LFR benchmark generates more *realistic* networks with community structures. The degree distribution may have a power-law distribution (with exponent  $t_1$  and degree cutoff  $k_{max}$ ). The size of the communities may also follow a power-law distribution (with exponent  $t_2$ )





**Fig. 4** The phase diagram of threshold model with uniformly distributed random seeds. Three example slices are taken from the contour plot (horizontal lines) and displayed in the right figure.  $N = 25,600$  with  $C = 160$  communities with 160 nodes each. The solid black line on the contour shows the MP (TL) results

$$\bar{q}_n^{(i)} = (1 - \mu)q_n^{(i)} + \frac{\mu}{(d - 1)} \sum_{j \neq i}^d q_n^{(j)}. \quad (7)$$

Here  $d$  is the number of communities and, as before,  $\mu$  represents the total fraction of inter-community bridges in the network. Also, in Eq. (5),  $\rho_0^{(i)} \neq 0$  for all  $i$ 's rather than just one. The equations can still be solved iteratively.

Figure 4 shows the results derived via the MP (TL) approach for a network with 25,600 nodes, 160 evenly sized communities, and seeds randomly distributed across the network. An optimal region emerges as in the previous multi-community cases.

We would like to note that the optimal region vanishes if each community has exactly same number of seeds; there is no dependence of  $\rho_\infty$  on  $\mu$ . The emergence of an optimal region critically depends on the existence of variability across communities. Individual communities show a sharp transition between inactivity and activity as the seed fraction  $\rho_o$  increases. As all nodes activate essentially simultaneously, the entire system can be regarded as a random network of supernodes, each representing a single community. The qualitative behavior is therefore the same as that of a random network with no communities. If there is sufficient variability across the communities, in terms of the number of seed nodes they contain then some communities will activate before others and the community structure will have a measurable effect, as shown in Fig. 4.

The effect of variability in the nodes' threshold was actually the focus of the original study of the linear threshold model by Granovetter [1]. He found that changes in the variance of the threshold distribution lead to qualitatively different spreading behavior even when the mean is kept constant. When variance is low, nodes have approximately the same threshold and, all other factors being the same, they get activated more or less simultaneously. Higher variance brings the existence

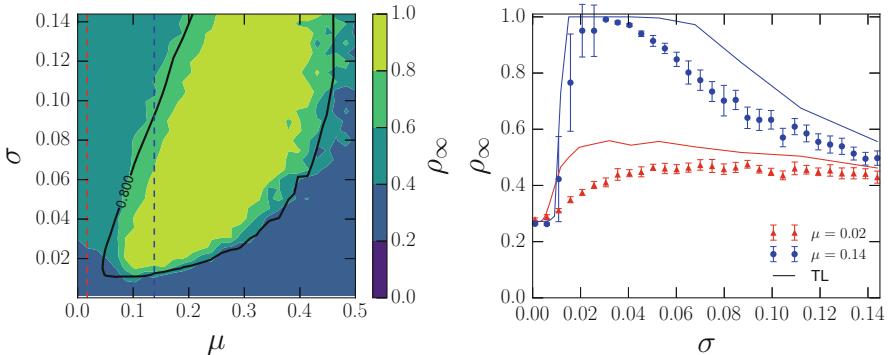
of a continuum spectrum from low to high threshold nodes. Low threshold are typically activated first and can help the activation of nodes with slightly higher threshold. In turn, these can contribute to activate even higher threshold nodes, and possibly generate a large size cascade. But if threshold variance is too high, the gap between low threshold and high threshold nodes is too large and the activation of the former is not sufficient to fill the gap in threshold.

Given the tendency of nodes in a community to activate simultaneously, it is possible to regard them as coarse-grained super-nodes whose threshold is effectively determined by the number of seeds they contain. This formulation provides similar insights as the Granovetter's study [1]. We investigated this idea by distributing seeds across communities according to a Beta distribution. We fix the  $\alpha$  and  $\beta$  parameters for the Beta distribution in such a way to maintain the expected value constant at  $\langle \rho_o \rangle = 0.19$  while the standard deviation ( $\sigma$ ) is varied.

Our experiments, as shown in Fig. 5 produce results qualitatively similar to those for Granovetter's model [1].

Specifically, Fig. 5 shows that two optimal behaviors emerge, one with respect to  $\sigma$  and one with respect to  $\mu$ . The peak along  $\sigma$  arises for exactly the reasons exposed above. A large cascade can be triggered for intermediate values of  $\sigma$ , when there is a continuum spectrum of effective activation thresholds across communities. Due to a cascading effect, increasing activity within the network makes it more likely to activate communities with fewer and fewer seeds. When  $\sigma$  is low communities have roughly the same number of seeds and none have enough seeds to fully activate unless the mean number of seeds is increased. At high  $\sigma$ , communities with many seeds activate, but they don't generate enough cumulative activity to activate the low seed communities.

The optimal region with respect to  $\mu$  arises for the same trade-off to those studied above. Low  $\mu$  implies strong connectivity inside single communities, but



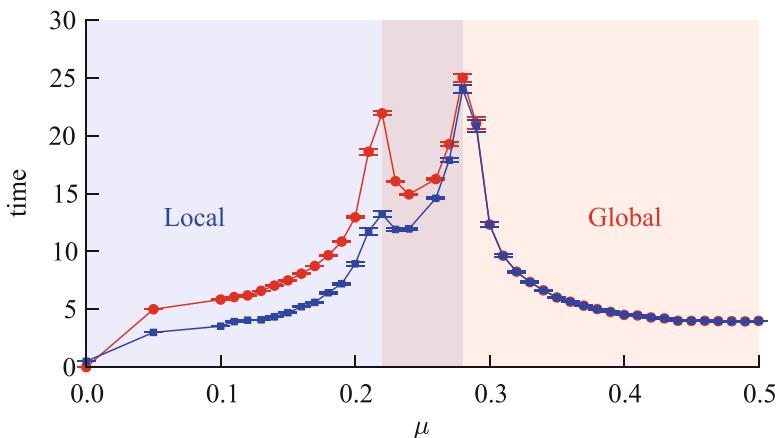
**Fig. 5** The phase diagram of threshold model with beta distributed random seeds. Two example slices are taken from the contour plot (vertical lines) and displayed in the right figure. The mean seed  $\langle \rho_o \rangle = 0.19$ . Numerical simulations were done with  $N = 25,600$  with  $C = 160$  communities with 160 nodes each. The solid black line on the contour shows the MP results

insufficient bridges to spread the activation signal externally. For high  $\mu$  there are bridges, but not sufficient internal connectivity in order to trigger the initial activation of a sufficiently large number of communities. An optimal balance is achieved at intermediate values of  $\mu$ .

## 5 Temporal Aspects of Optimal Modularity

*How fast* a contagion can spread is often as important as *how far* it can spread. Imagine, for example, the sudden availability of a prophylactic measure in the wake of a pandemic. The issue would then be not just whether this measure can spread broadly, but also whether it can spread sufficiently fast to effectively oppose the pandemic. Here we limit our study to the basic setting consisting of two communities with varying degree of modularity ( $\mu$ ) and only one seed community. We measure the total diffusion time: the number of time steps needed for the system to reach a steady state. We run a 1000 simulations (each with an independent network realization) and measure the mean  $\rho_\infty$  and total diffusion time. We also assume a uniform threshold ( $\theta = 0.4$ ).

Figure 6 demonstrates that, while  $\rho_\infty$  remains constant at its maximum value, the total diffusion time greatly varies. Close to either border of the optimal range, contagion significantly slows down, while the global spreading can happen fastest



**Fig. 6** Total diffusion time and optimal modularity. The blue symbols and line represent the total diffusion time in the community  $A$  (seed community), and the red symbols and line represent the total diffusion time in the community  $B$  (the other community). The optimal modularity range that allows global cascades is represented with a purple shade. The total diffusion time curve peaks at the two transition points, demonstrating that there exists a narrower range of  $\mu$  values where the global cascades happen, faster. The parameters for the simulation are:  $\rho_0 = 0.17$ ,  $\theta = 0.4$ ,  $N = 8192$ , and  $z = 20$

near the middle of the optimal modularity regime. When there are just enough bridges (the left border), the spreading from the seed community to the other community is slower than the case where there are more than just enough bridges to spare (center). Similarly, when there are just enough local cohesion (the right border), the local spreading produces just enough newly activated nodes to achieve global cascade, slowing down the spreading process.

## 6 Discussion

In this chapter, we have generalized the optimal modularity phenomena and studied its temporal aspect. We showed that many simplifying assumption made in the original study can be relaxed without disrupting the qualitative scenario that predicts a maximum in the fraction of active individuals for intermediate values of inter-community connectivity. In particular we considered the case of a large number of communities, with heterogeneous size, and nonuniform degree of initial activation.

Our experiment showed that our model behaves qualitatively same as one in which communities can be considered as super-nodes and are characterized by different threshold. This, in turn, may open the possibility to study very large system if one could devise a strategy to compute the effective parameters of a coarse grained model where communities are represented by single nodes. The interest in developing such “renormalization” techniques is not only theoretical: threshold models have found several applications to real-world problems, including the multi-scale modeling of brain networks [11] and of their activation dynamics in the brain [12].

## References

1. Granovetter M (1978) Threshold models of collective behavior. *Am J Sociol* 83(6):1420–1443
2. Watts DJ (2002) A simple model of global cascades on random networks. *Proc Natl Acad Sci* 99(9):5766
3. Gleeson JP (2008) Cascades on correlated and modular random networks. *Phys Rev E* 77(4):046117
4. Nematzadeh A, Ferrara E, Flammini A, Ahn Y-Y (2014) Optimal network modularity for information diffusion. *Phys Rev Lett* 113(8):088701
5. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proc Natl Acad Sci* 99(12):7821
6. Newman MEJ (2006) Modularity and community structure in networks. *Proc Natl Acad Sci* 103(23):8577–8582
7. Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194–1197
8. Weng L, Menczer F, Ahn Y-Y (2013) Virality prediction and community structure in social networks. *Sci Rep* 3:2522
9. Newman MEJ, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. *Phys Rev E* 64(2):026118

10. Lancichinetti A, Fortunato S, Radicchi F (2008) Benchmark graphs for testing community detection algorithms. *Phys Rev E* 78:046110
11. Sporns O, Chialvo DR, Kaiser M, Hilgetag CC (2004) Organization, development and function of complex brain networks. *Trends Cogn Sci* 8(9):418–425
12. Wang S-J, Hilgetag CC, Zhou C (2011) Sustained activity in hierarchical modular neural networks: self-organized criticality and oscillations. *Front Comput Neurosci* 5(30):1–14

# Probing Empirical Contact Networks by Simulation of Spreading Dynamics



Petter Holme

## 1 Introduction

Spreading processes<sup>1</sup> affect people at many levels. They are the basis of innovation processes [9] and spreading [1], they shape our opinions [11], lifestyles [14], and they are also part of the mechanisms giving us infectious disease [2, 25, 28]. Even though disease spreading is a bit special and not the primary topic of this book, much of the theory of the interaction of spreading and contact structure comes from epidemiology of infectious diseases. For this reason, we will mostly use disease spreading as our model spreading dynamics and leave it to the reader to draw the analogies to other phenomena.

At the time of writing “data science” is a buzzword. One key idea behind it is that we can understand much of the social world around us by analyzing the data we create—be it from the location traces of our smart-phones [90], transportation cards [78], etc. A special type of such data records contacts between pairs of people. By contact we will mean any kind of binary interaction where something can spread from one person to the other. It could be being in physical proximity, so that disease could spread, or following someone on a social media channel, so that information could spread. A contact network could thus be thought of as a list of pairs of individuals, annotated with the time, type, and locations of the interaction. In practice one usually does not have access to so much meta-information on

---

<sup>1</sup>In the social science and computer science literature these are commonly known as *diffusion* processes. In this chapter, we stick to the natural science convention (keeping “diffusion” for processes where the total mass or amount of whatever is spreading, or diffusing, is conserved).

P. Holme (✉)

Institute of Innovative Research, Tokyo Institute of Technology, Tokyo, Japan

e-mail: [holme@cns.pi.titech.ac.jp](mailto:holme@cns.pi.titech.ac.jp)

the contacts. In many cases one must settle with only the time and identities of the individuals (a temporal network [31, 35, 56]), or even just the identities (a static network) [6, 61]. On the other hand, several properties of the spreading are determined by the (temporal or static) network structure—the regularities making the network differing from a purely random one—and our understanding of how they shape the spreading dynamics is still incomplete.

One challenge for understanding how the structures of contact networks affect spreading is to be able to list and quantify the relevant structures. Structures are dependent, however, and this makes it a challenge even in the simplest case of static networks. Acquaintance networks are, for example, thought to contain many triangles [27]. The presence of triangles is thought to slow down spreading [85]. However, this does not necessarily mean that human friendships slow down spreading. In a simple network model where the density of triangles (a.k.a. clustering coefficient) is the only structure to control, a typical network would have one very densely connected core contributing with most of the triangles, not triangles distributed all over the network as empirical friendship networks have [12]. Thus, there are other constraints, or structures, present in the real networks that could potentially also affect spreading phenomena. An alternative approach to controlling the density of triangles would be to take empirical networks as the starting point and simulating disease spreading directly on these. To monitor the effect of triangles one could manipulate the original network, for example by randomly rewiring links [59]. Of course, one cannot isolate network structures completely—by rewiring the network, one could presumably change, e.g., the average path length as well. However, one would do that from a realistic part of the space of (temporal) networks. In addition, this approach gives insights about how the network itself acts as an infrastructure for the spreading process.

The problem of straightforward approaches to understanding the effects of network structure in real spreading processes becomes more severe the more information-rich network representation one uses. For temporal networks—where information about the time of contacts are included—this is particularly clear. It has been observed that human behavior has an intermittent, bursty behavior [5, 26]. Subsequently, authors noticed that fat-tailed interevent time distributions—the hallmark of bursty activity—slow down spreading [43, 60]. At the same time, other authors observed that simulated disease spreading was slowed down by randomizing the timing of contacts in some kinds of contact networks [70]. There must thus be other temporal structures also controlling disease spreading. In this work, we explore methods to understand the relationship between network structure and spreading dynamics that take empirical networks, rather than network models, as starting point.

In the remainder of this chapter, we will go through some of the empirical networks authors have used, typical models of spreading processes for the purpose, randomization methods, and similar techniques. Finally, we will discuss future prospects and relation to other methods.

## 2 Networks

In this section, we will discuss the empirical data available at the moment and some technical issues related to how to represent it mathematically or computationally.

### 2.1 Data Sources

#### 2.1.1 Proximity Networks

Human *proximity networks* have gained much attention recently. Such data sets records when, and sometimes where, persons are in contact. At least they contain identities of the people in contact and when the contacts happen. Typically these data sets sample people connected by some circumstance—workers in the same office [24, 79], at the same hospital [38, 52, 83], students in the same school [55, 72, 75, 91], visitors to an art gallery [82], conference attendants [39], etc. The time limits are typically set by the experiment and in most cases running throughout 1 day (when the school or office is open).

Researchers have been very creative in gathering proximity networks. One common method is to equip the participants with radio-frequency identification (RFID) sensors [7] which records proximity of a couple of meters. Notably, the organization Sociopatterns ([sociopatterns.org](http://sociopatterns.org)) provide many open access datasets. A similar performance to RFID sensors can be obtained by infrared [79] or wireless [65, 72] sensors. Another type of proximity measure is to use the Bluetooth channel of smart-phones. These typically record slightly more distant contacts (the order of 10 meters). Bluetooth-based studies typically run longer and are less constrained [18, 76, 77].

In addition to proximity recorded by sensors, researchers have used location information to infer who is close to whom at what time. Ref. [78] studies people sharing the same public transport; Ref. [91] uses a dataset of people connected to the same WiFi router. There is also a rather large field of studying patient flow within hospital systems, e.g. Refs. [16, 17, 52, 88] from the records of patients and healthcare workers. A contact in such networks corresponds to two persons being at the same ward at the same time.

Yet another kind of human proximity networks (perhaps different enough to constitute a stand-alone category) is sexual networks. Classic sexual network studies do not have the time of the contacts. The only large-scale temporal network of sexual contacts we are aware of is the prostitution data of Rocha et al. [69] where contacts with sex sellers are self-reported by the sex buyers at a web community.

Finally, although this book focuses on humans, we mention that proximity networks of animals have been studied fairly well. In particular, populations of livestock have been studied either as metapopulation networks (where one farm is one node and an animal transport between two farms is a contact) or as a temporal

network of individual animals where a contact represents being at the same farm at the same time. Livestock here could refer to either cattle [23, 74, 81] or swine [47]. In addition domesticated animals, researchers have also studied wild animals—zebras [50] and monkeys [15] by GPS traces, ants [13] by visual observation, and birds [64] from foraging records.

In the latter part of this chapter we will use some data set of this type as an example. In particular, we use several Sociopatterns data sets: *Conference* (participants of a computer science conference), *Hospital* (patients, doctors, and nurses of a hospital), *Office* (workers at the same office), *Primary* and *High School* (school students), *Gallery* (visitors to an art gallery), and families in rural Kenya (*Kenya*). Some of these data sets cover several days, which we treat separately. We also use one Bluetooth data set sampled among college students in USA (*Reality*) and a similar dataset from Romania (*Romania*) sampled with WiFi technology. Finally, we use one dataset based on a diary-style survey (*Diary*) and one from self-reported sexual contacts with escorts (*Prostitution*). Statistics and references to these data sets can be found in Table 1.

### 2.1.2 Communication Networks

Temporal networks of human communication are probably the largest class of systems modeled as temporal networks after proximity networks. One such type of data comes from call-data records of mobile phone operators [43, 48, 49]. These use lists who called whom, or who texted whom. Typically the data sets are restricted to one operator in one country. Another type of communication networks are e-mails sampled from the accounts of a group of people during a window of time [19, 20]. Yet another of this kind comes from messages at social media platforms such as Twitter [71, 73] or Internet communities [37, 40, 42, 57, 86]. A difference to proximity networks is that links in this category are naturally directed. (Later in this chapter, when we will compare networks of this kind to proximity networks and then treat contacts as undirected.)

Below we analyze one data set of wall-posts at Facebook (*Facebook*), one Facebook-like community for college students (*College*), one Internet dating service (*Dating*), and one film-discussion community (*Forum* for posts at a discussion forum, and *Messages* for direct, e-mail-like communication). We also study three data sets of e-mail communication *E-mail 1*, *2*, and *3*. A summary of these data sets and references can be found in Table 1.

## 2.2 Network Representations

The basic setting we are considering is a set  $V$  of  $N$  nodes (sometimes called *vertices*). For most purposes of this chapter, the nodes represent individual people. In a static network, or *graph* (emphasizing the mathematical representation rather than

**Table 1** Basic statistics of the empirical temporal networks

Data set	<i>N</i>	<i>C</i>	<i>T</i>	$\Delta t$	<i>M</i>	Ref.
<i>Conference</i>	113	20,818	2.50 day	20 s	2196	[39]
<i>Hospital</i>	75	32,424	96.5 h	20 s	1139	[83]
<i>Office</i>	92	9827	11.4 day	20 s	755	[24]
<i>Primary School 1</i>	236	60,623	8.64 h	20 s	5901	[75]
<i>Primary School 2</i>	238	65,150	8.58 h	20 s	5541	[75]
<i>High School 1</i>	312	28,780	4.99 h	20 s	2242	[55]
<i>High School 2</i>	310	47,338	8.99 h	20 s	2573	[55]
<i>High School 3</i>	303	40,174	8.99 h	20 s	2161	[55]
<i>High School 4</i>	295	37,279	8.99 h	20 s	2162	[55]
<i>High School 5</i>	299	34,937	8.99 h	20 s	2075	[55]
<i>Gallery 1</i>	200	5943	7.80 h	20 s	714	[82]
<i>Gallery 2</i>	204	6709	8.05 h	20 s	739	[82]
<i>Gallery 3</i>	186	5691	7.39 h	20 s	615	[82]
<i>Gallery 4</i>	211	7409	8.01 h	20 s	563	[82]
<i>Gallery 5</i>	215	7634	5.61 h	20 s	967	[82]
<i>Reality</i>	64	26,260	8.63 h	5 s	722	[18]
<i>Romania</i>	42	1,748,401	62.8 day	1 month	256	[65]
<i>Kenya</i>	52	2070	61 h	1 h	86	[45]
<i>Diary</i>	49	2143	418 day	1 day	345	[66]
<i>Prostitution</i>	16,730	50,632	6.00 year	1 day	39,044	[69]
<i>WiFi</i>	18,719	9,094,619	83.7 day	5 month	884,800	[91]
<i>Facebook</i>	45,813	855,542	1,561 day	1 s	183,412	[87]
<i>College</i>	1899	59,835	193 day	1 s	13,838	[62]
<i>Messages</i>	35,624	489,653	3018 day	1 s	94,768	[42]
<i>Forum</i>	7084	1,429,573	3141 day	1 s	138,144	[42]
<i>Dating</i>	29,341	529,890	512 day	1 s	115,684	[37]
<i>E-mail 1</i>	57,194	444,160	112 day	1 s	92,442	[19]
<i>E-mail 2</i>	3188	309,125	81 day	1 s	31,857	[20]
<i>E-mail 3</i>	986	332,334	526 day	1 s	16,064	[63]

*N* is the number of individuals; *C* is the number of contacts; *T* is the total sampling time;  $\Delta t$  is the time resolution of the data set, and *M* is the number of links in the projected static networks. One data set (*Romania*) was coarse-grained from second to minute resolution (we consider a pair with at least one contact (in the raw data) within a minute a contact)

the real system)  $G(V, E)$ , the nodes are connected pairwise by *M* links (sometimes called *edges*) *E*. In a temporal network the nodes are connected at specific times by *C* contacts (sometimes called *events*)—triples  $(i, j, t)$  showing that *i* interacted with *j* at time *t*. An alternative way of thinking about how time enters networks is to consider nodes as a sequence of static graphs  $\{G_t(V_t, E_t)\}_{t=1}^T$ , one for every discrete time step of the data. *T* is called the *sampling time*. This type of graph sequence is a special case of *multilayer* networks [10, 46]. Mathematically it is equivalent to sequences of contacts, but it does put other ideas into the mind of the

user. To be specific, thinking of the system as a sequence of graphs suggests that one can first apply static network theory to each time slice individually, then aggregate these results. This could be a powerful approach in many cases, but not if the time resolution is so high that the networks are mostly very fragmented (or perhaps even empty, as the case in, e.g., an e-mail network). Since paths in temporal networks need to follow the arrow of time, they are not transitive—the pairs  $(i, i')$  and  $(i', i'')$  can have contacts without  $i$  being able to influence  $i''$ . The reason is that all contacts between  $(i', i'')$  might have happened by the time the spreading has reached  $i''$ .

Many studies consider spreading processes in space. This is true not only for disease spreading [21], but the study of spreading of innovation (Ref. [58] is an important early such reference). In principle, space can be encoded into the contacts of a network. On the other hand, in cases one is not aware of the detailed contact structure, one can resort to spatial spreading models. Spatial information can be combined with a network representation [8] and such mixed approaches are efficient in modeling multi-scale human mobility patterns (and thus contact patterns) [4].

### 3 Spreading Dynamics

In this section we, discuss models of spreading phenomena that can be simulated on empirical contact sequences. It is not a complete review of the matter, but intended to make the latter discussion more concrete.

#### 3.1 Epidemic Spreading

The framework for modeling the spread of infections in a population is well established [2, 25]. The so-called *compartmental models* divide the population into states (classes, or compartments) with respect to the disease, and prescribe transition rules between these classes. The four most common states are: *susceptible* (S, individuals that can get the disease, but not spread it), *infectious* (I, who can spread the disease), *recovered* (R, who can neither get nor spread the disease), and *exposed* (E, who got the disease but can yet not infect others). The infection event typically happens between a susceptible and infectious individual. It is the only transition that requires two people to meet. Two canonical compartmental models are the SIR and SIS model. In SIR an S person can become I upon meeting an S, and an I will eventually become R. In SIS, I becomes S rather than R. There are some subtleties involved in how to implement the transitions. Mathematical epidemiology has traditionally implemented the transition from I (to R in the SIR model or S in the SIS model) as happening with a constant rate. In other words, all infected persons have the same chance of becoming uninfected every unit of time. This—*constant infection rate* (CIR) version—leads to an exponential distribution of the infection time, which is in contrast to observations [84], but simplifies the

calculations. An other approach—the *constant infection duration* (CID) version—is to model the duration of the infection as constant. As all infected nodes expose their neighbors the same amount of time, this simplifies some statistical analyses. It is also somewhat algorithmically more straightforward (but this should not be a ground for selecting the algorithm). In either of these cases, the SIR and SIS models have two parameter values. One controlling how easily a node gets infected. Another controlling how long the node stays infected. For the CIR version these are the infection and recovery rates. For the CID version they are the per-contact infection probability  $\lambda$  and disease duration  $\delta$ .

The second ingredient in epidemic modeling is a model or data of the contact patterns. In most approaches this part comes from a simple model—the simplest being that everyone has the same chance of meeting everyone else at every time, but, in particular, with the advent of network epidemiology [44], researchers have started to study more realistic contact patterns. One approach is to construct models of human contact patterns. For example, based on the observation that sexual networks have a power-law degree distribution, researchers have studied the transmission of sexually transmitted infections on model networks with such a degree distribution [51]. Another approach for increased realism in disease spreading studies is to simulate the spreading on data sets of empirical contacts [32, 34, 66, 70, 72]. As mentioned, this approach gives more than just better predictions—we can also use it to understand what structures of the contact sequence is important, and why.

### 3.2 Opinion and Information Spreading

Much of the previous section is true for modeling information and opinion spreading too. The main difference is that one cannot assume that such spreading is well-modeled by compartmental models. We have learned from studies of spreading in social media that individual behavior is very diverse and platform dependent. Not only do people have different activity levels, they could also follow completely different mechanisms [54, 71]. Sometimes authors make the distinction between *simple* and *complex contagion* [89]. The former are all types of spreading phenomena where the spreading can be modeled as a probabilistic event when an S meets an I, independent of the rest of the system. Complex contagion, on the other hand, can depend on more than a pairwise interaction: an opinion might need exposure from several different neighbors to spread from one vertex to another; a piece of information might spread slower with age; it could spread more easily between people that are similar to one another, etc. [89].

The simplest type of complex contagion are threshold models.<sup>2</sup> These assume that an individual adopts an idea when the exposure is over a threshold. What

---

<sup>2</sup>An even simpler type of opinion spreading model is the *voter model* [22]. This is a simple contagion model where random nodes copy the opinion of random neighbors.

“exposure” means is not trivial. It could be the number of different persons that one hear an opinion from; it could also be the number of times one has heard the opinion [3]. Furthermore, for temporal networks, old exposures are not as important as more recent ones [42, 80]. Authors have modeled this by counting exposures in a time window into the past [41] or assigning every contact with an exponentially decreasing importance metric [80].

## 4 Null Models, Randomizations, and Positional Comparisons

So far, we have discussed the kind of datasets available and different dynamic models of spreading phenomena. While running such simulations on the raw contact data can be interesting in its own right, it can be hard to generalize the results. As mentioned in the Introduction, one option is to compare the results to those expected from models. This approach has been a fruitful way for static networks but is challenging for more information-rich representations of the contact patterns. One reason is that it is hard to even name a reasonably complete set of simple structures in temporal networks (it is of course even harder to control them in a way such that the results are easy to interpret). An alternative approach is to draw the conclusions from comparisons. One way is to randomize some aspect of the real data and thereby destroy some particular structure. By comparing the spreading on the original and randomized networks, one can draw conclusions about the effects of the randomized structures. By successively randomizing less and less one can, in principle, home into the important structures. One may argue that this approach is only replacing the problem of listing fundamental structures, by the problem of listing structures to randomize. However, one is certain that going from the original data to the fully randomized data, one has removed all structure there is, and thus all structure that can play a role in the spreading (even though this procedure should be coarser than ideal).

References [30, 35] present several methods of randomization. In this chapter, we will exemplify with two: *Random times* (RT) and *Random links* (RL). For RT one replaces the timestamps of contacts with random times in the interval  $[0, T]$ , thus destroying several types of temporal structures including effects of: order of events, periodic changes in the overall activity, the turnover of individuals, etc. RT is thus a quite pervasive type of randomization, only conserving the number of contacts and the static network structure. One can regard RL as a topological counterpart to RT. For RT one replaces link by a link between two random nodes in the network. Thus one destroys all the topological structure, including the degree distribution.<sup>3</sup> With the results for the original and randomized networks at hand one

---

<sup>3</sup>Effectively one replaces the degree sequence by one drawn from a binomial distribution. For many applications one is rather interested in the topological structure other than the degree distribution, and would rather conserve the degree of the nodes [59], but to be able to compare the topological randomization to the temporal one, we use this definition.

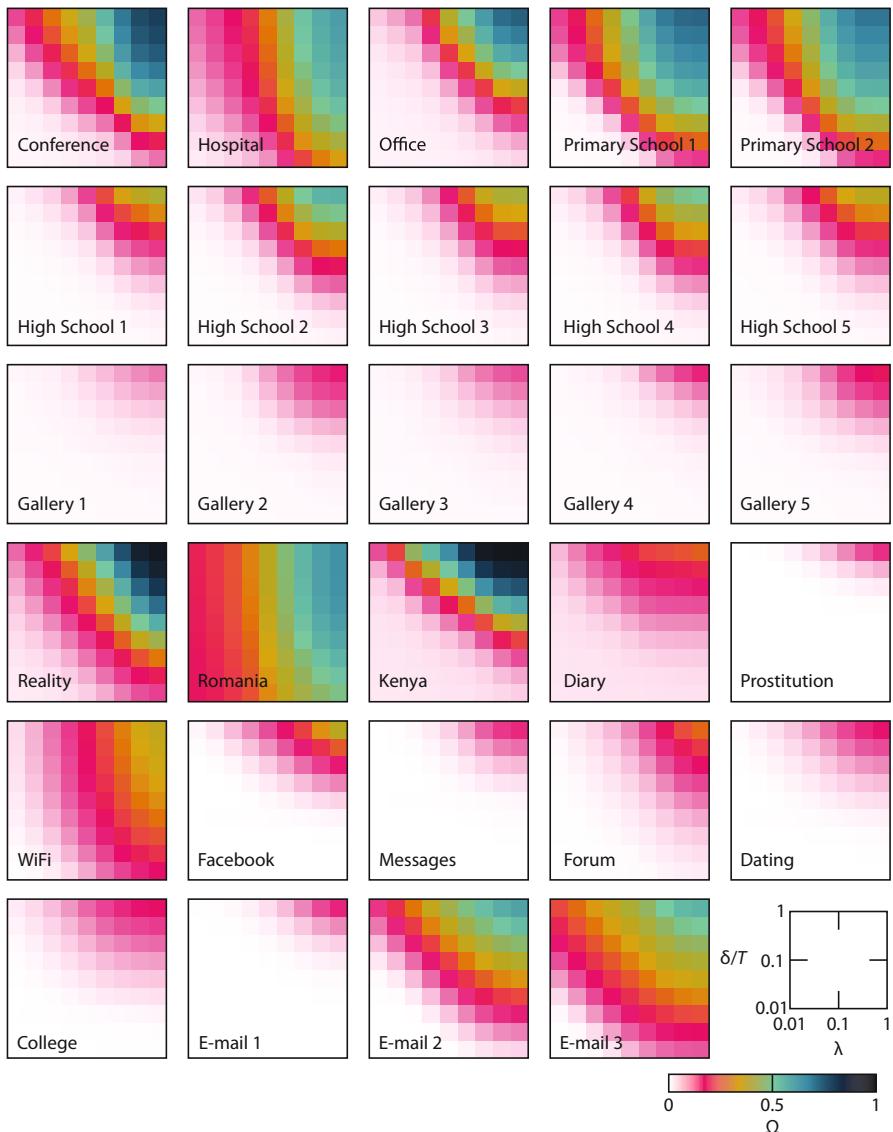
can see how the destroyed structure affects the spreading. This effect would typically depend on the parameter values of the spreading dynamics. To get interpretable results one typically averages over many randomized data sets. The good news is that temporal networks are typically “self-averaging” in the sense that fluctuations decrease with systems size. To move further into describing how the contact structure affects the spreading one can also include measurements of (temporal or static network structure). For example, Ref. [34] compares the discrepancies between two estimates of disease severity for different contact data sets. They correlate the discrepancies with measures—network descriptors—like the node and link burstiness [26], the fraction of nodes and links present throughout the sampling time, etc. From this analysis they can conclude that some types of discrepancies are more related to temporal structures, other to topological structures.

In addition to randomizing structure, one can learn about the structure of the contact network by comparing nodes and links within the same network. One can for example, compare spreading starting at different nodes and compare the average outbreak size, time to peak prevalence or time to extinction [29, 68]. Another approach would be to eliminate single nodes and links and study the changes of the mentioned quantities.

## 5 Example: SIR Model on Empirical Networks

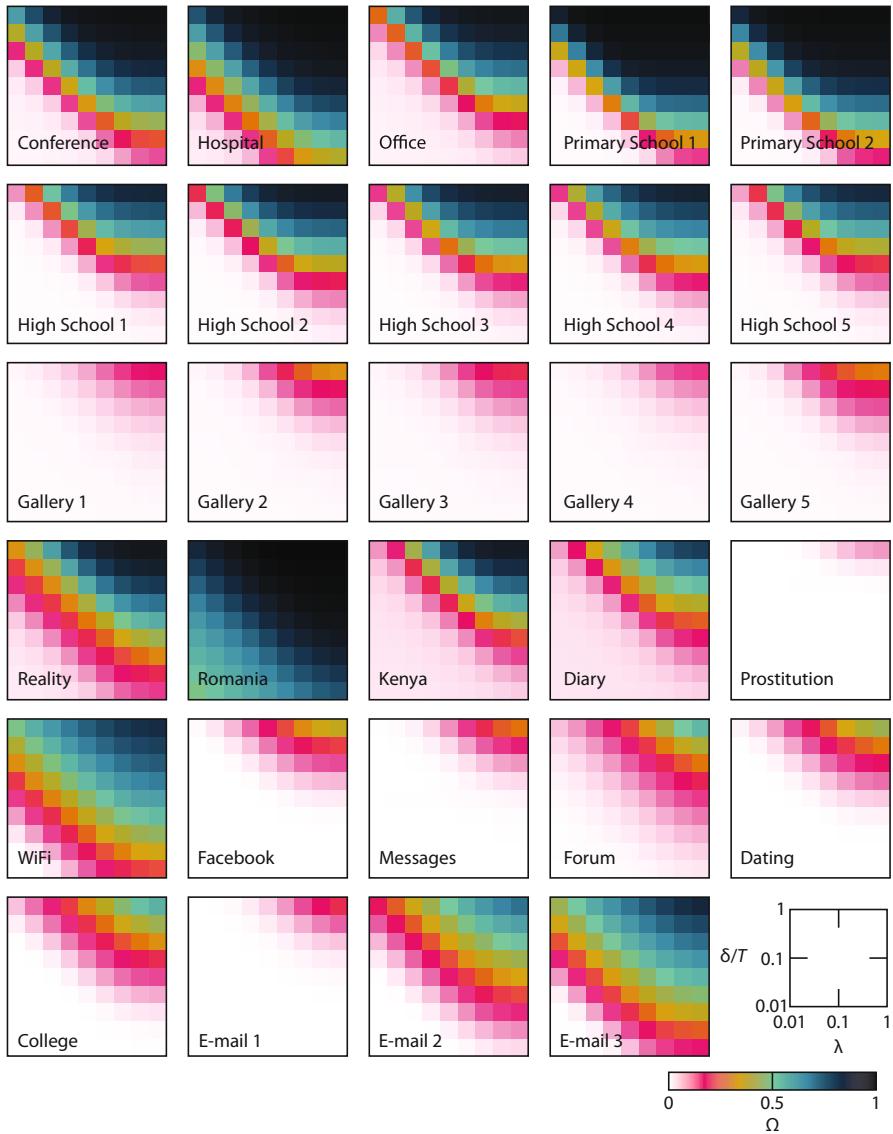
In this section, we will present an analysis along the lines outlined above for the 29 contact networks of Table 1. In Fig. 1, we show the average outbreak size  $\Omega$  (the fraction of recovered nodes at the end of the outbreak) in an SIR simulation. We use the CID version, so the two parameter values are the per-contact transmission probability  $\lambda$  and the disease duration  $\delta$ . The infection is started at one randomly chosen node at a random time between 0 and  $T$ . All data points are averaged over at least  $10^3$  outbreak runs per networks. In Figs. 2 and 3 we show plots of  $\Omega$  as a function of  $\lambda$  and  $\delta$  for the RT and RL randomizations, respectively. For these figures, we also average each value over 100 randomizations.

A first thing to notice in Fig. 1 is that  $\Omega$  is increasing with both  $\lambda$  and  $\delta$ . For some networks,  $\Omega$  reaches its maximal value 1, but for most it does not. For the *Gallery* data, *Prostitution* and the social media networks (*Facebook*, *Messages*, *Forum*, *Dating*, and *College*) there is a big overturn of agents—the individuals that are there in the beginning are not there in the end. (This is easy to imagine for the *Gallery* networks as a visitor to an art gallery would stay for a limited amount of time.) A small maximal  $\Omega$  can most easily be explained by the turnover of agents breaking many time-respecting paths, thus cutting many infection chains. Another explanation would be that most contacts happen at the beginning, so that by chance the network is very fragmented by the time a typical first infection event happens. Some of the networks are so dense that even for the lowest parameter values



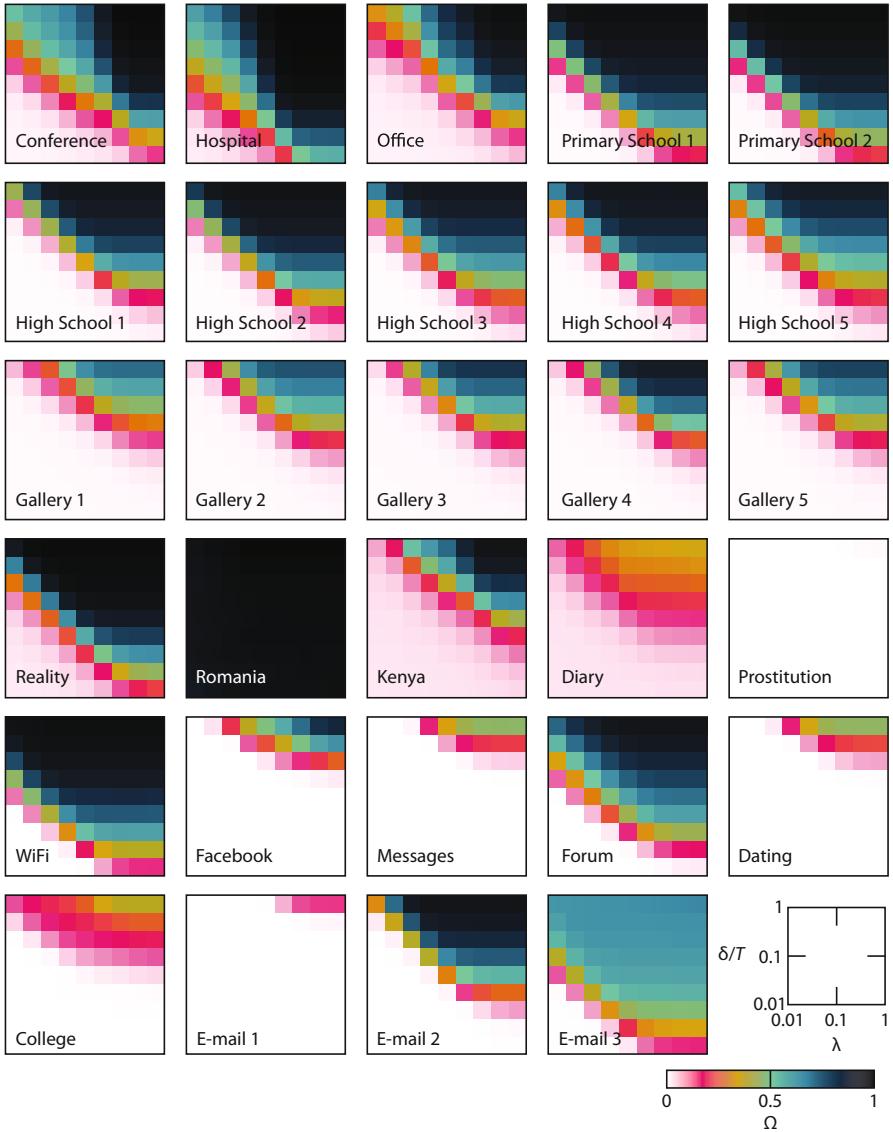
**Fig. 1** The average outbreak size for SIR epidemics on our data sets. The scales of the axes and colors are the same for all panels (as indicated in the legend)

( $\lambda = \delta/T = 0.01$ )  $\Omega$  is quite large. The most conspicuous example is perhaps *Romania* where both the minimum and maximum  $\Omega$  values are intermediate. At this point, it is worth noting that  $\lambda$  (unlike  $\delta$ ) should not be understood as a parameter that is unique for one disease. It must be defined in combination with the network



**Fig. 2** The average outbreak size for SIR epidemics data sets where the time stamps of contacts are replaced by random ones. Otherwise, the figure is the same as Fig. 1

representation—a more restrictive definition of a contact would correspond to a larger  $\lambda$  value [77]. Finally, we note that the data sets that come from the same setup (the five *High School* and *Gallery* networks look like each other—an indication that the used methods are not sensitive to occasional misinformation).



**Fig. 3** The average outbreak size for SIR epidemics data sets where the node identities are replaced by random ones. Otherwise, the figure is the same as Fig. 1

Next, we turn to the data with randomized time stamps.  $\Omega$  as a function of  $\lambda$  and  $\delta$  is displayed in Fig. 2. The effect of the randomization is different for different data sets. For several data sets  $\Omega$  increases, at least the maximal  $\Omega$ , or sometimes  $\Omega$  throughout the parameter space. The main exception is *Prostitution* where the maximal  $\Omega$  decreases upon randomization. Some other data sets—the *Gallery* data

and *E-mail 1* do not change. From this we understand that given the underlying contact network, and the number of contacts between each pair of nodes, the timing of the nodes can both speed up and slow down the disease spreading. Since bursty behavior is known to slow down spreading [43, 60] and the RT randomization makes contacts less bursty, we can understand that other temporal factors also determine the speed and scope of the spreading. We can also notice that the effects of RT randomization is largest for intermediate values of  $\Omega$  (of Fig. 1).

The outbreak sizes corresponding to Fig. 1 for topologically randomized datasets is shown in Fig. 3. The pattern from the RT plots of Fig. 2 remains, and is somewhat accentuated. For the RL randomized plots,  $\Omega$  reaches close to its maximum value  $\Omega = 1$  for most of the data sets (including *Gallery*, where this was not true for the RT randomization). For some other datasets—*Prostitution*, *Diary*, *E-mail 1*, and *E-mail 3*—the maximal  $\Omega$  decreases going from RT to RL. In summary, most datasets we have investigated have both temporal and topological structures that decrease the outbreak sizes. Why the opposite occasionally happens is an interesting and—at the moment of writing—not fully resolved problem. One observation is that all such data sets are fairly sparse in the sense that  $C/M$  is small, but other structures can separate these data sets from others even more clearly. One such structure is the average duration of a link [33]. Other quantities that describe the long-term evolution of the system could also work well.

## 6 Discussion

In this chapter, we have pointed at some ways one can analyze empirical datasets of human interaction by simulating spreading phenomena on top of them. We have discussed data sources, network representations, and some of the analysis techniques including some of the many randomization-based null-models available. The type of analysis we have outlined is not from a pipeline to handle contact data. Indeed, the analysis of temporal networks is not so developed or systematic as the ones for static networks. To understand how contact networks works, at the moment, one need to use different approaches at once (what we sketched in this chapter is one of them). Several papers that have used the same approach as this chapter do not stop after comparing the original and randomized networks, they continue to try to identify lower-level structures. For example, Refs. [32, 34] characterize the differences between SIR spreading on the original and randomized data, then perform a regression analysis to find which low-level structural descriptor has the highest explanatory power.

For the future, we hope it will be possible to formalize the ideas in the chapter to a more programmatic approach. In particular, it is probably possible to construct a flow-chart how to perform successive randomizations to identify the important structures for spreading on a particular data set. This would need to solve the question about how to randomize away arbitrary structures (or at least all structures that are easy to understand conceptually, thus contributing to our understanding of

network dynamics). For opinion spreading problems, a major challenge is to find an appropriate microscopic model of the spreading [89]. Reversely, one could argue that if the purpose of the dynamic system is not to understand social dynamics, but the underlying structure, then there more abstract dynamic systems could perhaps be useful. This approach has been used in, e.g., biochemistry where cellular automata-type symbolic dynamics have been used to explore metabolic networks [53]. In social networks, running prisoner's dilemma dynamics on empirical networks have been argued to say more about the network architecture than the stability of cooperation [36]. Using the Potts' model for community detection [67] is yet an example of a creative use of a seemingly unrelated model to explore network structure.

## References

1. Abrahamson E, Rosenkopf L (1997) *Organ Sci* 8(3):289
2. Anderson RM, May RM (1992) *Infectious diseases in humans*. Oxford University Press, Oxford
3. Backlund VP, Saramäki J, Pan RK (2014) *Phys Rev E* 89:062815
4. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A (2009) *Proc Natl Acad Sci U S A* 106(51):21484
5. Barabási AL (2005) *Nature* 435:907
6. Barabási AL (2015) *Network science*. Cambridge University Press, Cambridge
7. Barrat A, Cattuto C (2013) In: Holme P, Saramäki J (eds) *Temporal networks*. Springer, Berlin, pp 191–216
8. Barthélémy M (2011) *Phys Rep* 499(1):1
9. Bettencourt LMA, Lobo J, Helbing D, Kühnert C, West GB (2007) *Proc Natl Acad Sci U S A* 104(17):7301
10. Boccaletti S, Bianconi G, Criado R, del Genio CI, Gómez-Gardeñes J, Romance M, Sendiña-Nadal I, Wang Z, Zanin M (2014) *Phys Rep* 544(1):1
11. Borge-Holthoefer J, Rivero A, García I, Cauhé E, Ferrer A, Ferrer D, Francos D, Iñiguez D, Pérez MP, Ruiz G, Sanz F, Serrano F, Viñas C, Tarancón A, Moreno Y (2011) *PLoS One* 6(8):1
12. Burda Z, Jurkiewicz J, Krzywicki A (2004) *Phys Rev E* 69:026106
13. Charbonneau D, Blonder B, Dornhaus A (2013) In: Holme P, Saramäki J (eds) *Temporal networks*. Springer, Berlin, pp 217–244
14. Christakis NA, Fowler JH (2007) *N Engl J Med* 357(4):370
15. Crofoot MC, Rubenstein DI, Maiya AS, Berger-Wolf TY (2011) *Am J Primatol* 73(8):821
16. Donker T, Wallinga J, Grundmann H (2010) *PLOS Comput Biol* 6(3):1
17. Donker T, Wallinga J, Slack R, Grundmann H (2012) *PLoS One* 7(4):e35002
18. Eagle N, Pentland AS (2006) *Pers Ubiquit Comput* 10(4):255
19. Ebel H, Mielsch LI, Bornholdt S (2002) *Phys Rev E* 66:035103
20. Eckmann JP, Moses E, Sergi D (2004) *Proc Natl Acad Sci U S A* 101:14333
21. Elliott P, Wartenberg D (2004) *Environ Health Perspect* 112(9):998. <http://www.jstor.org/stable/3838101>
22. Fernández-Gracia J, Sucheki K, Ramasco JJ, Miguel MS, Eguíluz VM (2014) *Phys Rev Lett* 112:158701
23. Gates MC, Woolhouse MEJ (2015) *Epidemics* 12:11
24. Génoin M, Vestergaard CL, Fournet J, Panisson A, Bonmarin I, Barrat A (2015) *Netw Sci* 3:326
25. Giesecke J (2002) *Modern infectious disease epidemiology*, 2nd edn. Arnold, London

26. Goh KI, Barabási AL (2008) *Europhys Lett* 81(4):48002
27. Granovetter MS, Am J Soc 78:1360 (1973)
28. Hethcote HW (2000) *SIAM Rev* 32(4):599
29. Holme P (2013) *PLoS Comput Biol* 9:e1003142
30. Holme P (2015) *Eur Phys J B* 88:234
31. Holme P (2015) *Eur Phys J B* 88(9):1
32. Holme P (2016) *Phys Rev E* 64:022305
33. Holme P, Liljeros F (2014) *Sci Rep* 4:4999
34. Holme P, Masuda N (2015) *PLoS One* 10(3):e0120567
35. Holme P, Saramäki J (2012) *Phys Rep* 519(3):97
36. Holme P, Trusina A, Kim BJ, Minnhagen P (2003) *Phys Rev E* 68:030901
37. Holme P, Edling CR, Liljeros F (2004) *Soc. Networks* 26:155
38. Hornbeck T, Naylor D, Segre AM, Thomas G, Herman T, Polgreen PM (2012) *J Infect Dis* 206(10):1549–1557
39. Isella L, Stehlé J, Barrat A, Cattuto C, Pinton JF, van den Broeck W (2011) *J Theor Biol* 271:166
40. Jacobs AZ, Way SF, Ugander J, Clauset A (2015) Proceedings of the ACM web science conference
41. Karimi F, Holme P (2013) *Physica A* 392:3476
42. Karimi F, Ramenzoni VC, Holme P (2014) *Physica A* 414:263–273
43. Karsai M, Kivelä M, Pan RK, Kaski K, Kertész J, Barabási AL, Saramäki J (2011) *Phys Rev E* 83:025102
44. Keeling MJ, Eames KT (2005) *J R Soc Interface* 2(4):295
45. Kiti MC, Tizzoni M, Kinyanjui TM, Koech DC, Munywoki PK, Meriac M, Cappa L, Panisson A, Barrat A, Cattuto C, Nokes DJ (2016) *EPJ Data Sci.* 5(1):21
46. Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA (2014) *J Complex Netw* 2(3):203
47. Konschake M, Lentz HHK, Conraths FJ, Hövel P, Selhorst T (2013) *PLoS One* 8(2):e55223
48. Kovanen L, Kaski K, Kertész J, Saramäki J (2013) *Proc Natl Acad Sci U S A* 110(45):18070
49. Krings G, Karsai M, Bernhardsson S, Blondel VD, Saramäki J (2012) *EPJ Data Sci* 1:4
50. Lahiri M, Berger-Wolf TY (2007) IEEE symposium on computational intelligence and data mining, pp 35–42
51. Liljeros F, Edling CR, Amaral LAN (2003) *Microbes Infect* 5:189
52. Liljeros F, Giesecke J, Holme P (2007) *Math Popul Stud* 14:269
53. Marr C, Müller-Linow M, Hütt MT (2007) *Phys Rev E* 75:041917
54. Martino GD, Spina S (2015) *Physica A* 438:634
55. Mastrandrea R, Fournet J, Barrat A (2015) *PLOS One* 10(9):1
56. Masuda N, Lambiotte R (2016) A guide to temporal networks. World Scientific, Singapore
57. Mathiesen J, Angheluta L, Ahlgren PTH, Jensen MH (2013) *Proc Natl Acad Sci U S A* 110(43):17259
58. McVoy EC (1940) *Am Sociol Rev* 5(2):219
59. Milo R, Kashtan N, Itzkovitz S, Newman MEJ, Alon U (2003) On the uniform generation of random graphs with prescribed degree sequences. E-print cond-mat/0312028
60. Min B, Goh KI, Vazquez A (2011) *Phys Rev E* 83:036102
61. Newman MEJ (2010) Networks: an introduction. Oxford University Press, Oxford
62. Panzarasa P, Opsahl T, Carley KM (2009) *J Am Soc Inf Sci Technol* 60(5):911
63. Paranjape A, Benson AR, Leskovec J (2017) Proceedings of the tenth ACM international conference on web search and data mining. Association for Computing Machinery, New York, pp 601–610
64. Psorakis I, Roberts SJ, Rezek I, Sheldon BC (2012) *J R Soc Interface* 9(76):3055–3066
65. Radu-Corneliu M, Ciprian D, Fatos X (2012) Third international conference on emerging intelligent data and web technologies. IEEE Computer Society, Piscataway, NJ, pp 133–139
66. Read JM, Eames KTD, Edmunds WJ, J R Soc Interface 5(26):1001 (2008)
67. Reichardt J, Bornholdt S (2006) *Phys Rev E* 74:016110

68. Rocha LEC, Masuda N (2016) *Sci Rep* 6:31456
69. Rocha LEC, Liljeros F, Holme P (2010) *Proc Natl Acad Sc USA* 107:5706
70. Rocha LEC, Liljeros F, Holme P (2011) *PLoS Comput Biol* 7(3):e1001109
71. Romero DM, Meeder B, Kleinberg J (2011) Proceedings of the 20th international conference on world wide web. Association for Computing Machinery, New York, pp 695–704
72. Salathé M, Kazandjieva M, Lee JW, Levis P, Feldman MW, Jones JH (2010) *Proc Natl Acad Sci U S A* 107(51):22020
73. Sanli C, Lambiotte R (2015) *Front Phys* 3:79
74. Scholtes I, Wider N, Pfitzner R, Garas A, Tessone CJ, Schweitzer F (2014) *Nat Commun* 4:5024
75. Stehlé J, Voirin N, Barrat A, Cattuto C, Isella L, Pinton JF, Quaggiotto M, van den Broeck W, Réglis C, Lina B, Vanhems P (2011) *PLOS One* 6:e23176
76. Stopczynski A, Sekara V, Sapiezynski P, Cuttone A, Larsen JE, Lehmann S (2014) *PLoS One* 9(4):e95978
77. Stopczynski A, Pentland AS, Lehmann S (2015) Physical proximity and spreading in dynamic social networks. E-print: arXiv:1509.06530
78. Sun L, Axhausen KW, Lee DH, Huang X (2013) *Proc Natl Acad Sci U S A* 110(34):13774
79. Takaguchi T, Nakamura M, Sato N, Yano K, Masuda N (2011) *Phys Rev X* 1:011008
80. Takaguchi T, Masuda N, Holme P (2013) *PLoS One* 8:e68629
81. Valdano E, Poletto C, Giovannini A, Palma D, Savini L, Colizza V (2015) *PLoS Comput Biol* 11(3):e1004152
82. van den Broeck W, Quaggiotto M, Isella L, Barrat A, Cattuto C (2012) *Leonardo* 45:201
83. Vanhems P, Barrat A, Cattuto C, Pinton JF, Khanafer N, Réglis C, Kim BA, Comte B, Voirin N (2013) *PLOS One* 8:e73970
84. Vergu E, Busson H, Ezanno P (2010) *PLoS One* 5(2):1
85. Volz EM, Miller JC, Galvani A, Meyers L.A. (2011) *PLoS Comput Biol* 7(6):1
86. Villani A, Frigessi A, Liljeros F, Nordvik MK, de Blasio BF (2012) *PLoS One* 7(7):1
87. Viswanath B, Mislove A, Cha M, Gummadi KP (2009) Proceedings of the 2nd ACM workshop on online social networks. Association for Computing Machinery, New York, pp 37–42
88. Walker AS, Eyre DW, Wyllie DH, Dingle KE, Harding RM, O'Connor L, Griffiths D, Vaughan A, Finney J, Wilcox MH et al (2012) *PLoS Med* 9(2):e1001172
89. Weng L, Menczer F, Ahn YY (2013) *Sci Rep* 3:2522
90. Wiehe SE, Carroll AE, Liu GC, Haberkorn KL, Hoch SC, Wilson JS, Fortenberry J (2008) *Int J Health Geogr* 7(1):22
91. Zhang YQ, Li X, Xu J, Vasilakos A (2015) *IEEE Trans Syst Man Cybern* 45(2):214

# Theories for Influencer Identification in Complex Networks



Sen Pei, Flaviano Morone, and Hernán A. Makse

## 1 Introduction

In spreading processes of information, it is well known that certain individuals are more influential than others. In the field of information diffusion, it has been accepted that the ability of influencers to initiate a large-scale spreading is attributed to their privileged locations in the underlying social networks [41, 59, 71, 92]. Due to the direct relevance of influencer identification in such phenomena as viral marketing [46], innovation diffusion [81], behavior adoption [17], and epidemic spreading [69], the research on searching for influential spreaders in different settings is becoming increasingly important in recent years [71].

In the relative simple case of locating individual influencers, given the rich structural information encoded in nodes' location in the network, it is straightforward to measure the influence of a single node using centrality-based heuristics. Over the years, a growing number of predictors have been developed and routinely employed to rank single node's influence in spreading processes, among which the most widely used ones include number of connection [1], k-core [85], betweenness centrality [25], and PageRank [13], just to name a few. Beyond this non-interacting problem, a more challenging task is to identify a set of influencers to achieve maximal collective influence. Originally formulated in the context of viral marketing [80], collective influence maximization is in fact a core optimization problem in an array of important applications in various domains, ranging from cost-effective

---

S. Pei (✉)

Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY, USA  
e-mail: [sp3449@cumc.columbia.edu](mailto:sp3449@cumc.columbia.edu)

F. Morone · H. A. Makse

Levich Institute and Physics Department, City College of New York, New York, NY, USA  
e-mail: [hmakse@lev.ccny.cuny.edu](mailto:hmakse@lev.ccny.cuny.edu)

marketing in commercial promotion, optimal immunization in epidemic control, to strategic protection against targeted attacks on infrastructures. In addition to the topological complexity of network structure, collective influence maximization is further complicated by the entwined interactions between multiple spreaders, which renders the aforementioned centrality-based approaches invalid. As a result, it is required to treat the problem from a collective point of view to develop effective solutions [61].

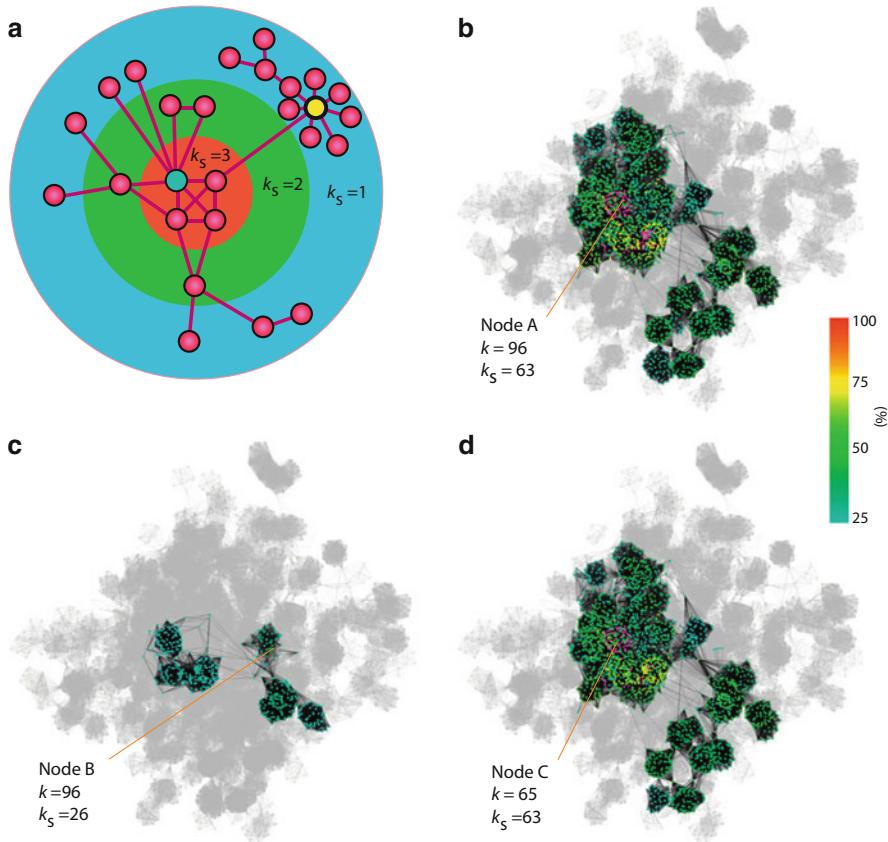
## 2 Finding Individual Influencers

In reality, many spreading phenomena are typically initiated by a single spreader. For instance, an epidemic outbreak in a local area is usually caused by the first infected person. For such processes, ranking the spreading capability of individual spreaders is of great significance in both accelerating and confining the diffusion.

### 2.1 Topological Measures

Intuitively, the nodes with large numbers of connections should have more influence on their direct neighbors. The disproportionate effect of highly-connected nodes, or hubs, on dynamical processes has been revealed in the early works on the vulnerability of scale-free networks [1, 23]. The targeted attack on a very small number of high-degree nodes will rapidly collapse the giant component of networks with heavy-tailed degree distribution. Compared with other more complex centrality measures, the computational burden of degree is almost negligible. Due to this, the simple degree centrality has been playing an important role in influencer identification. In implementation, the performance of high-degree ranking can be further enhanced by a simple adaptive calculation procedure, that is, recalculating the degree of remaining nodes after the removal of previously selected nodes.

An obvious drawback of degree centrality is that it only considers the number of direct neighbors. However, as indicated by empirical studies, most spreading phenomena are proceeded in a cascading fashion. Therefore, the ultimate influence of a single spreader is also affected by the global network structure. In realistic complex networks, high-degree nodes can appear at either the core area or the periphery region. This implies, the number of connections may not be a reliable indicator of influencers in real-world systems. Recently, Kitsak et al. confirmed this speculation through extensive simulations of susceptible-infected-recovered (SIR) and susceptible-infected-susceptible (SIS) dynamics on diverse real-world social networks [41]. In SIR model, a susceptible individual will become infected with a probability  $\beta$  upon contact with his/her infected neighbors, and infected population will recover with a probability  $\mu$  and become immune to the disease. In SIS model, the infection follows the same dynamics but infected persons will become



**Fig. 1** (a), A schematic diagram of k-shell decomposition. The two highlighted nodes (blue and yellow), although both with degree  $k = 8$ , are in different k-shells. (b–d), Infections starting from single nodes with same degree  $k = 96$  (A and B) can result in totally different outcomes. Whereas, infections originating from node C, located in the same k-shell of node A ( $k_s = 63$ ) but with a smaller degree, are quite similar to the spreading from node A. The colors indicate nodes' probability to be infected in SIR simulations with infection rate  $\beta = 0.035$  and recovery rate  $\mu = 1$ . Results are averaged over 10,000 realizations. Figure is adapted from Kitsak et al. [41]

susceptible again with a probability  $\mu$ . As shown in Fig. 1b–d, SIR spreading processes initiated by two hubs with the same degree could result in quite different infected population, depending on their global position in the network. In contrast, the k-core index, which distinguishes the network core and periphery, is a more reliable predictor of influence.

The k-core index is obtained by the k-shell decomposition in which nodes are iteratively pruned according to their remaining degree in the network (see Fig. 1a) [85]. Specifically, nodes with degree  $k = 1$  are first removed successively until there is no node left with one link. The removed nodes are assigned with

k-core index  $k_S = 1$ . Then we remove nodes with degree  $k = 2$  similarly and continue to prune higher k-shells until no node left in the network. In terms of computational complexity, the above decomposition process can be finished within  $O(M)$  operations, where  $M$  is the number of links [7]. Thus k-core ranking is feasible for large-scale complex networks encountered in big-data analysis.

As illustrated in Fig. 1a, the classification of k-core can be very different from that of degree. A hub with low k-core index is usually surrounded by many low-degree neighbors that limit the influence of the hub. On the contrary, nodes located in the core region, although may have moderate degree, are capable of generating large-scale spreading facilitated by their well-connected neighbors. In the case where recovered individuals do not develop immunity, infections would persist in the high k-core area. These findings challenge the previous predominate focus on the number of connections. The simple yet effective measure k-core has inspired several generalizations in consideration of the detailed local environment in the vicinity of high k-core nodes [50, 51, 54, 95].

Although k-core was found effective in SIR and SIS spreading dynamics, some studies indicate that it may not be a good predictor of influence for other spreading models. For instance, in rumor spreading model, Borge-Holthoefer and Moreno [11] showed that the spreading capabilities of the nodes did not depend on their k-core values. These contradictory results relying on the choice of specific spreading model necessitate more extensive empirical validation with real information flow [72].

Apart from the k-core index, another measure that takes into account the global network structure is eigenvector centrality [10, 79]. The reasoning behind the eigenvector centrality is that the influence of an individual is determined by the spreading capability of his/her neighbors. Starting from a uniform score assigned to each node, the scores propagate along the links until a steady state is reached. In calculation, each step of score propagation corresponds to a left multiplication of the adjacency matrix to the current score vector. This procedure is actually the power method to compute the principal eigenvalue of the adjacency matrix. As a result, the steady score vector is in fact proportional to the right eigenvector corresponding to the largest eigenvalue. Notice that, supposing the initial score of each node is one, the first step of iteration will recover the degree centrality.

Despite the wide application of eigenvector centrality, it was recently found that the scores could be localized at a few high degree nodes due to the repeated reflection of scores from their neighbors during the iteration. Martin et al. solved this problem by using the leading eigenvector of the Hashimoto Non-Backtracking (NB) matrix [56]. In NB matrix, the immediate backtracking paths  $i \rightarrow j$  and  $j \rightarrow i$  are not permissible [34], thus avoiding the heavy score accumulation caused by the recurrent one-step reflection. Recently, by mapping the SIR spreading process to bond percolation, Radicchi and Castellano proved that the NB centrality was an optimized predictor for single influencers in SIR model at criticality [76]. In the next section, we will see the important role of NB matrix in collective influence maximization and optimal percolation [61].

## 2.2 Dynamics-Based Measures

Beyond the above pure topological measures, a number of centralities are developed on the basis of specific assumptions on the spreading dynamics. In some classical centralities proposed in the field of social networks, much emphasis is put on the shortest path. Along this way, several renowned centralities were developed and widely accepted in social network ranking. For instance, the closeness centrality quantifies the shortest distance from a given node to all other reachable nodes in the network [84], while betweenness centrality measures the fraction of shortest paths cross through a certain individual between all node pairs [25]. A useful generalization of closeness centrality is the Katz centrality [39], which considers all possible paths in the network, but assigns a larger weight to shorter paths using a tunable parameter. In application, the applicability of these shortest-path-based centralities is limited by the high computational complexity of calculating the shortest paths between all pairs of nodes. As a result, they are more suitable for small or medium scale networks.

Another group of metrics are designed based on random walks. A famous random walk based centrality is PageRank [13]. As a revolutionary webpage ranking algorithm, PageRank mimics a random walk process along the directed hyperlinks. To avoid the random walker trapped in the dangled nodes, a jumping probability  $\alpha$  is introduced to allow the walker jump to a randomly chosen node. The PageRank score is the stationary probability of each node to be visited by the random walker, which can be calculated through iteration. In applications, the PageRank of a node  $i$  in a network can be calculated from  $p_t(i) = \frac{1-\alpha}{N} + \alpha \sum_j \frac{A_{ij} p_{t-1}(j)}{k_{\text{out}}(j)}$ , where  $k_{\text{out}}(j)$  is the number of outgoing links from node  $j$  and  $\alpha$  is the jumping probability. In a generalization called LeaderRank [53], a ground node is connected to all other nodes by additional bidirectional links. This procedure ensures the network to be strongly connected so that the convergence becomes faster.

In addition to the aforementioned centralities designed for general spreading processes, several measures are proposed aimed at specific dynamics, depending explicitly on model parameters. In these approaches, the development of measures is based on the equations depicting the dynamical process. Usually, the analysis of equations will naturally lead to the procedure of path counting in which the number of possible spreading paths is assessed. For instance, Klemm et al. developed a general framework to evaluate the dynamical importance (DI) of nodes in a series of dynamical processes [43]. The iterative calculation of DI centrality essentially counts the total number of arbitrarily long walks departing from each node. Another metric relying on possible spreading paths is the expected force (ExF) proposed by Lawyer [45]. To compute the expected force, all possible clusters of infected nodes after  $n$  transmission events starting from a given node are enumerated. Then the entropy of their cluster degree (i.e., number of outgoing links of the cluster, or infected-susceptible edges) is calculated as the expected force for each node.

The approaches introduced here are far from complete. A growing number of metrics and methods are continuously proposed in the active area of finding single influencers [52]. In designing effective methods for more complex spreading models, the basic principles behind these measures should be universal.

### 3 Finding Multiple Influencers

In spite of the great value of estimating individual nodes' influence with centralities, in a realistic situation, it is more relevant to understand spreading processes initiated by several spreaders. In applications such as viral marketing, it is expected that the spreaders can be coordinated in an optimal manner so that the final collective influence will be maximized. Although it sounds similar to the problem of locating single influencers, the collective influence maximization is in fact a fundamentally different and more difficult problem. In the seminal work of Kempe et al. [40], the influence maximization problems in both Independent Cascade Model (ICM) and Linear Threshold Model (LTM) were mapped to the NP-complete Vertex Cover problem. This implies, the influence maximization problem cannot be solved exactly within a polynomial time, leaving us the only choice of heuristic approach.

A straightforward idea to find multiple influencers is to select the top-ranked spreaders as individual seeds using centrality measures. However, this approach neglects the interactions and collective effect among spreaders. As demonstrated in SIR simulations, the selected spreaders have significant overlap in their influenced population [41]. Therefore, the set of influencers identified with centrality metrics are usually far from optimal. To solve this conundrum, it needs to be treated from a collective point of view [61].

#### 3.1 Optimal Percolation

We start our discussion from the percolation model point of view. As a well-studied dynamical process, percolation was shown to be closely related to spreading and immunization [16, 67, 70]. Percolation is a classical physical process in which nodes or links are randomly removed from a graph [86]. The critical quantity that is of particular interest is the fraction of nodes or links whose removal will collapse the giant component. It is well known that the size of giant component decreases continuously to zero as the number of removed nodes or links increases. In the pioneering works of Newman [67, 68], the class of SIR models were mapped to the percolation process for which the critical point of the continuous transition could be solved exactly.

In contrast to the studies focused on random removal, the problem of optimal percolation aims to find the minimal set of nodes that could guarantee the global connectivity of the network, or equivalently, dismantle the network if removed.

Morone and Makse showed that, mathematically, the optimization of spreading process following *exactly* the Linear Threshold Model with threshold  $k - 1$  ( $k$  is the degree of each node) can be mapped to the optimal percolation problem [61]. For this specific spreading model, finding the minimum number of seeds so that the information percolates the entire network is essentially equivalent to locating the optimal set of nodes in the optimal percolation problem. Similarly, the optimal immunization problem, dual of optimal spreading, can also be mapped to optimal percolation [61]. The relation between the cohesion of a network and influence spreading indicates that the most influential spreaders are the nodes that maintain the integrity of the network.

The collective influence theory for optimal percolation is developed based on the message passing equations of the percolation process. For a network with  $N$  nodes and  $M$  edges, suppose  $\mathbf{n} = (n_1, \dots, n_N)$  indicates whether node  $i$  is removed ( $n_i = 0$ ) or left ( $n_i = 1$ ) in the network. The total fraction of removed nodes is therefore  $q = 1 - \sum_{i=1}^N n_i/N$ . For a directed link from  $i$  to  $j$  ( $i \rightarrow j$ ), let  $v_{i \rightarrow j}$  denote the probability of node  $i$  belonging to the giant component  $G$  in the absence of node  $j$ . The evolution of  $v_{i \rightarrow j}$  satisfies the following self-consistent equation:

$$v_{i \rightarrow j} = n_i \left[ 1 - \prod_{k \in \partial i \setminus j} (1 - v_{k \rightarrow i}) \right], \quad (1)$$

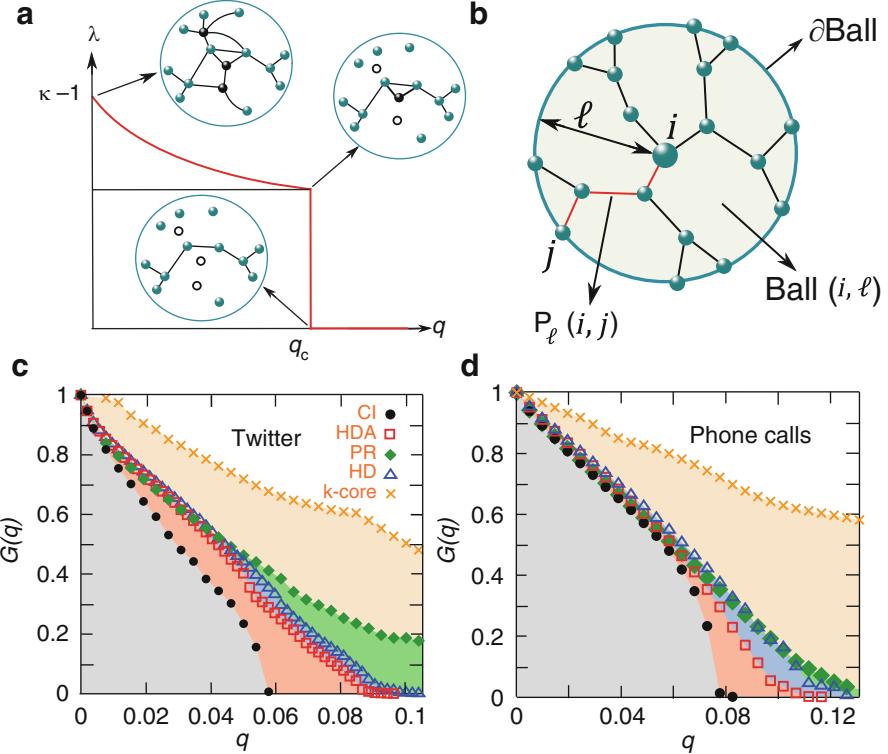
where  $\partial i \setminus j$  denotes the nearest neighbors of  $i$  excluding  $j$ . The final probability  $v_i$  of node  $i$  belonging to the giant component is then determined by  $v_{k \rightarrow i}$  ( $k \in \partial i$ ) through

$$v_i = n_i \left[ 1 - \prod_{k \in \partial i} (1 - v_{k \rightarrow i}) \right]. \quad (2)$$

The fraction of nodes in the giant component is then given by  $G(q) = \sum_{i=1}^N v_i/N$ .

For the continuous phase transition in percolation process, the stability of the zero solution  $G = 0$  is determined by the largest eigenvalue  $\lambda(\mathbf{n}; q)$  of the coupling matrix  $\mathcal{M}$  for the linearized Eq. (1) evaluated at  $\{v_{i \rightarrow j} = 0\}$  (see Fig. 2a). Concretely,  $\mathcal{M}$  is defined on the  $2M \times 2M$  directed links as  $\mathcal{M}_{k \rightarrow \ell, i \rightarrow j} \equiv \frac{\partial v_{i \rightarrow j}}{\partial v_{k \rightarrow \ell}}|_{\{v_{i \rightarrow j} = 0\}}$ . A simple calculation reveals that for locally-tree like random networks,  $\mathcal{M}$  is given in terms of the Non-Backtracking (NB) matrix  $\mathcal{B}$  [34] via  $\mathcal{M}_{k \rightarrow \ell, i \rightarrow j} = n_i \mathcal{B}_{k \rightarrow \ell, i \rightarrow j}$  in which  $\mathcal{B}_{k \rightarrow \ell, i \rightarrow j} = 1$  if  $\ell = i$  and  $j \neq k$ , and 0 otherwise.

To guarantee the stability of the solution  $\{v_{i \rightarrow j} = 0\}$ , it is required  $\lambda(\mathbf{n}; q) \leq 1$ . The optimal influence problem for a given  $q$  can be rephrased as finding the optimal configuration  $\mathbf{n}$  that minimizes the largest eigenvalue  $\lambda(\mathbf{n}; q)$ . As  $q$  approaches the optimal threshold  $q_c$ , there exist a decreasing number of configurations that satisfy  $\lambda(\mathbf{n}; q) \leq 1$ . At  $q_c$ , only one configuration  $\mathbf{n}^*$  exists such that  $\lambda(\mathbf{n}^*; q_c) = 1$ , and



**Fig. 2** (a) For  $q \geq q_c$ , the global minimum of the largest eigenvalue  $\lambda$  of the NB matrix over  $\mathbf{n}$  is 0. In this case,  $G = 0$  is stable, although there exist non-optimal configurations with  $\lambda > 1$  for which  $G > 0$ . For  $q < q_c$ , the minimum of the largest eigenvalue is always  $\lambda > 1$ . Therefore the solution  $G = 0$  is unstable and  $G > 0$ . At the optimal percolation transition, the minimum is at  $\mathbf{n}^*$  such that  $\lambda(\mathbf{n}^*, q_c) = 1$ . At  $q = 0$ ,  $\lambda = \kappa - 1$  where  $\kappa = \langle k^2 \rangle / \langle k \rangle$ . At  $\lambda = 1$ , the giant component is reduced to a tree plus one single loop. This loop is destroyed at the transition  $q_c$ , and  $\lambda$  abruptly falls to 0. (b) Ball( $i, \ell$ ) of radius  $\ell$  around node  $i$  is shown.  $\partial\text{Ball}$  is the set of nodes on the boundary. The highlighted route is the shortest path from  $i$  to  $j$ . (c and d), Giant component  $G(q)$  of Twitter ( $N = 469,014$ ) and Mobile phone network in Mexico ( $N = 1.4 \times 10^7$ ) computed using CI, high degree adaptive (HDA), PageRank (PR), high degree (HD), and k-core strategies. Figure is adapted from Morone et al. [61]

all other configurations will give  $\lambda(\mathbf{n}; q) > 1$ . The optimal configuration of  $Nq_c$  influencers  $\mathbf{n}^*$  is therefore obtained when the *minimum* of the largest eigenvalue satisfies  $\lambda(\mathbf{n}^*; q_c) = 1$ . In practice, the largest eigenvalue can be calculated by the power method (we leave out  $q$  in  $\lambda(\mathbf{n}; q)$ ):

$$\lambda(\mathbf{n}) = \lim_{\ell \rightarrow \infty} \left[ \frac{|\mathbf{w}_\ell(\mathbf{n})|}{|\mathbf{w}_0|} \right]^{1/\ell}. \quad (3)$$

Here  $|\mathbf{w}_\ell(\mathbf{n})|$  is the  $\ell$  iterations of  $\mathcal{M}$  on initial vector  $\mathbf{w}_0$ :  $|\mathbf{w}_\ell(\mathbf{n})| = |\mathcal{M}^\ell \mathbf{w}_0|$ . To find the best configuration of  $\mathbf{n}$ , we need to minimize the cost function  $|\mathbf{w}_\ell(\mathbf{n})|$  for a finite  $\ell$ . Through a proper simplification, we have an approximation of  $|\mathbf{w}_\ell(\mathbf{n})|^2$  of order  $1/N$  as

$$|\mathbf{w}_\ell(\mathbf{n})|^2 = \sum_{i=1}^N (k_i - 1) \sum_{j \in \partial \text{Ball}(i, 2\ell-1)} \left( \prod_{k \in \mathcal{P}_{2\ell-1}(i, j)} n_k \right) (k_j - 1), \quad (4)$$

in which  $\partial \text{Ball}(i, \ell)$  is the frontier of the ball of radius  $\ell$  in terms of shortest path centered around node  $i$ ,  $\mathcal{P}_\ell(i, j)$  is the shortest path of length  $\ell$  connecting  $i$  and  $j$ , and  $k_i$  is the degree of node  $i$ . See an example in Fig. 2b.

Based on the form of Eq. (4), an energy function for each configuration  $\mathbf{n}$  can be defined as follows:

$$E_\ell(\mathbf{n}) = \sum_{i=1}^N (k_i - 1) \sum_{j \in \partial \text{Ball}(i, \ell)} \left( \prod_{k \in \mathcal{P}_\ell(i, j)} n_k \right) (k_j - 1), \quad (5)$$

where  $E_\ell(\mathbf{n}) = |\mathbf{w}_{(\ell+1)/2}|^2$  for  $\ell$  odd and  $E_\ell(\mathbf{n}) = \langle \mathbf{w}_{\ell/2} | \mathcal{M} | \mathbf{w}_{\ell/2} \rangle$  for  $\ell$  even. For  $\ell = 1$ ,  $E_\ell(\mathbf{n})$  is exactly the energy function of an Ising model which can be optimized using the cavity method [57]. For  $\ell \geq 2$ , it becomes a hard optimization problem involving many-body interactions. To develop a scalable algorithm for big-data analysis, an adaptive method is proposed, which is essentially a greedy algorithm for minimizing the largest eigenvalue of the stability matrix  $\mathcal{M}$  for a given  $\ell$  in the form of Eq. (4). In fact, Eq. (5) can be rewritten as the sum of collective influence from single nodes:

$$E_\ell(\mathbf{n}) = \sum_{i=1}^N \text{CI}(i), \quad (6)$$

in which the collective influence (CI) of node  $i$  at length  $\ell$  is defined as:

$$\text{CI}_\ell(i) = (k_i - 1) \sum_{j \in \partial \text{Ball}(i, \ell)} (k_j - 1). \quad (7)$$

The main idea behind the CI algorithm is to remove the nodes that can cause largest decrease of energy function in Eq. (4). In each iteration of CI algorithm, the node with the largest CI value is deleted, after which the CI values for remaining nodes are recalculated. The adaptive removal continues until the giant component is fragmented, i.e.  $G(q) = 0$ . Notice that the procedure minimizes  $q_c$  but does not guarantee the minimization of  $G$  in the percolation phase  $G > 0$ . If we want to optimize the configuration for  $G(q) > 0$ , a reinsertion procedure is applied from the configuration at  $G(q) = 0$ . In practice, if we use a heap structure to find the node

with the largest CI and only update the nodes inside the  $(\ell+1)$ -radius ball around the removed node, the computational complexity of CI algorithm can achieve  $N \log(N)$  [62]. As a result, the CI algorithm is scalable for massively large-scale networks in modern social network analysis. For a Twitter network with 469,013 users (Fig. 2c) and a social network of  $1.4 \times 10^7$  mobile phone users in Mexico (Fig. 2d), CI algorithm finds a smaller set of influencers than simple scalable heuristics including high degree adaptive (HDA), PageRank (PR), high degree (HD), and k-core [61]. To apply CI algorithm to real-time influencer ranking, a Twitter search engine was developed at <http://www.kcore-analytics.com>. Notice that, for  $\ell = 0$ , CI algorithm degenerates to high-degree ranking. So degree can be interpreted as the zero-order approximation of CI in Eq. (7).

To guarantee the scalability of the algorithm, CI essentially takes an adaptive greedy approach. The performance of CI algorithm can be further improved by a simple extension of CI using the message passing framework for  $\ell \rightarrow \infty$ —the CI propagation algorithm (CI<sub>P</sub>) [62]. Remarkably, the CI propagation algorithm can reproduce the exact analytical threshold of optimal percolation for cubic random regular graphs [8]. Another belief-propagation variant of CI algorithm based on optimal immunization (CI<sub>BP</sub>) also has similar performance of CI<sub>P</sub> [62]. However, the improvement over CI algorithm is at the price of higher computational complexity  $O(N^2 \log(N))$ , which makes both CI<sub>P</sub> and CI<sub>BP</sub> unscalable.

Recent studies have shown that the optimal percolation problem is closely related to the optimal decycling problem, or minimum feedback vertex set (FVS) problem [38]. Using belief-propagation (BP) algorithms, the optimal percolation problem was solved in recent works [12, 65]. The result of BP algorithms was found better than CI algorithm. Another approach to the optimal destruction of networks makes use of the explosive percolation theory [22].

### 3.2 Independent Cascade Model

The percolation process is deterministic on a given network with a given seed set. An important class of spreading model with stochasticity is the independent cascade model (ICM) [42]. In these models, a node is infected or activated by its neighbors with a predefined probability independently. Frequently used independent cascade models include susceptible-infected (SI) model, susceptible-infected-susceptible (SIS) model, and susceptible-infected-removed (SIR) model. These models are widely adopted in modeling infectious disease outbreaks and information spreading in social networks [35, 41, 74, 87, 93, 94]. Therefore, it is of particular interest in relevant applications.

In the pioneering work of Kempe et al. [40], influence maximization was first formalized as a discrete optimization problem: For a given spreading process on a network and an integer  $k$ , how to find the optimal set of  $k$  seeds that could generate the largest influence. For a large class of ICM and LTM, the influence

maximization problem can be well approximated by a simple greedy strategy, with a provable approximation guarantee [40]. In the basic greedy algorithm, the seed set is obtained by repeatedly selecting the node that provides the largest marginal increase of influence at each time step. The performance guarantee is built on the submodular property of the influence function  $\sigma(S)$  [66], which is defined as the expected number of active nodes if the initial seed set is  $S$ . The influence function  $\sigma(\cdot)$  is submodular if the incremental influence of selecting a node  $u$  into a seed set  $S$  is no smaller than the incremental influence of selecting the same node into a larger set  $V$  containing  $S$ . That is,  $\sigma(S \cup \{u\}) - \sigma(S) \geq \sigma(V \cup \{u\}) - \sigma(V)$  for all nodes  $u$  and any sets  $S \subseteq V$ . Leveraging on the result of submodular function [66], the greedy algorithm is guaranteed to approximate the true optimal influence within a factor of  $1 - 1/e \approx 63\%$ , i.e.,  $\sigma(S) \geq (1 - 1/e)\sigma(S^*)$ , where  $S$  is the seed set obtained by the greedy algorithm and  $S^*$  is the true optimal seed set. Although the basic greedy algorithm is simple to implement and performance-guaranteed, it requires massive Monte Carlo simulations to estimate the marginal gain of each candidate node. Several works were proposed to improve the efficiency of greedy algorithm [19, 20, 30, 47].

While performance guaranteed, from an optimization point of view, the greedy algorithm may be stuck into local optimum. This drawback can be solved by a more sophisticated message passing approach. Altarelli et al. developed the message passing algorithms (both belief-propagation (BP) and max-sum (MS)) for the problem of optimal immunization for SIR and SIS model [4], which can be applied to general ICMs. From another point of view, the independent cascade model can be naturally mapped to a bond percolation. Hu et al. found that in a series of real-world networks, most SIR spreading would be restrained to a local area while global-scale spreading rarely occurs [37]. Using the bond percolation theory, a characteristic local length termed influence radius was revealed. They argue that the global spreading optimization problem in fact can be solved locally, with the knowledge of the local environment within the influence radius.

### 3.3 Linear Threshold Model

Compared with independent cascade model, linear threshold model is more complex in the sense that a node's state is collectively determined by its neighbors' state. In a typical instance of LTM, each node  $v$  is assigned with a threshold value  $\theta_v$  and each link  $(u, v)$  is assigned with a weight  $w(u, v)$ . During the cascade, a node is activated only if the sum of weights of its activated neighbors reaches the threshold value, i.e.  $\sum_{u \in \partial_v} w(u, v) \geq \theta_v$ . In the case where the weights and thresholds are drawn uniformly from the interval  $[0, 1]$ , LTM was proven to be submodular [40]. Therefore, the influence maximization in this class of LTM can be well approximated by the greedy strategy, as we introduced in the above section. However, even with the lazy forward update [47], the algorithm is still unscalable

for large networks. Chen et al. found a way to approximate the influence of a node in a local subgraph [21], and developed a scalable greedy algorithm. Goyal et al. [31] further improved this algorithm by considering more choices of paths.

The above greedy approach and its variants are applicable to LTM with submodular property. However, for the general class of LTM with fixed weight and threshold, it is not guaranteed to be submodular [40]. An important class of LTM that may not be submodular is defined as follows: A node  $i$  is activated only after a certain number  $m_i$  of its neighbors are activated. The choice of different threshold  $m_i$  can generate two qualitatively different cascade regimes with continuous and discontinuous phase transitions. For instance, in the special case of  $m_i = k_i - 1$  ( $k_i$  is the degree of node  $i$ ), a continuous phase transition of influence occurs as the seed set grows [61]. However, there also exist a wide class of LTM exhibiting a first-order, or discontinuous phase transition. In the case that seeds are selected randomly, the transition between these two regimes is explored in detail in the context of bootstrap percolation [9, 29] and a simple cascade model [91]. But these results are based on the typical dynamical properties starting from random initial conditions. For influence maximization with a special initial condition, the dynamical behavior should be deviated from the average ones. Altarelli et al. proposed a BP algorithm that could estimate statistical properties of nontypical trajectories and found the initial conditions that lead to cascading with desired properties [2]. To obtain the exact set of seeds, MS equations were derived by setting the inverse temperature  $\beta \rightarrow \infty$  in the energy function [3]. Extending the work under the assumption of replica symmetry, the theoretical limit of the minimal contagious set (the minimal seed set that can activate the entire graph) in random regular graphs is obtained using the cavity method with the effect of replica symmetry breaking [33].

In big-data analysis, an efficient and scalable algorithm designed for general LTM is needed. Starting from the message passing equations of LTM, generalized from Eq. (1) of percolation, a scalable algorithm named collective influence for threshold model (CI-TM) can be developed [75]. By iteratively solving the linearized message passing equations, the cascading process can be decomposed to separate components, each of which corresponds to the contribution made by a single seed. Interestingly, it is found the contribution of a seed is determined by the subcritical paths along which cascade propagates. In order to design a scalable algorithm, the node with the largest number of subcritical paths is recursively selected into the seed set. After each selection, the selected node and the subcritical paths attached to it are removed, and the status of the remaining nodes is recalculated. Making use of the heap structure, CI-TM algorithm can achieve the complexity of  $O(N \log N)$ . On one hand, computing  $CI - TM_\ell$  value for a given length  $\ell$  is equivalent to iteratively visiting subcritical neighbors of each node layer by layer within  $\ell$  radius. Because of the finite search radius, computing  $CI - TM_\ell$  for each node takes  $O(1)$  time. Initially, we have to calculate  $CI - TM_\ell$  for all nodes. However, during later adaptive calculation, there is no need to update  $CI - TM_\ell$  for all nodes. We only have to recalculate for nodes within  $\ell + 1$  steps from the removed vertices, which scales as  $O(1)$  compared to the network size as  $N \rightarrow \infty$  as shown in [62]. On the other hand, selecting the node with maximal CI-TM can

be realized by making use of the data structure of heap that takes  $O(\log N)$  time [62]. Therefore, the overall complexity of ranking  $N$  nodes is  $O(N \log N)$  even when we remove the top CI-TM nodes one by one. In both homogeneous and scale-free random networks, CI-TM achieves larger collective influence given the same number of seeds compared with other scalable approaches. This provides a practical method that can be applied to massively large-scale networks.

## 4 Applications of Influencer Identification

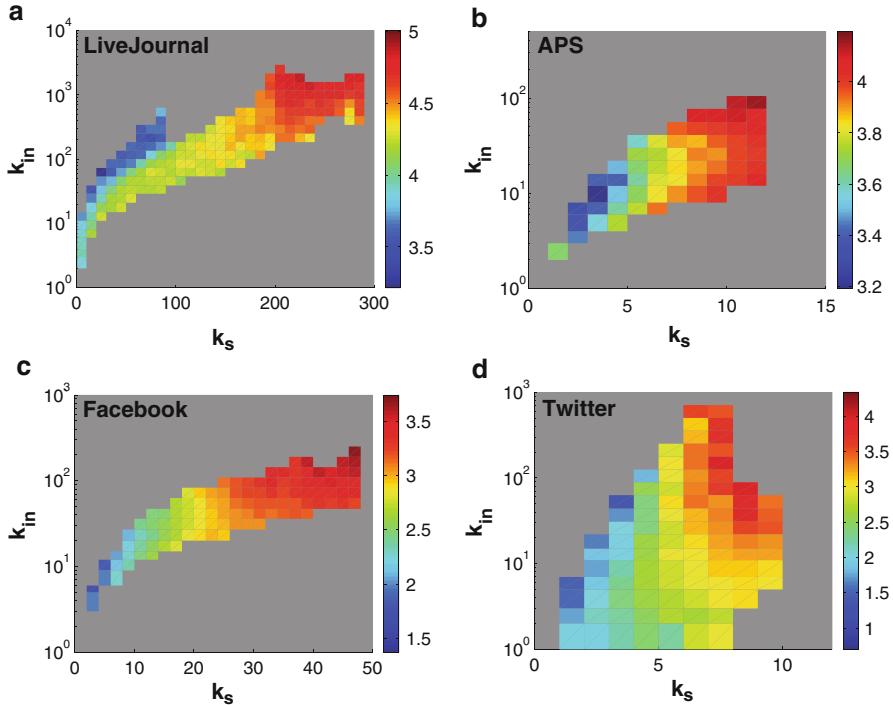
The problem of influencer identification is ubiquitous in a wide class of applications. So far, the theory of influencer identification has been applied to a number of important problems. In this section, we will introduce the application of influencer identification in three different areas: information diffusion, brain networks, and socioeconomic systems.

### 4.1 *Information Diffusion in Social Networks*

The most direct application of influencer identification is to maximize the information diffusion in social networks. In recent years, a huge number of research works have been performed aiming to relate users' spreading power to their locations, or personal features [58, 72, 89]. These works, mainly focusing on various types of online social networks including email communication [49], Facebook [60, 90], Twitter [6, 18, 44], and blogs sharing communities [5, 77], enrich our understanding of information diffusion in social networks.

A great challenge of developing effective predictors of influencers comes from the validation. In most of the previous works, the validation of proposed measures depends on modeling of information spreading in a given network. This approach, however, has led to several contradictory results on the best predictor of influence depending on the particular models [11, 41]. These models are built on simplified assumptions on human behavior [36] that neglect some of the most important features in real information diffusion [27], such as activity frequency [64, 83], behavior pattern [48, 73, 88], etc. Therefore, it is required to validate the various proposed predictors using empirical diffusion records in real-world social media.

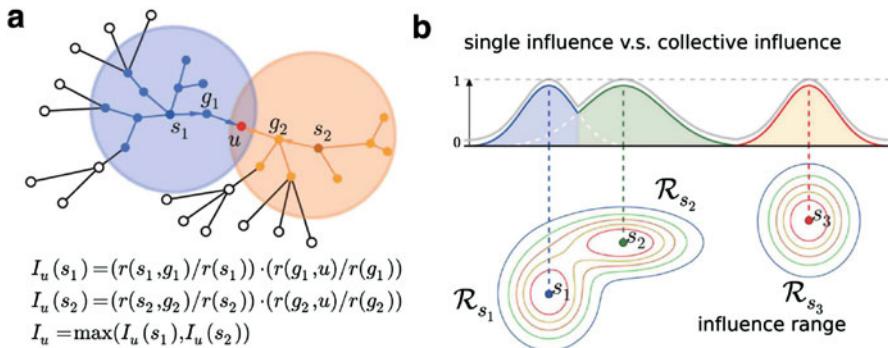
We first compare the performance of different predictors for single influencers [72]. Realistic information diffusion instances as well as the underlying social networks are collected in four dissimilar social platforms: a blog-sharing community LiveJournal, scientific journals of American Physical Society, an online social network Facebook, and microblog service Twitter. To determine the real influence of each node, a directed diffusion graph is first constructed for each system by combining all directed diffusion links together. Then starting from a source node



**Fig. 3** K-core predicts the average influence of spreading more reliably than in-degree. Logarithmic values (base 10) of the average size of influence region  $M(k_S, k_{in})$  when spreading originates from nodes with  $(k_S, k_{in})$  for LiveJournal (a), APS journals (b), Facebook (c), and Twitter (d) are shown. Figure is adapted from Pei et al. [72]

*i*, the total influence  $M_i$  of node *i* is computed by tracking the diffusion links layer by layer in a breadth-first-search (BFS) fashion. Once we get the realistic influence, it is convenient to compare the performance of different predictors, including degree, k-core, and PageRank. Specifically, we can calculate the average influence  $M(k_S, k_{in})$  for nodes with a given combination of k-core value  $k_S$  and in-degree  $k_{in}$ :  $M(k_S, k_{in}) = \sum_{i \in \Upsilon(k_S, k_{in})} M_i / N(k_S, k_{in})$ , where  $\Upsilon(k_S, k_{in})$  is the collection of users in the  $(k_S, k_{in})$  bin, and  $N(k_S, k_{in})$  is the size of this collection. In all the systems, it is consistently observed that nodes with fixed degree can have either large or small influence, while nodes located in the same k-core have similar influence (see Fig. 3). Thus the influence of nodes is more related to their global location in the network, indicated by their k-core values. The same conclusion is also obtained in the comparison with PageRank. K-core does not only predict the average influence better, but also recognize influencers more accurately. Although k-core is effective, it is too coarse to distinguish different nodes within same shells. In some cases, there may be millions of nodes in one shell.

We further investigate the identification of multiple influencers [89]. Again, we use the realistic diffusion instances in the above four platforms. However, the empirical data cannot be directly mapped to ideal multi-source spreading. Such ideal multi-source spreading instances in which spreaders send out the same piece of message at the same time rarely exist in reality. Even though we can find such instances, the initial spreaders are hardly the same as the set of nodes selected by CI or other heuristic strategies. To circumvent this difficulty, we can construct virtual multi-source spreading processes by leveraging the behavior patterns of users extracted from the data. Suppose  $n$  spreaders  $S = \{s_i | i = 1, 2, \dots, n, n = qN\}$  are activated at the beginning of the virtual process. The influence strength  $I_{g_1}(s)$  from seed  $s$  to its neighbor  $g_1$  depends on the tendency of  $g_1$  to receive information from  $s$ . Assume during the observation time,  $s$  has sent out  $r(s)$  pieces of messages and  $g_1$  has accepted  $r(s, g_1)$  of them. Then the influence strength can be approximated by  $I_{g_1}(s) = r(s, g_1)/r(s)$ . In subsequent spreading,  $g_1$  may affect its neighbor  $g_2 \neq s$  in the same manner. Following the spreading paths, we can acquire the influence strength  $s$  enforcing on its  $\ell$ -step neighbor  $g_\ell$ :  $I_{g_\ell}(s) = \prod_{k=1}^{\ell} r(g_{k-1}, g_k)/r(g_{k-1})$ , where  $g_0 = s$ . The collective influence  $I_u$  for node  $u$  imposed by the seed set  $S$  is therefore  $I_u = \max_{i=1}^n I_u(s_i)$ . See Fig. 4 for an example. Finally, summing up all the  $N$  nodes in the network, the collective influence of the spreaders imposed on the entire system is  $Q(q) = \sum_{u=1}^N I_u/N$ . Based on this virtual spreading process, we can evaluate the collective influence of the spreaders selected by different methods. In particular, we compare the influencers selected by collective influence algorithm (CI), adaptive high degree (HDA), high degree (HD), PageRank (PR), and k-core. In all the systems, CI consistently outperforms other ranking methods.



**Fig. 4** (a) Calculation of influence strength to node  $u$ . Suppose the maximum spreading layer is set as  $L = 2$  for two distinct seeds  $s_1$  and  $s_2$ . The collective influence enforcing to  $u$  is selected as the largest value of the strength  $I_u(s_1)$  and  $I_u(s_2)$ . (b) An illustration of single influence and collective influence. The three circle-like areas represent influence range  $R_{s_1}$ ,  $R_{s_2}$  and  $R_{s_3}$ , for different spreaders  $s_1$ ,  $s_2$ , and  $s_3$ . The contour lines show the levels of influence strength. The collective influence (grey curve) is obtained by combining single influence strengths of all spreaders. Figure is adapted from Teng et al. [89]

## 4.2 Collective Influence in Brain Networks

The human brain is a robust modular system interconnected as a Network of Networks (NoN) [15, 28, 78]. How this robustness emerges in a modular structure is an important question in many disciplines. Previous interdependent NoN models inspired by power grid are extremely fragile [14], thus cannot explain the observed robustness in brain networks. To reveal the mechanism beneath this robustness, a NoN model is proposed which can afford inter-link functionality and remain robust at the same time [63, 82].

In NoN system, the links are classified into two types: inter-modular links that represent the mutual dependencies between modules and intra-modular links that do not involve in the inter-modular dependencies. Denote  $\mathcal{S}(i)$  and  $\mathcal{F}(i)$  as the set of nodes connected to node  $i$  via intra-modular and inter-modular links, respectively. Suppose the variable state of node  $i$  is  $\sigma_i \in \{0, 1\}$  (inactive or active), and the external input to node  $i$  is  $n_i \in \{0, 1\}$  (no input or input). In the general activation model, the variable state is related to the input through  $\sigma_i = n_i \left[ 1 - \prod_{j \in \mathcal{F}(i)} (1 - n_j) \right]$ . That is, the node  $i$  is activated only if  $i$  receives the input ( $n_i = 1$ ) and at least one of its neighbors connected with inter-modular links receives the input. In a robust brain network, for typical input configuration  $\mathbf{n} = (n_1, \dots, n_N)$ , the giant (largest) component of the active nodes  $G$  with  $\sigma_i = 1$  should be globally connected. Therefore, the robustness of the brain network can be characterized by the critical value  $q_{rand} = 1 - \langle \mathbf{n} \rangle$  of zero inputs such that  $G(q_{rand}) = 0$ . Here the input configuration  $\mathbf{n}$  is sampled from a flat distribution. Ideally, the robust NoN should have no disconnected phase, with a large value of  $q_{rand}$  close to 1.

To explain both robustness and inter-link functionality of brain networks, a robust NoN (R-NoN) model is proposed [63]. Define  $\rho_{i \rightarrow j} \in \{0, 1\}$  as the message running along an intra-modular link  $i \rightarrow j$ ,  $\varphi_{i \rightarrow j} \in \{0, 1\}$  as the message running along an inter-modular link  $i \rightarrow j$ . The information flow follows the self-consistent equations

$$\rho_{i \rightarrow j} = \sigma_i \left[ 1 - \prod_{k \in \mathcal{S}(i) \setminus j} (1 - \rho_{k \rightarrow i}) \prod_{\ell \in \mathcal{F}(i)} (1 - \varphi_{\ell \rightarrow i}) \right], \quad (8)$$

$$\varphi_{i \rightarrow j} = \sigma_i \left[ 1 - \prod_{k \in \mathcal{S}(i)} (1 - \rho_{k \rightarrow i}) \prod_{\ell \in \mathcal{F}(i) \setminus j} (1 - \varphi_{\ell \rightarrow i}) \right]. \quad (9)$$

The physical meaning of the above equations is easy to be interpreted. For instance, in Eq. (8), a positive message  $\rho_{i \rightarrow j}$  is transmitted from  $i$  to  $j$  in the same module if node  $i$  is active  $\sigma_i = 1$  and if it receives at least one positive message from either a node  $k$  in the same module  $\rho_{k \rightarrow i} = 1$  or a node  $\ell$  in the other module  $\varphi_{\ell \rightarrow i} = 1$ .

Notice that, the logical OR is important since it is the basis of the robustness of R-NoN. The final probability of node  $i$  belonging to the largest active component  $G$  is

$$\rho_i = \sigma_i \left[ 1 - \prod_{k \in \mathcal{S}(i)} (1 - \rho_{k \rightarrow i}) \prod_{\ell \in \mathcal{T}(i)} (1 - \varphi_{\ell \rightarrow i}) \right]. \quad (10)$$

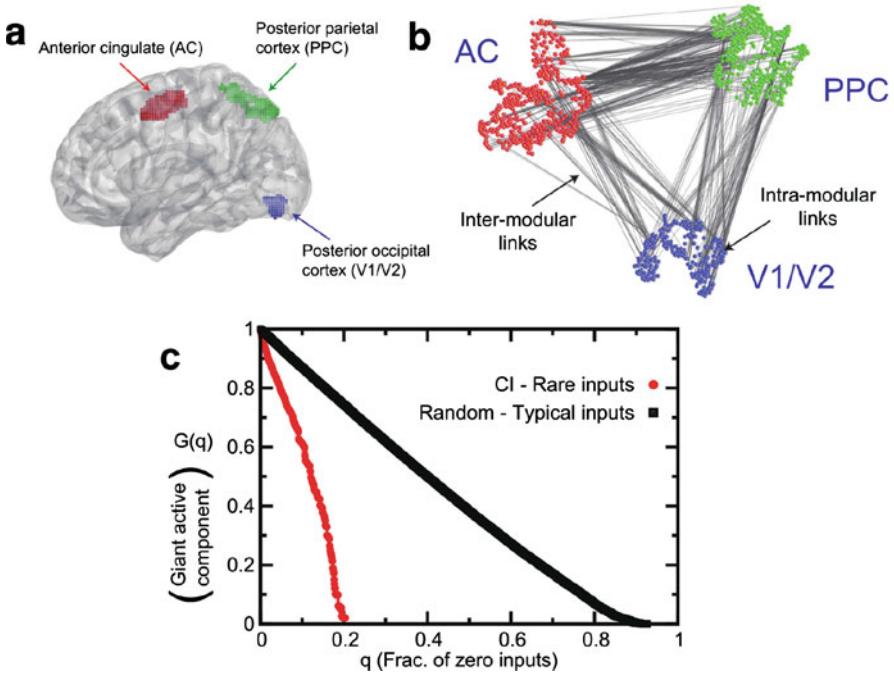
The size of  $G$  is therefore  $G = \langle \rho_i \rangle$ . In the R-NoN model, the system is robust since a node can be active  $\sigma_i = 1$  even it does not belong to  $G$ . This prevents catastrophic cascading effects in the catastrophic C-NoN model inspired by power grid failure [14]. In the C-NoN model, a node remains functional only if it belongs to the giant component in *both* networks. This implies the status of a node in one network is interdependent on its status in the other network. The fundamental difference between C-NoN and R-NoN is that, in C-NoN model, the size of  $G$  is computed through

$$\rho_i = \sigma_i \left[ 1 - \prod_{k \in \mathcal{S}(i)} (1 - \rho_{k \rightarrow i}) \right] \left[ 1 - \prod_{\ell \in \mathcal{T}(i)} (1 - \varphi_{\ell \rightarrow i}) \right]. \quad (11)$$

So the logical OR in Eq. (10) is replaced by the logical AND in C-NoN. This stricter condition makes the system extremely sensitive to small perturbations. In synthetic NoN made of ER and SF random graphs, it is found the percolation threshold  $q_{rand}$  of R-NoN model is close to 1. On the contrary, the C-NoN model has threshold  $q_{rand}$  close to 0. This indicates that the two models indeed capture two different phenomena.

After exploring the behavior of R-NoN model under typical inputs, it is required to study the response to rare events targeting the influencers in the brain networks. Rare malfunction of nodes in the brain network that targets influencers may interrupt the global communication in the brain, which have been conjectured be responsible for certain neurological disorders. Or conversely, activating the influencers would optimally broadcast information to the entire network. Therefore, it is important to predict the location of the most influential nodes involved in information processing in the brain. To find the minimal fraction of nodes  $q_{infl}$  in the brain network whose removal would optimally fragment the giant component, the R-NoN model is mapped to the optimal percolation. The collective influence of nodes is calculated by minimizing the largest eigenvalue of the modified NB matrix. Particularly, the collective influence of node  $i$  is given by

$$\text{CI}_\ell(i) = z_i \sum_{j \in \partial \text{Ball}(i, \ell)} z_j + \sum_{j \in \mathcal{T}(i): k_j^{\text{out}}=1} z_j \sum_{m \in \partial \text{Ball}(j, \ell)} z_m, \quad (12)$$



**Fig. 5** (a) Spatial location of the three main modules (AC, PPC, and V1/V2) in the 3NoN. (b) Topology of the 3NoN. Inter-links and intra-links are displayed. (c) Size of the largest active cluster  $G(q)$  as a function  $q$  of the nodes with  $n_i = 0$  following CI optimization (red curve,  $\ell=3$ ) and random states (black curve, random percolation). Figure is adapted from Morone et al. [63]

where  $z_i \equiv k_i^{\text{in}} + k_i^{\text{out}} - 1$ . The first term is the node-centric contribution, which presents in the single network case of optimal percolation, while the second term is the node-eccentric contribution, which is a new feature of the brain NoN.

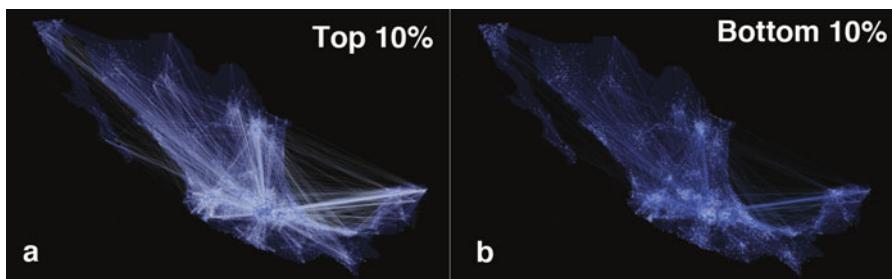
Applying the R-NoN model and collective influence theory to real brain networks, it is possible to obtain the collective influence map of brain NoN. The brain network is constructed from the functional magnetic resonance imaging (fMRI) data of the experiment of stimulus driven attention [26, 28, 63]. In the experiment, each subject performs a dual visual-auditory task when receiving a visual stimulus and an auditory pitch simultaneously. This experiment requires the deployment of high level control modules in the brain, thus captures the role of dependency inter-modular connections. In the obtained brain network (see Fig. 5a, b), it is observed that the system is robust with large threshold  $q_{rand} \approx 0.9$ . While the minimal set of influencers only requires  $q_{infl} \approx 0.2$  fraction of nodes (see Fig. 5c). Using the CI-map of the brain network, it is confirmed that control is deployed from the higher level module (Anterior cingulate) towards certain strategic locations in the lower ones (posterior parietal cortex, posterior occipital cortex). Moreover, the coarse-grain of the NoN to top CI nodes can predict the strategic areas in the brain.

### 4.3 Financial Status in Socioeconomic Systems

It has long been recognized that the pattern of individuals' social connection in society can affect people's financial status [32]. However, how to quantify the relationship between the location of an individual in social network and his/her economic wellness remains an open question. Despite that the effect of network diversity on economic development has been tested in the community level [24], inference of people's financial status from social network centralities or metrics in individual level is still needed. The difficulty of such investigation comes from the lack of empirical data containing both individual's financial information and pattern of social ties.

To find a reliable social network predictor of people's financial status, a massively large social network of the mobile and residential communication in Mexico containing  $1.10 \times 10^8$  users together with financial banking data are analyzed [55]. With this dataset, it is possible to precisely cross-correlate the financial information of a person with his/her location in the communication network at the country level. Particularly, the financial status of individuals is reflected by their credit limit. In the analysis of the  $5.02 \times 10^5$  bank clients identified in the phone call network, the top 10% and bottom 10% individuals present completely different communication pattern (see Fig. 6). Richer people maintain more active and diverse links, some connecting to remote locations and forming tightly linked "rich clubs."

To characterize the affluent people with network metrics, several centralities that are feasible for large-scale networks are compared, including degree, PageRank, k-core, and collective influence (CI). In the communication network, these four metrics are correlated. Therefore, they all show correlations with financial status when age is controlled. Among them, both k-core and CI capture the strong correlation with credit line with a  $R^2$  value of 0.96 and 0.93, respectively. However, CI is more preferable since it satisfies both, a strong correlation and a high resolution. According to the definition of CI, top CI nodes are surrounded by hubs hierarchically. This is exactly the structure of ego-centric network of the top 1% wealthy people.



**Fig. 6** (a and b) Visualization of communication activity of population in the top 10% and bottom 10% total credit limit classes. Figure is adapted from Luo et al. [55]

The performance of predictions can be further enhanced by considering the factor of age. An age-network combined metric  $ANC = \alpha Age + (1 - \alpha)CI$  with  $\alpha = 0.5$  can achieve a correlation with  $R^2 = 0.99$ . Moreover, it is able to identify 70% high credit individuals at the highest earner level. To validate the effectiveness, a real social marketing campaign was performed. Specifically, text messages inviting new credit card clients were sent to 656,944 people selected by their high CI values in the social network. Meanwhile, the same message was sent to a control group of 48,000 individuals selected randomly. The response rate, measured by the fraction of recipients who requested the product, is augmented by threefold in the top influencers identified by CI compared with the random control group.

The same analysis was also applied to individuals' diversity of links [24]. The diversity of an individual can be measured by the diversity ratio  $DR = W_{out}/W_{in}$ , i.e., the ratio of total communication events with people in other communities  $W_{out}$  and within the same community  $W_{in}$ . The correlation between DR and CI is weak so they should reflect different aspects of network structure. In comparison with financial data, the age-diversity composite  $ADC = \alpha Age + (1 - \alpha)DR$  ( $\alpha = 0.5$ ) well correlates with people's financial status. These evidences indicate that both CI and DR are effective predictors of people's financial situation in an individual level. This finding has a great practical value in relevant applications, for instance, social marketing campaigns.

**Acknowledgements** We acknowledge funding from NIH-NIBIB 1R01EB022720, NIH-NCI U54CA137788 / U54CA132378 and NSF-IIS 1515022.

## References

- Albert R, Jeong H, Barabási AL (2000) Error and attack tolerance of complex networks. *Nature* 406(6794):378–382
- Altarelli F, Braunstein A, Dall'Asta L, Zecchina R (2013) Large deviations of cascade processes on graphs. *Phys Rev E* 87(6):062115
- Altarelli F, Braunstein A, Dall'Asta L, Zecchina R (2013) Optimizing spread dynamics on graphs by message passing. *J Stat Mech: Theory Exp* 2013(09):P09011
- Altarelli F, Braunstein A, Dall'Asta L, Wakeling JR, Zecchina R (2014) Containing epidemic outbreaks by message-passing techniques. *Phys Rev X* 4(2):021024
- Backstrom L, Huttenlocher D, Kleinberg J, Lan X (2006) Group formation in large social networks: membership, growth, and evolution. In: Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining. Association for Computing Machinery, New York, pp 44–54
- Bakshy E, Hofman JM, Mason WA, Watts DJ (2011) Everyone's an influencer: quantifying influence on twitter. In: Proceeding of the 4th ACM international conference on web search and data mining. Association for Computing Machinery, New York, pp 65–74
- Batagelj V, Zaversnik M (2003) An  $O(m)$  algorithm for cores decomposition of networks. *arXiv preprint cs/0310049*
- Bau S, Wormald NC, Zhou S (2002) Decycling numbers of random regular graphs. *Random Struct Algoritm* 21(3–4):397–413

9. Baxter GJ, Dorogovtsev SN, Goltsev AV, Mendes JF (2010) Bootstrap percolation on complex networks. *Phys Rev E* 82(1):011103
10. Bonacich P (1972) Factoring and weighting approaches to status scores and clique identification. *J Math Socio* 2(1):113–120
11. Borge-Holthoefer J, Moreno Y (2012) Absence of influential spreaders in rumor dynamics. *Phys Rev E* 85(2):026116
12. Braunstein A, Dall'Asta L, Semerjian G, Zdeborová L (2016) Network dismantling. *Proc Natl Acad Sci U S A* 113(44):12368–12373
13. Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. *Comput Netw ISDN Syst* 30(1):107–117
14. Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S (2010) Catastrophic cascade of failures in interdependent networks. *Nature* 464(7291):1025–1028
15. Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10(3):186–198
16. Callaway DS, Newman ME, Strogatz SH, Watts DJ (2000) Network robustness and fragility: percolation on random graphs. *Phys Rev Lett* 85(25):5468
17. Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194–1197
18. Cha M, Haddadi H, Benevenuto F, Gummadi PK (2010) Measuring user influence in twitter: the million follower fallacy. In: Proceeding of the 4th international AAAI conference on weblogs and social media 10(10–17):30
19. Chen W, Wang Y, Yang S (2009) Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining. Association for Computing Machinery, New York, pp 199–208
20. Chen W, Wang C, Wang Y (2010) Scalable influence maximization for prevalent viral marketing in large-scale social networks. In: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining. Association for Computing Machinery, New York, pp 1029–1038
21. Chen W, Yuan Y, Zhang L (2010) Scalable influence maximization in social networks under the linear threshold model. In: 2010 IEEE 10th international conference on data mining (ICDM). IEEE, Los Alamitos, CA, pp 88–97
22. Clusella P, Grassberger P, Pérez-Reche FJ, Politi A (2016) Immunization and targeted destruction of networks using explosive percolation. *Phys Rev Lett* 117(20):208301
23. Cohen R, Erez K, Ben-Avraham D, Havlin S (2001) Breakdown of the internet under intentional attack. *Phys Rev Lett* 86(16):3682
24. Eagle N, Macy M, Claxton R (2010) Network diversity and economic development. *Science* 328(5981):1029–1031
25. Freeman LC (1978) Centrality in social networks conceptual clarification. *Soc Netw* 1(3):215–239
26. Gallos LK, Sigman M, Makse HA (2007) The conundrum of functional brain networks: small-world efficiency or fractal modularity. *Front Psychol* 3:123
27. Gallos LK, Song C, Makse HA (2008) Scaling of degree correlations and its influence on diffusion in scale-free networks. *Phys Rev Lett* 100(24):248701
28. Gallos LK, Makse HA, Sigman M (2012) A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. *Proc Natl Acad Sci U S A* 109(8):2825–2830
29. Goltsev AV, Dorogovtsev SN, Mendes JFF (2006) k-core (bootstrap) percolation on complex networks: critical phenomena and nonlocal effects. *Phys Rev E* 73(5):056101
30. Goyal A, Lu W, Lakshmanan LV (2011) Celf++: optimizing the greedy algorithm for influence maximization in social networks. In: Proceedings of the 20th international conference on world wide web. Association for Computing Machinery, New York, pp 47–48
31. Goyal A, Lu W, Lakshmanan LV (2011) Simpath: an efficient algorithm for influence maximization under the linear threshold model. In: 2011 IEEE 11th international conference on data mining (ICDM). IEEE, Los Alamitos, CA, pp 211–220

32. Granovetter MS (1973) The strength of weak ties. *Am J Sociol* 78(6):1360–1380
33. Guggiola A, Semerjian G (2015) Minimal contagious sets in random regular graphs. *J Stat Phys* 158(2):300–358
34. Hashimoto KI (1989) Zeta functions of finite graphs and representations of p-adic groups. *Adv Stud Pure Math* 15:211–280
35. Hethcote HW (2000) The mathematics of infectious diseases. *SIAM Rev* 42(4):599–653
36. Hu Y, Havlin S, Makse HA (2014) Conditions for viral influence spreading through multiplex correlated social networks. *Phys Rev X* 4(2):021031
37. Hu Y, Ji S, Feng L, Havlin S, Jin Y (2015) Optimizing locally the spread of influence in large scale online social networks. *arXiv preprint arXiv:1509.03484*
38. Karp RM (1972) Reducibility among combinatorial problems. In: Complexity of computer computations. Springer, Berlin, pp 85–103
39. Katz L (1953) A new status index derived from sociometric analysis. *Psychometrika* 18(1): 39–43
40. Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. In: Proceedings of the 9th ACM SIGKDD international conference on knowledge discovery and data mining. Association for Computing Machinery, New York, pp 137–146
41. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA (2010) Identification of influential spreaders in complex networks. *Nat Phys* 6(11):888–893
42. Kleinberg J (2007) Cascading behavior in networks: algorithmic and economic issues. *Algorithmic Game Theory* 24:613–632
43. Klemm K, Serrano M, Eguiluz VM, Miguel MS (2012) A measure of individual role in collective dynamics. *Sci Rep* 2:292
44. Kwak H, Lee C, Park H, Moon S (2010) What is twitter, a social network or a news media? In: Proceeding of the 19th ACM international conference on world wide web. Association for Computing Machinery, New York, pp 591–600
45. Lawyer G (2015) Understanding the influence of all nodes in a network. *Sci Rep* 5:8665
46. Leskovec J, Adamic LA, Huberman BA (2007) The dynamics of viral marketing. *ACM Trans Web* 1(1):5
47. Leskovec J, Krause A, Guestrin C, Faloutsos C, VanBriesen J, Glance N (2007) Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining. Association for Computing Machinery, New York, pp 420–429
48. Li W, Tang S, Pei S, Yan S, Jiang S, Teng X, Zheng Z (2014) The rumor diffusion process with emerging independent spreaders in complex networks. *Physica A* 397:121–128
49. Liben-Nowell D, Kleinberg J (2008) Tracing information flow on a global scale using internet chain-letter data. *Proc Natl Acad Sci U S A* 105(12):4633–4638
50. Liu Y, Tang M, Zhou T, Do Y (2015) Core-like groups result in invalidation of identifying super-spreader by k-shell decomposition. *Sci Rep* 5:9602
51. Liu Y, Tang M, Zhou T, Do Y (2015) Improving the accuracy of the k-shell method by removing redundant links-from a perspective of spreading dynamics. *Sci Rep* 5:13172
52. Lü L, Chen D, Ren XL, Zhang QM, Zhang YC, Zhou T (2016) Vital nodes identification in complex networks. *Phys Rep* 650:1–63
53. Lü L, Zhang YC, Yeung CH, Zhou T (2011) Leaders in social networks, the delicious case. *PLoS One* 6(6):e21202
54. Lü L, Zhou T, Zhang QM, Stanley HE (2016) The h-index of a network node and its relation to degree and coreness. *Nat Commun* 7:10168
55. Luo S, Morone F, Sarraute C, Makse HA (2017) Inferring personal financial status from social network location. *Nat Commun* 8:15227
56. Martin T, Zhang X, Newman M (2014) Localization and centrality in networks. *Phys Rev E* 90(5):052808
57. Mézard M, Parisi G (2003) The cavity method at zero temperature. *J Stat Phys* 111(1):1–34
58. Min B, Liljeros F, Makse HA (2015) Finding influential spreaders from human activity beyond network location. *PLoS One* 10(8):e0136831

59. Min B, Morone F, Makse HA (2016) Searching for influencers in big-data complex networks. In: Bunde A, Caro J, Karger J, Vogl G (eds) *Diffusive spreading in nature, technology and society*. Springer, Cham
60. Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B: Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM conference on internet measurement. Association for Computing Machinery, New York, pp 29–42
61. Morone F, Makse HA (2015) Influence maximization in complex networks through optimal percolation. *Nature* 524:65–68
62. Morone F, Min B, Bo L, Mari R, Makse HA (2016) Collective influence algorithm to find influencers via optimal percolation in massively large social media. *Sci Rep* 6:30062
63. Morone F, Roth K, Min B, Stanley HE, Makse HA (2017) A model of brain activation predicts the neural collective influence map of the human brain. *Proc Natl Acad Sci U S A* 114(15):3849–3854
64. Muchnik L, Pei S, Parra LC, Reis SD, Andrade Jr, JS, Havlin S, Makse HA (2013) Origins of power-law degree distribution in the heterogeneity of human activity in social networks. *Sci Rep* 3:1783
65. Mugisha S, Zhou HJ (2016) Identifying optimal targets of network attack by belief propagation. *Phys Rev E* 94(1):012305
66. Nemhauser GL, Wolsey LA, Fisher ML (1978) An analysis of approximations for maximizing submodular set functions—I. *Math Program* 14(1):265–294
67. Newman ME (2002) Spread of epidemic disease on networks. *Phys Rev E* 66(1):016128
68. Newman ME, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. *Phys Rev E* 64(2):026118
69. Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. *Phys Rev Lett* 86(14):3200
70. Pastor-Satorras R, Vespignani A (2002) Immunization of complex networks. *Phys Rev E* 65(3):036104
71. Pei S, Makse HA (2013) Spreading dynamics in complex networks. *J Stat Mech: Theory Exp* 2013(12):P12002
72. Pei S, Muchnik L, Andrade Jr JS, Zheng Z, Makse HA (2014) Searching for superspreaders of information in real-world social media. *Sci Rep* 4:5547
73. Pei S, Muchnik L, Tang S, Zheng Z, Makse HA (2015) Exploring the complex pattern of information spreading in online blog communities. *PLoS One* 10(5):e0126894
74. Pei S, Tang S, Zheng Z (2015) Detecting the influence of spreading in social networks with excitable sensor networks. *PLoS One* 10(5):e0124848
75. Pei S, Teng X, Shaman J, Morone F, Makse HA (2017) Efficient collective influence maximization in threshold models of behavior cascading with first-order transitions. *Sci Rep* 7:45240
76. Radicchi F, Castellano C (2016) Leveraging percolation theory to single out influential spreaders in networks. *Phys Rev E* 93(6):062314
77. Ramos M, Shao J, Reis SD, Anteneodo C, Andrade Jr JS, Havlin S, Makse HA (2015) How does public opinion become extreme? *Sci Rep* 5:10032
78. Reis SD, Hu Y, Babino A, Andrade Jr JS, Canals S, Sigman M, Makse HA (2014) Avoiding catastrophic failure in correlated networks of networks. *Nat Phys* 10(10):762–767
79. Restrepo JG, Ott E, Hunt BR (2006) Characterizing the dynamical importance of network nodes and links. *Phys Rev Lett* 97(9):094102
80. Richardson M, Domingos P (2002) Mining knowledge-sharing sites for viral marketing. In: Proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining. Association for Computing Machinery, New York, pp 61–70
81. Rogers EM (2010) *Diffusion of innovations*. Simon and Schuster, London
82. Roth K, Morone F, Min B, Makse HA (2017) Emergence of robustness in networks of networks. *Phys Rev E* 95(6):062308

83. Rybski D, Buldyrev SV, Havlin S, Liljeros F, Makse HA (2012) Communication activity in a social network: relation between long-term correlations and inter-event clustering. *Sci Rep* 2:560
84. Sabidussi G (1966) The centrality index of a graph. *Psychometrika* 31(4):581–603
85. Seidman SB (1983) Network structure and minimum degree. *Soc Netw* 5(3):269–287
86. Stauffer D, Aharony A (1994) Introduction to percolation theory. CRC press, Boca Raton
87. Tang S, Teng X, Pei S, Yan S, Zheng Z (2015) Identification of highly susceptible individuals in complex networks. *Physica A* 432:363–372
88. Teng X, Yan S, Tang S, Pei S, Li W, Zheng Z (2014) Individual behavior and social wealth in the spatial public goods game. *Physica A* 402:141–149
89. Teng X, Pei S, Morone F, Makse HA (2016) Collective influence of multiple spreaders evaluated by tracing real information flow in large-scale social networks. *Sci Rep* 6:36043
90. Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in facebook. In: Proceedings of the 2nd ACM workshop on online social networks. Association for Computing Machinery, New York, pp 37–42
91. Watts DJ (2002) A simple model of global cascades on random networks. *Proc Natl Acad Sci U S A* 99(9):5766–5771
92. Watts DJ, Dodds PS (2007) Influentials, networks, and public opinion formation. *J Constr Res* 34(4):441–458
93. Yan S, Tang S, Pei S, Jiang S, Zheng Z (2014) Dynamical immunization strategy for seasonal epidemics. *Phys Rev E* 90(2):022808
94. Yan S, Tang S, Fang W, Pei S, Zheng Z (2015) Global and local targeted immunization in networks with community structure. *J Stat Mech: Theory Exp* 2015(8):P08010
95. Zeng A, Zhang CJ (2013) Ranking spreaders by decomposing complex networks. *Phys Lett A* 377(14):1031–1035

## **Part III**

# **Observational Studies**

# Service Adoption Spreading in Online Social Networks



Gerardo Iñiguez, Zhongyuan Ruan, Kimmo Kaski, János Kertész,  
and Márton Karsai

## 1 Introduction

A human society abounds with examples of collective patterns of behaviour that arise due to the correlated decisions of a large number of individuals. This is evidenced in the spread of religious beliefs and political movements, in the behavioural, cultural, and opinion shifts in a population, in the adoption of technological and medical innovations, in the rise of popularity of political and media figures, in the growth of bubbles in financial markets, and in the use of products and online services. All of these phenomena tend to evolve similarly over time, as they start with individuals that independently from their peers and due to external influence such as mass media take the risk by adopting a certain behaviour [1, 2]. Then, these processes continue as friends, colleagues, and acquaintances observe such individuals and engage with the same behaviour, therefore participating in a spreading process throughout society [3].

---

G. Iñiguez

Institute for Research in Applied Mathematics and Systems, National Autonomous University of Mexico, México, DF, Mexico

Department of Computer Science, Aalto University School of Science, Aalto, Finland  
e-mail: [gerardo.iniguez@aalto.fi](mailto:gerardo.iniguez@aalto.fi)

Z. Ruan · J. Kertész

Center for Network Science, Central European University, Budapest, Hungary  
e-mail: [KerteszJ@ceu.edu](mailto:KerteszJ@ceu.edu)

K. Kaski

Department of Computer Science, Aalto University School of Science, Aalto, Finland  
e-mail: [kimmo.kaski@aalto.fi](mailto:kimmo.kaski@aalto.fi)

M. Karsai (✉)

Univ de Lyon, ENS de Lyon, INRIA, CNRS, UMR 5668, IXXI, Lyon, France  
e-mail: [marton.karsai@ens-lyon.fr](mailto:marton.karsai@ens-lyon.fr)

The way ideas, products, and behaviour spread throughout a population over time, commonly known as *innovation diffusion*, was first observed empirically in the mid twentieth century by the likes of Rogers [4] and Bass [5]. In the following decades, many mathematical models were introduced with the goal of identifying mechanisms by which behaviour diffuses through society [6–8]. One of the first (and arguably simplest) is the Bass model for forecasting sales of new consumer durables [5], which characterises the diffusion of innovation as a process of contagion initiated by some external influence [2] (e.g. mass communication, news media) and promoted by internal, social influence [9] (via word-of-mouth, viral marketing, etc.). The model assumes a homogeneous population of adopters and it predicts that aggregated sales data has an s-shaped pattern as a function of time [6, 10].

Despite the success of the Bass model and similar diffusion-like models to capture qualitatively the temporal behaviour of adoption processes, macroscopic models only provide empirical generalisations based on the behaviour of society as a whole (by means of aggregated data on adoption rates, for example). Hence, these models do not take into account individual heterogeneities and the complex structure and dynamics of social processes [11]. In other words, since the same macro-level behaviour may arise from several individual-level mechanisms (like learning, externalities, or contagion), it is difficult for these models to assess what mechanisms are actually responsible for large-scale spreading phenomena [12–14]. In order to overcome this issue, agent-based diffusion models consider behavioural heterogeneities, networked social interactions [15, 16], and decision-making processes based on the cognitive capacities of individuals [17–19]. Then, behaviour at the level of society emerges dynamically from the interplay between network structure and the actions of people. This microscopic approach allows for the modelling of varying behaviour across individuals, while recognising that social interactions and interpersonal communication are essential in determining adoption [1, 20].

Under the network approach, the Bass model is an archetypal example of *simple contagion* [21] where, akin to the transmission of a disease, information and individuals' willingness to adopt may propagate with exposure to a single person engaging in some particular behaviour. However, when adoption turns out costly, risky or controversial, the spread of ideas and products often requires social reinforcement and exposure to several sources, a phenomenon usually called *complex contagion* [22, 23]. The requirement of multiple interactions for adoption was first implemented theoretically by Granovetter via behavioural thresholds, namely “*the number or proportion of others who must make one decision before a given actor does so*” [6]. Following this idea various agent-based network models have been introduced and analysed by Watts and others [1, 24–31] in order to understand the properties of threshold-driven social contagion.

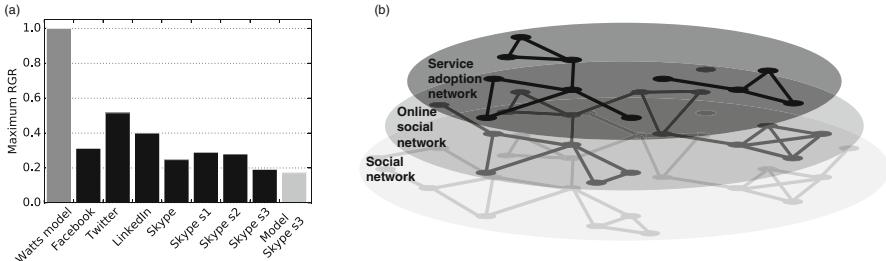
Despite the allure of social influence as the reason behind innovation diffusion, it is more challenging to identify causal mechanisms in adoption spreading than in biological contagion, since the same empirical, large-scale observations may be obtained as effects of social influence [32], homophily [33], or the environment. For example, collective adoption patterns may appear as a consequence of homophilic

structural correlations, where interacting individuals adopt due to their similar interests and not due to actual social influence [23]. Hence distinguishing between the effects of social influence and homophily at the individual level remains a challenge [34, 35]. Moreover, regarding the particular role social influence may have in adoption spreading, several assumptions have been proposed about its functional dependency on the number of adopters necessary to influence an individual. While Granovetter and others [6, 24] suggest a simple linear dependency, as observed in some large techno-social systems [19], Latané [36] argues for non-linear effects that have been demonstrated empirically by online experiments at different scales [9, 37, 38].

Perhaps one of the most intriguing features of threshold-driven social contagion is its ability to capture what Watts calls the *robust yet fragile* nature of complex systems [24]. This means that a population may be robust and disregard many ideas and products, but suddenly exhibit fast system-wide adoption patterns known as *behavioural cascades*. While homophily suggests that adoption behaviour is only seemingly correlated, and simple contagion implies that external influence always induces global adoption in a connected population, complex contagion captures the additional feature that large cascades of behavioural patterns tend to happen only rarely, and may be triggered by actions at the individual level that are indistinguishable from the rest. Indeed, behavioural cascades are rare but potentially disrupting social spreading phenomena, where collective patterns of exposure arise through reinforcement as a consequence of small initial perturbations [39]. Examples include the rapid emergence of political and grass-root movements [40–42], or the fast spreading of information [12, 27, 43–48] and behavioural patterns [49]. Moreover, cascades may appear in both online [50–54] and offline [55] social environments.

The characterisation [12, 13, 56–59] and modelling [24, 60–63] of behavioural cascades have received a lot of attention in the past and provide some understanding of the causal mechanisms and structure of empirical and synthetic cascades on various types of networks [64–67]. However, these studies fail in addressing the temporal dynamics of the emerging cascades, which may vary among empirical examples of social contagion. In other words, previous works do not answer why real-world cascades may evolve either slowly or rapidly over time. In contrast to the cases of rapid cascading mentioned above, the propagation of products in social networks is typically slower, with adoption spreading gradually, even if it is driven by threshold mechanisms and may eventually cover a large fraction of the total population [19]. This slow behaviour characterises the adoption of online services such as Facebook, Twitter, LinkedIn, and Skype (Fig. 1a), since their yearly maximum relative growth rate of cumulative adoption [68] is lower than in the case of rapid cascades, as suggested in standard models of threshold-driven social contagion like the Watts threshold (WT) model [24].

In this chapter we review recent works [69, 70] focusing on the empirical characterisation and mathematical modelling of the slow, threshold-driven spreading of service adoption in online social networks, particularly in the case of Skype. We first provide empirical evidence of the distribution of individual adoption thresholds



**Fig. 1** The speed and layers of online service adoption. **(a)** Yearly maximum relative growth rate (RGR) of cumulative adoptions obtained by taking the maximum of the yearly adoption rate (yearly count of adoptions) normalised by the final observed number of adoptions of a given service. We show it for several online social-communication services [68] (black bars), including three paid Skype services (s1—“subscription”, s2—“voicemail”, and s3—“buy credit”). The dark grey bar corresponds to a rapid cascade of adoption as suggested by the Watts threshold model, while the light grey bar is the prediction of our model for Skype s3. **(b)** Schematic layer structure of online service adoption systems. The lowest layer represents a real, offline social network; the middle layer corresponds to any online social network; and the top layer is the adoption of a service within the social network. As an advantage in this study we have full knowledge about the Skype online social network in this multi-layer structure, while we follow a paid service spreading on the online network. This figure is adopted from Ref. [70] and it is licensed under Creative Commons Attribution 4.0 International Licences

and other structural and dynamical features of the worldwide Skype adoption cluster. We then show how to incorporate the observed structural and threshold heterogeneities into a dynamical threshold model where multiple individuals may adopt spontaneously (i.e. firstly among their acquaintances). We find that if the fraction of users who reject to adopt a product or idea in the model is large, the system enters a quenched state where the evolution and structure of the global adoption cluster is very similar to our observations of services within Skype. Model calculations and the analysis of the real social contagion process suggest that the evolving structure of an adoption cluster differs radically from previous expectations [24], since it is triggered by several spontaneous adoptions arriving at a constant rate. Furthermore, the stable adopters (who initially resist exposure) are actually responsible for the emergence of global social adoption.

## 2 Empirical Observations

In order to observe service adoption dynamics we analyse an example of an online diffusion process, where we have access to individual service adoption events as well as the underlying social network. Our aim is to identify the crucial mechanisms necessary to consider in models of complex contagion to match them better with reality, and define a model that incorporates these mechanisms and captures the possible dynamics leading to the emergence of real-world global cascades.

To fully understand service adoption processes on online social structures, we need to keep in mind some of their proxy characteristics. People of a society constitute a social network by being connected with ties of several kinds that are maintained in various ways. However, and despite their recent popularity, online social systems are not capable of mapping the entire social network as offline, occasionally maintained, temporary, or ill-favoured social ties may remain invisible in such systems. Therefore, these networks provide only a proxy sample of the real social structure (Fig. 1b), with important but also insignificant social ties present. Moreover, data available for social network studies commonly arrives as a sample of a larger online social system, which unavoidably leads to observational biases. In addition, connections in an online social structure cannot precisely assign the flow of direct social influence among the connected individuals, only the possibility of it. Finally, just like real social networks, online social systems evolve over time via the creation and dissolution of social ties or by nodes entering or leaving the system. Due to all these limitations it is rather challenging to make unbiased observations about any unfolding dynamical processes, without making some assumption about the underlying online social systems.

In our study we use the social network of one of the largest voice-over-internet providers in the world, the network of Skype, which actually copes well with the limitations listed above. It maps all connections in the Skype network without sampling, thus it provides us with a complete, unbiased map of the underlying social network, maintaining the diffusion of services available only for registered users in the network. This network evolves as a function of time via adoption, churning, and link creation dynamics. We have shown in an earlier study [19] that while rates of these actions increase considerably with time, the adoption processes can be well characterised by the net rate of the actual number of users. We also found that while spontaneous adoptions and churning evolve with a constant rate, the probability of peer-pressured adoptions corresponds linearly to the strength of social influence, giving rise to a non-linear dynamics at the system level, which enables its modelling as a complex contagion process.

In our study we concentrate on the adoption dynamics of a paid service that unfolds over the Skype social network (Fig. 1b). Since this adoption process evolves in a considerably faster time-scale than the underpinning social network, we can validly assume a time-scale separation. Thus, from here on we consider the network structure to be static, which may give us a good first approximation while concentrating on the adoption dynamics unfolding on its fabric. To identify the effects of social influence in our empirical system we also present a null model study (Sect. 2.4).

## 2.1 Data Description

In our social network nodes represent users and edges between pairs of users exist if they are in each other's contact lists. A user's contact list is composed of *friends*.

If user  $u$  wants to add another user  $v$  to his/her contact list,  $u$  sends  $v$  a contact request, and the edge is established at the moment  $v$  approves the request (or not, if the contact request is rejected). For the purpose of our study we use the largest connected component of the aggregated free Skype service network, which was recorded from September 2003 to November 2011 (i.e. over 99 months) and contains roughly 4.4 billion links and 510 million registered users worldwide [71]. The data is fully anonymised and considers only confirmed connections between users after the removal of spammers and blocked nodes.

To study an example of service adoption dynamics we follow the purchases of the “buy credit” paid service for 89 months starting from 2004. Data includes the time of first payment of each adopting user, an individual and conscious action that tracks adoption behaviour. Note that other examples about the adoption dynamics of similar services are presented in [70].

## 2.2 Degree and Threshold Heterogeneities

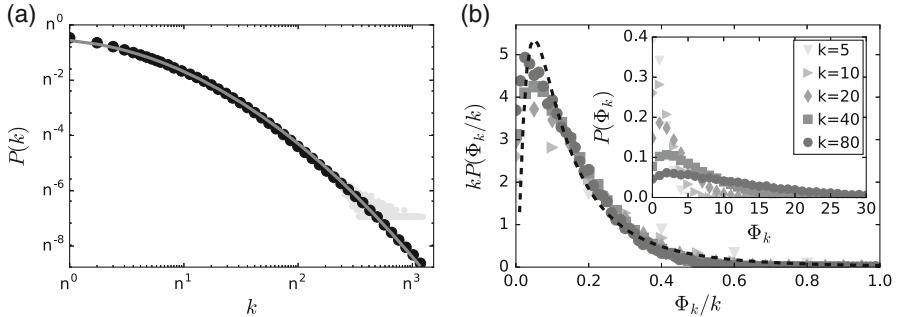
In his seminal work on modelling adoption cascades [24], Watts identified two structural characteristics that control the emergence of collective adoption cascades. One is the distribution  $P(k)$  of degrees (i.e. number of neighbours of a node), with average  $z = \langle k \rangle$ , and the other is the distribution  $P(\varphi)$  of adoption thresholds (with average  $w = \langle \varphi \rangle$ ), defined as the necessary fraction of exposed neighbours that triggers the adoption of an individual under study, or central ego.

Degree heterogeneities have been in the focus of network science for a while now, and a broad degree distribution  $P(k)$  is one of the main characteristics of complex networks [72, 73]. This distribution has been usually described as a power-law, but a log-normal fit has often turned out to work better [74]. The latter is the case with our data:

$$P(k) \propto k^{-1} \exp[-(\ln k - \mu_D)^2 / (2\sigma_D^2)], \quad (1)$$

where the best fit is obtained with  $k \geq k_{\min}$  and parameters  $\mu_D = 1.2$ ,  $\sigma_D = 1.39$  and  $k_{\min} = 1$  (Fig. 1a), giving an average degree  $z = 8.56$ .

It is a challenging task to quantify individual adoption thresholds, as their observation simultaneously requires information about the underlying network structure and the dynamical adoption process evolving on top. Therefore, besides measuring the number  $k$  of friends of an ego in the Skype social network (already needed for the degree distribution), for  $k$ -degree users at the time of their adoption we measure the number  $\Phi_k$  of their neighbours who have adopted the service earlier, i.e. the integer threshold [57]. To our knowledge, this is the first detailed study measuring the number of adopting neighbours of adopters in an empirical setting. The obtained distribution  $P(\Phi_k)$  for varying  $k$  is shown in the inset of Fig. 2b. The importance of our empirical findings is amplified by the observation that these distributions can be scaled together when using the fractional threshold variable



**Fig. 2** Degree and threshold heterogeneities. (a) Degree distribution  $P(k)$  of the Skype network (light/dark grey circles for raw/binned data) on a double log-scale with arbitrary base  $n$ .  $P(k)$  is fitted with a log-normal distribution (see text) with parameters  $\mu_D = 1.2$  and  $\sigma_D = 1.39$ , and average  $z = 8.56$  (grey line). (b) Distribution  $P(\Phi_k)$  of integer thresholds  $\Phi_k$  for several degree groups in Skype s3 (inset). By using  $P(\Phi_k, k) = kP(\Phi_k/k)$ , these curves collapse into a master curve approximated by a log-normal function (dashed line in main panel) with parameters  $\mu_T = -2$  and  $\sigma_T = 1$ , as constrained by the average threshold  $w = 0.19$ . This figure is adopted from Ref. [70] and it is licensed under Creative Commons Attribution 4.0 International Licences

$\varphi = \Phi_k/k$ , i.e. the fraction of adopting neighbours at the time of adoption (Fig. 2b main panel). Thus, in a discussion of whether the number or the ratio of adopting neighbours matters in behavioural adoption [22, 24], our results give strong support to the latter.

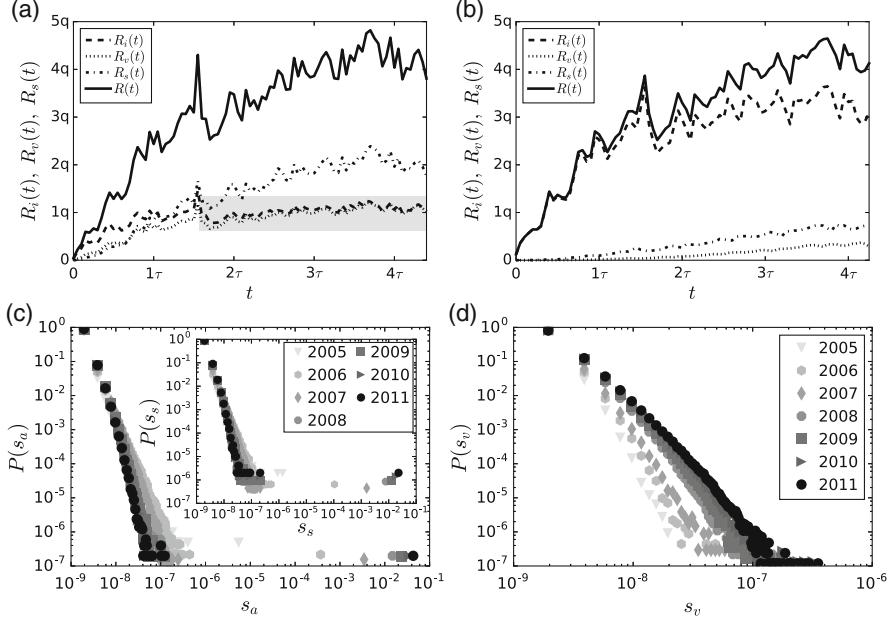
Using fractional thresholds and the relationship  $P(\Phi_k, k) = kP(\Phi_k/k)$ , all distributions collapse to a master curve, which is once again well-approximated by a log-normal function of the following form,

$$P(\Phi_k/k) = P(\varphi) \propto \varphi^{-1} \exp[-(\ln \varphi - \mu_T)^2/(2\sigma_T^2)], \quad (2)$$

with parameters  $\mu_T = -2$  and  $\sigma_T = 1$  as constrained by the average threshold  $w = 0.19$  [70]. These empirical observations, in addition to the broad degree distribution, provide quantitative description of the heterogeneous nature of adoption thresholds.

### 2.3 Dynamics and Structure of Adoption Cascades

Since we know the complete structure of the online social network, as well as the first time of service usage for all adopters, we can follow the temporal evolution of the adoption dynamics. By counting the number of adopting neighbours of an ego, we identify innovators ( $\Phi_k = 0$ ), and vulnerable ( $\Phi_k = 1$ ) or stable ( $\Phi_k > 1$ ) nodes, in accordance with the categorisation of Watts [24]. As we show in Fig. 3a, the adoption rates for these categories behave rather differently from previous suggestions [24]. First, there is not only one seed but an increasing fraction of innovators in the system who, after an initial period, adopt approximately at a



**Fig. 3** The dynamics and structure of adoption cascades. **(a)** Adoption rate of innovators [ $R_i(t)$ ], vulnerable nodes [ $R_v(t)$ ], and stable nodes [ $R_s(t)$ ], as well as the net service adoption rate [ $R(t)$ ], where the rates are measured with a 1-month time window, and  $q$  and  $\tau$  are arbitrary constants. The shaded area indicates the regime where innovators adopt approximately with constant rate. **(b)** Null model rates where times of adoption are randomly shuffled. **(c)** Empirical connected-component size distribution at different times for the adoption [ $P(s_a)$ , main panel] and stable adoption [ $P(s_s)$ , inset] networks, with  $s_a$  and  $s_s$  relative to system size. **(d)** Empirical connected-component size distribution  $P(s_v)$  for the relative size of innovator-induced vulnerable trees at different times. This figure is adopted from Ref. [70] and it is licensed under Creative Commons Attribution 4.0 International Licences

constant rate (denoted by the grey shaded area in Fig. 3a). Second, vulnerable nodes adopt approximately with the same rate as innovators, which suggests a strong correlation between these types of adoption. This stationary behaviour is rather surprising as environmental effects, like competition or marketing campaigns, could potentially influence the adoption dynamics. On the other hand, the overall adoption process accelerates due to the increasing rate of stable adoptions induced by social influence.

To better understand how innovation spreads throughout the social network, we take a closer look at the internal structure of the service adoption process. To do so, we consider individual adoption times and construct an evolving adoption network, where links exist between users who have adopted the service before time  $t$  and are connected in the social network underneath. In order to avoid the effect of instantaneous group adoptions (evidently not driven by social influence), we only consider links between connected nodes whose adoption did not happen at the same time. This way links in the adoption graph indicate ties where social

influence among individuals could have existed. By observing the evolution of the adoption network, we are interested in its connectedness and its composition of sub-components of adopters of different kinds.

The size distribution  $P(s_a)$  of connected components in the adoption network shows the emergence of a giant percolating component over time (Fig. 3c main panel), along with several other small clusters. Moreover, after decomposition we observe that the giant cluster builds up from several innovator seeds that induce small vulnerable trees locally (Fig. 3d), each with small depth [12, 70, 75]. At the same time the stable adoption network (considering connections between all stable adopters at the time) has a giant connected component, indicating the emergence of a percolating stable cluster with size comparable to the largest adoption cluster (Fig. 3c, inset). These observations suggest a scenario for the evolution of the global adoption component where multiple innovators adopt at different times and trigger local vulnerable trees, which in turn induce a percolating component of the connected stable nodes holding the global adoption cluster together. Consequently, in the structure of the adoption network primary triggering effects are important only locally, while external and secondary triggering mechanisms seem to be responsible for the emergence of global-scale adoption.

Despite this expansion dynamics and connected structure of the service adoption network, we need to take a closer look at spurious effects, which could potentially induce the observed behaviour. First, during our analysis we assume that the adoption process is exclusively driven by social influence, without any direct information about the presence of the influence itself. One can argue that the observed phenomena is simply explained by homophily, i.e. by frequent links between people who are both interested in the given service and who would adopt independently from each other. Second, the service reaches less than 6% of the total number of active Skype users over a period of 7 years [71]. Since this adopting minority is connected within a giant adopting cluster, it may indicate local effects of social influence but also raises the question about the role of non-adopting users. Finally, we observe that the giant adoption cluster evolves over several years, which could simply be the consequence of individual decisions of users to wait to adopt the service even after their threshold has been reached. In the following we further investigate these questions to better understand the adoption process. First we present a null model study to underline the overall effects of social influence as compared to homophily; we also perform a time re-scaling experiment to explore the role of waiting times on the global adoption dynamics; and finally we propose a dynamical threshold model [69, 70], which helps us understand the role of multiple innovators and non-adopters in the unfolding of the service adoption processes.

## 2.4 Social Influence vs. Homophily

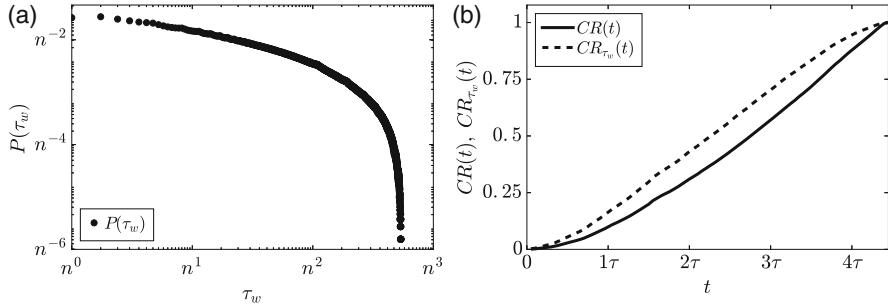
Studies of social contagion phenomena assume that social influence is responsible for the correlated adoption of connected people. However, an alternative explanation for the observed correlated adoption patterns is homophily: a link creation

mechanism by which similar egos get connected in a social structure. In the latter case, the correlated adoption of a connected group of people would be explained by their similarity and not necessarily due to social influence. Homophily and influence are two processes that may simultaneously play a role during the adoption process. Nevertheless, distinguishing between them on the individual level is very challenging using our or any similar datasets [34, 35]. Fortunately, at the system level one may identify which process is dominant in the empirical data. To do that we first need to elaborate on the differences between these two processes.

Influence-driven adoption of an ego may take place once one or more of its neighbours have adopted, since then their actions may influence the decision of the central ego. Consequently, the time ordering of adoptions of the ego and its neighbours matters. Homophily-driven adoption is, however, different. Homophily drives social tie formation such that similar people tend to be connected in the social structure. In this case connected people may adopt because they have similar interests, but the time ordering of their adoptions would not matter. Therefore, it is valid to assume that adoption could evolve in clusters due to homophily, but these adoptions would appear in a more-or-less random order.

To test this hypothesis we define a null model where we take the adoption times of users and shuffle them randomly among all adopting egos. This way a randomly selected time is assigned to each adopter, while the adoption rate and the final set of adopters remain the same. Moreover, this procedure only destroys correlations between adoption events induced by social influence, but keeps the social network structure and node degrees unchanged. In this way, during the null model process the same egos appear as adopters, but the rates of adoption may in principle change (or not), corresponding to social influence (or homophily) as a dominant factor during the adoption process. If adoption is mostly driven by homophily, the rates of adoption would not change considerably beyond statistical fluctuations. On the other hand, if social influence plays a role in the process, rates of adoption in the null model should be very different from the empirical curves, implying that the time ordering of events matters in the adoption process. In this case, the rate of innovators should be higher than in the empirical data, since nodes that are in the adoption cluster originally without being directly connected would have a greater chance to appear as innovators, due to a random adoption time that is not conditional to the time ordering of the adopting neighbours.

After calculating the adoption rates of different user groups in the shuffled null model, we observe the latter situation (Fig. 3b): the rate of innovators becomes dominant, while the rates of stable and vulnerable adoptions drop considerably as they appear only by chance. This suggests that the temporal ordering of adoption events matters a lot in the evolution of the observed adoption patterns, and thus social influence may play a strong role here. Of course one cannot decide whether influence is solely driving the process or homophily has some impact on it; in reality it probably does to some extent. However, we can use this null model measure to demonstrate the presence and importance of the mechanism of influence during the adoption process.



**Fig. 4** The waiting time distribution and its effect on the adoption process. **(a)** Distribution  $P(\tau_w)$  of times between the last adoption in the egocentric network of an individual and his/her own adoption. **(b)** Cumulative adoption rates before and after the removal of waiting times [ $CR(t)$  and  $CR_{\tau_w}(t)$ , respectively].  $n$  and  $\tau$  are arbitrary constant values. This figure is adopted from Ref. [70] and it is licensed under Creative Commons Attribution 4.0 International Licences

## 2.5 Waiting Time of Adoption

As we mentioned earlier, one reason behind the slow evolution of the adoption process could be due to the time users wait after their personal adoption threshold is reached and before adopting the service. This lag in adoption can be due to individual characteristics, or can come from the fact that social influence does not spread instantaneously (as commonly assumed in threshold models). This waiting time  $\tau_w$  can be estimated by measuring the time difference between the last adoption in a user's egocentric network and the time of his/her adoption. This time is  $\tau_w = 0$  by definition for innovators, but  $\tau_w$  can take any positive value for vulnerable and stable adopters up to the length of the observation period.

We find that waiting times are broadly distributed for adopters in our dataset (Fig. 4a), meaning that many users adopt the service shortly after their personal threshold is reached, but a considerable fraction waits long before adopting the service. This heterogeneous nature of waiting times may be a key element behind the observed adoption dynamics. One way to figure out its effect on the speed of cascade evolution is by removing them. We can extract the waiting time from the adoption time of adopters and assign a rescaled adoption time for each of them. The rescaled adoption time of a user is the last time when his/her fraction of adopting neighbours changed and the adoption threshold was (hypothetically) reached. After this procedure, we can calculate a new adoption rate function by using the rescaled adoption times and compare this rate to the original. From Fig. 4b we conclude that although adoption becomes faster, the rescaled adoption dynamics is still not rapid. On the contrary, it suggests that the rescaled adoption dynamics is still very slow and quite similar to the original. Consequently, waiting times cannot explain the observed slow dynamics of adoption.

Note that long waiting times can have a further effect on the measured dynamics. After the “real” threshold of a user is reached and he/she waits to adopt, some

neighbours may adopt the product. Hence all observed measures are in this sense “effective”: observed thresholds are larger or equal than real thresholds; the innovator rate is smaller or equal; the vulnerable and stable rates will be larger or equal; and waiting times will be shorter or equal than the real values. Consequently the process may actually be faster than that we observe in Fig. 4b after removing the effective waiting times. However, this bias becomes important only after the majority of individuals in the social network has adopted the service and the spontaneous emergence of adopting neighbours becomes more frequent. As the fraction of adopters in our dataset is always less than 6% [71], we expect minor effects of this observational bias on our measurements.

### 3 Modelling Social Contagion

In order to understand better the possible microscopic mechanisms behind the empirical observations of online service adoption described previously, we introduce and analyse two agent-based network models of threshold-driven social contagion. First we discuss the WT model as originally proposed by Watts [24], and secondly an extended, dynamical threshold model devised by us [69, 70], where both multiple innovators and non-adopters have a role in social contagion.

#### 3.1 The Watts Model

Under the complex contagion hypothesis by Granovetter, Centola, and others [6, 22], social contagion may be modelled as a binary-state process evolving in a network and driven by a threshold mechanism. In this framework individuals are represented by agents or network nodes, each in either a susceptible (0) or adopter (1) state, while the influence by an agent is achieved by transferring information via social ties. Nodes are connected in a network with degree distribution  $P(k)$  and average degree  $z = \langle k \rangle$ . Moreover, each node has an individual threshold  $\varphi \in [0, 1]$  drawn from a distribution  $P(\varphi)$  with average  $w = \langle \varphi \rangle$ . The threshold  $\varphi$  determines the minimum fraction of exposed neighbours that triggers adoption, capturing the resistance of an individual against engaging in a given behaviour. Hence, in case a node has  $m$  adopting neighbours and  $m \geq k\varphi$  (the so-called *threshold rule*), it switches state from 0 to 1 and remains so until the end of the dynamics. In his seminal paper about threshold dynamics [24], Watts classified nodes into three categories based on their threshold and degree: He first identified *innovator* nodes that spontaneously change state to 1 and therefore start the spreading process. Such nodes have a trivial threshold  $\varphi = 0$ . Then there are nodes with threshold  $0 < \varphi \leq 1/k$ , called *vulnerable*, which need one adopting neighbour before their own adoption. Finally, there are more resilient nodes with threshold  $\varphi > 1/k$ , known as *stable*, representing individuals in need of strong social influence to follow the actions of their acquaintances.

In the WT model [24], small perturbations (like the spontaneous adoption of a single seed node) can trigger network-wide cascading patterns. However, their emergence is subject to the following *cascade condition*: the innovator seed has to be linked to a percolating vulnerable cluster, which adopts immediately afterwards and further triggers a *global cascade* (i.e. a set of adopters larger than a fixed fraction of a finite network, or a nonzero fraction of adopters in an infinite network). The cascade condition is satisfied if the network is inside a bounded regime in  $(w, z)$ -space [24]. When considering a vanishingly small innovator seed and a configuration-model network [72] [i.e., by ignoring structural correlations in the social network and characterising it solely by its degree distribution  $P(k)$ ], a generating function approach allows us to write the cascade condition as

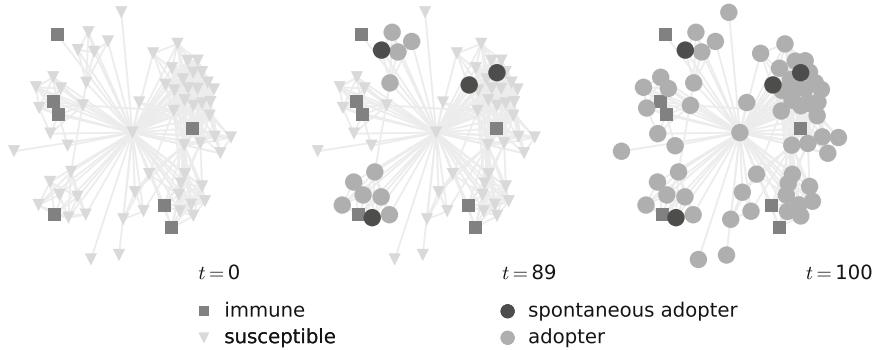
$$\sum_k \frac{k}{z} (k-1) P(k) f(k, 1) > 1, \quad (3)$$

where  $f(k, 1) = C(1/k)$  is the probability that a randomly-selected node with degree  $k$  is vulnerable, and  $C$  is the cumulative distribution function of  $P(\varphi)$ . More generally,  $f(k, m)$  (for  $m = 0, \dots, k$ ) is also known as a response function of the monotone binary dynamics defining the WT model [23, 69].

As Eq. (3) shows, the cascade regime depends on degree and threshold heterogeneities [24] and may change its shape if several innovators start the process [61]. In addition, while models with more sophisticated functional forms of social influence may be introduced [36, 76], the original assumption proposed by Watts and Granovetter seems to be sufficient to interpret our observations.

### 3.2 Dynamical Threshold Model with Immune Nodes

Our modelling framework is an extension to the WT model and similar threshold dynamics on networks, studied by Watts, Gleeson, Singh, and others, where all the nodes are initially susceptible and innovators are only introduced as an initial seed of arbitrary size [24, 30, 61, 62]. Apart from the above discussed threshold rule and motivated by the empirical observations in the spread of online services within Skype, our model considers two additional features, namely that (1) a fraction  $r$  of “immune” nodes never adopts, indicating a lack of interest in the online service, and that (2) due to external influence, susceptible nodes adopt the service spontaneously (i.e. become innovators) throughout the time with constant rate  $p_n$ , rather than only at the beginning of the dynamics. In this way, the dynamical evolution of the system is completely determined by the online social network, the distribution  $P(\varphi)$  of thresholds, and the parameters  $r$  and  $p_n$  (Fig. 5). For the sake of simplicity, we consider a configuration-model network and statistical independence between degrees and thresholds [57, 78, 79]. We remark that the somewhat similar concepts of “stubborn nodes”, mimicking individuals’ resistance against adoption [80, 81], and “global nodes”, capturing adoption driven by external effects [82], have also

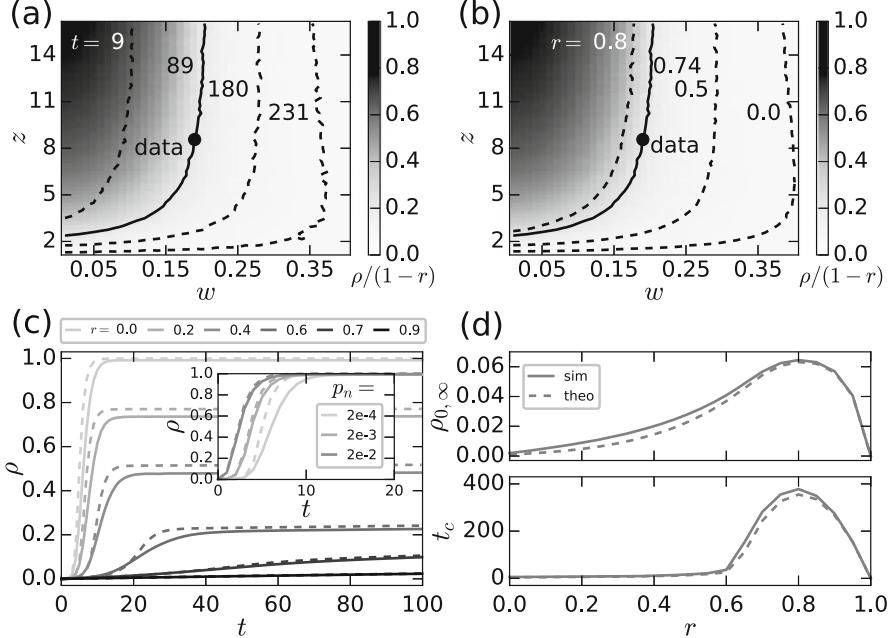


**Fig. 5** Immune individuals in social contagion. Numerical simulation of our dynamical threshold model in an empirical network, with a single adoption threshold  $\varphi = 0.2$  for all the nodes, rate of spontaneous adopters  $p_n = 0.0005$ , and fraction of immune nodes  $r = 0.1$ . The network is an ego sample of Facebook friendships with size  $N = 96$  and average degree  $z = 10.63$  [77]. The network shows how susceptible nodes adopt spontaneously with rate  $p_n$ , or after a fraction  $\varphi$  of their neighbours has adopted, while immune nodes never adopt. Reprinted figure with permission from Ruan et al. [69]. Copyright 2018 by the American Physical Society

been considered in threshold models and show a rich variety of effects on cascading behaviour.

As we show in the Appendix, our threshold model [69, 70] can be studied analytically by extending the framework of approximate master equations (AMEs) for monotone binary-state dynamics recently developed by Gleeson [57, 78, 79], where the transition rate between susceptible and adoption states only depends on the number  $m$  of network neighbours that have already adopted. We may also implement the model numerically via a Monte Carlo simulation in a network of size  $N$ , with a log-normal degree distribution and a log-normal threshold distribution as observed empirically in the case of Skype. Hence we can explore the behaviour of the fractions of adopters and innovators in the network,  $\rho$  and  $\rho_0$ , as a function of  $z$ ,  $w$ ,  $p_n$  and  $r$ , both in the numerical simulation and in the theoretical approximation given by Eqs. (9) and (12) (see Appendix). For  $p_n > 0$  some nodes adopt spontaneously as time passes by, leading to a frozen state characterised by the final fraction of adopters  $\rho(\infty) = 1 - r$ . However, the time needed to reach such a state depends heavily on the distribution of degrees and thresholds, as indicated by a region of large adoption ( $\rho \approx 1 - r$ ) that grows in  $(w, z)$ -space with time (contour lines in Fig. 6a). If we fix the time in the dynamics and vary the fraction of immune nodes instead, this region shrinks as  $r$  increases (contour lines in Fig. 6b). In other words, the set of networks (defined by their average degree and threshold) that allow the spread of adoption is larger at later times in the dynamics, or when the fraction of immune nodes is small. When both  $t$  and  $r$  are fixed, the normalised fraction of adopters  $\rho/(1 - r)$  gradually decreases for less connected networks with larger thresholds (surface plot in Fig. 6a, b).

Both numerical simulations and analytical approximations show how the dynamics of spreading changes by introducing immune individuals in the social network.



**Fig. 6** A dynamical threshold model for the adoption of online services. **(a, b)** Surface plot of the normalised fraction of adopters  $\rho/(1-r)$  in  $(w, z)$ -space, for  $r = 0.73$  and  $t = 89$ . Contour lines signal the parameter values for which 20% of non-immune nodes have adopted, for fixed  $r$  and varying time **(a)**, and for fixed time and varying  $r$  **(b)**. The continuous contour line and dot indicate parameter values of the last observation of Skype s3. A regime of maximal adoption ( $\rho \approx 1-r$ ) grows as time goes by, and shrinks for larger  $r$ . **(c)** Time series of the fraction of adopters  $\rho$  for fixed  $p_n = 0.00019$  and varying  $r$  (main), and for fixed  $r = 0$  and varying  $p_n$  (inset). These curves are well approximated by the solution of Eq. (9) for  $k_0 = 3$ ,  $k_{M-1} = 150$  and  $M = 25$  (dashed lines). The dynamics is clearly faster for larger  $p_n$  values. As  $r$  increases, the system enters a regime where the dynamics is slowed down and adopters are mostly innovators. **(d)** Final fraction of innovators  $\rho_{0,\infty}$  and the time  $t_c$  when 50% of non-immune nodes have adopted as a function of  $r$ , both simulated and theoretical. The crossover to a regime of slow adoption is characterised by a maximal fraction of innovators and time  $t_c$ . Unless otherwise stated,  $p_n = 0.00019$  and we use  $N = 10^4$ ,  $\mu_D = 1.09$ ,  $\sigma_D = 1.39$ ,  $k_{\min} = 1$ ,  $\mu_T = -2$ , and  $\sigma_T = 1$  to obtain  $z = 8.56$  and  $w = 0.19$  as in Skype s3. The difference in  $\mu_D$  between data and model is due to finite-size effects. Numerical results are averaged over  $10^2$  **(a, b)** and  $10^3$  **(c, d)** realisations. This figure is adopted from Ref. [70] and it is licensed under Creative Commons Attribution 4.0 International Licences

For  $r \approx 0$ , the adoption cascade appears sooner for larger  $p_n$ , since this parameter regulates how quickly we reach the critical fraction of innovators necessary to trigger a cascade of fast adoption throughout all susceptible nodes (Fig. 6c, inset). Yet as we increase  $r$  above a critical value  $r_c$  (and thus introduce random quenching), the system enters a regime where rapid cascades disappear and adoption is slowed down, since stable nodes have more immune neighbours and it is difficult to fulfil their threshold condition. The crossover between these fast and slow regimes is gradual, as seen in the shape of  $\rho$  for increasing  $r$  (Fig. 6c, main panel). We may

identify  $r_c$  in various ways: by the maximum in both the final fraction of innovators  $\rho_{0,\infty} = \rho_0(\infty)$  and the critical time  $t_c$  when  $\rho = (1 - r)/2$  (Fig. 6d), or as the  $r$  value where the inflection point in  $\rho$  disappears. These measures indicate  $r_c \approx 0.8$  for parameter values calibrated with Skype data. All global properties of the dynamics (like the functional dependence of  $\rho$  and  $\rho_0$ ) are very well approximated by the solution of Eqs. (9) and (12) (dashed lines in Fig. 6c, d). Indeed, the AME framework is able to capture the shape of the  $\rho$  time series, the crossover between regimes of fast and slow adoption, as well as the maximum in  $\rho_{0,\infty}$  and  $t_c$ .

In the simplified case of an Erdős-Rényi random graph as the underlying social network, the crossover between fast and slow regimes of spreading may also be characterised by a percolation-type transition in the asymptotic limit ( $t \rightarrow \infty$ ) of the size distribution  $P(s)$  of *induced* adoption clusters, i.e. connected components of adopters disregarding innovators [69]. For early times  $P(s)$  includes small induced clusters only, which in turn indicates that a larger fraction of spontaneous adopters is crucial for global spreading in the absence of a percolating vulnerable component. However, for late times the behaviour of  $P(s)$  differs between regimes: in the regime of fast spreading the distribution becomes bimodal due to the appearance of a global cluster of induced adopters, while in the slow regime it remains unimodal until the end of dynamics.

Finally, in the extreme case of  $p_n = 0$  (corresponding to the WT model with immune nodes), the reduced AME system of Eq. (9) can be used to derive a cascade condition and thus give insight into the dynamics of spreading in the presence of immune individuals [23, 69]. Equation (9) has an equilibrium point for the initial condition  $(\rho(0), v(0)) = (0, 0)$ . If this equilibrium point is linearly unstable, the perturbation of a single innovator seed may move the dynamical system away from equilibrium and create a global cascade. A linear stability analysis shows that this condition is equivalent to

$$(1 - r) \sum_k \frac{k}{z} (k - 1) P(k) f(k, 1) > 1, \quad (4)$$

where  $f(k, 1) = C(1/k)$  implements the response of a non-immune node of degree  $k$  to one adopting neighbour, and  $C$  is the cumulative distribution function of  $P(\varphi)$  (for non-immune nodes with  $c > 0$ ). When  $r = 0$ , Eq. (4) reduces trivially to the cascade condition of the original WT model in Eq. (3). This shows that the shape of the cascade regime can be obtained either by using generating functions in percolation theory or by performing a stability analysis of the AMEs.

## 4 Validation

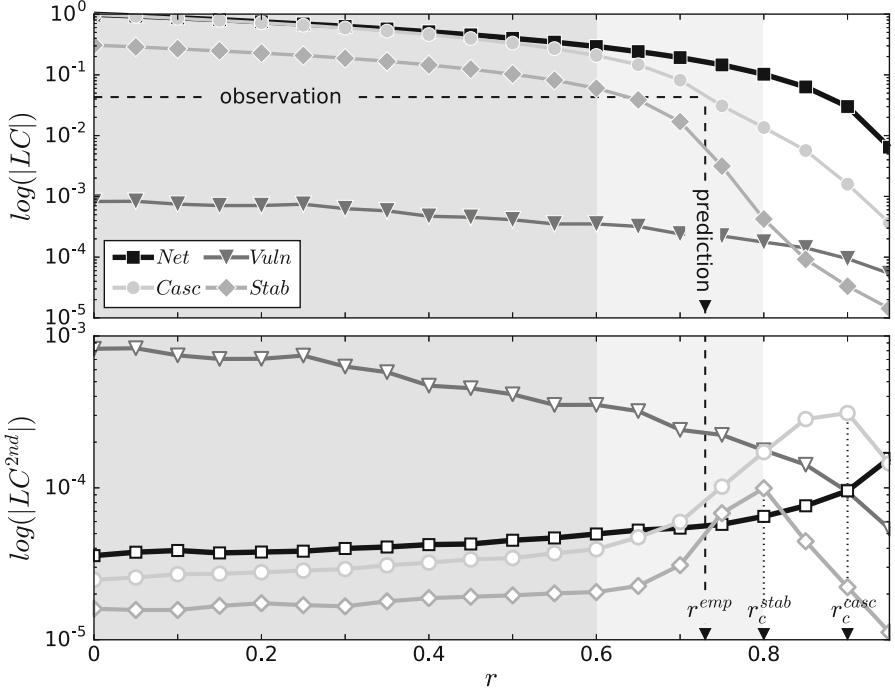
As demonstrated above, our model provides insight on the role of innovators and immune nodes in controlling the speed of the adoption process. However, in empirical datasets information about the fraction of non-adopters is usually not

available, which makes it difficult to predict the future dynamics of service adoption. Here we use our modelling framework to perform data-driven simulations with parameters determined from Skype for two reasons: (a) to estimate the fraction  $r$  of immune nodes in the real system; and (b) to validate our modelling as compared to real data.

To set up our data-driven simulations we use the Skype data to directly determine all model parameters, apart from the fraction  $r$  of immune nodes. As we already discussed, the best approximation of the degree distribution of the real network is a log-normal function (Eq. (1)) with parameters  $\mu_D = 1.2$ ,  $\sigma_D = 1.39$ , minimum degree  $k_{\min} = 1$  and average degree  $z = 8.56$ . To account for finite-size effects in the model results for low  $N$ , we decrease  $\mu_D$  slightly to obtain the same value of  $z$  as in the real network. We also observe in Fig. 2b that the threshold distribution of each degree group collapses into a master curve after normalisation by using the scaling relation  $P(\Phi_k, k) = kP(\Phi_k/k)$ . This master curve can be well-approximated by the log-normal distribution shown in Eq. (2), with parameters  $\mu_T = -2$  and  $\sigma_T = 1$  as determined by the empirical average threshold  $w = 0.19$  and standard deviation 0.233. We estimate a rate of innovators  $p_n = 0.00019$  by fitting a constant function to  $R_i(t)$  for  $t > 2\tau$  (Fig. 3a). The fit to  $p_n$  also matches the time-scale of a Monte Carlo iteration in the model to 1 month. To model the observed dynamics and explore the effect of immune nodes, we use a configuration-model network [72] with log-normal degree and threshold distributions and  $p_n$  as the constant rate of innovators, all determined from the empirical data. Model results in Fig. 7 (and Fig. 8) are averaged over 100 networks of size  $N = 10^5$  ( $10^6$ ) after  $T = 89$  iterations, matching the length of the observation period in Skype.

As a function of  $r$ , the underlying and adoption networks pass through three percolation-type phase transitions. First, the appearance of immune nodes (for increasing  $r$ ) can be considered as a removal process of nodes available for adoption from the underlying network structure. After the appearance of a critical fraction of immune nodes,  $r_c^{\text{net}}$ , the effective network structure available for adoption will be fragmented and will consist of small components only, limiting the size of the largest adoption cluster possible. Second,  $r$  also controls the size of the emergent adoption cascades evolving on top of the network structure. While for small  $r$  the adoption network is connected into a large component, for larger  $r$  cascades cannot evolve since there are not enough nodes to fulfil the threshold condition of susceptible stable nodes, even if the underlying network is still connected. The transition point between these two phases of the adoption network is located at  $r_c^{\text{casc}} \leq r_c^{\text{net}}$ , limited from above by the critical point  $r_c^{\text{net}}$ . Finally, we observe from the empirical data and model results that the adoption network is held together by a large connected component of stable nodes. Consequently, for increasing  $r$  the stable adoption network goes through a percolation transition as well, with a critical point  $r_c^{\text{stab}} \leq r_c^{\text{casc}} \leq r_c^{\text{net}}$ .

To characterise these percolation phase transitions we compute the average size of the largest ( $LC$ ) and second largest ( $LC^{2\text{nd}}$ ) connected components (Fig. 7). We measure these quantities for the underlying network, and for the stable, vulnerable and global adoption networks, as a function of the fraction of immune nodes  $r$ .



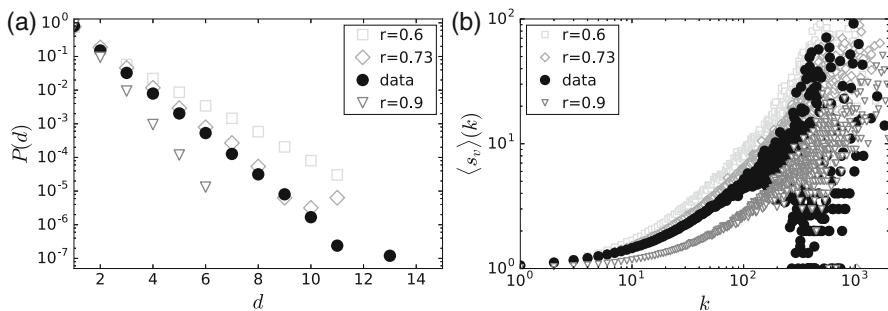
**Fig. 7** Empirical cluster statistics and simulation results. Average size of the largest ( $LC$ , upper panel) and the 2nd largest ( $LC^{2nd}$ , lower panel) components of the model network (“Net”, squares), adoption network (“Casc”, circles), stable network (“Stab”, diamonds), and induced vulnerable trees (“Vuln”, triangles) as a function of the fraction  $r$  of immune nodes. Dashed lines show the observed relative size of the real  $LC$  of the adopter network in 2011 (Fig. 3c) and the predicted  $r^{emp}$  value. Dotted lines on the lower panel indicate the critical percolation points for the full ( $r_c^{casc}$ ) and stable ( $r_c^{stab}$ ) adoption networks. This figure is adopted from Ref. [70] and it is licensed under Creative Commons Attribution 4.0 International Licences

After  $T = 89$  iterations (matching the length of the real observation period), we identify three regimes of the dynamics: if  $0 < r < 0.6$  (dark-shaded area) the spreading process is very rapid and evolves as a global cascade, which reaches most of the nodes of the shrinking susceptible network in a few iteration steps. About 10% of adopters are connected in a percolating stable cluster, while vulnerable components remain very small in accordance with empirical observations. In the crossover regime  $0.6 < r < 0.8$  (light-shaded area), the adoption process slows down considerably (Fig. 7, upper panel), as stable adoptions become less likely due to the quenching effect of immune nodes. The adoption process becomes the slowest at  $r_c^{stab} = 0.8$  when the percolating stable cluster falls apart, as demonstrated by a peak in the corresponding  $LC^{2nd}$  curve in Fig. 7 (diamonds in lower panel). Finally, around  $r_c^{casc} = 0.9$  the adoption network becomes fragmented and no global cascade takes place. Since the underlying network has a broad degree distribution, it is robust against random node removal processes [72]. That is why

its critical percolation point  $r_c^{\text{net}}$  appears after 95% or more nodes are immune. Note that similar calculations for another service have been presented before [70] with qualitatively the same results, but with the crossover regime shifted towards larger  $r$  due to different parameter values of the model process.

We can use these calculations to estimate the only unknown parameter, namely the fraction  $r$  of immune nodes in Skype, by matching the relative size of the largest component ( $LC_{\text{Net}}$ ) between real and model adoption networks at time  $T$ . Empirically, this value is the relative size  $s_a^{LC} \simeq 0.043$  corresponding to the last point on the right-hand side of the distribution for 2011 in Fig. 3c (main panel). Matching this relative size with the simulation results (see the observation line in Fig. 7 upper panel), we find that it corresponds to  $r^{\text{emp}} = 0.73$  (prediction line in Fig. 7), suggesting that the real adoption process lies in the crossover regime. In other words, large adoption cascades could potentially evolve in Skype but with reduced speed, as 73% of users might not be interested in adopting a service within the network.

To test the validity of the predicted  $r^{\text{emp}}$  value we perform three different calculations. First we measure the maximum relative growth rate of cumulative adoptions and find a good match between model and data (see Skype s3 and Model Skype s3 in Fig. 1). In other words, the model correctly estimates the speed of the adoption process. Second, we measure the distribution  $P(d)$  of the depths of induced vulnerable trees (Fig. 8a). Vulnerable trees evolve with a shallow structure in the empirical and model processes. After measuring the distribution  $P(d)$  for various  $r$  values below, above and at  $r^{\text{emp}}$ , we find that the distribution corresponding to the predicted  $r^{\text{emp}}$  value fits the best with the empirical data. Finally, in order to verify earlier theoretical suggestions [61], we look at the correlation  $\langle s_v \rangle(k)$  between the degree of innovators and the average size of vulnerable trees induced by them (Fig. 8b). Similar to the distribution  $P(d)$ , we perform this measurement on the



**Fig. 8** Additional empirical cluster statistics and simulation results. (a) Distribution  $P(d)$  of depths of induced vulnerable trees in both data and model for several  $r$  values, showing a good fit with the data for  $r = 0.73$ . The difference in the tail is due to finite-size effects. (b) Correlation  $\langle s_v \rangle(k)$  between innovator degree and average size of vulnerable trees in both data and model with the same  $r$  values as in (a). Model calculations correspond to networks of size  $N = 10^6$  and are averaged over  $10^2$  realisations. This figure is adopted from Ref. [70] and it is licensed under Creative Commons Attribution 4.0 International Licences

real data and in the model for  $r = 0.6$  and  $0.9$ , as well as for the predicted value  $r_{emp} = 0.73$ . We find a strong positive correlation in the data, explained partially by degree heterogeneities in the underlying social network, but surprisingly well emulated by the model as well. More importantly, although this quantity appears to scale with  $r$ , the estimated  $r$  value fits the empirical data remarkably well, thus validating our estimation method for  $r$  based on a matching of relative component sizes.

## 5 Conclusion and Future Directions

The analysis and modelling of the diffusion of services and innovations is a long-standing scientific challenge, with recent developments built on large digital datasets registering adoption processes in a society with a large population. Due to these advancements we are currently at the position to simultaneously observe various types of adoption processes and the underlying social structure. Individual-level observations of social and adoption behaviour are crucial in identifying the mechanisms that fuel collective patterns of rapid or slow adoption cascades. In this chapter, using one of the first datasets of this kind, we observe the worldwide spread of an online service in the techno-social communication network of Skype. First we provide novel empirical evidence about heterogeneous adoption thresholds and non-linear dynamics of the adoption process. We have also identified two additional components necessary to introduce into the modelling of product adoption, namely (a) a constant flow of innovators, which may induce rapid adoption cascades even if the system is initially out of the cascading regime, and (b) a fraction of immune nodes that forces the system into a quenched state where adoption slows down. These features are responsible for a critical structure of empirical adoption components that radically differs from previous theoretical expectations. We incorporate these mechanisms into a threshold model that, despite containing several simplifying assumptions, successfully recovers and predicts real-world adoption scenarios such as the spreading of Skype services.

Our aim in this chapter has been to provide empirical observations as well as methods and tools to model the dynamics of social contagion phenomena, with the hope that it will foster thoughts for future research. One possible direction is the observation of the reported structure and evolution of the global adoption cluster in other systems similar to the ones studied in [12, 13, 41, 43, 44, 75]. Other promising directions are the consideration of structural homophilic or assortative correlations, the evolving nature of the underpinning social network with timely created and dissolved social ties (as studied in [19]), and the effects of interpersonal influence or leader-follower mechanisms on the social contagion process. We hope that our results provide a direction for data-driven modelling of these phenomena, and serve as a scholarly example in future studies of the dynamics of service adoption processes.

**Acknowledgements** The results presented in this chapter are adapted from [69, 70] and were obtained in collaboration with Riivo Kikas. The authors gratefully acknowledge the support of M. Dumas, A. Saabas, and A. Dumitras from STACC and Microsoft/Skype Labs. GI acknowledges a Visiting Fellowship from the Aalto Science Institute. JK and ZR were supported by FP7 317532 Multiplex and JK by H2020 FETPROACT-GSS CIMPLEX 641191. KK is supported by the Academy of Finland's project COSDYN project, No. 276439 and EU HORIZON 2020 FET Open RIA IBSEN project No. 662725.

## Appendix: Analytical Treatment of the Model

Our threshold model [69, 70] may be studied analytically by extending the AME framework for monotone binary-state dynamics [57, 78, 79], where the transition rate between susceptible and adoption states only depends on the number  $m$  of network neighbours that have already adopted. We describe a node by the property vector  $\mathbf{k} = (k, c)$ , where  $k = k_0, k_1, \dots, k_{M-1}$  is its degree and  $c = 0, 1, \dots, M$  its type, i.e.  $c = 0$  is the type of the fraction  $r$  of immune nodes, while  $c \neq 0$  is the type of all non-immune nodes that have threshold  $\varphi_c$ . In this way,  $P(\varphi)$  is substituted by the discrete distribution of types  $P(c)$  (for  $c > 0$ ). The integer  $M$  is the maximum number of degrees (or non-zero types) considered in the AME framework, which can be increased to improve the accuracy of the analytical approximation at the expense of speed in its numerical computation.

We characterise the static social network by the extended distribution  $P(\mathbf{k})$ , where  $P(\mathbf{k}) = rP(k)$  for  $c = 0$  and  $P(\mathbf{k}) = (1 - r)P(k)P(c)$  for  $c > 0$ . Non-immune and susceptible nodes with property vector  $\mathbf{k}$  adopt spontaneously with a constant rate  $p_n$ , otherwise they adopt only if a fraction  $\varphi_c$  of their  $k$  neighbours has adopted before. These rules are condensed into the probability  $F_{\mathbf{k},m}dt$  that a node will adopt within a small time interval  $dt$ , given that  $m$  of its neighbours are already adopters,

$$F_{\mathbf{k},m} = \begin{cases} p_r & \text{if } m < k\varphi_c \\ 1 & \text{if } m \geq k\varphi_c \end{cases}, \quad \forall m \text{ and } k, c \neq 0, \quad (5)$$

with  $F_{(k,0),m} = 0 \forall k, m$  and  $F_{(0,c),0} = p_r \forall c \neq 0$  (for immune and isolated nodes, respectively). The rescaled rate  $p_r = p_n/(1 - r)$  (with  $p_r = 1$  for  $p_n > 1 - r$ ) is necessary if we wish to obtain a rate  $p_n$  of innovators for early times of the dynamics, regardless of the value of  $r$ .

The dynamics of adoption is well described by an AME for the fraction  $s_{\mathbf{k},m}(t)$  of  $\mathbf{k}$ -nodes that are susceptible at time  $t$  and have  $m = 0, \dots, k$  adopting neighbours [23, 78, 79],

$$\dot{s}_{\mathbf{k},m} = -F_{\mathbf{k},m}s_{\mathbf{k},m} - \beta_s(k - m)s_{\mathbf{k},m} + \beta_s(k - m + 1)s_{\mathbf{k},m-1}, \quad (6)$$

where

$$\beta_s(t) = \frac{\sum_{\mathbf{k}} P(\mathbf{k}) \sum_m (k-m) F_{\mathbf{k},m} s_{\mathbf{k},m}(t)}{\sum_{\mathbf{k}} P(\mathbf{k}) \sum_m (k-m) s_{\mathbf{k},m}(t)}, \quad (7)$$

and the sum is over all the degrees and types, i.e.  $\sum_{\mathbf{k}} \bullet = \sum_k \sum_c \bullet$ . To reduce the dimensionality of Eq. (6), we consider the ansatz

$$s_{\mathbf{k},m}(t) = B_{k,m}[\nu(t)] e^{-p_r t} \quad \text{for } m < k\varphi_c \text{ and } c \neq 0, \quad (8)$$

with  $\nu(t)$  the probability that a randomly-chosen neighbour of a susceptible node is an adopter.

Introducing the ansatz of Eq. (8) into the AME system of Eq. (6) leads to the condition  $\dot{\nu} = \beta_s(1 - \nu)$ . With some algebra, the AMEs for our dynamical threshold model are reduced to the pair of ordinary differential equations

$$\dot{\rho} = h(\nu, t) - \rho, \quad (9a)$$

$$\dot{\nu} = g(\nu, t) - \nu, \quad (9b)$$

where  $\rho(t) = 1 - \sum_{\mathbf{k}} P(\mathbf{k}) \sum_m s_{\mathbf{k},m}(t)$  is the fraction of adopters in the network, and the initial conditions are  $\rho(0) = \nu(0) = 0$ . Here,

$$h = (1 - r) \left[ f_t + (1 - f_t) \sum_{\mathbf{k}|c \neq 0} P(k) P(c) \sum_{m \geq k\varphi_c} B_{k,m}(\nu) \right], \quad (10)$$

and

$$g = (1 - r) \left[ f_t + (1 - f_t) \sum_{\mathbf{k}|c \neq 0} \frac{k}{z} P(k) P(c) \sum_{m \geq k\varphi_c} B_{k-1,m}(\nu) \right], \quad (11)$$

where  $f_t = 1 - (1 - p_r)e^{-p_r t}$ , and  $B_{k,m}(\nu) = \binom{k}{m} \nu^m (1 - \nu)^{k-m}$  is the binomial distribution. The fraction of adopters  $\rho$  is then obtained by solving Eq. (9) numerically. Since the susceptible nodes adopt spontaneously with rate  $p_n$ , the fraction of innovators  $\rho_0(t)$  in the network is given by

$$\rho_0(t) = p_r \int_0^t [1 - r - \rho(\tau)] d\tau. \quad (12)$$

## References

1. Valente TW (1996) Social network thresholds in the diffusion of innovations. *Soc Networks* 18:69–89
2. Toole JL, Cha M, González MC (2012) Modeling the adoption of innovations in the presence of geographic and media influences. *PLoS One* 7:e29528
3. Kleinberg J (2007) Cascading behavior in networks: algorithmic and economic issues. In: Nisan N et al (eds) *Algorithmic game theory*. Cambridge University Press, Cambridge
4. Rogers EM (2003) Diffusion of innovations. Simon and Schuster, London
5. Bass FM (1969) A new product growth for model consumer durables. *Manag Sci* 15:215–227
6. Granovetter M (1978) Threshold models of collective behavior. *Am J Sociol* 83:1420–1443
7. Schelling TC (1969) Models of segregation. *Am Econ Rev* 59:488–493
8. Axelrod R (1997) The dissemination of culture. *J Confl Resolut* 41:203–226
9. Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329:1194–1197
10. Zhang H, Vorobeychik Y (2016) Empirically grounded agent-based models of innovation diffusion: a critical review. *Eprint arXiv:1608.08517*
11. Kiesling E, Günther M, Stummer C, Wakolbinger LM (2012) Agent-based simulation of innovation diffusion: a review. *Cent Eur J Oper Res* 20:183–230
12. Goel S, Watts DJ, Goldstein DG (2012) The structure of online diffusion networks. Association for Computing Machinery, New York, pp 623–638
13. Borge-Holthoefer J, Baños RA, González-Bailón S, Moreno Y (2013) Cascading behaviour in complex socio-technical networks. *J Complex Net* 1:1–22
14. Goel S, Anderson A, Hofman J, Watts DJ (2015) The structural virality of online diffusion. *Manage Sci* 62:180–196
15. Castellano C, Fortunato S, Loreto V (2009) Statistical physics of social dynamics. *Rev Mod Phys* 81:591–646
16. Bakshy E, Rosenn I, Marlow C, Adamic L (2012) The role of social networks in information diffusion. Association for Computing Machinery, New York, pp 519–528
17. Holt CA (2006) Markets, games, strategic behavior. Addison Wesley, Boston
18. Bikhchandani S, Hirshleifer D, Welch I (1992) A theory of fads, fashion, custom, and cultural change as informational cascades. *J Polit Econ* 100:992–1026
19. Karsai M, Ifíñez G, Kaski K, Kertész J (2014) Complex contagion process in spreading of online innovation. *J R Soc Interface* 11:20140694
20. Romero DM, Meeder B, Kleinberg J (2011) Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. Association for Computing Machinery, New York, pp 695–704
21. Barrat A, Barthélémy M, Vespignani V (2008) *Dynamical processes on complex networks*. Cambridge University Press, Cambridge
22. Centola D, Macy M (2007) Complex contagions and the weakness of long ties. *Am J Sociol* 113:702–734
23. Porter MA, Gleeson JP (2016) *Dynamical systems on networks: a tutorial*. Springer International Publishing, New York
24. Watts DJ (2002) A simple model of global cascades on random networks. *Proc Natl Acad Sci U S A* 99:5766–5771
25. Handjani S (1997) Survival of threshold contact processes. *J Theor Probab* 10:737–746
26. Neill DB (2005) Cascade effects in heterogeneous populations. *Ration Soc* 17:191–241
27. Watts DJ, Dodds PS (2007) Influentials, networks, and public opinion formation. *J Consum Res* 34:441–458
28. Melnik S, Ward JA, Gleeson JP, Porter MA (2013) Multi-stage complex contagions. *Chaos* 23:013124
29. Gómez V, Kappen HJ, Kaltenbrunner A (2010) Modeling the structure and evolution of discussion cascades. Association for Computing Machinery, New York, pp 181–190

30. Karampourniotis PD, Sreenivasan S, Szymanski BK, Korniss G (2015) The impact of heterogeneous thresholds on social contagion with multiple initiators *PLoS One* 10:e0143020
31. Miller JC (2015) Complex contagions and hybrid phase transitions. *J Complex Net* 4:201–223
32. Onnela J-P, Reed-Tsochas F (2010) Spontaneous emergence of social influence in online systems. *Proc Natl Acad Sci U S A* 107:18375–18380
33. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. *Ann Rev Sociol* 27:415–444
34. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci U S A* 106:21544–21549
35. Shalizi CR, Thomas AC (2011) Homophily and contagion are generically confounded in observational social network studies. *Sociol Methods Res* 40:211–239
36. Latané B (1981) The psychology of social impact. *Am Psychol* 36(4):343–356
37. Centola D (2011) An experimental study of homophily in the adoption of health behavior. *Science* 334:1269–1272
38. Suri S, Watts DJ (2011) Cooperation and contagion in web-based, networked public goods experiments. *PLoS One* 6:e16836
39. Motter AE, Yang Y (2017) The unfolding and control of network cascades. *Phys Today* 70:32–39
40. González-Bailón S, Borge-Holthoefer J, Rivero A, Moreno Y (2011) The dynamics of protest recruitment through an online network. *Sci Rep* 1:197
41. Borge-Holthoefer J et al (2011) Structural and dynamical patterns on online social networks: The Spanish May 15th movement as a case study. *PLoS One* 6:e23883
42. Ellis CJ, Fender J (2011) Information cascades and revolutionary regime transitions. *Econ J* 121:763–792
43. Dow PA, Adamic LA, Friggeri A (2013) The anatomy of large Facebook cascades. *AAAI*, Boston, MA, pp 145–154
44. Gruh D, Guha R, Nowell DL, Tomkins A (2004) Information diffusion through blogspace. Association for Computing Machinery, New York, pp 491–501
45. Baños RA, Borge-Holthoefer J, Moreno Y (2013) The role of hidden influentials in the diffusion of online information cascades. *EPJ Data Sci* 2:6
46. Hale HE (2013) Regime change cascades: what we have learned from the 1848 revolutions to the 2011 Arab uprisings. *Annu Rev Polit Sci* 16:331–353
47. Leskovec J, Singh A, Kleinberg J (2006) Patterns of influence in a recommendation network, Singapore, pp 380–389
48. Leskovec J, Adamic LA, Huberman BA (2007) The dynamics of viral marketing, vol 1. Association for Computing Machinery, New York, p 5
49. Fowler JH, Christakis NA (2009) Cooperative behavior cascades in human social networks. *Proc Natl Acad Sci U S A* 107:5334–5338
50. Leskovec J, McGlohon M, Faloutsos C, Glance N, Hurst M (2007) Patterns of cascading behavior in large blog graphs, Philadelphia, PA, pp 551–556
51. Duan W, Gu B, Whinston AB (2009) Informational cascades and software adoption on the internet: an empirical investigation. *MIS Q* 33:23–48
52. Bond RM et al (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489:295–298
53. Hui C, Tyshchuk Y, Wallace WA, Magdon-Ismail M, Goldberg M (2012) Information cascades in social media in response to a crisis: a preliminary model and a case study. Association for Computing Machinery, New York, pp 653–656
54. Hodas NO, Lerman K (2014) The simple rules of social contagion. *Sci Rep* 4:4343
55. Green B, Horel T, Papachristos AV (2017) Modeling contagion through social networks to explain and predict gunshot violence in Chicago, 2006 to 2014. *JAMA Intern Med* 177(3):326–333
56. Hackett A, Gleeson JP (2013) Cascades on clique-based graphs. *Phys Rev E* 87:062801
57. Gleeson JP (2008) Cascades on correlated and modular random networks. *Phys Rev E* 77:046117

58. Brummitt CD, D'Souza RM, Leicht EA (2011) Suppressing cascades of load in interdependent networks. *Proc Natl Acad Sci U S A* 109:E680–E689
59. Ghosh R, Lerman K (2010) A framework for quantitative analysis of cascades on networks. Association for Computing Machinery, New York, pp 665–674
60. Hurd TR, Gleeson JP (2013) On Watts' cascade model with random link weights. *J Complex Net* 1:25–43
61. Singh P, Sreenivasan S, Szymanski BK, Korniss G (2013) Threshold-limited spreading in social networks with multiple initiators. *Sci Rep* 3:2330
62. Gleeson JP, Cahalane DJ (2007) Seed size strongly affects cascades on random networks. *Phys Rev E* 75:050101(R)
63. Gleeson JP, Cellai D, Onnela JP, Porter MA, Reed-Tsochas F (2014) A simple generative model of collective online behavior. *Proc Natl Acad Sci U S A* 111:10411–10415
64. Yağan O, Gligor V (2012) Analysis of complex contagions in random multiplex networks. *Phys Rev E* 86:036103
65. Brummitt CD, Kobayashi T (2015) Cascades in multiplex financial networks with debts of different seniority. *Phys Rev E* 91:062813
66. Karimi F, Holme P (2013) Threshold model of cascades in empirical temporal networks. *Physica A* 392:16
67. Backlund V-P, Saramäki J, Pan RK (2014) Effects of temporal correlations on cascades: threshold models on temporal networks. *Phys Rev E* 89:062815
68. White DS (2013) Social media growth 2006 to 2012. Accessed 29 Jan 2015
69. Ruan Z, Iñiguez G, Karsai M, Kertész J (2015) Kinetics of social contagion. *Phys Rev Lett* 115:218702
70. Karsai M, Iñiguez G, Kikas R, Kaski K, Kertész J (2016) Local cascades induced global contagion: How heterogeneous thresholds, exogenous effects, and unconcerned behaviour govern online adoption spreading. *Sci Rep* 6:27178
71. Morrissey RC, Goldman ND, Kennedy KP (2011) Skype S.A. United States Security Registration Statement, Amendment 3, Reg.No. 333-168646. Accessed 14 Oct 2014
72. Newman MEJ (2010) Networks: an introduction Oxford University Press, Oxford
73. Barabási A-L (2016) Network science. Cambridge University Press, Cambridge
74. Mitzenmacher M (2004) A brief history of generative models for power law and lognormal distributions. *Int Math* 1:226–251
75. Bakshy E, Hofman JM, Mason WA, Watts DJ (2011) Everyone's an influencer: quantifying influence on Twitter. Association for Computing Machinery, New York, pp 65–74
76. Dodds PS, Watts DJ (2004) Universal behavior in a generalized model of contagion. *Phys Rev Lett* 92:218701
77. Leskovec J, McAuley JJ (2012) Learning to discover social circles in ego networks. In: Pereira F et al (eds) Advances in neural information processing systems. Curran Associates, Red Hook
78. Gleeson JP (2013) Binary-state dynamics on complex networks: Pair approximation and beyond. *Phys Rev X* 3:021004
79. Gleeson JP (2011) High-accuracy approximation of binary-state dynamics on networks. *Phys Rev Lett* 107:068701
80. Brummitt CD, Lee K-M, Goh K-I (2012) Multiplexity-facilitated cascades in networks. *Phys Rev E* 85:045102(R)
81. Lee K-M, Brummitt CD, Goh K-I (2014) Threshold cascades with response heterogeneity in multiplex networks. *Phys Rev E* 90:062816
82. Kobayashi T (2015) Trend-driven information cascades on random networks. *Phys Rev E* 92:062823

# Misinformation Spreading on Facebook



Fabiana Zollo and Walter Quattrociocchi

## 1 Introduction

The rapid advance of the Internet and web technologies facilitated global communications all over the world, allowing news and information to spread rapidly and intensively. These changes led up to the formation of a new scenario, where people actively participate in both contents' production and diffusion, without the mediation of journalists or experts in the field. The emergence of such a wide, heterogeneous (and disintermediated) mass of information sources may affect contents' quality and the mechanisms behind the formation of public opinion [29, 32, 49]. Indeed, despite the enthusiastic rhetoric about collective intelligence [35], unsubstantiated or untruthful rumors reverberate on social media, contributing to the alarming phenomenon of misinformation. Since 2013, the World Economic Forum (WEF) has been placing the global danger of massive digital misinformation at the core of other technological and geopolitical risks, ranging from terrorism, to cyber attacks, up to the failure of global governance [26]. People are misinformed when they hold beliefs neglecting factual evidence, and misinformation may influence public opinion negatively. Empirical investigations have showed that, in general, people tend to resist facts, holding inaccurate factual beliefs confidently [31]. Moreover, corrections frequently fail to reduce misperceptions [39] and often act as a *backfire effect*.

Thus, beyond its undoubted benefits, a hyperconnected world may allow the viral spread of misleading information, which may have serious real-word consequences.

---

F. Zollo  
Ca' Foscari University of Venice, Venice, Italy  
e-mail: [fabiana.zollo@unive.it](mailto:fabiana.zollo@unive.it)

W. Quattrociocchi (✉)  
IMT School for Advanced Studies, Lucca, Italy  
e-mail: [walter.quattrociocchi@imtlucca.it](mailto:walter.quattrociocchi@imtlucca.it)

In that direction, examples are numerous. Inadequate health policies in South Africa led to more than 300,000 unnecessary AIDS deaths [37], however the events were exacerbated by AIDS *denialists*, who state that HIV is inoffensive and that antiretroviral drugs cause, rather than treat, AIDS. Similar considerations may be extended to the Ebola outbreak in west Africa: after the death of two people having drunk salt water, the World Health Organisation (WHO) had to restate that all rumors about hypothetical cures or practices were false and that their use could be dangerous [14]. Or again, the American case of Jade Helm 15, a military training exercise which took place in multiple US states, but turned out to be perceived as a conspiracy plot aiming at imposing martial law, to the extent that Texas Gov. Greg Abbott ordered the State Guard to monitor the operations.

Certainly, such a scenario represents a florid environment for digital wildfires, especially when combined with functional illiteracy, information overload, and *confirmation bias*—i.e., the tendency to seek, select, and interpret information coherently with one's system of beliefs [38]. On the Internet people can access always more extreme versions of their own opinions. In this way, the benefits coming from the exposure to different points of views can be dramatically reduced [34]. Individuals, and the groups that they form, may move to a more extreme point in the same direction indicated by their own preexisting beliefs; indeed, when people discuss with many like-minded others, their views become more extreme [46]. First evidences of social contagion and misperception induced by social groups emerged in the famous experiment conducted by Solomon Asch in 1955 [7]. The task of the participants was very simple: they had to match a certain line placed on a white card with the corresponding one (i.e., having the same length) among three other lines placed on another white card. The subject was one of the eight people taking part to the test, but was unaware that the others were there as part of the research. The experiment consisted of three different rounds. In the first two rounds everyone provided the right (and quite obvious) answer. In the third round some group members matched the reference line to the shorter or longer one on the second card, introducing the so-called *unexpected disturbance* [28]. Normally subjects erred less than 1% of the time; but in the third case they erred 36.8% of the time [4]. Another relevant study was conducted by James Stoner, who identified the so-called *risky shift* [45]. In the experiment people were first asked to study twelve different problems and provide their solution; after that, they had to take a final decision together, as a group. Out of thirteen groups, twelve repeatedly showed a pattern towards greater risk-taking.

Misinformation, as well of rumor spreading, deals with these and several other aspects of social dynamics. However, adoption and contagion are often illustrated under the oversimplified metaphor of the *virus*: ideas spread by “contact” and people “infected” become active spreaders in the contagion process. We believe that such a metaphor is misleading, unless we consider that the receptor of such a virus is complex and articulated. Indeed, the adoption of ideas and behaviors deals with a multitude of cognitive dimensions, such as intentionality, trust, social norms, and confirmation bias. Hence, simplistic models adapted from mathematical epidemiology are not enough to understand social contagion. It is crucial to focus

on such relevant research questions by using methods and applying tools that go beyond the pure, descriptive statistics of big data. In our view, such a challenge can be addressed by implementing a cross-methodological, interdisciplinary approach which takes advantage of both the question-framing capabilities of social sciences and the experimental and quantitative tools of hard sciences.

## 2 Outline

The chapter is structured as follows. In Sect. 3 we provide the background of our research work, as well as tools and methodology adopted; in Sect. 4 we describe the datasets; in Sect. 5 we discuss the dynamics behind information consumption and the existence of echo chambers on both the Italian and the US Facebook; in Sect. 6 we show how confirmation bias dominates information spreading; in Sect. 7 we focus on users' interaction with paradoxical and satirical information (trolls), while in Sect. 8 we analyze users' response to debunking attempts. In Sect. 9 we target the emotional dynamics inside and across echo chambers. Finally, we draw our conclusions in Sect. 10.

## 3 Background and Research Methodology

In 2009 a paper on Science [33] proclaims the birth of the Computational Social Science (CSS), an emerging research field aiming at studying massive social phenomena quantitatively, by means of a multidisciplinary approach based on Computer Science, Statistics, and Social Sciences. Since CSS benefits from the large availability of data from online social networks, it is attracting researchers in ever-increasing numbers as it allows for the study of mass social dynamics at an unprecedented level of resolution. Recent studies have pointed out several important results ranging from social contagion [6, 36, 48] up to information diffusion [2, 8], passing through the virality of false claims [15, 21]. A wide literature branch is also devoted to understanding the spread of rumors and behaviors by focusing on structural properties of social networks to determine the way in which news spread in social networks, what makes messages go viral, and what are the characteristics of users who help spread such information [13, 15, 21, 48]. Several works investigated how social media can shape and influence the public sphere [1, 9, 17, 18], and efforts to contrast misinformation spreading range from algorithmic-based solutions up to tailored communication strategies [5, 16, 25, 42–44].

Along this path, important issues have been raised around the emergence of the *echo chambers*, enclosed systems where users are exposed only to information coherent with their own system of beliefs [47]. Many argue that such a phenomenon is directly related to the algorithms used to rank contents [40]. Speaking of this, Facebook research scientists quantified exactly how much individuals can

be exposed to ideologically diverse news and information on social media [9], finding that individual's choice about contents has an effect stronger than that of Facebook's News Feed algorithm in limiting the exposure to cross-cutting content. Undoubtedly, selective exposure to specific contents facilitates the aggregation of users in echo chambers, wherein external and contradicting versions are ignored [30]. Moreover, the lack of experts mediating the production and diffusion of content may encourage speculations, rumors, and mistrust, especially on complex issues. Pages about conspiracy theories, chem-trails, reptilians, or the link between vaccines and autism, proliferate on social networks, promoting alternative narratives often in contrast to mainstream content. Thus, misinformation online is pervasive and difficult to correct. To face the issue, several algorithmic-driven solutions have been proposed both by Google and Facebook [20, 23], that joined other major corporations to provide solutions to the problem and try to guide users through the digital information ecosystem [27]. Simultaneously, it has also been observed the rapid spread of blogs and pages devoted to debunk false claims, namely *debunkers*.

Moreover, the diffusion of unreliable content may lead to confuse unverified stories with their satirical counterparts. Indeed, it has been noticed the proliferation of satirical, wacky imitations of conspiracy theses. In this regard, there is a large community of people, known as *trolls*, behind the creation of Facebook pages aimed at diffusing caricatural and paradoxical contents mimicking conspiracy news. Their activities range from controversial comments and satirical posts, to the fabrication of purely fictitious statements, heavily unrealistic and sarcastic. According to Poe's law [3], without a blatant display of humor, it is impossible to create a parody of extremism or fundamentalism that someone won't mistake for the real thing. Hence, trolls are often accepted as realistic sources of information and, sometimes, their memes become viral and are used as evidence in online debates from real political activists. As an example, we report one of the most popular memes in Italy:

Italian Senate voted and accepted (257 in favor, 165 abstained) a law proposed by Senator Cirenga aimed at providing politicians with a 134 Billion fund to help them find a job in case of defeat in the next political competition.

It would be easy to verify that the text contains at least three false statements: (1) Senator Cirenga does not exist and has never been elected in the Italian Parliament, (2) the total number of votes is higher than the maximum possible number of voters, and (3) the amount of the fund corresponds to more than 10% of Italian GDP. Indeed, the bill is false and such a meme was created by a troll page. Nonetheless, on the wave of public discontent against Italian policy-makers, it quickly became viral, obtaining about 35K shares in less than 1 month. Nowadays, it is still one of the most popular arguments used by protesters manifesting all over Italian cities.

Such a scenario makes crucial the quantitative understanding of the social determinants related to content selection, news consumption, and beliefs formation and revision. In this essay, we focus on a collection of works [10–12, 19, 50, 51] aiming at characterizing the role of confirmation bias in viral processes online. We want to investigate the cognitive determinants behind misinformation and rumor spreading by accounting for users' behavior on different and specific narratives. In particular, we define the domain of our analysis by identifying two well-distinct

narratives: (a) conspiracy and (b) scientific information sources. Notice that we do not focus on the quality or the truth value of information, but rather on its verifiability. While producers of scientific information as well as data, methods, and outcomes are readily identifiable and available, the origins of conspiracy theories are often unknown and their content is strongly disengaged from mainstream society and sharply divergent from recommended practices.

Thus, we first analyze users' interaction with Facebook pages belonging to such distinct narratives on a time span of 5 years (2010–2014), in both the Italian and the US context. Then, we measure users' response to (1) information consistent with one's narrative, (2) troll contents, and (3) dissenting information e.g., debunking attempts.

## 4 Datasets

We identify two main categories of pages: conspiracy news—i.e., pages promoting contents *neglected* by mainstream media—and science news. The first category includes all pages diffusing conspiracy information (i.e., pages that disseminate controversial information, most often lacking supporting evidence and sometimes contradictory of the official news). Pages like *I don't trust the government*, *Awakening America*, or *Awakened Citizen* promote heterogeneous contents ranging from aliens, chem-trails, geocentrism, up to the causal relation between vaccinations and homosexuality. The second category is that of scientific dissemination and includes institutions, organizations, scientific press having the main mission to diffuse scientific knowledge. For example, pages like *Science*, *Science Daily*, and *Nature* are active in diffusing posts about the most recent scientific advances. Finally, we identify two additional categories of pages:

1. Troll: sarcastic, paradoxical messages mocking conspiracy thinking (for the Italian dataset);
2. Debunking: information aiming at correcting false conspiracy theories and untruthful rumors circulating online (for the US dataset).

To produce our datasets, we built a large atlas of Facebook public pages with the assistance of several groups (*Skeptic Forum*, *Skeptical spectacles*, *Butac*, *Protesi di Complotto*), which helped in labelling and sorting both conspiracy and scientific sources. To validate the list, all pages have then been manually checked by looking at their self-description and the type of promoted content. The exact breakdowns of the Italian and US Facebook datasets are reported in Tables 1 and 2, respectively. The entire data collection process is performed exclusively by means of the Facebook Graph API [24], which is publicly available and can be used through one's personal Facebook user account. We used only public available data (users with privacy restrictions are not included in our dataset). Data was downloaded from public Facebook pages that are public entities. Users' content contributing to such entities is also public unless users' privacy settings specify otherwise and in that case it is

**Table 1** Breakdown of the Italian Facebook dataset

	Science	Conspiracy	Troll
Pages	34	39	2
Posts	62,705	208,591	4,709
Likes	2,505,399	6,659,382	40,341
Comments	180,918	836,591	58,686
Likers	332,357	864,047	15,209
Commenters	53,438	226,534	43,102

**Table 2** Breakdown of the US Facebook dataset

	Science	Conspiracy	Debunking
Pages	83	330	66
Posts	262,815	369,420	47,780
Likes	453,966,494	145,388,117	3,986,922
Comments	22,093,692	8,304,644	429,204
Likers	39,854,663	19,386,131	702,122
Commenters	7,223,473	3,166,726	118,996

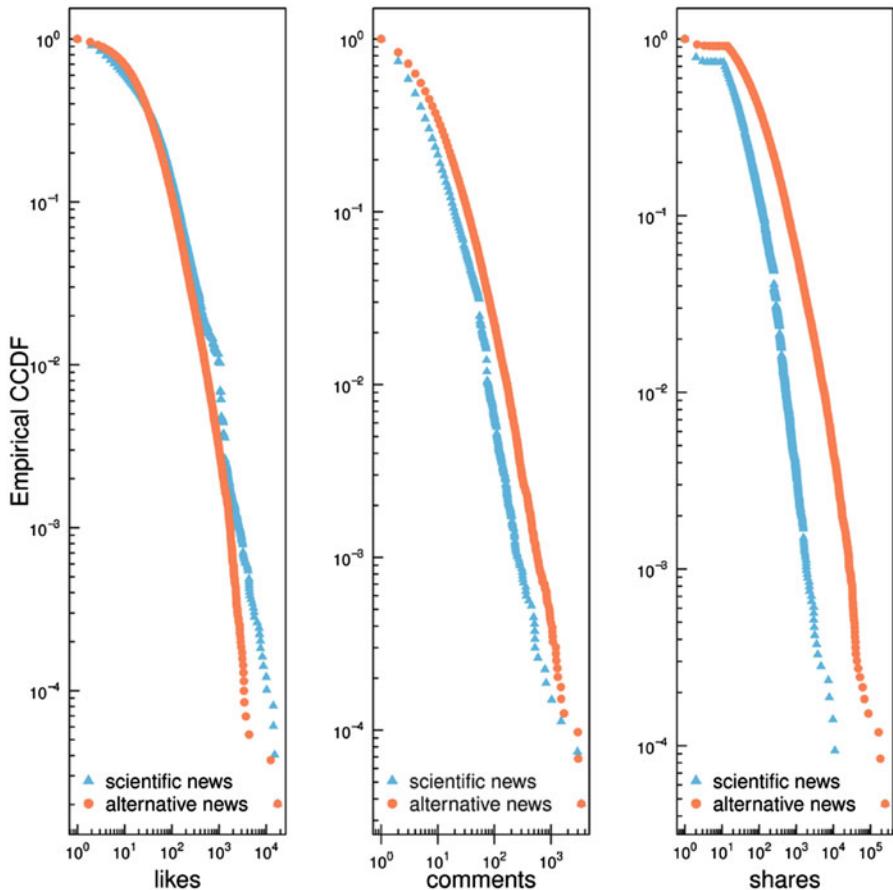
not available to us. When allowed by users’ privacy specifications, we accessed public personal information. However, in our study we used fully anonymized and aggregated data. We abided by the terms, conditions, and privacy policies of Facebook.

## 5 Echo Chambers

### 5.1 Attention Patterns

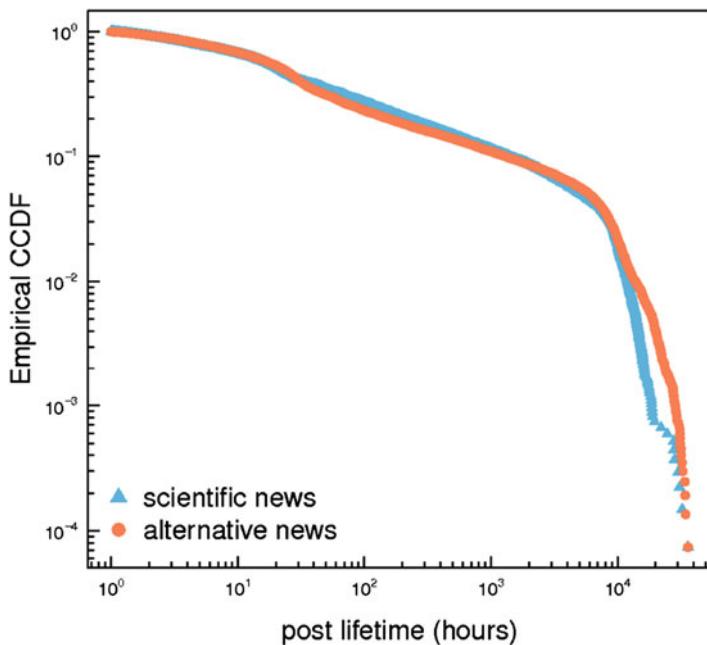
We start our discussion by analyzing how information gets consumed by users in both the Italian [10–12] and the US Facebook [50]. As a first step, we focus on users’ actions allowed by Facebook’s interaction paradigm i.e., likes, comments, and shares. Each action has a particular meaning [22]: while a *like* represents a positive feedback to the post, a *share* expresses the desire to increase the visibility of a given information; finally, a *comment* is the way in which the debate takes form around the topic of the post. Also, we consider the lifetime of a post (respectively, a user) i.e., the temporal distance between the first and last comment to the post (respectively, of the user). We also define the persistence of a post (respectively, a user) as the Kaplan-Meier estimates of survival functions by accounting for the lifetime of the post (respectively, the user).

Figure 1 shows the empirical Complementary Cumulative Distribution Functions (CCDFs) of users’ activity on posts grouped by category on the Italian Facebook. We may notice that distributions of likes, comments, and shares are all heavy-tailed. To further investigate users’ consumption patterns, in Fig. 2 we also plot the CCDF of the posts’ lifetime, observing that distinct kinds of contents show a comparable lifetime.

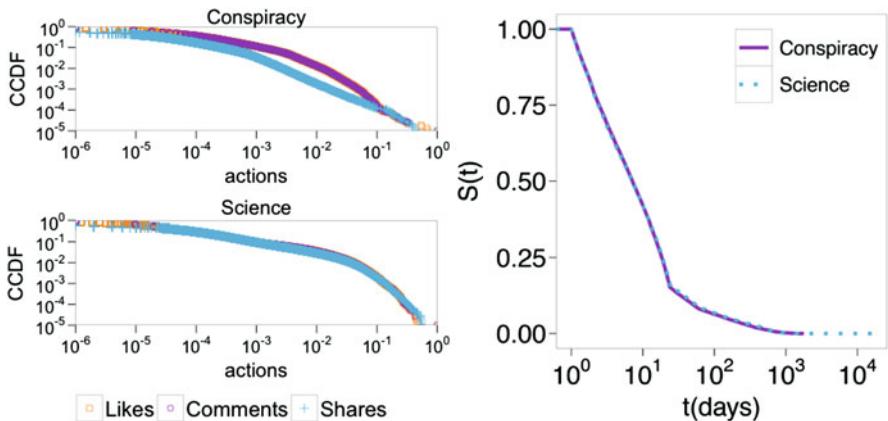


**Fig. 1** ITALIAN FACEBOOK. Empirical complementary cumulative distribution functions (CCDFs) of users' activity (likes, comments and shares) on posts grouped by category. Distributions denote heavy-tailed consumption patterns

As for the US Facebook, the distribution of the number of likes, comments, and shares on posts belonging to both scientific and conspiracy news is shown in the left panel of Fig. 3. As seen from the plots, all distributions are heavy-tailed—i.e, they are best fitted by power laws and possess similar scaling parameters. In the right panel of Fig. 3, we plot the Kaplan-Meier estimates of survival functions of posts grouped by category. To further characterize differences between the survival functions, we perform the Peto and Peto [41] test to detect whether there is a statistically significant difference between the two survival functions. Since we obtain a  $p$ -value of 0.944, we can state that there are not significant statistical differences between posts' survival functions on both science and conspiracy news. Thus, posts' persistence in the two categories is similar also in the US case.



**Fig. 2** ITALIAN FACEBOOK. Empirical CCDF, grouped by category, of the posts' lifetime i.e., the temporal distance (in hours) between the first and last comment. Lifetime is similar for both categories



**Fig. 3** US FACEBOOK **Left:** Complementary cumulative distribution functions (CCDFs) of the number of likes, comments, and shares received by posts belonging to conspiracy (*top*) and scientific (*bottom*) news. **Right:** Kaplan-Meier estimates of survival functions of posts belonging to conspiracy and scientific news. Error bars are on the order of the size of the symbols

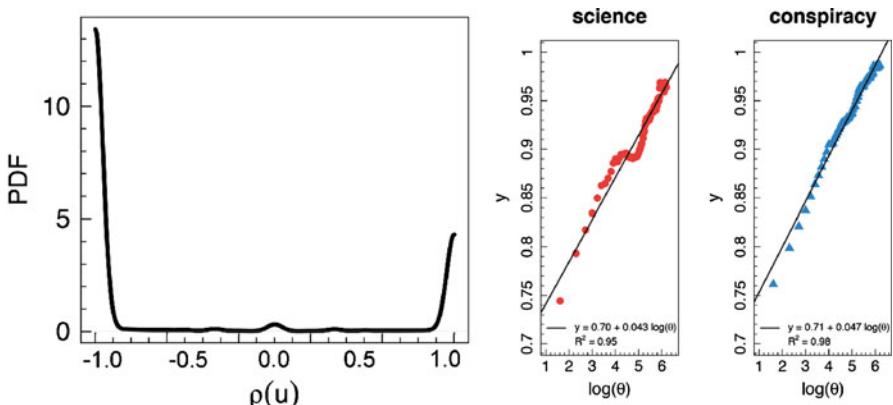
Summarizing, our findings show that distinct kinds of information (science, conspiracy) are consumed in a comparable way. However, when considering the correlation between couples of actions, we find that users of conspiracy pages are more prone to both share and like a post, denoting a higher level of commitment [10]. Conspiracy users are more willing to contribute to a wide diffusion of their topics of interest, according to their belief that such information is intentionally neglected by mainstream media.

## 5.2 *Polarization*

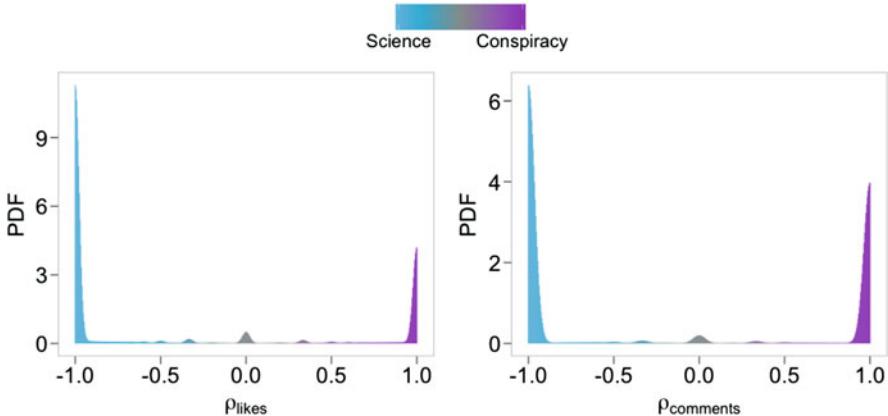
We now want to understand if users' engagement with a specific kind of content can become a good proxy to detect groups of users sharing the same system of beliefs i.e., echo chambers. Assume that a user  $u$  has performed  $x$  and  $y$  likes (comments) on scientific and conspiracy posts, respectively, and let  $\rho(u) = (y - x)/(y + x)$ . Thus, we say that user  $u$  is polarized towards science if  $\rho(u) \leq -0.95$ , while she is towards conspiracy if  $\rho(u) \geq 0.95$  user  $u$  is polarized towards conspiracy.

In Fig. 4 we show the Probability Density Function (PDF) of users' polarization on the Italian Facebook. We observe a sharply peaked bimodal distribution where the vast majority of users is polarized either towards science ( $\rho(u) \sim 1$ ) or conspiracy ( $\rho(u) \sim -1$ ). Hence, most of likers can be divided into two groups of users, those *polarized towards science* and those *polarized towards conspiracy* news.

Let us consider now the fraction of friends  $y$  of a user  $u$  sharing the same polarization of  $u$ . We define the *engagement*  $\theta(u)$  of a user  $u$  as her liking activity



**Fig. 4** ITALIAN FACEBOOK **Left:** Probability density function (PDF) of users' polarization. Notice the strong bimodality of the distribution, with two sharp peaks localized at  $-1 \lesssim \rho(u) \lesssim -0.95$  (conspiracy users) and at  $0.95 \lesssim \rho(u) \lesssim 1$  (science users). **Right:** Fraction of polarized neighbors as a function of the engagement  $\theta$  for both science (left) and conspiracy (right) users



**Fig. 5** US FACEBOOK Probability Density Functions (PDFs) of the polarization of all users computed both on likes (left) and comments (right)

normalized with respect to the total number of likes in our dataset. We find that the more a user is active on her narrative, the more she is surrounded by friends sharing the same attitude. Such a pattern is shown in the right panels of Fig. 4. Hence, social interactions of Facebook users are driven by homophily: users not only tend to be very polarized, but they also tend to be linked to users with similar preferences. Indeed, in both right panels of Fig. 4 we can observe that for polarized users the fraction of friends with the same polarization is very high ( $\gtrsim 0.75$ ) and grows with the engagement.

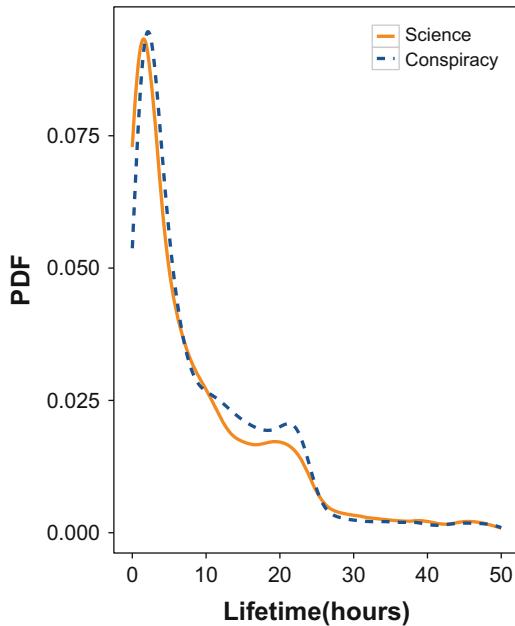
Similar patterns can be observed on the US Facebook. In Fig. 5 we show that the PDF for the polarization of all users is sharply bimodal here as well, with most having ( $\rho(u) \sim -1$ ) or ( $\rho(u) \sim 1$ ). Thus, most users may be divided into two main groups, those *polarized towards science* and those *polarized towards conspiracy*. The same pattern holds if we look at polarization based on comments rather than on likes.

In summary, our results confirm the existence of echo chambers on both the Italian and the US Facebook. Indeed, contents related to distinct narratives aggregate users into distinct, polarized communities, where users interact with like-minded people sharing their own system of beliefs.

## 6 Information Spreading and Cascades

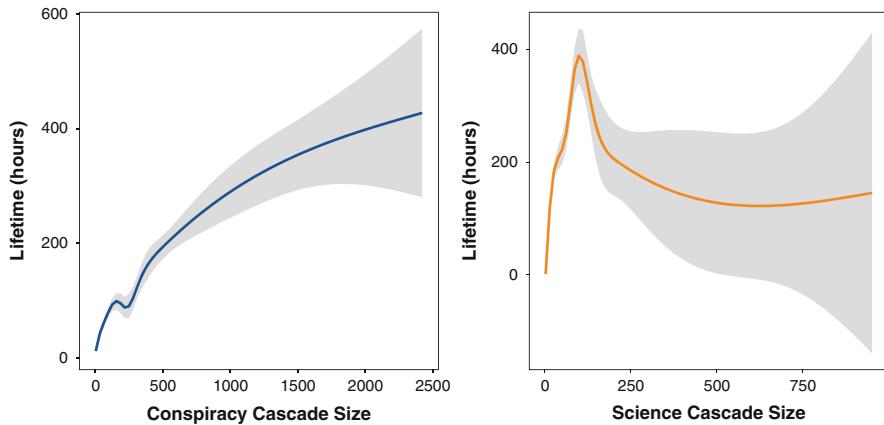
In this section we show how confirmation bias dominates viral processes of information diffusion and that the size of the (mis)information cascades may be approximated by the size of the echo chamber [19]. We begin our analysis by characterizing the statistical signature of cascades according to the narrative

**Fig. 6 ITALIAN FACEBOOK**  
 Probability Density Function (PDF) of lifetime computed on science news and conspiracy theories, where the lifetime is here computed as the temporal distance (in hours) between the first and last share of a post. Both categories show a similar behavior, with a peak in the first 2 h and another around 20 h



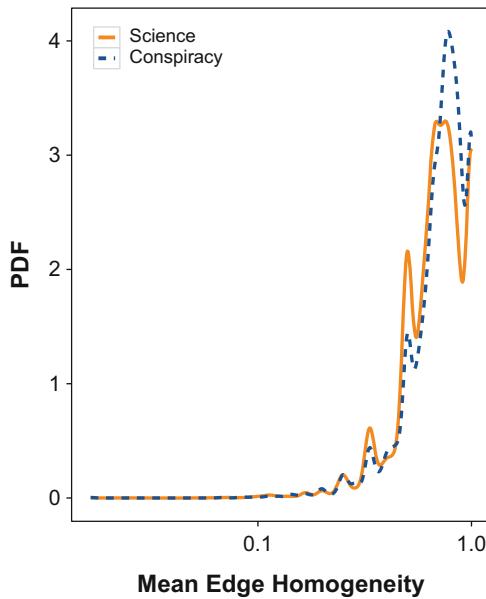
(science or conspiracy). Figure 6 shows the PDF of the cascade lifetime for both science and conspiracy. We compute the lifetime as the time (in hours) elapsed between the first and the last share of the post. In both categories we find a first peak at approximately 1–2 h and a second peak at approximately 20 h, denoting that the temporal sharing patterns are similar, independently of the narrative. We also find that a significant percentage of the information spreads rapidly (24.42% of the science news and 20.76% of the conspiracy rumors diffuse in less than 2 h, and 39.45% of science news and 40.78% of conspiracy theories in less than 5 h). Only 26.82% of the diffusion of science news and 17.79% of conspiracy lasts more than 1 day.

In Fig. 7 we show the lifetime as a function of the cascade's size, i.e. the number of users sharing the post. For science news we observe a peak in the lifetime corresponding to a cascade's size value of  $\approx 200$ ; moreover, the variability of the lifetime grows with the cascades' sizes, and higher cascade's size values correspond to high lifetime variability. For conspiracy-related contents, lifetime variability increases with cascade's size, and for highest values we observe a variability of the lifetime 50% around the average values. Such results suggest that news assimilation differs according to the categories. Science information is usually assimilated (i.e., it reaches a higher level of diffusion) quickly. A longer lifetime does not necessarily correspond to a higher level of interest, but possibly to a prolonged discussion within a specialized group of experts. Conversely, conspiracy rumors are assimilated more slowly and show a positive relation between lifetime and size; long-lived posts tend to be discussed by larger communities.



**Fig. 7** ITALIAN FACEBOOK Lifetime as a function of the cascade's size for conspiracy news (left) and science news (right). We observe a contents-driven differentiation in the sharing patterns. For conspiracy the lifetime grows with the size, while for science news there is a peak in the lifetime around a value of the size equal to 200, and a higher variability in the lifetime for larger cascades

**Fig. 8** ITALIAN FACEBOOK Mean edge homogeneity for science (solid orange) and conspiracy (dashed blue) news. The mean value of edge homogeneity on the whole sharing cascades is always greater or equal to zero



Finally, Fig. 8 shows that the majority of links between consecutively sharing users is homogeneous, i.e. both users share the same polarization and, hence, belong to the same echo chamber. In particular, the average edge homogeneity value of all the observed sharing cascades is always greater than or equal to zero, suggesting

that information spreading occurs mainly inside homogeneous clusters in which all users share the same polarization. Thus, contents tend to circulate only inside the echo chambers.

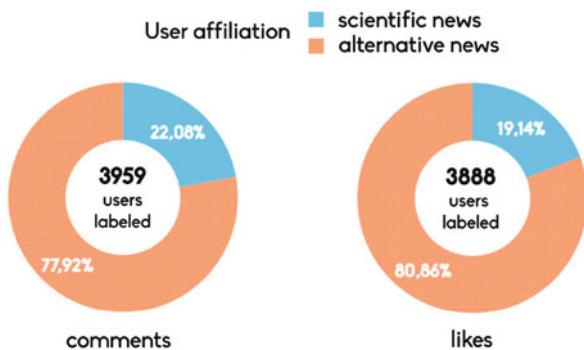
Summarizing, we found that cascades' dynamics differ, although consumption patterns on science and conspiracy pages are similar. Indeed, selective exposure is the primary driver of contents' diffusion and generates the formation of echo chambers, each with its own cascades' dynamics.

## 7 Response to Paradoxical Information

We have showed that users tend to aggregate around preferred contents shaping well-separated and polarized communities. Our hypothesis is that users' exposure to unsubstantiated claims may affect their selection criteria and increase their attitude to interact with false information. Thus, in this section we want to test how polarized users interact with information that is deliberately false i.e., troll posts, which are paradoxical imitations of conspiracy contents [10]. Such posts diffuse clearly dubious claims, such as the undisclosed news that infinite energy has been finally discovered, or that a new lamp made of actinides (e.g., plutonium and uranium) will finally solve the lack of energy with less impact on the environment, or that chemical analysis reveal that chem-trails contain *sildenafil citratum* (sold as the brand name Viagra).

Figure 9 shows how polarized users of both categories interact with troll posts in terms of comments and likes on the Italian Facebook. Our findings show that users usually exposed to conspiracy claims are more likely to jump the credulity barrier: indeed, conspiracy users are more active in both liking and commenting troll posts. Thus, even when information is deliberately false and framed with a satirical purpose, its conformity with the conspiracy narrative transforms it into credible content for members of the conspiracy echo chamber. Evidently, confirmation bias plays a crucial role in content selection.

**Fig. 9** ITALIAN FACEBOOK  
Percentage of comments and  
likes on troll posts from users  
polarized towards science  
(light blue) and conspiracy  
(orange)



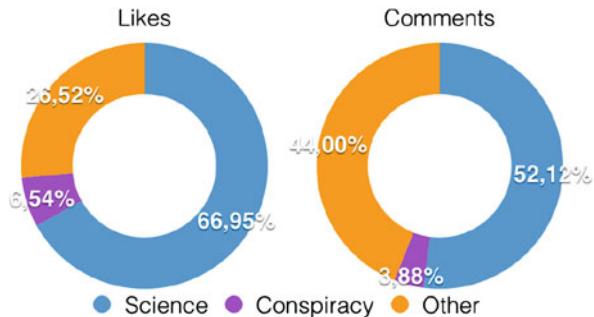
## 8 Response to Dissenting Information

Debunking pages on Facebook strive to contrast misinformation spreading by providing fact-checked information to specific topics. However, if confirmation bias plays a pivotal role in selection criteria, then debunking is likely to sound to conspiracy users such as information dissenting from their preferred narrative. In this section, our aim is to study and analyze users' behavior w.r.t. debunking contents on the US Facebook [50].

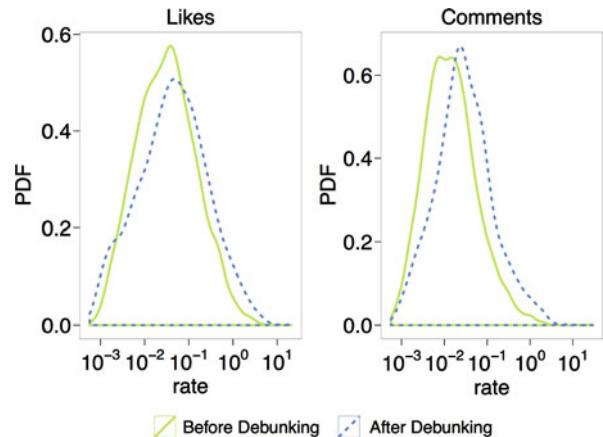
As a first step, we show how debunking posts get liked and commented according to users' polarization. Figure 10 shows how users' activity is distributed on debunking posts: left (respectively, right) panel shows the proportions of likes (respectively, comments) left by users polarized towards science, users polarized towards conspiracy, and not polarized users. We notice that the majority of both likes and comments is left by users polarized towards science (respectively, 66.95% and 52.12%), while only a small minority is made by users polarized towards conspiracy (respectively, 6.54% and 3.88%). Indeed, the first interesting result is that the biggest consumer of debunking information is the scientific echo chamber. Out of 9,790,906 polarized conspiracy users, just 117,736 interacted with debunking posts—i.e., commented a debunking post at least once.

Hence, debunking posts remain mainly confined within the scientific echo chamber and only few users usually exposed to unsubstantiated claims actively interact with the corrections. Dissenting information is mainly ignored. However, in our scenario few users belonging to the conspiracy echo chamber do interact with debunking information. We now wonder about the effect of such an interaction. Therefore, we perform a comparative analysis between users' behavior before and after they first comment on a debunking post. Figure 11 shows the liking and commenting rates—i.e., the average number of likes (or comments) on conspiracy posts per day—before and after the first interaction with debunking. We can observe that users' liking and commenting rates increase after the interaction. Thus, their activity in the conspiracy echo chamber is reinforced. In practice, debunking attempts are acting as a backfire effect.

**Fig. 10** US FACEBOOK  
Proportions of likes (*left*) and  
comments (*right*) left by  
users polarized towards  
science, users polarized  
towards conspiracy, and  
not polarized users



**Fig. 11** US FACEBOOK Rate—i.e., average number, over time, of likes (*left*) and comments (*right*) on conspiracy posts of users who interacted with debunking posts

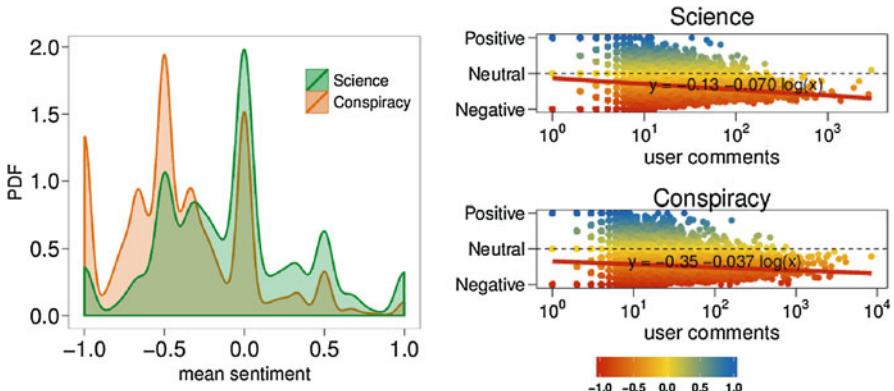


## 9 Emotional Dynamics

In this section, we aim at analyzing the emotional dynamics inside and across echo chambers. In particular, we apply sentiment analysis techniques to the comments of our Facebook Italian dataset, and study the aggregated sentiment with respect to scientific and conspiracy-like information [51]. The sentiment analysis is based on a supervised machine learning approach, where we first annotate a substantial sample of comments, and then build a Support Vector Machine (SVM) classification model. The model is then applied to associate each comment with one sentiment value: negative, neutral, or positive. The sentiment is intended to express the emotional attitude of Facebook users when posting comments.

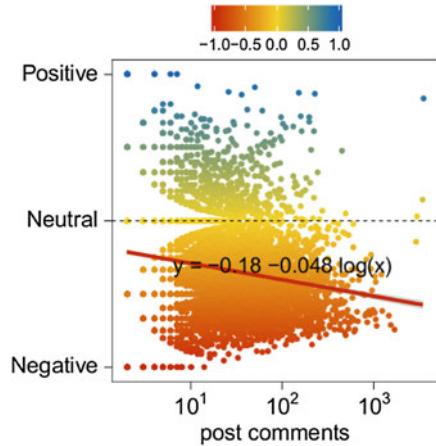
To further investigate the dynamics behind users' polarization, we now study how the sentiment changes w.r.t. users' engagement in their own echo chamber. In the left panel of Fig. 12, we show the PDF of the mean sentiment of polarized users with at least two comments. We may observe an overall negativity, more evident on the conspiracy side. When looking at the sentiment as a function of the number of comments of the user, we find that the more active a polarized user is, the more she tends towards negative values, both on science and conspiracy posts. Such results are shown in the right panel of Fig. 12, where the sentiment has been regressed w.r.t. the logarithm of the number of comments. Interestingly, the sentiment of science users decreases faster than that of conspiracy users.

We now want to investigate the emotional dynamics when such polarized (and negative-minded) users meet together. To this aim, we pick all the posts representing the arena where the debate between science and conspiracy users takes place. In particular, we select all the posts commented at least once by both a user polarized on science and a user polarized on conspiracy. We find 7751



**Fig. 12** ITALIAN FACEBOOK **Left:** Probability Density Function (PDF) of the mean sentiment of polarized users having commented at least twice, where  $-1$  corresponds to negative sentiment,  $0$  to neutral and  $1$  to positive. **Right:** Average sentiment of polarized users as a function of their number of comments. Negative (respectively, neutral, positive) sentiment is denoted by red (respectively, yellow, blue) color. The sentiment has been regressed w.r.t. the logarithm of the number of comments

**Fig. 13** US FACEBOOK  
Aggregated sentiment of posts as a function of their number of comments.  
Negative (respectively, neutral, positive) sentiment is denoted by red (respectively, yellow, blue) color



such posts (out of 315,567), reinforcing the fact that the two communities are strictly separated and do not often interact with one another. Then, we analyze how the sentiment changes when the number of comments of the post increases i.e., when the discussion becomes longer. Figure 13 shows the aggregated sentiment of such posts as a function of their number of comments. Clearly, as the number of comments increases—i.e., the discussion becomes longer—the sentiment is always more negative. Therefore, we may conclude that the length of the discussion does affect the negativity of the sentiment.

## 10 Conclusions

We investigated how information related to two very distinct narratives—i.e., scientific and conspiracy news—gets consumed and shapes communities on Facebook. For both the Italian and the US scenario, we showed the emergence of two well-separated and polarized groups—i.e., echo chambers—where users interact with like-minded people sharing the same system of beliefs. We found that users are extremely focused and self-contained on their specific narrative. Such a highly polarized structure facilitates the reinforcement and contents' selection by confirmation bias. Moreover, we observed that social interactions of Facebook users are driven by homophily: users not only tend to be very polarized, but they also tend to be linked to users with similar preferences. According to our results, confirmation bias dominates viral processes of information diffusion. Also, we found that the size of misinformation cascades may be approximated by the same size of the echo chamber.

Furthermore, by measuring the response to the injection of false information (parodic imitations of alternative stories), we observed that users prominently interacting with alternative information sources—i.e. more exposed to unsubstantiated claims—are more prone to interact with intentional and parodic false claims. Thus, our findings suggest that conspiracy users are more likely to jump the credulity barrier: even when information is deliberately false and framed with a satirical purpose, its conformity with the conspiracy narrative transforms it into credible content for members of the conspiracy echo chamber.

Then, we investigated users' response to dissenting information. By analyzing the effectiveness of debunking on conspiracy users on the US Facebook, we found that scientific echo chamber is the biggest consumer of debunking posts. Indeed, only few users usually active in the conspiracy echo chamber interact with debunking information and, in the latter case, their activity in the conspiracy echo chamber increases after the interaction, rather than decreasing. Thus, debunking attempts are acting as a backfire effect.

Finally, we focused on the emotional dynamics inside and between the two echo chambers, finding that the sentiment of users on science and conspiracy pages tends to be negative, and is more and more negative when the discussion becomes longer or users' activity on the social network increase. In particular, the discussion degenerates when the two polarized communities interact with one another.

Our findings provide insights about the determinants of polarization and the evolution of core narratives on online debating, suggesting that fact-checking is not working as expected. As long as there are no immediate solutions to functional illiteracy, information overload and confirmation bias will continue dominating social dynamics online. In such a context, misinformation risk and its consequences will remain significant. To contrast misinformation spreading, we need to smooth polarization. To this aim, understanding how core narratives behind different echo chambers evolve is crucial and could allow to design more efficient communication strategies that account for users' cognitive determinants behind these kind of mechanisms.

**Acknowledgements** This work is based on co-authored material. We thank Aris Anagnostopoulos, Alessandro Bessi, Guido Caldarelli, Michela Del Vicario, Shlomo Havlin, Igor Mozetič, Petra Kralj Novak, Fabio Petroni, Antonio Scala, Louis Shekhtman, H. Eugene Stanley, and Brian Uzzi.

## References

1. Adamic LA, Glance N (2005) The political blogosphere and the 2004 US election: divided they blog. In: Proceedings of the 3rd international workshop on Link discovery. Association for Computing Machinery, New York, pp 36–43
2. Adar E, Zhang L, Adamic LA (2004) Lukose RM: implicit structure and the dynamics of blogsphere. In: Workshop on the weblogging ecosystem, vol 13, pp 16989–16995
3. Aikin SF (2013) Poe’s law, group polarization, and argumentative failure in religious and political discourse. *Soc Semiot* 23(3):301–317
4. Akerlof GA, Yellen JL, Katz ML (1996) An analysis of out-of-wedlock childbearing in the United States. *Q J Econ* 111:277–317
5. AlMansour AA, Brankovic L, Iliopoulos CS (2014) A model for recalibrating credibility in different contexts and languages—a twitter case study. *Int J Digit Inf Wirel Commun* 4(1): 53–62
6. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci* 106(51):21544–21549
7. Asch SE (1955) Opinions and social pressure. *Readings About Soc Anim* 193:17–26
8. Bakshy E, Hofman JM, Mason WA, Watts DJ (2011) Everyone’s an influencer: quantifying influence on Twitter. In: Proceedings of the fourth ACM international conference on web search and data mining. Association for Computing Machinery, New York, pp 65–74
9. Bakshy E, Messing S, Adamic LA (2015) Exposure to ideologically diverse news and opinion on facebook. *Science* 348(6239):1130–1132
10. Bessi A, Coletto M, Davidescu GA, Scala A, Caldarelli G, Quattrociocchi W (2015) Science vs conspiracy: collective narratives in the age of misinformation. *PLoS One* 10(2):e0118,093
11. Bessi A, Petroni F, Del Vicario M, Zollo F, Anagnostopoulos A, Scala A, Caldarelli G, Quattrociocchi W (2015) Viral misinformation: the role of homophily and polarization. In: Proceedings of the 24th international conference on world wide web. Association for Computing Machinery, New York, pp 355–356
12. Bessi A, Petroni F, Del Vicario M, Zollo F, Anagnostopoulos A, Scala A, Caldarelli G, Quattrociocchi W (2016) Homophily and polarization in the age of misinformation. *Eur Phys J Spec Top* 225(10):2047–2059
13. Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194–1197
14. Centre WHOM (2014) Ebola: Experimental therapies and rumoured remedies. *Situation Assessment* (2014). <http://www.who.int/mediacentre/news/ebola/15-august-2014/en/>
15. Cheng J, Adamic L, Dow PA, Kleinberg JM, Leskovec J (2014) Can cascades be predicted? In: Proceedings of the 23rd international conference on world wide web. International World Wide Web Conferences Steering Committee, Canton of Geneva, pp 925–936
16. Ciampaglia GL, Shiralkar P, Rocha LM, Bollen J, Menczer F, Flammini A (2015) Computational fact checking from knowledge networks. *PLoS One* 10(6):e0128193
17. Conover M, Ratkiewicz J, Francisco MR, Gonçalves B, Menczer F, Flammini A (2011) Political polarization on twitter. *ICWSM* 133:89–96
18. Conover MD, Goncalves B, Ratkiewicz J, Flammini A, Menczer F (2011) Predicting the political alignment of twitter users. In: 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing, pp 192–199. <https://doi.org/10.1109/PASSAT/SocialCom.2011.34>

19. Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, Stanley HE, Quattrociocchi W (2016) The spreading of misinformation online. *Proc Natl Acad Sci* 113(3):554–559
20. Dong XL, Gabrilovich E, Murphy K, Dang V, Horn W, Lugaresi C, Sun S, Zhang W (2015) Knowledge-based trust: estimating the trustworthiness of web sources. *Proc VLDB Endowment* 8(9):938–949
21. Dow PA, Adamic LA, Frigeri A (2013) The anatomy of large Facebook cascades. In: ICWSM
22. Ellison NB, Steinfield C, Lampe C (2007) The benefits of facebook “friends:” social capital and college students’ use of online social network sites. *J Comput-Mediat Commun* 12(4): 1143–1168
23. Erich O, Udi W: News feed fyi: Showing fewer hoaxes (2015). <http://newsroom.fb.com/news/2015/01/news-feed-fyi-showing-fewer-hoaxes/>
24. Facebook: Using the graph api. Website (2017). <https://developers.facebook.com/docs/graph-api/using-graph-api/>. Last checked: 24.02.2017
25. Gupta A, Kumaraguru P, Castillo C, Meier P (2014) Tweetcred: real-time credibility assessment of content on twitter. In: Social informatics. Springer, Berlin, pp 228–243.
26. Howell WL (2013) Digital wildfires in a hyperconnected world. Tech. Rep. Global Risks 2013, World Economic Forum
27. Jenni Sargent MD (2017) First draft coalition. Website. <https://firstdraftnews.com>
28. Kahan DM (1997) Social influence, social meaning, and deterrence. *Virginia Law Rev* 83: 349–395
29. Katz E, Lazarsfeld PF (1970) Personal influence, the part played by people in the flow of mass communications. Transaction Publishers, New Brunswick
30. Knobloch-Westerwick S (2012) Selective exposure and reinforcement of attitudes and partisanship before a presidential election. *J Commun* 62(4):628–642
31. Kuklinski JH, Quirk PJ, Jerit J, Schwieder D, Rich RF (2000) Misinformation and the currency of democratic citizenship. *J Polit* 62(3):790–816
32. Lazarsfeld PF, Berelson B, Gaudet H (1968) The peoples choice: how the voter makes up his mind in a presidential campaign. Columbia University Press, New York
33. Lazer D, Pentland AS, Adamic L, Aral S, Barabasi AL, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, et al (2009) Life in the network: the coming age of computational social science. *Science* (New York, NY) 323(5915):721
34. Lessig L (2009) Code: and other laws of cyberspace. [ReadHowYouWant.com](http://ReadHowYouWant.com)
35. Levy P (1999) Collective intelligence: Mankind’s emerging world in cyberspace. Perseus Publishing, Cambridge
36. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. *Annu Rev Sociol* 27:415–444
37. Moore JP (2009) The dangers of denying HIV. *Nature* 459(7244):168–168
38. Nickerson RS (1998) Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol* 2(2):175
39. Nyhan B, Reifler J (2010) When corrections fail: the persistence of political misperceptions. *Polit Behav* 32(2):303–330
40. Pariser E (2011) The filter bubble: what the Internet is hiding from you. Penguin, London
41. Peto R, Peto J (1972) Asymptotically efficient rank invariant test procedures. *J R Stat Soc Ser A* 135:185–207
42. Qazvinian V, Rosengren E, Radev DR, Mei Q (2011) Rumor has it: identifying misinformation in microblogs. In: Proceedings of the conference on empirical methods in natural language processing. Association for Computational Linguistics, New Brunswick, pp 1589–1599
43. Ratkiewicz J, Conover M, Meiss M, Gonçalves B, Flammini A, Menczer F (2011) Detecting and tracking political abuse in social media. In: ICWSM
44. Resnick P, Carton S, Park S, Shen Y, Zeffer N (2014) RumorLens: a system for analyzing the impact of rumors and corrections in social media. In: Proceedings of the computational journalism conference
45. Stoner JA (1968) Risky and cautious shifts in group decisions: the influence of widely held values. *J Exp Soc Psychol* 4(4):442–459

46. Sunstein CR (2002) The law of group polarization. *J Polit Philos* 10(2):175–195
47. Sunstein CR (2009) Republic.com 2.0. Princeton University Press, Princeton, NJ
48. Ugander J, Backstrom L, Marlow C, Kleinberg J (2012) Structural diversity in social contagion. *Proc Natl Acad Sci* 109(16):5962–5966
49. Watts DJ, Dodds PS (2007) Influentials, networks, and public opinion formation. *J Consum Res* 34(4):441–458
50. Zollo F, Bessi A, Del Vicario M, Scala A, Caldarelli G, Shekhtman L, Havlin S, Quattrociocchi W (2015) Debunking in a world of tribes. arXiv preprint arXiv:1510.04267
51. Zollo F, Novak PK, Del Vicario M, Bessi A, Mozetič I, Scala A, Caldarelli G, Quattrociocchi W (2015) Emotional dynamics in the age of misinformation. *PLoS One* 10(9):e0138740

# Scalable Detection of Viral Memes from Diffusion Patterns



Pik-Mai Hui, Lilian Weng, Alireza Sahami Shirazi, Yong-Yeol Ahn, and Filippo Menczer

## 1 Introduction

A *meme* is a distinct piece of information that replicates among people, like biological genes replicating through reproduction [1]. Memes resemble infectious diseases, in the sense that both travel through social ties between people [2, 3]. As blooming online social media services facilitate online social interactions, they also change how memes spread through society. Most importantly, social media platforms such as Facebook, Google Plus, Twitter, and Tumblr connect billions of users into a network that can spread a meme to the whole world instantly. At the same time, these services allow us to directly observe and study the spreading of memes and user behaviors by recording detailed data about user activities.

A vast number of memes are created every day. However, only a tiny fraction goes viral. This raises the most fundamental question in information diffusion research: *What makes something viral?* This question has attracted attention across disciplines including marketing and advertisement, as well as machine learning and network science. One shall agree that the question is meaningful but too broad. Here we focus on a more specific and well-defined question: *How can we predict the virality of a meme early?*

There are roughly two general approaches to the problem of meme virality prediction: time series analysis and feature-based classification. What follows in this chapter focuses on feature-based classification [4–6]. Readers who are interested in the approach of time series analysis are referred to a different literature [7–10].

---

P.-M. Hui (✉) · L. Weng · Y.-Y. Ahn · F. Menczer

Center for Complex Networks and Systems Research, School of Informatics and Computing,  
Indiana University, Bloomington, IN, USA  
e-mail: [huip@indiana.edu](mailto:huip@indiana.edu)

A. Sahami Shirazi  
Yahoo!, Sunnyvale, CA, USA

The feature-based classification approach aims to discover distinguishing features of viral memes and to apply supervised machine learning techniques using these features. As in standard feature-based machine learning problems, a general saying is *garbage in, garbage out*, implying that if inputs to a model are not informative, its output will neither be meaningful. Therefore the most critical step is to identify and extract useful features from datasets at hand.

We study a set of useful features from our theoretical and conceptual understanding of network structure and social information diffusion processes. In particular, we discuss the features of the diffusion patterns based on dense subgroups (communities) in underlying networks. We will demonstrate that diffusion pattern can be extracted at scale, which preserves its strength in virality prediction in two massive datasets from Twitter and Tumblr.

## 2 What Makes It Viral?

Although we do not address this question directly, understanding the potential reasons why memes go viral is nevertheless crucial for identifying useful features and for any discussion about viral memes. From literature we identify three key aspects of viral spreading, namely innate attractiveness of memes, user characteristics, and properties of the underlying social network. Motivated readers are recommended to query the references for more details on these aspects of virality.

### 2.1 Innate Attractiveness

The innate appeal of a meme may be the most basic factor contributing to its virality. It is intuitive that users are more likely to reshare memes with better “quality.” Quality can be defined in different contexts. For example, Berger and Milkman studied the emotional constituents in news articles and their impact on the articles’ virality. They find that news articles that actively evoke arousal become more viral later on [11]. Many studies presuppose virality as an intrinsic trait of memes. Since a meme is represented by its content, it justifies the search for content features that correlate with quality. For one, Guerini et al. characterized various aspects of virality and how they indicate the future virality of text-based content [12].

Although innate attractiveness is an intuitive explanation of virality, it does not paint the whole picture. The attractiveness of a meme is highly dependent on many contextual features, such as other existing memes and the culture of surrounding population. Studies have also demonstrated that quality alone does not explain virality well. In fact, agent-based simulation showed that highly skewed distribution

of meme popularity can arise even if we do not assume any difference in innate quality of memes [13]. Moreover, the success of online content, such as songs from online music downloads and social news filtering, depends significantly on provided social cues [14, 15]. This suggests that factors other than innate quality, such as visibility and reachability of the memes, may as well contribute to virality.

## 2.2 User Characteristics

The importance of social influence leads us to the concept of influencers and the roles of user characteristics in general. Although there are seemingly countless memes available, the scarcity of user effort in consuming information leads to limited individual *attention* in any social networks. Similar to biological organisms (and genes) striving for resources to reproduce, all memes strive for the attention of people. Since user consumes meme at a limited rate, only the memes that are seen within a short time period have a chance to propagate. Memes originating at an isolated location in the social network may not have any chance to spread because no one can see them in the first place. Such memes quickly go extinct in the system. Meanwhile, a meme that happens to be reshared by a user with many followers will have a significantly higher chance to reproduce across the followers' minds.

When user  $B$  reads user  $A$ 's post, the likelihood of user  $B$  resharing the information depends on his/her evaluation of user  $A$ . That is, the influence that one exerts on others varies across the actors. Content by a well-respected celebrity such as a founder of a famous organization naturally generates a stronger influence on others than that by a normal person, despite that they are two copies of the same content. In addition, each user has a specific set of topical interests. Some care more about global politics and wars in the Middle-East, while others may only want to know about new French recipes. Since users consume and share information according to their own interests, it is more likely for meme to spread between users with similar interests, when one shares and one consumes closely relevant contents. These effects are further exacerbated by a combination of limited user attention and abundant supply of memes. Weng et al. showed that limited individual attention in the competition among memes induces strong heterogeneity in meme popularity and longevity [13]. In deciding which meme to consume, each user prioritizes based on their interests and this alters meme popularity [16].

In other words, the spreading of viral memes favors users of specific characteristics. We call them influential users. Many methods have been proposed for quantifying user influence and identifying these influential users. In general, these methods use relevant observables of user characteristics, such as high degree or retweetability [17, 18], topical similarity [19–21], information forwarding activity [18, 22], or size of cascades [23, 24], to infer the strength of user influence over other.

### 2.3 Properties of Underlying Social Network

The characteristics of social ties in the underlying social network, through which memes spread, also affect the success of memes. Strong and homophilous ties are considered more effective than weak ties for spreading messages [25], while weak ties are thought of as transmitting novel information [26]. These theories are commonly used in viral marketing and consumer studies, where researchers actively apply network approaches to analyze and model local and global patterns of social network structure [27–29]. In addition, the existence of hubs, namely nodes with extremely large degree, is known to affect the persistence of infections, the distribution of cascade sizes, and the vulnerability of the system [30, 31]. Intuitively, hubs provide pathways through which memes can teleport to distant parts of the network instantly, facilitating the development of meme popularity on the whole network.

Another important network structure feature in most social networks is the presence of dense subgraphs called *communities* [32–35]. Communities are characterized by internal cohesion (more internal edges than expected) and external isolation (fewer outgoing edges than expected). While communities naturally constrain information flow across their borders, they may be necessary for providing initial critical mass before a meme can spread broadly [36]. In addition, the theory of complex contagion [37–41] suggests that we may expect an even stronger constraining effect from community structure [4]. Therefore, information extracted from the network structure and early spreading patterns is valuable to predict the virality of a meme. Further discussion on extracting features from community structures of social networks follows in a later section.

## 3 Data and Methods

In this section we present details of the datasets used in our experiments, and explain the methods we applied to extract network communities and to predict virality. We begin with a brief introduction to the online social media platforms from which our data was collected, and the networks that we constructed using each of these platforms.

### 3.1 Social Media Platforms

Online social media platforms enable people to share information and subscribe to updates from other users. The information can be of any type, ranging from short text messages and blog posts, to images and video clips. On these platforms, users typically choose others to whom they pay attention by “following” them. Most platforms also provide users with multiple mechanisms of information sharing, which serve different purposes.

Twitter is one of the most popular social media platforms. On Twitter, users post short messages called *tweets*. Between a pair of users ( $u, v$ ), we consider three main types of interactions: (1)  $u$  can *follow*  $v$  to subscribe to tweets from  $v$ ; (2)  $u$  can *retweet*  $v$ 's messages to re-broadcast them to  $u$ 's followers; and (3)  $u$  can *mention*  $v$ 's screen name in tweets by using the “@” symbol (e.g., '@potus'). Users can also explicitly attach indexable topic identifiers to a tweet by using *hashtags*, terms with the “#” symbol as a prefix (e.g., #news).

Tumblr is another popular social networking and microblogging platform supported by Yahoo! since 2013, hosting hundreds of millions of monthly active users and blogs. Tumblr features many functions similar to Twitter, such as hashtags, resharing, liking, and replies.

On both Twitter and Tumblr, hashtags can be used to operationalize the concept of memes, thanks to multiple characteristics of hashtags that accord with the definition of a meme [1]. First, hashtags are concretely defined by user consensus and uniquely identifiable through searches; second, hashtags reproduce through imitation by users; third, hashtags mutate, compete, and dominate in the same system over time. For example, #ows rapidly suppressed several similar hashtags to become the reference label for the Occupy Wall Street movement among hundreds of thousands of people who participated in related public discourse [42]. The usage of hashtags also makes the application of our methods straightforward and our findings easily comparable to results based on other platforms.

### 3.2 Community Detection

Communities contain rich information about the structure of a social network. These communities can be extracted by applying different algorithms. The results in this chapter are based on communities detected by two methods, namely InfoMap [33] and Louvain's method [43]. We have chosen these two methods, based on contrasting principles, to evaluate the robustness of the results under different choices of community detection algorithm. InfoMap and Louvain's method optimize for different objective functions and are therefore expected to produce distinct results, particularly regarding community size and resolution [44]. Another difference is that InfoMap considers the direction of edges, while Louvain's method treats all edges as undirected. Therefore the results may provide insight about the usefulness of edge directionality as signals for virality prediction.

Nowadays, the sizes of online social networks and the volume of information traffic on them are so large that analysis requires distributed storage and computing environments. Algorithms running on single computers do not scale well to such large networks, say with tens of million nodes. Additionally, moving large volumes of data stored on different storage nodes to a single machine is costly. Although the original implementations of the InfoMap and Louvain's algorithms were not designed for parallel computation, distributed implementations of these algorithms have been developed to better utilize resources in multiple-machine

clusters. These scalable methods optimize execution speed and resource efficiency without sacrificing accuracy. We use *distributed Louvain* [45] and *RelaxMap* [46], parallel implementations of Louvain's and InfoMap methods, respectively, to extract communities from large Twitter and Tumblr networks.

### 3.3 Twitter Information-Sharing Network

In prior work, virality was predicted using community features extracted from a Twitter follower network [5]. While constructing such a follower network is desirable, it poses some challenges. Some social media platforms, such as Facebook, regard friendship data as private, and therefore do not make it available for research. Furthermore, collecting complete follower information among many users can be forbiddingly expensive. The APIs provided by popular online social platforms restrict the rate at which such data can be queried without payment, making even moderate-size experiment difficult. This motivates an alternative approach.

We can extract communities based on an information-sharing network rather than a follower network. The links in such a network represent how memes spread through, e.g., retweets and replies. This can be used as a proxy for the social network that captures the process of meme diffusion. Since people typically retweet messages from users they follow, an information-sharing network has a significant overlap with the follower network. Let us consider two networks constructed in this fashion, using high-volume streams of Twitter and Tumblr posts.

In our experiment, the Twitter information-sharing network is constructed using a 10% sample of public tweet stream. The tweets used in our study are from July to September 2015 (Table 1). We divide the collected tweets into two temporal parts: a one-month *observation period* followed by a two-month *experiment period*.

In the observation period we collect existing hashtags and information-sharing activities. These activities are used to construct a directed information-sharing network. Each edge in the network is formed by retweets and mentions of one user by another, and is weighted by the frequency of information flow from source to destination user. When user A is retweeted by user B, or when user A mentions user B, information flows from A to B. Communities in this network are extracted by RelaxMap and Louvain's algorithms. To reduce noise, only the largest weakly-connected component of network is used in community detection.

In the experiment period, we consider only newly-born hashtags, which did not occur in the observation period. Each new hashtags is tracked for a period of 30 days, starting from its first occurrence. If a hashtag first occurs within 30 days of the end of the experiment period, so that we do not have 30 days of data in the experiment period, we do not consider it in our study. For each tracked hashtag, we record the sequence of users who share it (adopters).

This setup has some desirable properties. Since the networks are constructed using only information from the observation period and evaluation is done strictly

over content in the experiment period, there is no information leak between training and evaluation. Moreover, every hashtag in the evaluation is observed for exactly 30 days after its first use, avoiding a bias against late hashtags.

In summary, tweets from the observation period are used to construct the directed network from which communities are extracted. The experiment period is used to construct meme adoption histories and run the prediction experiments.

### 3.4 Tumblr Information-Sharing Network

We also collected posts from the Tumblr firehose, a database with the complete history of user posts. On Tumblr, a user can create and own multiple blogs with one account. Tumblr identifies the same user posting in distinct blogs as different *persona*. However, each user is identified by one primary blog while reacting to posts from other users, such as when replying and liking posts. Therefore we consider a user's primary blog as their identity. We focus on text posts, excluding other types of content such as pictures and video clips.

We divided this dataset the same way we did for the Twitter network (Table 1). A directed network is constructed by scanning all text posts in November 2015 (the observation period), and its largest weekly connected component is used to extract communities. An edge is generated when a user likes or replies to a post by another user, and edges are weighted by the frequencies of interaction. Edges are directed from user A to user B when B likes or replies to posts by A. Text posts in December 2015 and January 2016 (the experiment period) were collected to run the predictions.

The Tumblr dataset contains a very diverse set of hashtags. Tumblr hashtags are case sensitive, can contain spaces and emoji, and have no length limit. As a result, they can be very long (full sentences) and have duplication, for instance “Cute cat” and “cute\_cat.” To limit the noise caused from these degenerate cases, we filtered out hashtags that are longer than 20 characters and trimmed all emoji, common phrase separators (space, underscore, etc.) and repeated expressions, then lowercased all characters.

**Table 1** Information and basic statistics about the network datasets in the study

	Twitter	Tumblr
Type of edge	Retweets and mentions	Replies and likes
Observation period	2015-07	2015-11
Experiment period	2015-08/09	2015-12/2016-01
# Nodes	29,224,842	19,701,097
# Edges	169,685,133	711,573,645

## 4 Network Community Features

In this section we present the features extracted from the networks. The features are a subset of the ones used in our prior work [5]. In particular, we focus on features that are motivated by the community structure of the underlying social networks. These network features are computed based on the locations of the first  $n$  adopters of each hashtag, where the parameter  $n$  is set to be a relatively small number compared to the final number of tweets generated by viral hashtags. In our experiment,  $n = 25$ .

Let us start by defining a few key concepts and mathematical notations. Some of the information is mentioned in previous sections, but is included below for the sake of completeness.

**Definition 1 (Meme)** We consider each hashtag  $h$  as a meme. The popularity of meme  $h$  is quantified by the number of adopters.  $A(h)$  is the set of all adopters who posted about  $h$  and  $A_n(h) \subseteq A(h)$  is the set of early adopters who posted at least one of the first  $n$  posts. We define the *popularity* of  $h$  as  $|A(h)|$ .

**Definition 2 (Adopter Sequence)** For a given meme  $h$ , we consider the sequence of meme adopters,  $\langle a_1^h, a_2^h, \dots, a_n^h \rangle$ , where  $a_i^h \in A(h)$  is the creator of the  $i$ -th post containing  $h$ . A user may appear multiple times in the sequence if the user posts about  $h$  more than once.

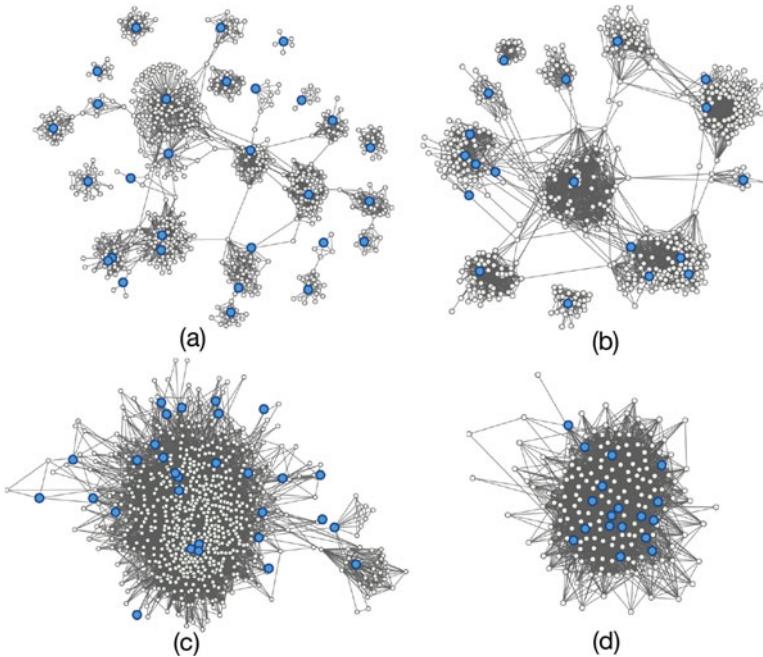
**Definition 3 (Community)** A community  $c \in \mathcal{C}$  is a dense subgraph of nodes (users) in the network. Given information about which nodes belong to which communities,  $A(h|c)$  is the set of adopters of a meme  $h$  in community  $c$ .  $A_n(h|c)$  is the similar set that only considers the first  $n$  relevant tweets.  $C(h)$  denotes the *infected communities* of  $h$ , which are communities with at least one tweet containing  $h$ . Similarly, the infected communities with early posts are  $C_n(h)$ .

Community structure is useful in predicting meme virality because of how memes travel among users who are socially connected. This process is commonly called *social contagion*. It has been argued that social contagions are *complex* contagions, in contrast to *simple* contagions like epidemic spreading. To explain the connection between complex contagion and community structure in the context of social network analysis, we note that complex contagion is known to possess two distinctive characteristics:

*Social reinforcement.* Until a certain point, each additional exposure drastically increases the probability of adoption [47–49].

*Homophily.* Social relationships are more likely to be formed between people who share certain characteristics, captured in the sayings “birds of a feather flock together” and “similarity breeds connection” [50, 51].

Community structure has been shown to help quantify the strength of both social reinforcement and homophily by the following mechanisms [4]. First, dense



**Fig. 1** Visualizations of diffusion patterns of viral (a, b) and non-viral (c, d) memes on Twitter. Early adopters among the first 30 tweets (in blue) and their neighbors in the same communities are shown. Each node represents a user and each link indicates the reciprocal follow relationship between two users. Figure reproduced with permission [5]

connectivity inside a community increases the chances of multiple exposures, thus enhancing the contagion that is sensitive to social reinforcement. Second, groups with similar tastes naturally establish more edges among them, forming communities. Therefore members of the same community are more likely to share similar interests. We thus expect that, if these two effects are strong, communities will facilitate the internal circulation of memes while preventing diffusion across communities, causing strong concentration or low community diversity.

Unpopular memes tend to be concentrated in a small number of communities, while a few viral memes have high community diversity, spreading widely across communities like epidemic outbreaks [4]. This can be explained by trapping of information flow in communities. Viral memes are able to breach the borders of communities and out-survive other memes. Therefore, features that quantify the community diversity should help predict future meme virality. As an illustration, Fig. 1 is a visualization of the early diffusion patterns of a few memes based on the first 30 tweets, #TheWorseFeeling and #IAdmit clearly exhibit more community diversity than non-viral memes, e.g. #ProperBand and #FollowFool.

Based on the above analysis, we define a key feature of diffusion patterns based on community structure as follows:

**Definition 4 (Adopter Entropy,  $H_n^A(h)$ )** The measurement of entropy describes how adopters of a given meme are scattered or concentrated across communities. Large entropy indicates low concentration or high diffusion diversity:

$$H_n^A(h) = - \sum_{c \in C(h)} \frac{|A_n(h|c)|}{|A_n(h)|} \log \frac{|A_n(h|c)|}{|A_n(h)|}.$$

## 5 Experiment

Let us present the details of our experiment on virality prediction using the diffusion features extracted from the network community structure. We first define a virality prediction task. We will show that diffusion diversity is a strong predictor of virality.

### 5.1 Task Specification

Each new hashtag is associated with a series of adopters within the experiment period. We only compute features using the positions of the first  $n = 25$  adopters in the network. Our method therefore requires that a new hashtag has been used at least 25 times within the experiment period.

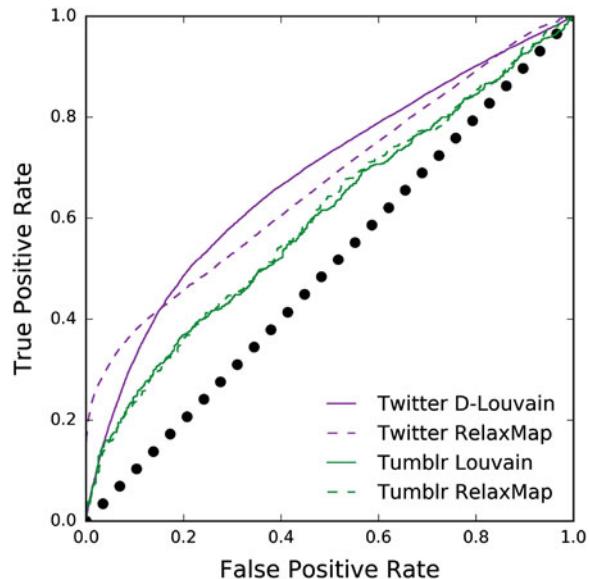
Meme popularity exhibits a broad and skewed distribution, as observed in many previous studies [13, 52]. Our key questions are whether the diffusion diversity feature based on community structure provides a predictive signal, and whether this signal is informative at the large scales of our information-sharing networks. The following recipe defines a meme virality prediction task:

1. Each hashtag is given either viral (1) or not (0) as its ground-truth class; the most-frequent 50% of the inspected hashtags within a month of usage are defined as viral.
2. All hashtags are ranked by adopter entropy  $H_n^A(h)$ , from the highest to the lowest.
3. The top 50% of hashtags based on the ranking in step 2 are predicted as viral.
4. Receiver Operating Characteristic (ROC) curve and the corresponding Area Under the Curve (AUC) are used to evaluate prediction accuracy.

We note that this balanced binary classification task is simpler than the more realistic scenario in which only a small fraction of memes go viral.

ROC curves are drawn by first ranking the scores of the hashtags, then evaluating each sample point as a true positive or false positive in the ranked order. If the true positive data points are among the top ranks, the curve will bounce up, hence the

**Fig. 2** A plot of ROC curves using diffusion diversity (adopter entropy) as the ranking criterion. Different curves correspond to different information-sharing networks and community detection algorithms



**Table 2** Prediction accuracy (AUC) from evaluation on each of the datasets

	Twitter	Tumblr
RelaxMap	0.67	0.60
D-Louvain	0.68	0.60

AUC will be close to one. On the other hand, if false positive sample points are ranked high, the AUC will be close to zero. A random ranking will spread true and false positives evenly, and therefore yield an AUC close to 0.5.

## 5.2 Evaluation

The ROC curves in Fig. 2 and AUC values in Table 2 show that community entropy of adopters  $H_n^A(h)$  alone provides a useful signal in predicting which memes will go viral in large-scale social media. The AUC values around 0.7 and 0.6 for Twitter and Tumblr networks, respectively, represent significant improvements upon the random baseline. Naturally, the results could be improved further by combining entropy with other features in the literature [5, 6].

The RelaxMap and distributed Louvain's methods perform similarly on the same data. Recall that Louvain's method ignores the direction of edges, while RelaxMap does not—InfoMap is based on directed random walks. To investigate the contribution of edge directionality, we ran RelaxMap on an undirected version of the Twitter information-sharing network. This was done by adding weights for reciprocal edges, similarly to the way this is done in the distributed implementation of Louvain's method. The resulting AUC is not significantly different from the

random baseline. This suggests that RelaxMap makes use of both weights and directionality of the edges while extracting communities, and this affects the signal we use for virality prediction.

The diffusion patterns are informative in the prediction task on both Twitter and Tumblr platforms. Despite the simplicity of the task, the results of our evaluation demonstrate that for meme virality prediction, diffusion patterns are robust against source platforms and network construction, and scale up to very large networks.

Compared to Twitter, virality prediction in Tumblr seems to be much more challenging. The difficulty may be attributed to different ways in which the platform is used and the data is collected. First, hashtags on Tumblr tend to be used differently due to the lack of strong limitations on the set of characters. People use hashtags with more characters and diverse types of expression styles, such as irony and sarcasm. As the possible space of hashtags grows, it becomes less clear if the assumption of hashtags as proxies of memes is appropriate. Further, unlike Twitter, Tumblr encourages users to create blog posts without length limitation, giving rise to distinct meme consumption and diffusion patterns.

Another potential difference between the two platforms is the sampling of posts in the Twitter stream, which is biased toward active users who are responsible for most of the tweets. The Tumblr firehose includes barely active and less predictable users.

## 6 Conclusion

In this chapter, we explore the question of virality of online content and its prediction on large social media platforms. We summarize three perspectives on driving factors of virality—innate attractiveness of the content, user characteristics, and the network structure of the underlying social network. We present a simple, yet effective community feature that captures the diffusion patterns of memes in the network. We show that the communities, from which the entropy feature is derived, can be extracted in large-scale information-sharing networks such as Twitter and Tumblr. We also find that diffusion diversity provides a predictive signal across platforms.

There are multiple future directions for this line of research. A noteworthy challenge in deploying the methods in any real-time system is the computational complexity of updating the required features as the social network evolves. Although community structures can be assumed to be fairly stable over time, it is unclear for how long this assumption of static network holds. Consensus clustering [53] could be applied to explore this question.

Another potential direction is to investigate the effect of groups with different characteristics, for instance cultures, religions, and genders, on meme consumption. There has been little work on feature-based models that are aware of group-level characteristics. One can imagine that a meme will gain attention in a particular group while being ignored in others. If early adopters of the meme are in relevant groups

of users who are motivated to share it, the meme is more likely to go viral. Such content-aware approach, accompanied with powerful community features, may lead to the development of more powerful prediction algorithms.

## References

1. Dawkins R (1989) *The selfish gene*. Oxford University Press, Oxford
2. Daley DJ, Kendall DG (1964) Epidemics and rumours. *Nature* 204(4963):1118
3. Goffman W, Newill VA (1964) Generalization of epidemic theory: an application to the transmission of ideas. *Nature* 204:225–228
4. Weng L, Menczer F, Ahn Y-Y (2013) Virality prediction and community structure in social networks. *Sci Rep* 3:2522
5. Weng L, Menczer F, Ahn Y-Y (2014) Predicting successful memes using network and community structure. In: Proceedings of eighth international AAAI conference on weblogs and social media (ICWSM)
6. Cheng J, Adamic L, Dow A, Kleinberg J, Leskovec J (2014) Can cascades be predicted? In: Proceedings of the international world-wide web conference (WWW)
7. Jamali S, Rangwala H (2009) Digging digg: comment mining, popularity prediction, and social network analysis. In: Proceedings of the international conference on web information systems and mining (WISM), pp 32–38
8. Asur S, Huberman BA, Szabo G, Wang C (2011) Trends in social media: persistence and decay. In: Proceedings of the international conference on weblogs and social media (ICWSM)
9. Yang J, Leskovec J (2011) Patterns of temporal variation in online media. In: Proceedings of the ACM international conference on web search and data mining (WSDM), pp 177–186
10. Nikolov S (2012). Trend or no trend: a novel nonparametric method for classifying time series. Technical report. MIT, Cambridge
11. Berger J, Milkman KL (2009), What makes online content viral? *J Market Res* 49(2):192–205
12. Guerini M, Strapparava C, Özbal G (2011) Exploring text virality in social networks. In: Proceedings of the AAAI international conference on weblogs and social media (ICWSM), pp 506–509
13. Weng L, Flammini A, Vespignani A, Menczer F (2012) Competition among memes in a world with limited attention. *Sci Rep* 2:335
14. Salganik M, Dodds P, Watts D (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856
15. Muchnik L, Aral S, Taylor SJ (2013) Social influence bias: a randomized experiment. *Science* 341(6146):647–651
16. Yang L, Sun T, Mei Q (2012) We know what @you #tag: does the dual role affect hashtag adoption? In: Proceedings of the international world-wide web conference (WWW), pp 261–270
17. Cha M, Haddadi H, Benevenuto F, Gummadi KP (2010) Measuring user influence in twitter: the million follower fallacy. In: Proceedings of the international AAAI conference on weblogs and social media (ICWSM), pp 10–17
18. Suh B, Hong L, Pirolli P, Chi EH (2010) Want to be retweeted? Large scale analytics on factors impacting retweet in twitter network. In: Proceedings of the IEEE international conference on social computing, pp 177–184
19. Tang J, Sun J, Wang C, Yang Z (2009) Social influence analysis in large-scale networks. In: Proceedings of the ACM international conference on knowledge discovery and data mining (KDD)
20. Weng J, Lim E-P, Jiang J, He Q (2010) Twitterrank: finding topic-sensitive influential twitterers. In: Proceedings of the ACM international conference on web search and data mining (WSDM)

21. Weng L, Menczer F (2015) Topicality and impact in social media: diverse messages, focused messengers. *PLoS One* 10(2):e0118410
22. Romero DM, Galuba W, Asur S, Huberman BA (2011) Influence and passivity in social media. In: Proceedings of the international world wide web conference (Companion Volume), pp 113–114
23. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA (2010) Identification of influential spreaders in complex networks. *Nat Phys* 6(11):888–893
24. Bakshy E, Mason WA, Hofman JM, Watts DJ (2011) Everyone's an influencer: quantifying influence on twitter. In: Proceedings of the ACM international conference on web search and data mining (WSDM), pp 65–74
25. Brown J, Reingen P (1987) Social ties and word-of-mouth referral behavior. *J Consum Res* 14(3):350–362
26. Granovetter MS (1973) The strength of weak ties. *Am J Sociol* 78(6):1360–1380
27. Leskovec J, Adamic L, Huberman B (2007) The dynamics of viral marketing. *ACM Trans Web* 1(1):1–39
28. Mason WA, Jones A, Goldstone RL (2008) Propagation of innovations in networked groups. *J Exp Psychol Gen* 137(3):422
29. Aral S, Walker D (2011) Creating social contagion through viral product design: a randomized trial of peer influence in networks. *Manag Sci* 57(9):1623–1639
30. Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. *Phys Rev Lett* 86:3200–3203
31. DJ Watts (2002) A simple model of global cascades on random networks. *Proc Natl Acad Sci* 99(9):5766–5771
32. Newman MEJ (2006) Modularity and community structure in networks. *Proc Natl Acad Sci* 103(23):8577–8582
33. Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. *Proc Natl Acad Sci* 105(4):1118–1123
34. Ahn Y-Y, Bagrow J, Lehmann S (2010) Link communities reveal multiscale complexity in networks. *Nature* 466(7307):761–764
35. Fortunato S (2010) Community detection in graphs. *Phys Rep* 486(3):75–174
36. Nematzadeh A, Ferrara E, Flammini A, Ahn Y-Y (2014) Optimal network modularity for information diffusion. *Phys Rev Lett* 113(8):088701
37. Granovetter M (1978) Threshold models of collective behavior. *Am J Sociol* 83(6):1420–1443
38. Schelling TC (1971) Dynamic models of segregation. *J Math Sociol* 1(2):143–186
39. Centola D, Macy M (2007) Complex contagions and the weakness of long ties 1. *Am J Sociol* 113(3):702–734
40. Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194–1197
41. Watts DJ (2002) A simple model of global cascades on random networks. *Proc Natl Acad Sci* 99(9):5766–5771
42. Conover MD, Ferrara E, Menczer F, Flammini A (2013) The digital evolution of occupy wall street. *PLoS One* 8(3):e55957
43. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008(10):P10008
44. Fortunato S, Barthelemy M (2007) Resolution limit in community detection. *Proc Natl Acad Sci* 104(1):36–41
45. Sotera. Spark-distributed-louvain-modularity. <https://github.com/Sotera/spark-distributed-louvain-modularity>
46. Bae S-H, Halperin D, West J, Rosvall M, Howe B (2013) Scalable flow-based community detection for large-scale network analysis. In: IEEE 13th international conference on Data mining workshops (ICDMW), 2013. IEEE, Piscataway, pp 303–310
47. Bakshy E, Karrer B, Adamic L (2009) Social influence and the diffusion of user-created content. In: Proceedings of the ACM conference on electronic commerce, pp 325–334

48. Romero DM, Meeder B, Kleinberg J (2011) Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In: Proceedings of the international world-wide web conference (WWW)
49. Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194–1197
50. McPherson M, Lovin L, Cook J (2001) Birds of a feather: homophily in social networks. *Annu Rev Sociol* 27(1):415–444
51. Centola D (2011) An experimental study of homophily in the adoption of health behavior. *Science* 334(6060):1269–1272
52. Lerman K, Ghosh R (2010) Information contagion: an empirical study of the spread of news on digg and twitter social networks. In: Proceedings of the international AAAI conference on weblogs and social media (ICWSM), pp 90–97
53. Lancichinetti A, Fortunato S (2012) Consensus clustering in complex networks. *Sci Rep* 2:336

# Attention on Weak Ties in Social and Communication Networks



Lilian Weng, Márton Karsai, Nicola Perra, Filippo Menczer,  
and Alessandro Flammini

## 1 Introduction

With the aid of Internet technologies we can easily communicate with essentially anybody in the world at any time. Social media platforms, for example, provide inexpensive opportunities of creating and maintaining social connections and of broadcasting and gathering information through these connections [1]. In fact, the huge amount of information that we create and exchange exceeds our capacity to consume it [2, 3] and increases the competition among ideas for our collective attention [4–6]. As a result, our interactions are steered more than ever before by the “economy of attention” [7, 8]. As Simon predicted:

“What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.” [7]

---

L. Weng  
Affirm Inc., San Francisco, CA, USA

M. Karsai  
Univ Lyon, ENS de Lyon, Inria, CNRS, UCB Lyon 1, LIP UMR 5668, IXXI, Lyon, France  
e-mail: [marton.karsai@ens-lyon.fr](mailto:marton.karsai@ens-lyon.fr)

N. Perra  
Centre for Business Networks Analysis, University of Greenwich, London, UK  
e-mail: [n.perra@greenwich.ac.uk](mailto:n.perra@greenwich.ac.uk)

F. Menczer · A. Flammini (✉)  
Center for Complex Networks and Systems Research, School of Informatics and Computing,  
Indiana University, Bloomington, IN, USA  
e-mail: [aflammin@indiana.edu](mailto:aflammin@indiana.edu)

Attention has thus become a valuable resource to be spent parsimoniously. Here we investigate how individuals allocate attention to different classes of social connections.

In the seminal paper “The strength of weak ties,” Granovetter [9] defines the *strength* of social ties as proportional to the size of the shared social circles of connected individuals. The more common friends two individuals have, the stronger is the tie between them. We adopt this same definition here. In the *weak tie hypothesis*, he postulates that social ties of different strength play distinct roles in the dynamics of social structure and information sharing [9, 10]. In particular, weak ties do not carry as much communication as strong ties do, but they often act as bridges between communities, and thus as important channels for novel information otherwise unavailable in close social circles.

There is a vast literature supporting the idea that weak ties play an important role in spreading novel information across communities [11–15]. This body of work, however, is not concerned with the nature and importance of the information exchanged across ties, and in particular does not confirm (or disprove) the second of Granovetter’s hypotheses, namely that weak ties carry “important” information. One major aim of this chapter is to address this second, more subtle, point by measuring the *attention* that users pay to information exchanged on ties of different strength.

Specifically, here we address two questions:

1. How is the *intensity* of communication related to the strength of a social tie?
2. How is *attention* differently allocated among strong and weak ties?

Answering these two questions leads us to naturally discriminate between ties of different strength and the kind of interactions they represent. In particular we study how social exchange and information gathering interactions are typically related to the strength of the ties. We investigate these questions using three large-scale networks describing different types of human interactions: information sharing in online social media, cell phone calls, and email exchanges.

The first question can be quantitatively addressed by measuring the *strength* of a social tie as the size of the neighborhood shared by two connected agents. Our results, in agreement with previous studies (e.g., by Onnela et al. [13]), confirm the first of the weak tie hypothesis: the largest fraction of interactions do happen on strong ties while weak ties carry much less traffic [9, 13]. We then focus on the second of Granovetter’s hypotheses by examining the role of *attention* and its relationship with tie strength. We propose to use attention as a proxy for the importance of the information exchanged across a tie. Attention is here defined as the fraction of an individual’s activities that is devoted to a particular tie. We study how attention changes as a function of the strength of ties, and examine how it is distributed among the user’s ties to either access information or maintain social connections. Interestingly, we find that only very weak or very strong ties attract a good amount of attention, implying two potentially competing trends. On one hand, people frequently interact with strong ties to satisfy their social needs. On the other hand, people look for information through weak ties, as suggested by

both Granovetter's and Simon's work. The former activity assigns more attention to strong ties, while the latter prefers weak ones. While these observations hold across all the datasets we examine, the relative magnitude of the two tendencies depends on the specific network functionality.

## 2 Related Work

Motivated by Granovetter's work, many empirical studies explored the role of weak ties in social networks mostly by surveys or interviews, and found support for the weak tie hypothesis [11, 12, 16–19]. Brown and Reingen [11] found an important bridging function of weak ties in word-of-mouth referral behavior, allowing information to travel from one distinct subgroup of referral actors to another. Levin and Cross [12] investigated dyadic social ties in transferring useful knowledge. They found that strong ties lead to the reception of useful knowledge more than weak ties, but weak ties benefit knowledge transmission when the trustworthiness is controlled. Gilbert and Karahalios [14] tested several dimensions of tie strength on social media and revealed that both intensity of communication and intimate language are strong indicators of relationship closeness. Strong ties are also believed to provide greater emotional support [20, 21] and to be more influential [11, 15, 22], while weak ties provide novel information and connect us to opportunities outside our immediate circles [9, 23, 24].

Advances in technology have lowered the cost of communication, information production and consumption, and social link formation, creating unprecedented opportunities to study social interactions through massive digital traces [25, 26]. However, only a handful of studies have leveraged recently available large-scale data to explore the weak tie hypothesis. Onnela et al. [13] analyzed a mobile call network and showed that individuals in clusters tend to communicate more, while weak ties, acting as bridges between clusters, have less traffic. Bakshy et al. [15] found that on Facebook, strong ties are individually more influential in propagating information (external URLs) compared to weak ties. However, the greater number of weak ties collectively contribute to a larger influence in aggregate [27]. Weak ties also play a dominant role in slowing down information spreading in temporal networks, due to their special topological bottleneck position and limited communication frequencies [28–31]. The presence of strong and weak ties has been recently linked also to the opposite effect. In fact, the concentration of interactions between strong ties facilitates classes of contagion processes characterized by endemic states such as Susceptible-Infected-Susceptible (SIS) processes [32].

The body of empirical work referenced above includes both small experiments conducted in controlled settings and “big data” approaches. As an introduction to the work presented here, it is important to stress the different advantages that these two approaches bring to the study of weak and strong ties. Big data approaches have obviously the advantage of scale, and, often, of addressing questions in the

wild. Their major weakness is that they provide much less control on the nature of specific social ties and of information exchanged. Here we try to overcome this limitation by adopting attention as a proxy for the importance of the information exchanged and as a tool to infer the nature of a tie.

### 3 Datasets and Network Representation

We consider three very different datasets. The basic statistics of each network are summarized in Table 1.

Twitter network. Twitter is a micro-blogging platform used by many millions of people to broadcast short messages through social connections. Users can subscribe to (or “follow”) people they deem interesting to automatically receive the information they produce. The collection of all “follow” connections forms the *follower network*. In the follower network, each node  $i \in V$  represents a user and a directed link  $(i, j) \in E$  is drawn between nodes  $i$  and  $j$  if user  $i$  follows  $j$ . In such a directed link, we call  $i$  the *source* node and  $j$  the *target* (but note that information travels in the opposite direction). Users post short messages (“tweets”), which may be reposted (“retweeted”) by their followers. We define the weight of a link  $(i, j)$  as the number of times that  $i$  retweets  $j$ .

Twitter allows for other forms of interaction, such as direct mentions of specific users. While these could alternatively be used to define edge weights, mentions are typically used in discussions and do not necessarily indicate replies to previous tweets. Retweets provide a more direct measure of the extent to which a user  $i$  pays attention to information broadcast by  $j$ .

We collected about 934 millions tweets, 150 millions of which were retweets, from a 10% sample of the public tweets provided by the Twitter streaming API.<sup>1</sup> The information about following connections is gathered for a randomly sampled subset of creators of the collected tweets through the Twitter follower API.<sup>2</sup>

Phone call network. The mobile phone call dataset records about 487 millions call events during 120 days with one second resolution. The dataset was recorded by

**Table 1** Statistics of three network datasets

Network name	# Nodes	# Links	% Mutual links	Weight	Duration
Twitter	628,916	44,611,893	64%	# reposts	Mar–Apr 2012
Cell phones	6,101,641	19,013,221	61%	# calls	120 days
Email	86,818	359,817	16%	# messages	Sep 1999–Feb 2002

Note that a link  $(i, j)$  is deemed mutual if both  $(i, j)$  and  $(j, i)$  exist in the network

<sup>1</sup><https://developer.twitter.com/en/docs/tweets/sample-realtime/overview/decahose>.

<sup>2</sup><https://developer.twitter.com/en/docs/accounts-and-users/follow-search-get-users/api-reference/get-followers-ids>.

a single operator with 20% market share in an undisclosed European country.<sup>3</sup> This dataset naturally leads to a social network where nodes represent users, and a direct edge  $(i, j) \in E$  is present if target user  $j$  has received at least one call from source user  $i$ . The weight of each tie represents the number of calls.

Enron email network. The Enron email network records 246,391 emails exchanged inside the Enron corporation. An edge  $(i, j) \in E$  is established if there is at least one email from source user  $i$  to target user  $j$ , as  $i$  directs individual attention to  $j$  intentionally. The weight of an edge is the number of emails from  $i$  to  $j$ . The Enron email corpus was made publicly available during the legal investigation concerning the Enron corporation [33].

## 4 Tie Strength, Weight, and Attention

### 4.1 Tie Strength

In line with Granovetter's hypothesis, we measure *tie strength*—the closeness between two connected users  $i$  and  $j$ —as the Jaccard coefficient between their friend sets [9, 13]:

$$O_{ij} = \frac{|N_i \cap N_j|}{|N_i \cup N_j \setminus \{i, j\}|} \quad (1)$$

where  $N_i$  and  $N_j$  are the sets of neighbors of  $i$  and  $j$ , respectively:

$$N_i = \{u \mid (i, u) \in E \vee (u, i) \in E\}. \quad (2)$$

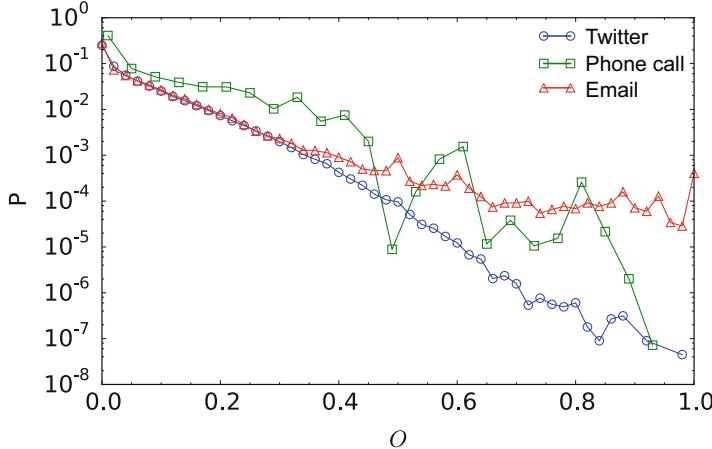
In measuring the strength of a tie according to this definition, we ignore the direction of links. Although considering direction might convey a more nuanced interpretation of the notion of strength itself, it would require introducing an additional hypothesis not directly testable, which we prefer to avoid in this study. Link direction is obviously important when one is concerned with the flow of information, therefore we will consider it later when we examine the information and attention flows.

In the subsequent discussion we also refer to tie strength as *link overlap*. In Fig. 1 we plot the probability distribution of link overlap in the three datasets. All of them present fast (exponential) decay: most ties are weak with little overlap, while only a very small fraction of ties are strong.

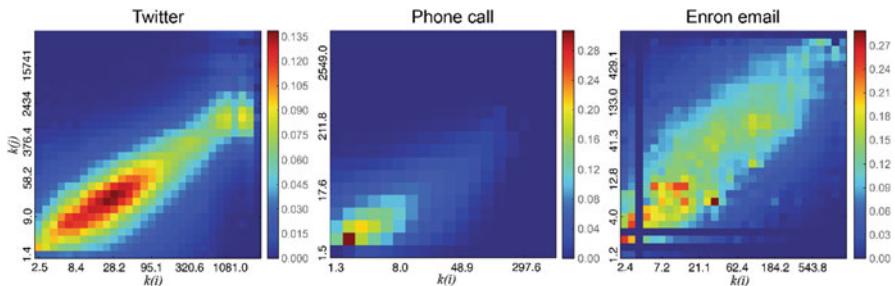
The heat maps in Fig. 2 show tie strength as a function of the degrees of the two nodes connected by the link. In Twitter, high link overlap is more likely to appear

---

<sup>3</sup>A statement about the ethical use of this dataset was issued by Northeastern University's Institutional Review Board.



**Fig. 1** Distribution of link overlap. We plot the probability distributions of link overlap for the three datasets

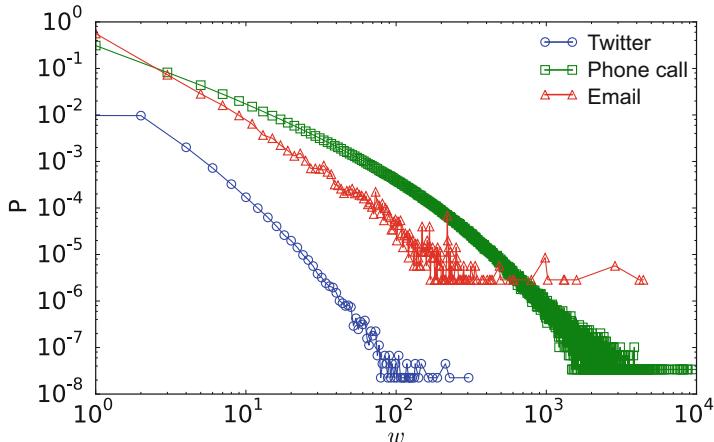


**Fig. 2** Tie strength as a function of the degree. Heat maps of link overlap of an edge  $(i, j)$  as a function of degree  $k(i)$  of the source node  $i$  and degree  $k(j)$  of the target node  $j$  in Twitter, cell phone network and Enron email network. Degrees are plotted using logarithmic bins. The color of each cell represents the average link overlap of all the edges that fall into that bin given the degrees of the target and source nodes. Note that the degree is the sum of in-degree and out-degree, i.e. the number of neighbors of a given node irrespective of direction

between two nodes with similar degrees; in the cell phone call network, ties between users with fewer contacts tend to have higher overlap; in the Enron network, people with similar numbers of email contacts are more likely to have overlapping contact groups.

## 4.2 Weight

The intensity of communication on a tie  $(i, j)$  is quantified by the total number of times that  $i$  retweets, calls, or emails  $j$ , denoted as link weight  $w_{ij}$ . Figure 3

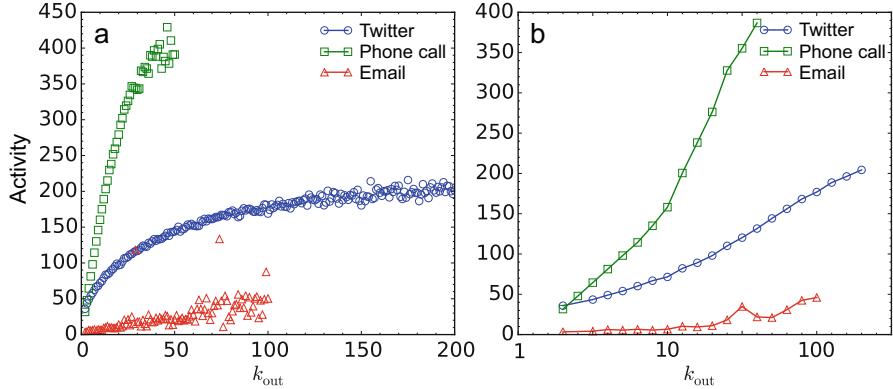


**Fig. 3** Distribution of link weight. We plot the probability distributions of link weights for the three datasets

shows broad distributions of link weights, suggesting that in all three networks, the majority of links carries little traffic but a significant minority supports extremely high volumes of interactions.

### 4.3 Attention

As we mentioned earlier, we propose to use *attention* toward a social contact as a proxy for the importance of information provided by that contact. Attention is therefore a key notion in the present analysis. In principle we would like to have a quantity that measures the amount of cognitive resources that an individual invests in interacting with other individuals. A good proxy could be time spent on the specific “platform” but this information is not available in our data. A second alternative would be the activity of the users (e.g., the tweets produced) but this could yield an artificially low value for users who mostly consume information. A third, computationally convenient alternative is to link attention to the number of friends a user has in the social network. It is reasonable to expect that the cognitive resources spent in maintaining social relationships is, on average, an increasing function of the degree of a node, up until a saturation limit compatible with the finite attention of individuals [2, 3, 5, 6, 34–37], and after which attention should remain essentially constant. There is a considerable amount of empirical work that supports this hypothesis. Romero et al. [38] showed that the probability of adopting (and therefore paying attention to) a hashtag exhibits this qualitative behavior when plotted vs. the number of times the user is exposed to the hashtag—and therefore, on average, the number of friends. Hodas and Lerman [6] found an analogous result for the probability of retweeting a URL. Kwak et al. [39] observed the same



**Fig. 4** Average activity (the number of tweets, calls, or emails) of individuals with a given out-degree on (a) linear and (b) logarithmic scales in three networks. We track users with up to  $k_{\text{out}} = 200$  in the Twitter network,  $k_{\text{out}} = 50$  in the phone call network, and  $k_{\text{out}} = 100$  in the Enron email network to avoid the noise caused by scarcity of data points. More than 92% of users in the Twitter network have  $k_{\text{out}} \leq 200$ , more than 99% of users in the phone call network have  $k_{\text{out}} \leq 50$ , and more than 92% of users in the email network have  $k_{\text{out}} \leq 100$

qualitative behavior between user activity and both number of followers and friends on Twitter. These studies together suggest that different proxies of attention behave in a qualitatively similar fashion when considered as functions of the degree of the user, i.e., a relatively quick growth for small values of the degree, followed by a saturation or a very slow growth regime.

We find support for this general behavior in our datasets as well. Indeed, Fig. 4a illustrates how activity (tweets, phone calls, emails) grows as a function of out-degree (people one follows, calls, or emails). In general, we observe that the activity of an individual grows nonlinearly with out-degree; it can be approximated by a linear dependence in logarithmic scale (Fig. 4b).

To capture this qualitative behavior we define the total attention of user  $i$  as

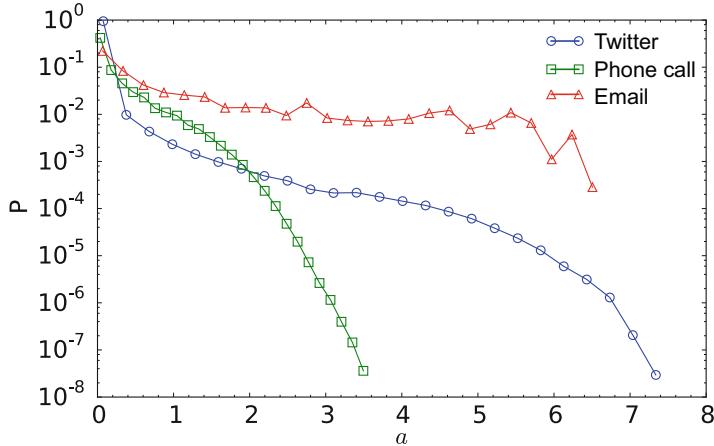
$$a(i) = \alpha \log k_{\text{out}}(i) \quad (3)$$

and without loss of generality, we set  $\alpha = 1$ .

Next, we assume that the fraction of attention devoted by user  $i$  to user  $j$ ,  $a_{ij}$  is proportional to the weight  $w_{ij}$  of link  $(i, j)$ . We thus obtain:

$$a_{ij} = a(i) \cdot \frac{w_{ij}}{\sum_{u \in N_i^{\text{out}}} w_{iu}} = \log k_{\text{out}}(i) \cdot \frac{w_{ij}}{\sum_{u \in N_i^{\text{out}}} w_{iu}} \quad (4)$$

where  $N_i^{\text{out}} = \{u \mid (i, u) \in E\}$ . Unlike tie strength, attention considers direction, because it flows from the source to the target and only depends on the actions of the source. Attention has a narrow distribution in all three datasets, as shown in Fig. 5.



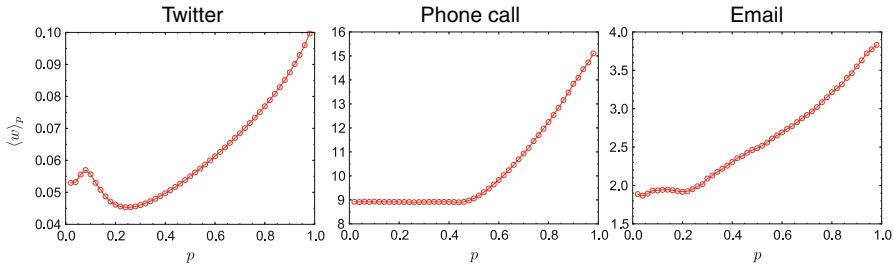
**Fig. 5** Distribution of link attention. We plot the probability distributions of link attention for the three datasets

## 5 Weak Ties Hypothesis and the Role of Attention

The weak tie hypothesis maintains that strong ties carry the majority of interactions, while weak ties act as bridges between communities and are crucial channels for transferring important or novel information. If this is true, we expect that users pay more attention to information received through a weak tie. In the present section we test such hypothesis by measuring how attention is allocated across strong and weak ties. The use of attention as a proxy for importance allows us to overcome the difficulty of defining and empirically measuring the elusive notions of importance or novelty of a piece of information.

### 5.1 Traffic on Strong Ties

As a first step, we aim to confirm that strong ties carry more traffic. To this end we plot the average link weight versus overlap. More precisely, following Onnela et al. [13], we define the average weight  $\langle w \rangle_p$  over the fraction  $p$  of weakest ties (links with lowest overlap), and plot it as a function of  $p$ . As shown in Fig. 6, the average link weights in the three datasets increase as a function of tie strength. Strong ties carry more traffic than weak ties, confirming that people tend to communicate more with close friends, or others with very similar social circles. The observed pattern is consistent with the weak tie hypothesis and with several previous empirical studies [13, 16, 40–43]. The plateaus of the average curves for the weakest ties are due to links with zero overlap. These are quite common: 5.5% of links in Twitter, 40% in the cell phone network, and 23.6% in the Enron email network connect nodes without common neighbors.



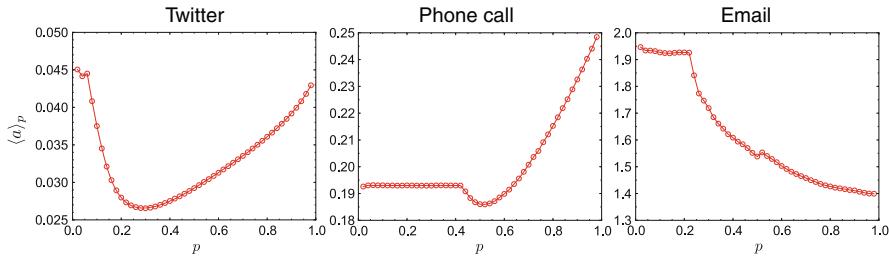
**Fig. 6** Average weight  $\langle w \rangle_p$  of the fraction  $p$  of weakest ties versus  $p$ . Weak links have low overlap (on the left of the  $x$  axis) while strong links have high overlap (on the right)

It is important to stress the diversity of the datasets considered; they reflect the usage of communication media with different purposes, governed by different norms. Despite such differences in usage patterns, the networks corresponding to the three platforms exhibit consistent characteristics. In Twitter, the result implies that users are more likely to adopt and repost messages from neighbors with similar social circles. In the phone call network, people tend to call more frequently individuals with very similar contact lists. In the email network, people working in the same or close divisions of the corporation and thus sharing many common coworkers have more email exchanges. The emerging picture in such diverse networks provides strong evidence for the generality of the first part of Granovetter's weak tie theory.

## 5.2 Attention on Weak Ties

The second part of the weak tie hypothesis states that weak ties function as key communication channels in the social network by conveying important information that one is unlikely to discover through strong ties. Removing a strong tie is unlikely to have a significant effect on our access to information generated in our circle of friends, as alternative contacts could provide the same information. On the other hand, the removal of a weak tie could prevent us from being exposed to information from another community, to which the weak tie provides a bridge. This intuition suggests that more attention could be devoted to information received through weak ties.

Let us compute the average link attention  $\langle a \rangle_p$  over the fraction  $p$  of weakest ties (links with lowest overlap), and plot it as a function of  $p$ . While the three datasets show the same qualitative behavior in link weights (Fig. 6), they exhibit crucial differences in the allocation of attention versus tie strength, as reported in Fig. 7.



**Fig. 7** Average link attention  $\langle a \rangle_p$  of the fraction  $p$  of weakest ties versus  $p$ . Weak ties have low overlap (on the left of the  $x$  axis) while strong ties have high overlap (on the right). The flat portion of the phone call curve for low  $p$  corresponds to a high number of links with zero overlap, i.e., connecting nodes with no common neighbors

The attention curve is U-shaped in the Twitter network—a positive correlation between attention and overlap for strong ties but a negative correlation for weak ties suggests that people are likely to allocate much attention on both very weak and very strong ties. The U-shape is less evident in the phone call network. Weak ties acquire attention slightly more than intermediate ties while the majority of attention is assigned to strong connections.

However, the trend is reversed in the Enron email network, where weak ties are dominant in attracting attention and there is a negative correlation between the amount of attention per tie and its strength.

A possible interpretation for the observed U-shaped attention curves in Twitter and phone data stems from two coexisting trends: on one hand, people are actively maintaining their social relationships by frequent interactions with close friends, so that strong ties capture much attention; on the other hand, people are paying attention to novel and useful information from weak ties. We can argue that a *typical* user pays attention to *both* weak and strong ties. Some users may pay attention to their strongest ties while others may pay attention to their weakest ties. It is conceivable that both tendencies coexist. In the aggregate, attention is split between weak and strong ties.

In Twitter, people follow close friends (strong ties) as well as other important information sources (weak ties). Hence we can observe a combined effect of the attention allocation toward both ends of the tie strength spectrum. It seems plausible for the phone call network to be more driven by social interactions. People often call their closest friends, accounting for the greater attention toward strong ties. Calls to weak ties, such as consumer service hotlines, command attention but are much less common. In contrast, the email exchanges in the Enron dataset happen within a corporation and therefore we presume the network to be information-driven. The tendency for maintaining social relationships in such a network is hardly expected, consistently with the little attention observed on strong ties. This interpretation of the attention patterns, driven by the distinction between information-driven and social-driven communication, is further explored in the next section.

## 6 Social and Informational Links

Attention concentrates on either very weak or very strong ties, as seen in Fig. 7. We conjecture that this observed pattern may originate from the coexistence of two different, potentially competing, communication needs: maintaining social bonds and acquiring novel information. Let us first look into the different types of links in the three networks that might account for these two distinct tendencies.

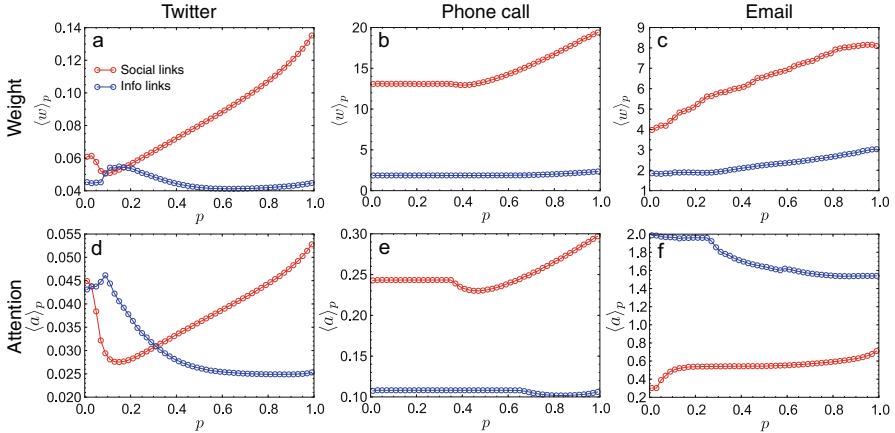
Micro-blogging systems like Twitter, Tumblr, Weibo, and Google+ have several fundamental differences from offline social networks. These systems are designed for efficient information sharing, not only for maintaining mutual friendships. People may establish directed connections unilaterally, and therefore links do not necessarily represent relationships of mutual trust or reciprocal friendship. Many users in micro-blogging platforms follow unknown but interesting others, such as musicians, politicians, technology experts, news sources, and brands. Owing to this special mechanism in micro-blogging systems, Huberman et al. [44] distinguished friends from followers based on the number of reply and mention interactions and pointed out that most traffic is conveyed by an underlying social networks of reciprocal friends.

A similar phenomenon can be found in the phone call network. Real-world friends frequently talk to each other on phone and the interactions are usually intensive, mutual, and long-lasting. Meanwhile, business hotlines and customer services get calls from individual callers on an occasional basis, and the ties between them are expected to be weak and non-mutual.

In the Enron email network, most messages are supposed to be business- or information-driven, and therefore the social activity is weaker than in the Twitter or call networks. The number of exchanges on a tie may still be dependent on how much overlap two individuals have at work, and these routine email exchanges are more likely to go through both directions. However, cross-division communication on a weak tie, though maybe not mutual (i.e., an announcement from the board), is expected to be more crucial and of higher priority, thus attracting more attention.

The social relationship between real-world friends is expected to be different from one between unknown people or coworkers (i.e., a Twitter user following a celebrity, a consumer calling a business hotline, or two coworkers with no personal contact). The former reflects existing social ties, while the latter represents information gathering. We therefore refer to these two classes of connections as *social links* and *informational links*, respectively.

We consider *mutual* links as social and *unilateral* ones (i.e., unreciprocated Twitter followers, phone calls, and emails) as informational [9, 13, 28, 44]. Let us compare the use of these classes of connections by separately computing average link weight and attention as a function of link overlap for social and informational links, respectively. As shown in Fig. 8, we observe clear distinctions between the two types of links in terms of the allocation of both traffic and attention. More importantly, the distinctions provide us with an interpretation of the different distributions of attention observed in the three networks (Fig. 7).



**Fig. 8** Social links versus information links in terms of weight and attention allocation in three networks. In panels (a), (b), and (c) we plot the average link weight  $\langle w \rangle_p$  of the fraction  $p$  of weakest ties versus  $p$ . In panels (d), (e), and (f) we plot the average link attention  $\langle a \rangle_p$  of the fraction  $p$  of weakest ties versus  $p$

Let us start with a discussion of link weights in Fig. 8a–c. In all three networks, social links have larger weights than informational ones, irrespective of tie strengths. Their average weights increase with tie strength. The average weights of informational links, instead, do not display a robust dependence on tie strength. In Fig. 8d–f we display the attention distributions on social ties of different nature. Among *social* links, strong ties attract more attention than weak ones. Among *informational* links, weak ties are more appealing with regard to attention.

Furthermore, considering that links with zero overlap play a special topological role—a perfect bridge<sup>4</sup> connecting distant groups—we expect to see more zero-overlap ties among informational links than among social ties. In Twitter, 7.5% of informational links have zero overlap, compared with 4.4% of social links; in the phone call network, about 65% of informational links have zero overlap versus about 40% of social links; this effect is the strongest in the email network, where 27.5% of informational links have zero overlap as opposed to 4.1% of social links.

The distinctions between informational and social links in terms of attention allocation help us interpret the difference between the patterns observed in Fig. 7. The Twitter network allows users to maintain social contacts and information sources at the same time, and the volume of attention on social and informational links is comparable. The phone call network is more commonly used for social purposes, so informational links only win little attention overall. The email exchanges in the Enron corporate network are designed for gaining information and processing business issues, making information links dominant. In fact the

<sup>4</sup>Note that in our calculation, leaf nodes (with only one out-link) are removed.

Enron email network only contains 16% social (mutual) links, compared to 64% and 61% in Twitter and phone call networks, as shown in Table 1. When we aggregate attention across both classes of links (Fig. 7), the increasing attention toward strong ties is explained by social interactions, while the higher attention toward weak ties originates from informational links. In the Twitter and phone call networks, the combined effects of the two classes of ties lead to the U-shaped attention profiles. In the email network, the predominance of informational links is consistent with the monotonically decreasing attention with increasing tie strength.

## 7 Conclusion

This chapter aimed to verify the two different aspects of the weak tie hypothesis [9] on three large empirical networks. We found that the large majority of interactions are indeed localized among strong ties. We then studied the fraction of an individual's attention directed towards a neighbor to quantify the importance of a social connection with respect to information diffusion. Interestingly we found that while strong ties do carry more traffic, weak ties succeed in attracting attention similar to or even more than strong ties.

We hypothesize that the extent to which weak ties acquire attention can be explained by two distinct link roles, whose prevalence is network dependent. By distinguishing between social and informational links based on reciprocity, we found evidence supporting our interpretation that people interact along strong ties due to their social relationships, while looking for novel information through weak ties. In systems used for information-driven communication, such as a corporate email network, informational links are dominant, explaining higher attention toward weak ties. In systems designed for social communication, such as mobile phones, social links yield more attention and explain the importance of strong ties; however, a portion of traffic is devoted to information seeking, and so we also observe a weaker increase of attention toward weak ties. Finally, microblogs have dual social and informational purposes, explaining the non-monotonic pattern of attention versus tie strength.

Inferring the nature and purpose of a social link from its “usage” is challenging, but could lead to improved ranking algorithms to prioritize social media content. This work aims to be a step in this direction.

While many studies have confirmed the first part of Granovetter’s hypothesis, namely that strong ties receive more traffic in social networks, our analysis provides empirical evidence and a quantitative interpretation of the second part of Granovetter’s theory, i.e., that weak ties are more important for information gathering. Until now, studies in this direction have been hampered by a lack of operational definitions of attention or importance, as well as by limits in the availability of social and communication network data that would allow one to measure these quantities at a large scale. As additional datasets of this kind become available, they will enable further refinements in our understanding of the relationships between strength, attention, and importance of social links.

**Acknowledgements** We would like to thank Albert-László Barabási for the mobile phone cell dataset used in this research, Twitter for providing public streaming data, and the Enron Email Analysis Project at UC Berkeley for cleaning up and sharing the Enron email dataset. MK acknowledges support from LABEX MiLyon. This work was partially funded by NSF grant CCF-1101743 and the James S. McDonnell Foundation.

## References

1. Weng L, Ratkiewicz J, Perra N, Gonçalves B, Castillo C, Bonchi F, Schifanella R, Menczer F, Flammini A (2013) The role of information diffusion in the evolution of social networks. In: Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining (KDD), pp 356–364
2. Dunbar RIM (1998) The social brain hypothesis. *Evol Anthropol* 9(10):178–190
3. Gonçalves B, Perra N, Vespignani A (2011) Modeling users' activity on twitter networks: validation of Dunbar's number. *PLoS One* 6(8):e22656
4. Backstrom L, Bakshy E, Kleinberg J, Lento T, Rosenn I (2011) Center of attention: how facebook users allocate attention across friends. In: Proceedings of the AAAI international conference on weblogs and social media (ICWSM), pp 1–8
5. Weng L, Flammini A, Vespignani A, Menczer F (2012) Competition among memes in a world with limited attention. *Nat Sci Rep* 2:335
6. Hodas NO, Lerman K (2012) How visibility and divided attention constrain social contagion. In: Proceedings of the ASE/IEEE international conference on social computing, p 249–257
7. Simon H (1971) Designing organizations for an information-rich world. In: Greenberger M (ed) Computers, communication, and the public interest, vol 72. The Johns Hopkins Press, Baltimore, pp 37–52
8. Davenport TH, Beck JC (2001) The attention economy: understanding the new currency of business. Harvard Business School Press, Boston
9. Granovetter M (1973) The strength of weak ties. *Am J Sociol* 78(6):1
10. Granovetter M (1995) Getting a job: a study of contacts and careers. University of Chicago Press, Chicago
11. Brown J, Reingen P (1987) Social ties and word-of-mouth referral behavior. *J Consum Res* 14(3):350–362
12. Levin DZ, Cross R (2004) The strength of weak ties you can trust: the mediating role of trust in effective knowledge transfer. *Manag Sci* 50(11):1477–1490
13. Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási A-L (2007) Structure and tie strengths in mobile communication networks. *Proc Natl Acad Sci (PNAS)* 104(18):7332–7336
14. Gilbert E, Karahalios K (2009) Predicting tie strength with social media. In: Proceedings of the ACM international conference on human factors in computing systems (CHI), pp 211–220
15. Bakshy E, Rosenn I, Marlow C, Adamic L (2012) The role of social networks in information diffusion. In: Proceedings of the ACM international conference world wide web (WWW), pp 519–528
16. Friedkin N (1980) A test of structural features of granovetter's strength of weak ties theory. *Soc Netw* 2(4):411–422
17. Lin N, Ensel WM, Vaughn JC (1981) Social resources and strength of ties: structural factors in occupational status attainment. *Am Sociol Rev* 46:393–405
18. Granovetter M (1983) The strength of weak ties: a network theory revisited. *Sociol Theory* 1(1):201–233
19. Nelson RE (1989) The strength of strong ties: social networks and intergroup conflict in organizations. *Acad Manag J* 32(2):377–401
20. Haythornthwaite C, Wellman B (1998) Work, friendship, and media use for information exchange in a networked organization. *J Am Soc Inf Sci* 49(12):1101–1114

21. Wellman B, Wortley S (1990) Different strokes from different folks: community ties and social support. *Am J Sociol* 96(3):558–588
22. Bond RM, Fariss CJ, Jones JJ, Kramer ADI, Marlow C, Settle JE, Fowler JH (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415):295–298
23. Putnam RD (2001) *Bowling alone: the collapse and revival of American community*. Simon and Schuster, New York
24. Burt RS (2009) *Structural holes: the social structure of competition*. Harvard University Press, Boston
25. Lazer D, Pentland A, Adamic L, Aral S, Barabási A-L, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, Jebara T, King G, Macy M, Roy D, Alstoy MV (2009) Computational social science. *Science* 323(5915):721–723
26. Vespignani A (2009) Predicting the behavior of techno-social systems. *Science* 325(5939):425–428
27. Meo PD, Ferrara E, Fiumara G, Provetti A (2014) On facebook, most ties are weak. *Commun. ACM* 57(11):78–84
28. Karsai M, Kivelä M, Pan RK, Kaski K, Kertész J, Barabási A-L, Saramäki J (2011) Small but slow world: how network topology and burstiness slow down spreading. *Phys Rev E* 83(2):025102
29. Miritello G, Moro E, Lara R (2011) Dynamical strength of social ties in information spreading. *Phys Rev E* 83(4):045102
30. Karsai M, Perra N, Vespignani A (2014) Time varying networks and the weakness of strong ties. *Sci Rep* 4:4001
31. Ubaldi E, Perra N, Karsai M, Vezzani A, Burioni R, Vespignani A (2016) Asymptotic theory of time-varying social networks with heterogeneous activity and tie allocation. *Sci Rep* 6:35724
32. Sun K, Baronchelli A, Perra N (2015) Contrasting effects of strong ties on sir and sis processes in temporal networks. *Eur Phys J B* 88(12):1–8
33. Klimt B, Yang Y (2004) The Enron corpus: a new dataset for email classification research. In: *Proceedings of the European conference on machine learning (ECML)*, pp 217–226
34. Miritello G, Moro E, Lara R, Martínez-López R, Belchamber J, Roberts SGB, Dunbar RIM (2013) Time as a limited resource: communication strategy in mobile phone networks. *Soc Netw* 35(1):89–95
35. Stiller J, Dunbar RIM (2007) Perspective-taking and memory capacity predict social network size. *Soc Netw* 29(1):93–104
36. Baronchelli A, Ferrer-i Cancho R, Pastor-Satorras R, Chater N, Christiansen MH (2013) Networks in cognitive science. *Trends Cogn Sci* 17(7):348–360
37. Arnaboldi V, Conti M, Passarella A, Dunbar R (2013) Dynamics of personal social relationships in online social networks: a study on twitter. In: *Proceedings of the first ACM conference on online social networks*. ACM, New York, pp 15–26
38. Romero DM, Meeder B, Kleinberg J (2011) Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In: *Proceedings of the ACM international conference on world wide web (WWW)*, pp 695–704
39. Kwak H, Lee C, Park H, Moon S (2010) What is twitter, a social network or a news media? In: *Proceedings of the ACM international conference on world wide web (WWW)*, pp 591–600
40. Onnela J-P, Saramäki J, Hyvönen J, Szabó G, De Menezes MA, Kaski K, Barabási A-L, Kertész J (2007) Analysis of a large-scale weighted network of one-to-one human communication. *New J Phys* 9(6):179
41. Cheng X-Q, Ren F-X, Shen H-W, Zhang Z-K, Zhou T (2010) Bridgeness: a local index on edge significance in maintaining global connectivity. *J Stat Mech Theory Exp* 2010(10):P10011
42. Grabowicz PA, Ramasco JJ, Moro E, Pujol JM, Eguiluz VM (2012) Social features of online networks: the strength of intermediary ties in online social media. *PLoS One* 7(1):e29358
43. Pajevic S, Plenz D (2012) The organization of strong links in complex networks. *Nat Phys* 8(5):429–436
44. Huberman B, Romero D, Wu F (2009) Social networks that matter: twitter under the microscope. *First Monday* 14(1):8

# Measuring Social Spam and the Effect of Bots on Information Diffusion in Social Media



Emilio Ferrara

## 1 Introduction

Social media have received widespread recognition as enablers of modern society communication [14, 18, 55, 56, 58], as a tool to democratize discussion about politics [2, 10, 15, 25, 26, 61, 90] and social issues [9, 22, 23, 40, 41, 81, 84], and even as an effective system to respond to crises and emergencies [39, 57, 78, 91, 92].

The benefits of the rise to popularity of social media are hard to quantify, as they touch billions of people every day, all over the world. However, as early as 2006, concerns have been raised regarding the possibility of manipulating public opinion through social media [44]. Particularly problematic can be the fact that social media have proved effective in influencing individuals, their beliefs and behaviors [7, 17, 33, 54, 67]. These concerns have been later proved well grounded by several scientific studies, which highlighted a variety of manipulation strategies and related contexts where such forms of abuse can take place [27, 30, 32, 45, 66, 72, 73, 86].

One way to manipulate social media is by using social bots, algorithmically-controlled accounts that emulate the activity of human users but operate at much higher pace (e.g., automatically producing content or engaging in social interactions), while successfully keeping their robotic identity undisclosed [36, 46, 65, 85].

Evidence of the adoption of social media bots to attempt manipulating political communication dates back nearly a decade: during the 2010 U.S. midterm elections, social bots were employed to support some candidates and smear others, by

---

E. Ferrara (✉)

University of Southern California, Information Sciences Institute, Los Angeles, CA, USA  
e-mail: [emiliofe@usc.edu](mailto:emiliofe@usc.edu)

injecting thousands of tweets pointing to websites with fake news [71]. The research community reported another similar case around the time of the 2010 Massachusetts special election [66]. Campaigns of this type are sometimes referred to as astroturf or Twitter bombs. Unfortunately, most of the times, it has proven impossible to determine who's behind these types of operations [11, 36, 53]. Governments, organizations, and other entities with sufficient resources can obtain the technological capabilities to deploy thousands of social bots and use them to their advantage, either to support or to attack particular political figures or candidates.

Bots have been used in other contexts too, most prominently for social spamming and social phishing purposes [48, 50, 69, 74, 82, 83, 89]. A large body of scientific literature covers the challenges related to detecting social spam [38, 63, 94], spam bots [12, 59, 60, 76], fake reviews [69], etc. Differently from traditional Internet spam, distributed via email or mailing lists, social spam proliferates in online platforms, and bots have been extensively used to make its diffusion more effective. Although much work has been devoted to characterize and detect social spam campaigns or spam bots, the interplay between these two, and in particular the effect of spam bots on the diffusion of spam in social media, has not received much attention.

## 1.1 Contributions of This Chapter

This chapter aims at investigating both the directions of social bots influence on political discussion and spam bots influence in social spam campaigns. In particular, we will be concerned with measuring the role and effects of bots in social media information spreading dynamics. The scope and contributions of this chapter are therefore threefold:

- We will first review how social bots, and in particular Twitter bots, are created, how they operate, and what are the challenges in detecting them (see Sect. 2). The literature discussed here will be mostly aligned with a recent review paper we published on *Communications of the ACM* [36].
- We will then discuss how social bots have been used during the 2016 US Presidential Election to sway the discussion around the presidential candidates, and to frame agendas and messages attaching particular sentiments. This review (see Sect. 3.1) will be based on results we recently published [11].
- Then, we will propose novel analysis of the effects of social spam bots on the diffusion of social spam campaigns and promotional content on Twitter (see Sect. 3.2). We will investigate the differences between traditional spammers and social spam bots, provide a characterization of their most typical features, and describe their effect of the diffusion of social spam on Twitter.

## 2 What Social Bots Are and How They Operate

### 2.1 How to Create a Social Spam Bot

In the early days of online social media, over one decade ago, creating a bot was not a simple task: a skilled programmer would need to sift through various platforms' documentation to create a software capable of automatically interfacing with the platform and operate functions in a human-like manner. For example, in 2009, we spent significant amounts of efforts to create a simple bot that would navigate Facebook pages and extract basic publicly-available social network information [16]: that required the application of sophisticated Web scripting techniques [35] in conjunction with a trial-and-error approach to deal with the Web platform infrastructure. Similar efforts have been reported for other such type of early endeavors [4, 20].

These days, the landscape has completely changed: indeed, it has become increasingly simpler to deploy social bots, so that, in some cases, no coding skills are required to set up accounts that perform simple automated activities: tech blogs often post tutorials and ready-to-go tools for this purposes. Various source codes for sophisticated social media bots can be found online as well, ready to be customized and optimized by the more technically-savvy users [53].

We inspected some of the readily-available Twitter bot-making tools and this is a (non-comprehensive) list of capabilities they provide:

- Search Twitter for phrases/hashtags/keywords and automatically retweet them;
- Automatically reply to tweets that meet a certain criteria;
- Automatically follow any users that tweet something with a specific hashtag, keyword, or phrase;
- Automatically follow back any users that have followed the bot;
- Automatically follow any users that follow a specified user;
- Automatically add users tweeting about something to public lists;
- Search Google (and other engines) for articles/news according to specific criteria and post them, or link them in automatic replies to other users;
- Automatically aggregating public sentiment on certain topics of discussion;
- Buffer and post tweets automatically.

Most of these bots can run within cloud services or infrastructures like Amazon Web Services (AWS) or Heroku, making it more difficult to block them when they violate the Terms of Service of the platform where they are deployed.

Finally, a very recent trend is that of providing Bot-As-A-Service (BaaS): companies like RoboLike<sup>1</sup> provide “Easy-to-use Instagram/Twitter auto bots” performing certain automatic activities for a monthly price. Advanced conversational

---

<sup>1</sup>RoboLike: <https://roboLike.com/>.

bots powered by sophisticated Artificial Intelligence are provided by companies like ChatBots.io that allow anyone to “Add a bot to services like Twitter, Hubot, Facebook, Skype, Twilio, and more”.<sup>2</sup>

## 2.2 How to Detect Social Bots

The detection of social bots in online social media platform has proven a challenging task. For this reason, it has attracted a lot of attention from the computing research community. Even DARPA became interested to the point that a DARPA Challenge was organized, namely the 2016 DARPA Twitter Bot Detection [77]: over one dozen academic and industry teams participated, with University of Maryland, University of Southern California, and Indiana University topping the challenge.

For these reasons, the literature on social bot detection has become very extensive. We tried to summarize the most relevant approaches in a survey paper recently appeared on *Communications of the ACM* [36]: we refer the interested reader to that review for a deeper analysis of this problem.

In our review, we proposed a simple taxonomy to divide the social bot detection approaches proposed in literature into three classes: (1) bot detection systems based on social network information; (2) system based on crowd-sourcing and leveraging human intelligence; (3) machine learning methods based on the identification of highly-revealing features that discriminate between bots and humans. In the following, we report some examples of these three classes.

### 2.2.1 Graph-Based Social Bot Detection

Social bot detection has been framed as an adversarial setting [6]: an adversary may control multiple social bots to impersonate different identities and infiltrate a system. Proposed detection strategies often rely on examining the structure of a social graph, and assume that bot accounts exhibit a small number of links to legitimate users, connecting mostly to other bots. This feature is exploited to identify densely interconnected groups of bots. Yet, a wise attacker may counterfeit the connectivity of the controlled bot accounts; this strategy would make the attack invisible to these detection methods. To address this shortcoming, some systems also employ the paradigm of *innocent by association*: an account interacting with a legitimate user is considered itself legitimate. Unfortunately, the effectiveness of such detection strategies is bound by the behavioral assumption that legitimate users refuse to interact with unknown accounts. This was proven unrealistic by various experiments [13, 29, 76]. On other platforms like Twitter and Tumblr, connecting and interacting with strangers is one of the main features. In these circumstances, the

---

<sup>2</sup>Pandora bot: <https://developer.pandorabots.com/>.

innocent-by-association paradigm yields high false positive rates. Moreover, real-world platforms may contain many mixed groups of legitimate users who fall prey of some bots [6], and sophisticated bots may succeed in large-scale infiltration making it impossible to detect them solely from network structure information. Despite its high false-positive rate, social network information can complement other sources of information to improve prediction accuracy, as demonstrated by prior work [36].

### 2.2.2 Crowd-Sourcing Social Bot Detection

Some authors suggested crowd-sourcing social bot detection, assuming that it would be a simple task for humans to evaluate an account's behavior and to observe emerging patterns and anomalies associated with bots [88]. Using data from Facebook and Renren (a popular Chinese online social network), the authors tested the efficacy of human detectors, using both expert annotators and workers hired online. Although this strategy exhibited a near-zero false positive rate, it has proven unfeasible for several reasons: for existing platform with large user bases, like Facebook and Twitter, manually verify millions of suspicious accounts has a prohibitive cost; even if large social network companies could afford to hire teams of analysts for this purpose [75], such cost might not be sustainable for small social networks in their early stages; finally, exposing personal information to online workers for annotation would raise privacy issue [28].

### 2.2.3 Feature-Based Social Bot Detection

Encoding behavioral patterns into features, in conjunction with machine learning techniques to learn the signature of human and bot behavior, may be the most popular bot detection strategy. One example of feature-based system is represented by *Bot or Not*: released in 2014, and constantly updated, this was the first Twitter bot detection tool to be made publicly available [24].<sup>3</sup> *Bot or Not* implements a detection algorithm relying upon highly-predictive features capturing a variety of suspicious behaviors to separate social bots from humans. The system employs off-the-shelf supervised learning algorithms trained with examples of both humans and bots behaviors. In addition to the classification results, *Bot or Not* provides a variety of interactive visualizations that yield insights on the features exploited by the system. We will later describe how we used *Bot or Not* for our studies.

Bots are continuously changing and evolving: the analysis of the highly-predictive behaviors that feature-based detection systems can detect may reveal interesting patterns and provide unique opportunities to understand how to discriminate between bots and humans. User meta-data are considered among the most predictive features and the most interpretable ones [46, 88]: we can suggest few

---

<sup>3</sup><http://Truthy.indiana.edu/botornot>.

rules of thumb to infer whether an account is likely a bot, by comparing its meta-data with that of legitimate users. Further work, however, will be needed to detect sophisticated strategies exhibiting a mixture of humans and social bots features (sometimes referred to as *cryptobots*). Detecting these bots, or hacked accounts [93], is currently impossible for feature-based systems. Recent studies suggested that some advanced social bots may no longer aim at mimicking human behavior, but rather at misdirecting attention to irrelevant information [1]: such *smoke screening* strategies, requiring high degree of coordination among bots, can also escape feature-based detection systems.

### 3 Applications and Case Studies

In the following, we present two case studies. We first study the use of social bots in the context of the 2016 US Presidential Election (cf. Sect. 3.1). The results we present are based on recently published work [11]. Then, we discuss new results on the effect of bots on the diffusion of social media spam (cf. Sect. 3.2).

#### 3.1 Case Study 1: Political Campaigns

In the introduction of this chapter, we discussed at length the widespread abuse of social media platforms. In the context of political campaigns, one could try to boost the popularity of a candidate, for example by creating the impression that there is an organic support behind that candidate; however, the apparent support can be artificially generated by means of orchestrated campaigns. This phenomenon is commonly referred to as *astroturf*, and it has long-lasting roots, starting from offline campaigns [62], and evolving, during more recent times, into various forms of Internet [52] and social media [72] campaigns. We report our study of social media astroturf in the context of the 2016 US Presidential Election next, with a special focus on the role of social bots. We discuss data collection first, then we go over the employed bot detection and sentiment analysis approaches. The case study concludes with some discussion of the insights our analysis yielded.

##### 3.1.1 Data Collection

We manually crafted a list of hashtags and keywords related to the 2016 US Presidential Election. The list was compiled so that to contain a roughly equal number of hashtags/keywords associated with each major presidential candidate: we selected 23 terms in total, including 5 terms specifically for the Republican Party nominee Donald Trump (#donaldtrump, #trump2016, #neverhillary, #trump-pence16, #trump), 4 terms for the Democratic Party nominee Hillary Clinton

(#hillaryclinton, #imwithher, #nevertrump, #hillary), and several terms relative to the four presidential debates. The full list of search terms is reported in our paper [11]. By querying the Twitter Search API at regular intervals of 10s, continuously and without interruptions in three periods between September 16 and October 21, 2016, we collected a large dataset constituted by 20.7 million tweets posted by nearly 2.8 million distinct users. We used the Twitter Search API<sup>4</sup> to obtain all tweets that contain the search terms, posted during the data collection period, rather than a sample of unfiltered tweets: this avoids incurring in the issues reported in the literature related to collecting sample data from the Twitter Stream API<sup>5</sup> instead [68].

### 3.1.2 Bot Detection

Determining whether either human or a bot controls a social media account has proven a very challenging task [36, 77]. Our prior efforts produced an openly accessible solution called Bot Or Not [24], consisting of a Python API<sup>6</sup> and a Website.<sup>7</sup> As we briefly discussed earlier, Bot Or Not is a machine-learning framework that extracts and analyzes a set of over one thousand features, spanning content and network structure, temporal activity, user profile data, and sentiment analysis to produce a score that suggests the likelihood that the inspected account is indeed a social bot. Extensive analysis revealed that the two most important classes of feature to detect bots are, maybe unsurprisingly, the metadata and usage statistics associated with the user accounts.

The following indicators provide the strongest signals to separate bots from humans: (1) whether the public Twitter profile looks like the default one or it is customized (it requires some human efforts to customize the profile, therefore bots are more likely to exhibit the default profile setting); (2) absence of geographical metadata (humans often use smartphones and the Twitter iPhone/Android App, which records as digital footprint the physical location of the mobile device); (3) and activity statistics such as total number of tweets and frequency of posting (bots exhibit incessant activity and excessive amounts of tweets), proportion of retweets over original tweets (bots retweet contents much more frequently than generating new tweets), proportion of followers over followees (bots usually have less followers and more followees), account creation date (bots are more likely to have recently-created accounts), randomness of the username (bots are likely to have randomly-generated usernames). We point the reader interested in further technical details to our prior work [24, 36].

---

<sup>4</sup>Twitter Search API: <https://dev.twitter.com/rest/public/search>.

<sup>5</sup>Twitter Stream API: <https://dev.twitter.com/streaming/overview>.

<sup>6</sup>Bot or Not Python API: <https://github.com/truthy/botornot-python>.

<sup>7</sup>Bot or Not Website: <https://truthy.indiana.edu/botornot/>.

Bot Or Not has been trained with thousands of instances of social bots, from simple to sophisticated, and an accuracy of above 95% [24]. Typically, Bot Or Not yields likelihood scores above 50% only for accounts that look suspicious to a scrupulous analysis. We adopted the Python Bot Or Not API to systematically inspect the most active users in our dataset. The Python Bot Or Not API queries the Twitter API to extract the most 300 tweets and all the publicly available account metadata, and feed this features to an ensemble of machine learning classifiers, which produce a bot score. To label accounts as bots, we use the 50% threshold—which has proven effective in prior studies [24, 36]—an account is considered to be a bot if the bot score is above 0.5.

Since the Python Bot Or Not API incurs in the query limitations imposed by the Twitter API, it would have been impossible to test all the 2.78 million accounts. Therefore, we tested the top 50 thousand accounts ranked by activity volume. Although these top 50 thousand users account for roughly only 2% of the entire population, it is worth noting that they are responsible for producing over 12.6 million tweets, which is about 60% of the total conversation. This choice gives us sufficient statistical power to extrapolate the distribution of bots and humans for the entire population without the need to test accounts that are only marginally involved in the conversation. Out of the top 50 thousand accounts, Bot Or Not assigned a bot score greater than the established 0.5 threshold, and therefore classified as likely bots, to a total of 7183 users, responsible for 2,330,252 tweets. A total of 40,163 users (responsible for 10.3 million tweets) were labeled as humans. Bot Or Not labeled the remainder 2654 users as unknown/undecided, either because their scores does not significantly diverge from the classification threshold of 0.5, or because the accounts have been suspended/deleted. Even if all the 2654 users were bots, and Twitter suspended their accounts for violating the terms of service, this would suggest that roughly 70% of the total bot population (the remainder 7183 accounts) was still active on the platform at the time of our verification. By extrapolating for the entire population, we estimate the presence of at least 400 thousand bots, accounting for roughly 15% of the total Twitter population active in the U.S. presidential election discussion, and responsible for about 3.8 million tweets, roughly 19% of the total volume. Additional statistics are summarized in our paper [11].

### 3.1.3 Sentiment Analysis

To understand how bots and humans discuss about the presidential candidates we will rely upon sentiment analysis. To attach a sentiment score to the tweets in our dataset, we used SentiStrength [80]. SentiStrength is a sentiment analysis algorithm which has been specifically designed to annotate social media data. This design choice provides some desirable advantages: first, it is optimized to annotate short, informal texts, like tweets, that contain abbreviations, slang, and other non-orthodox language features; second, SentiStrength employs additional linguistic rules for negations, amplifications, booster words, emoticons, spelling corrections, etc. Applications of SentiStrength to social media data found it particularly effective

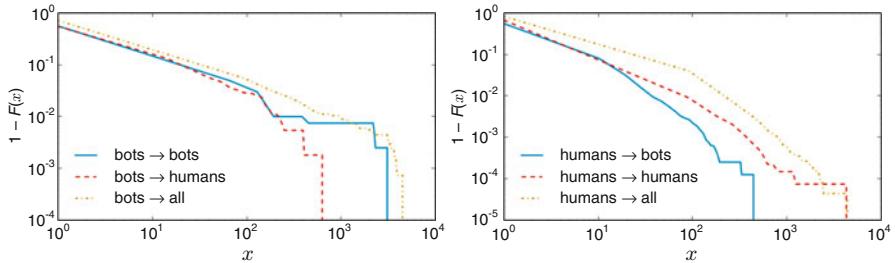
at capturing positive and negative emotions with, respectively, 60.6% and 72.8% accuracy [79]. We tested it extensively and also used it in prior studies to validate the effect of sentiment on the diffusion of information in social media [33]. The algorithm assigns to each tweet  $t$  a positive  $P^+(t)$  and negative  $P^-(t)$  polarity score, both ranging between 1 (neutral) and 5 (strongly positive/negative). Starting from the polarity scores, we capture the emotional dimension of each tweet  $t$  with one single measure, the sentiment score  $S(t)$ , defined as the difference between positive and negative polarity scores:  $S(t) = P^+(t) - P^-(t)$ . The above-defined score ranges between  $-4$  and  $+4$ . The negative extreme indicates a strongly negative tweet, and occurs when  $P^+(t) = 1$  and  $P^-(t) = 5$ . Vice-versa, the positive extreme identifies a strongly positive tweet labeled with  $P^+(t) = 5$  and  $P^-(t) = 1$ . In the case  $P^+(t) = P^-(t)$ —positive and negative sentiment scores for a tweet  $t$  are the same—the sentiment  $S(t) = 0$  of tweet  $t$  is considered as neutral (note that the neutral class represents the majority, by construction, since it contains all tweets that have equal number of positive and negative words, as well as all tweets with no sentiment-labeled terms).

### 3.1.4 Partisanship and Supporting Activity

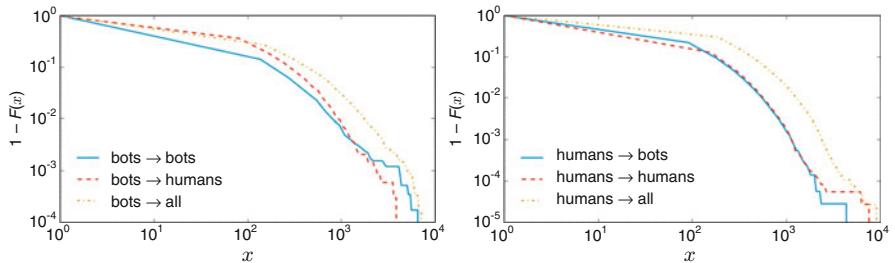
We next inferred the partisanship of the users in our dataset. We used the five Trump-supporting hashtags (#donaldtrump, #trump2016, #neverhillary, #trumppence16, #trump) and the four Clinton-supporting (#hillaryclinton, #imwithher, #nevertrump, #hillary) to attribute partisanship. In detail, we employed a simple heuristics based on hashtag adoption: for each user, we calculated the top ten hashtags that appear in the tweets posted by that user. If the majority of hashtags support one particular candidate, we assigned the given user to that political faction (Clinton- or Trump-supporter). This is a very strict and conservative partisanship assignment, likely less prone to misclassification that may be yield by automatic machine-learning techniques not based on manual validation, e.g., [21]. Our procedure yielded a small, high-confidence, annotated dataset constituted by 7112 Clinton supporters (590 bots and 6522 humans) and 17,202 Trump supporters (1867 bots and 15,335 humans).

### 3.1.5 Analytic Insight 1: Human vs. Bot Engagement

Figures 1 and 2 show the Complementary Cumulative Distribution Functions (CCDFs) of the interactions respectively replies and retweets, initiated by bot and human users. Each plot disaggregates the interactions in three categories: (1) within group (for example, bot–bot, or human–human); (2) across groups (e.g., bot–human, or human–bot); and, (3) total (i.e., bot-all and human-all). Both figures exhibit broad distributions typical of social media activity. What interestingly emerges from contrasting the two figures is that humans are engaging in replies interactions significantly more (one order of magnitude difference) with other humans than with bots (see right panel of Fig. 1). Conversely, bots fail to substantially engage humans and end up interacting via replies with other bots significantly more than



**Fig. 1** Complementary cumulative distribution function (CCDF) of replies interactions generated by bots (left) and humans (right) (published in Bessi and Ferrara, 2016 [11])

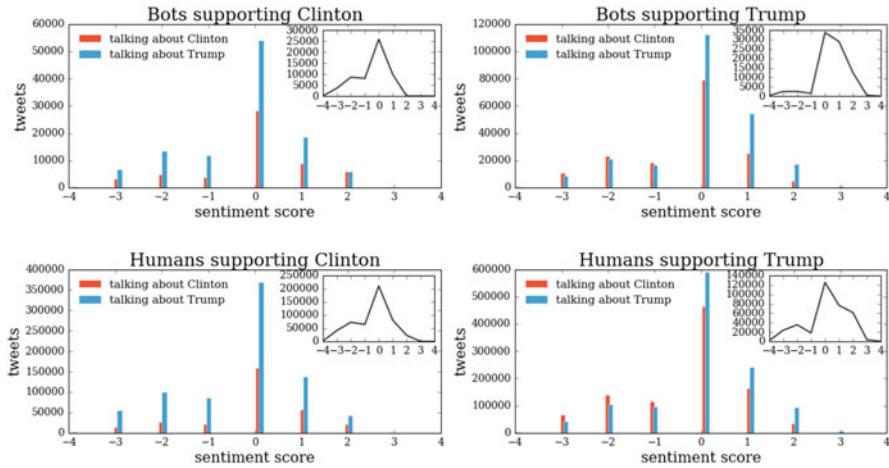


**Fig. 2** Complementary cumulative distribution function (CCDF) of retweets interactions generated by bots (left) and humans (right) (published in Bessi and Ferrara [11])

with humans. Given that bots by design are intended to engage in interactions with humans, our observation goes against what we would have intuitively expected—similar paradoxes have been highlighted in our prior work [36]. One intuitive explanation to this phenomenon is that bots that are not sophisticated enough, cannot produce engaging-enough questions to foster meaningful discussions with humans. Figure 2, however, demonstrates that rebroadcasting is a much more effective channel of information spreading: there is no significant difference in the amounts of retweets that humans generate by rebroadcasting content produced by other humans or by bots. In fact, humans and bots retweet each other substantially at the same rate. This suggests that bots are being very effective at spreading information in the human population, which could have some nefarious consequences in the cases when humans fail at verifying the correctness and accuracy of such information and information sources.

### 3.1.6 Analytic Insight 2: Human vs. Bot Sentiment

To further understand how social media users (both bots and humans) are talking about the two presidential candidates, we explore the sentiment that the tweets convey. To this purpose, we rely upon sentiment analysis and in particular on *SentiStrength*. Figure 3 shows four panels: the top two panels illustrate the sentiment



**Fig. 3** Distributions of the sentiment of bots (top) and humans (bottom) supporting the two presidential candidates. The main histograms show the disaggregated volumes of tweets talking about the two candidates separately, while the insets show the absolute value of the difference between them (published in Bessi and Ferrara [11])

of the tweets produced by the bots, while the bottom two panels show the same information for tweets generated by humans. Furthermore, the two left panels show the support to Hillary Clinton (respectively by bots and humans), whereas the two right panels show the support to Donald Trump (respectively by bots and humans). The main histograms in each panel show the volume of tweets about Clinton or Trump, separately, whereas the insets show the difference between the two (this to illustrate the disproportion in support of the candidate of one's factions, as opposed to the other candidate). What appears evident from contrasting the left and right panels is that, on average, the tweets produced by Trump's supporters are significantly more positive than that of Clinton's supporters, regardless of whether the source is human or bot. If we focus on Trump's bot supporters, we note that they generate almost no negative tweets; they indeed produce the most positive set of tweets in the entire dataset—a very significant fraction of these non-negative bot-generated tweets (about 200,000 or nearly two-third of the total) are in support of Donald Trump. This generates a stream of support that is at staggering odds with respect to the overall negative tone that characterizes the 2016 presidential election campaigns. The fact that bots produce systematically more positive content in support of a candidate can bias the perception of the individuals exposed to it, suggesting that there exists an organic, grassroots support for a given candidate, while in reality it is all artificially generated. Some interesting insights emerge also from the analysis of Clinton's supporters: on average, human-generated tweets show slightly more positive sentiment toward the candidate than the bot-generated ones. Overall, a more natural distribution of tweets' sentiment emerges from the two groups of bots and human supporters, with a roughly equal number of positive and negative tweets being present in the pro-Clinton discussion. To further understand

these dynamics, we manually analyzed two hashtags, namely #NeverTrump and #NeverHillary, as emblematic examples of campaigns explicitly devoted to target the candidate of one's opposing political leaning. The hashtag #NeverTrump, used by supporters of the Democratic Candidate Hillary Clinton, accrued 105,906 positive tweets, and 118,661 negative ones, roughly an equal split; on the other hand, the hashtag #NeverHillary pushed by Trump's supporters generated significantly more negative tweets (204,418) than positive ones (171,877). The paper [11] reports various examples of tweets generated by bots, and the candidate they support. A final consideration emerges when contrasting the pro-Clinton and pro-Trump factions: the former focuses much more on their candidate, with a significant number of tweets referring to Clinton. Conversely, pro-Trump supporters (humans and bots) devote a significant number of tweets to their opponent: in fact, the majority of negative tweets generated by both humans and bots are addressing Hillary Clinton.

### 3.2 Case Study 2: Social Spam Campaigns

In the second part of this chapter, we study social spam campaigns. The widespread use of social media makes them an ideal target as a vector to diffuse spam campaigns. Indeed, spam has evolved, moving away from traditional vectors like emails and mailinglists [43], due to the increasing effectiveness of email spam filters, and migrating to social platforms like social media [19, 38, 94] and digital marketplaces [51, 64, 70], etc. In the former scenario, the use of bots has been documented to generate artificial promotional campaigns, to advertise dubious products (whose sale is sometimes illicit), etc. In the latter, bots are exploited to generate and diffuse fake product reviews. Next, we study social media spam, focusing on the effects of social bots in the diffusion of spam campaigns on Twitter. We first discuss social spam data collection, then introduce a tool named *dynamical activity-connectivity map* we recently proposed to study the mechanisms of influence in social media. We conclude studying spam campaigns' sentiment and its interplay with bots' efficacy.

#### 3.2.1 Data Collection

Similarly to the political discussion scenario, we manually crafted a list of hashtags and keywords to collect our data. We focused on the tobacco-related discussion, and in particular electronic cigarettes. We identified this case study by noticing how spam seems to be a pervasive presence in this topic of discussion on Twitter [5]. The list included over one hundred terms covering nicotine-related products (e.g., *tobacco*, *cigar*, *cigarettes*, etc.), electronic cigarettes (multiple variants like *ecig*, *e-cig*, *ecigs*, *e-cigs*, *e-cigarette*, *ecigarette*, etc.), vaping products (e.g., *vape*, *ehookah*, *ejuices*, *eliquids*, etc.), popular vaping brands (e.g., *green smoke*, *eversmoke*,

etc.), health-related terms (e.g., *second-hand smoke*, *second-hand vape*), health campaigns terms (e.g., *still blowing smoke*, *not blowing smoke*, *tobacco free kids*, etc.), and more. We queried the Search API at regular intervals from January 1 to September 30, 2015 and collected a large dataset constituted by over 9 million unique tweets.

### 3.2.2 Spam Detection

Detecting social spam has proven a challenging and tedious task. The lack of a rigorous definition of what spam is makes detection a complex problem. Although various detection techniques have been proposed in the machine learning literature, they carry some limitations: they are either outdated, being trained and tested on early (2008–2010) Twitter spam data [12, 59, 60, 76], or overly-specific to detect certain types of campaigns [37, 38, 63, 94]. The first limitation becomes a problem due to the fact that bots evolve, becoming increasingly sophisticated thus rendering detection less effective if training data is not current; the latter issue hinders the applicability of detection systems to a broader range of problem domains.

For the reasons above, to detect spam campaigns in our data and separate legitimate tobacco-related discussion from social spam, we implemented a novel strategy. We first performed traditional data cleaning operations on the texts of the tweets in our dataset, namely removing stop-words and punctuation, then tokenizing and stemming the terms. Afterwards, we elaborated the following iterative three-stages detection procedure:

1. We generated a list of keywords appearing in the tweets, ranked by frequency.
2. Then, two independent human annotators manually identified and labeled keywords associated to spam campaigns appearing in the list of the top 250 most common keywords (to provide contextual information, the annotators had access to the full text of some example tweets where such keywords occur).
3. Finally, all tweets containing spam-associated keywords are moved into a separate repository that we will call *spam dataset*; the iterative process then restarts. It is worth noting that, at each next iteration of the algorithm, the ranked list of keywords changes because the spam keywords identified at stage 2 are removed.

The process ended when the list of top 250 most common keywords did not contain any spam-associated term. This yielded a manually-curated list of 87 spam keywords,<sup>8</sup> that appear in the *spam dataset* accounting for 3.06M unique tweets posted by over 850 thousand distinct users. Of these users, about 74K posted more than one tweet. We will focus our attention, for the rest of our analysis, on these 74K active spammers.

---

<sup>8</sup>The combination of the top 250 non-spam keywords, plus the 87 spam keywords, accounts for over 90% of all tweets in the original dataset.

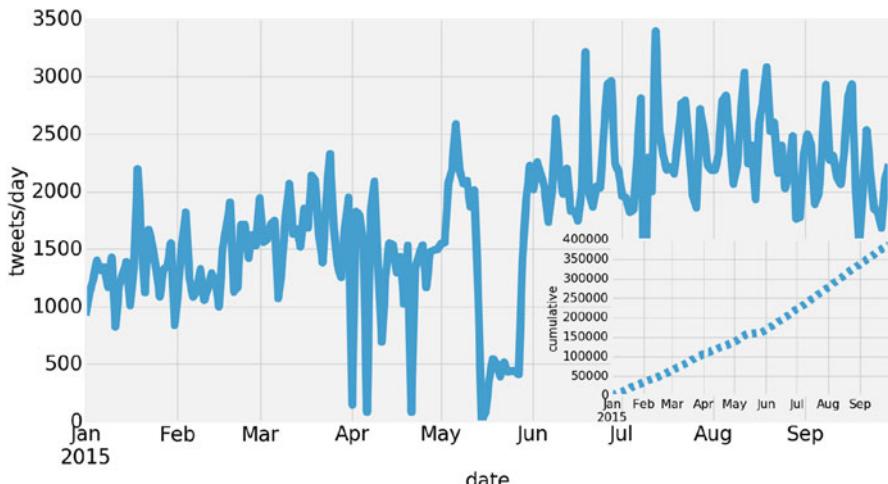
The top ten most recurring spam keywords, in order of frequency, are: *win*, *dvd*, *movies*, *giveaway*, *deals*, *horror*, *bluray*, *ebay*, *gameofthrones*, *movie*. Manual inspection of the 87 keywords suggests that three main types of social media spam campaigns occur in this scenario:

- Tobacco-related product promotions (sales, coupons, discount codes, etc.);
- Tobacco-unrelated product promotions (sales, coupons, discount codes, etc.), in particular related to entertainment products (dvd, music, books, etc.);
- Topic-hijacking campaigns, i.e., spam that includes tobacco-related keywords to attract the attention of users to tweets related to completely different topics, including movies and TV shows (keywords like *gameofthrones*, *fiftyshades*, *hungergames*, *celebs*, *ageofultron*, *insurgent*, and many others), and offline news events (e.g., *charlestonshooting*, *ericgarner*).

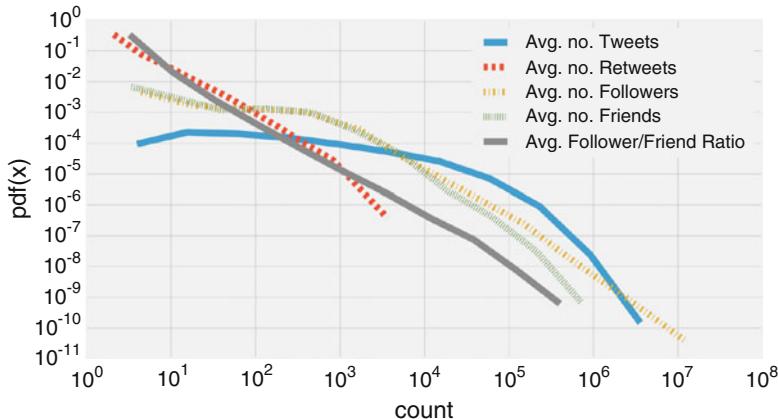
The phenomenon of Twitter hashtag hijacking has been documented extensively [19, 42, 47, 49]. In the following analysis, we do not make a specific distinction between different types of spam campaigns. However, in the future, we will try to determine whether campaign types, as well as different scopes and intents lead to different social spam dynamics.

### 3.2.3 Descriptive Data Statistics

Our initial exploratory analysis aims at highlighting the temporal dynamics of social spam production. Figure 4 shows the timeline of the volume of spam tweets per day in our dataset. Overall, we can note a mild upward trend over the course of



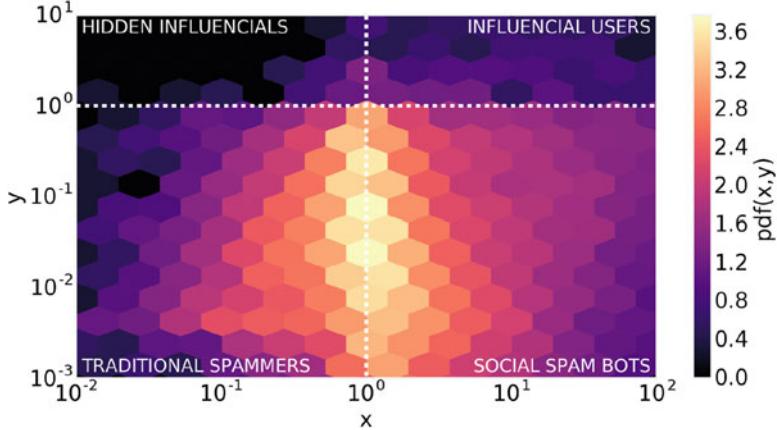
**Fig. 4** Timeline of the volume of spam tweets per day during the observation period. The inset shows the cumulative count. A few drops visible in April and May are associated with Twitter data collection service outages



**Fig. 5** Distributions of the average number of tweets, retweets, followers, friends, and follower vs friend ratio of the users in our spam dataset

the 9 months of observation. By the end of the year, the volume of tweets per day is roughly twice that of the beginning. This growth suggests the effectiveness of social spam in the tobacco-related context: if ineffective, the cost associated with running social spam campaigns would outweigh their benefits and therefore we would observe declining trends.

After assessing that social spam was “alive and well” during our analysis period, we moved forward to provide a statistical characterization of the actors therein involved: the Twitter spammers. Figure 5 shows the distribution of the average number of posted tweets, obtained retweets, number of followers and friends, and follower vs. friend ratio, for the set of users in our spam dataset. The averages are calculated across the 9-month observation period. A few observations are in order. Firstly, although all distributions exhibit the heavy tails typical of social networks [3, 8], some are significantly different from others. For example, the distribution of posted tweets is somewhat unexpected; if compared with the distribution of obtained retweets, which exhibits the typical power-law like behavior (i.e., a truncated straight line in the log-log plot of Fig. 5), the distribution of posted tweets appears anomalous. In particular, it appears that there is roughly the same probability of observing accounts with a number of posted tweets that spans from a few to over ten thousands: this is represented by the nearly-flat slope of the blue solid curve in the regime  $10 \leq x < 10^4$ . After that point, the probability decreases very rapidly. This unusual behavior is commonly linked to the activity of social bots. Their activity, however, does not catch up with the lack of influence they are typically characterized by, and therefore the amount of average retweets that most of these accounts receive is orders of magnitude lesser than the amount of tweets they post. Concluding, both the friends and follower distribution exhibit uncommon shapes, suggesting the presence of two different regimes, one for  $10 \leq x < 10^3$  and one for  $x \geq 10^3$ . The slope in the former regime is nearly flat, whereas in the



**Fig. 6** Dynamical activity-connectivity map of the users in our dataset. The  $x$  axis represents the proportional variation of followers/friends for each user over the accounted time period. The  $y$  axis represents the proportional variation of received/posted tweets of each user over the time period

latter both distributions decay with more typical heavy tails suggesting the presence of accounts with a very large number of friends and followers, another interesting behavior associated with two types of users: influential individuals, or social bots. Next, we study in detail the relation between activity and connectivity patterns.

### 3.2.4 Dynamical Activity-Connectivity Maps

The analysis above was static: taking the average values of the five features above made the results oblivious of the temporal dynamics of activity and connectivity as they unfold over the observation time. We now plan to investigate what effect the progression of activity levels of a user has on their connectivity evolution (and viceversa). In Fig. 6 we provide a *Dynamical Activity-Connectivity map*: we recently introduced this type of maps [31, 84] as dynamic variants of the map proposed by Gonzalez-Bailon and collaborators—see Figure 4 in the paper titled *Broadcasters and Hidden Influentials in Online Protest Diffusion* [41].

Figure 6 shows the probability density of users in the two-dimensional space where the  $x$ -axis represents the growth of network connectivity, and the  $y$ -axis conveys the messaging activity rate. For a given user  $u$ ,  $x_u$  and  $y_u$  are here defined as

$$x_u = \frac{1 + \delta f_u}{1 + \delta F_u} \quad \text{and} \quad y_u = \frac{1 + \delta r t_u}{1 + \delta t_u}.$$

We use the notations  $f_u$  and  $F_u$  to identify the number of followers and friends, respectively, of a user  $u$ . The variations of followers and friends of user  $u$  over a period of time  $t$  are thus defined as  $\delta f_u = \frac{f_{u\max} - f_{u\min}}{t}$  and  $\delta F_u = \frac{F_{u\max} - F_{u\min}}{t}$ ; the length of time  $t$  is defined as the number of days of  $u$ 's activity, measured from

registration to last observed activity (this varies from user to user). Finally, the variations of received retweets, and posted tweets, are defined as  $\delta rt_u = \frac{rt_u^{\max} - rt_u^{\min}}{t}$  and  $\delta t_u = \frac{t_u^{\max} - t_u^{\min}}{t}$ , respectively, where  $rt_u$  and  $t_u$  are the number of obtained retweets and posted tweets by user  $u$  during the period of activity  $t$ .

All values are added to the unit to avoid zero-divisions and to allow for logarithmic scaling (i.e., in those cases where the variation is zero). The “heat” (the color intensity) in the map represents the joint probability density  $pdf(x, y)$  for users with given values of  $x$  and  $y$ . The plot also introduce a bin normalization to account for the logarithmic binning.

The *Dynamical Activity-Connectivity map* we conceived is interpreted as follows: the bulk of the joint probability density mass should be observed in the neighborhood of  $(1, 1)$ , as the majority of accounts would usually exhibit a comparable variation along the two dimensions. That would be in line with what all previous social media studies where this type of map was employed reported [31, 41, 84]. However, the results Fig. 6 shows are unprecedented: we hypothesize that this is due to the spam dynamics characterizing this dataset. Let us discuss the two dimensions of *connectivity growth* and *activity rate* separately.

The *connectivity growth* is captured by the  $x$  axis and, in our case, ranges roughly between  $10^{-2}$  and  $10^2$ . Users for which  $x > 1$  (i.e.,  $10^0$ ) are those with a followership that grows much faster than the rate at which these users are following others. In other words, they are acquiring social network popularity (followers) at a fast-paced rate. Note that, if a user is acquiring many followers quickly, but s/he is also following many users at a similar rate, the value of  $x$  will be near 1. This is a good property of our measure because it is common strategy on social media platforms, especially among bots [11, 36], to indiscriminately follow others in order to seek for reciprocal followings. Our Dynamical Activity-Connectivity map will discriminate users with fast-growing followings, who will appear in the right-hand side of the map, from those who adopt that type of reciprocity-seeking strategy. The former group can be associated with highly popular users with a fast-paced followership growth. According to Gonzalez-Bailon and collaborators [41] this category is composed by two groups: *influential users* and *information broadcasters*, depending on their activity rates. Values of  $x < 1$  indicate users who follow others at a rate higher than that they are being followed; they fall in the left-hand side of the map. According to Gonzalez-Bailon and collaborators, these are mostly the *common users*, although the so-called *hidden influentials* also sit in this *low-connectivity* regime.

As for what concerns the  $y$  axis, it measures the *activity rate*, i.e., the rate at which a user receives retweets versus how frequently s/he tweets. Users with values of  $y > 1$  are those who receive systematically more retweets with respect to how frequently they tweet. This group of users can be referred to as *influentials*, i.e., those who are referred to significantly more frequently than others in the conversation; they fall in the upper region of the map, and according to Gonzalez-Bailon et al., depending on their connectivity growth can be divided in influential ( $x > 1$ ) and hidden influential ( $x < 1$ ) users. Conversely, users with values of  $y < 1$  are those who post exceedingly more tweets than the retweets they receive. This group

would generally represent the common-user behavior ( $x < 1$ ), although information broadcasters ( $x > 1$ ) also exhibit the same *low-activity* rate. These users fall in the lower region of the map.

Now that a reading of dynamical activity-connectivity maps has been provided, we can proceed with interpreting Fig. 6: the bottom-left quadrant reports the most common users, those with both activity and connectivity growth lesser than 1. In our case, we identify these accounts as traditional spammers. Manual validation of some of these accounts revealed that they employ simple automatic posting strategies, thus they generate a very large number of tweets, but they never attract other users' attention and thus they are rarely retweeted. We identified over 27K such accounts.

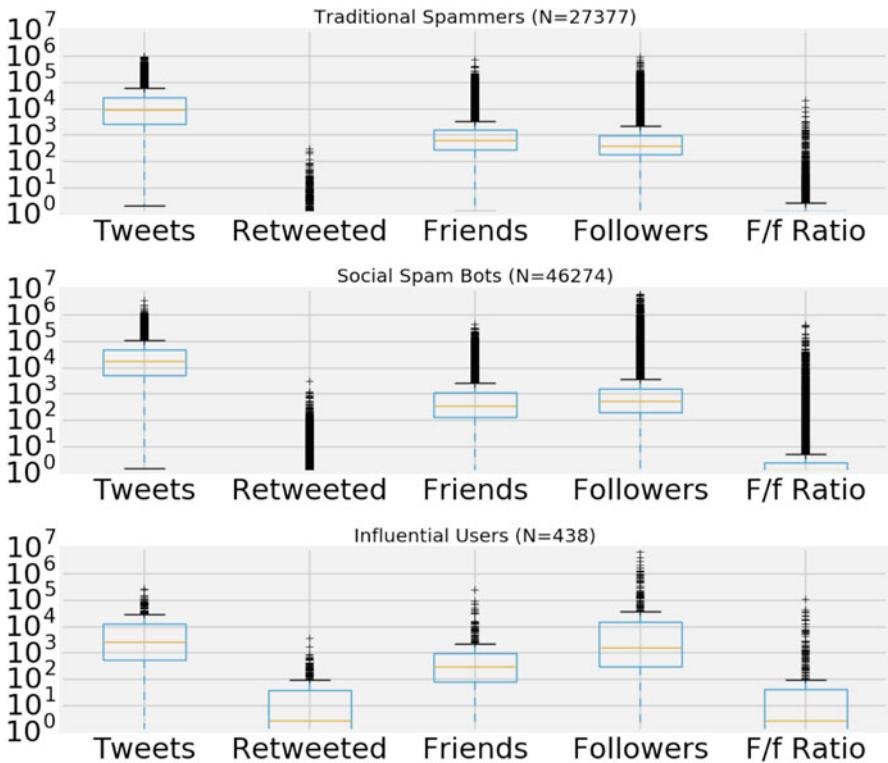
Conversely, the upper-right quadrant reports users with the higher connectivity growth and activity rates. These are influential accounts: they systematically attract other users' attention by receiving lots of retweets compared with how often they tweet, and their followings grow at a very fast pace. Influential users are quite rare in this context, and in fact we identified only 438 users according to our method. Manual inspection of all these users revealed that our technique correctly detects influential users which are not bots: accounts in this category include official accounts of movies and TV shows (e.g., *Avengers*, *CaptainAmerica*, *Divergent*, *GameOfThrones*, etc.), and various official accounts of tobacco-related sellers.

Lastly, social spam bots sit in the bottom-right quadrant. Differently from traditional spammers, their connectivity growth is much more similar to that of influential accounts. Their followings increase at a pace higher than their following others. They still produce disproportionately more tweets than the retweets they receive, but their embeddedness in the social network looks somewhat effective. Further analysis reveals that many of these spam bots tend to reciprocate followings to external users (accounts not present in the spam dataset) but also tend to follow each other; this coordinated behavior gives the appearance of network influence. We identified over 46K social spammers, the majority class by far in our spam dataset. Finally, we detected only 47 hidden influentials, too few to warrant further analysis.

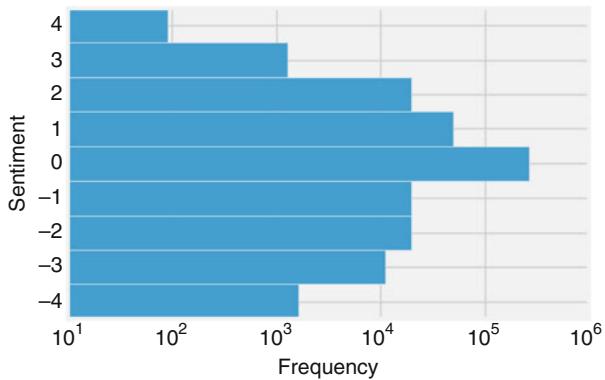
Figure 7 provides a different view on the five features characterizing the users in the three classes. As opposed to spammers, influential users receive significantly more attention (retweets), significantly more followers than friends (thus a much higher followers/friends ratio), and on average post one order of magnitude fewer tweets than bots. Concluding, the only significant difference between traditional spammers and social spam bots is their social network: social bots exhibit more followers than friends on average; the vice versa is true for traditional spam bots.

### 3.2.5 The Interplay Between Sentiment of Spam Bots

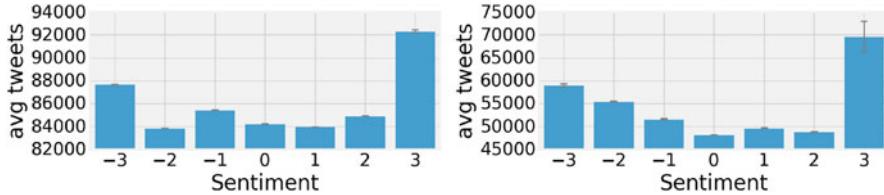
We conclude our analysis with a high-level investigation of the interplay between spam sentiment and spam bot characteristics. We applied the same Sentiment Analysis technique, i.e., *SentiStrength*, as in the previous case study, to our spam dataset. Figure 8 shows the distribution of sentiment scores for the tweets in our



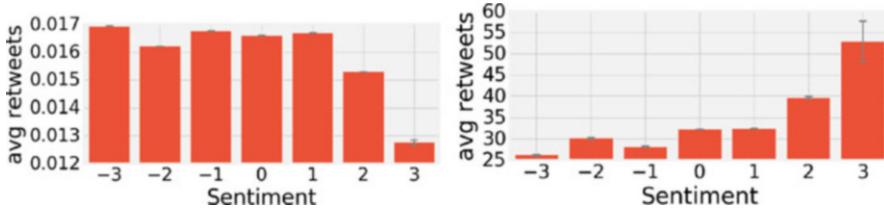
**Fig. 7** Box plot of the distributions of posted tweets, obtained retweets, number of friends and followers, and follower/friend ratio for the main three classes of users in our spam dataset



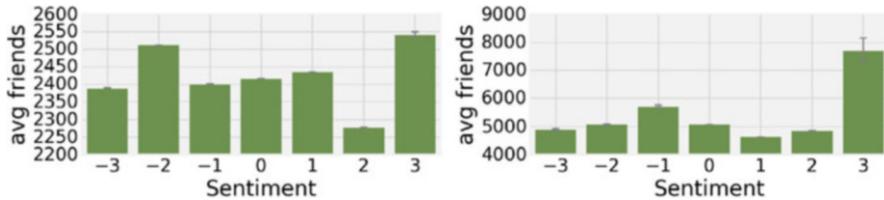
**Fig. 8** Distribution of tweet sentiment scores (SentiStrength) in the spam dataset



**Fig. 9** Average number of tweets posted as a function of tweet’s sentiment, calculated only on tweets retweeted at most once (left) and on those that have been retweeted more than once (right)



**Fig. 10** Average number of obtained retweets as a function of sentiment, calculated only on tweets retweeted at most once (left) and on those that have been retweeted more than once (right)



**Fig. 11** Average number of user friends as a function of sentiment, calculated only on tweets retweeted at most once (left) and on those that have been retweeted more than once (right)

corpus. The distribution exhibits its typical peak around zero [34, 79]. However, in contrast with respect to previous findings on Twitter sentiment obtained using SentiStrength [34], the distribution in the spam dataset appears skewed toward negativity. In particular, roughly one order of magnitude more strongly negative tweets ( $S \leq -3$ ) appear than strongly positive ones ( $S \geq 3$ ).

Worth noting, this dataset is significantly smaller and topically biased (i.e., it covers only spam) than the comprehensive Twitter dataset we previously studied [34]: we hypothesize that some correlation may exist between this atypical sentiment distribution and the role of spam bots.

To this purpose, in Figs. 9, 10 and 11 we plotted four features we used to characterize the bots (i.e., *number of posted tweets*, *obtained retweets*, *friends*, and *followers*). All figures report error bars (obtain hardly noticeable) that convey the standard error of the sampled average feature distributions. We will use them for diagnostic purpose, i.e., to highlight anomalies in spam dynamics with respect to

organic social media sentiment [34]. Given the exiguous number of tweets with extremely positive or negative sentiment (i.e.,  $S = 4$  or  $S = -4$ ), next we will limit our analysis to values of sentiment in the range  $-3 \leq S \leq 3$ .

The interpretations of the bar plots in Figs. 9, 10 and 11 is the following: given a fixed value of sentiment  $x$ , then  $y$  is the average value of the selected feature for all tweets with sentiment equal to  $x$ . Plots on the left are for the subset of tweets retweeted at most once; plots on the right are for tweets retweeted more than once. The separation is carried out to address the issue of activity heterogeneity highlighted before (cf. Fig. 5) and is necessary to avoid problems like the *Simpson Paradox* [87].

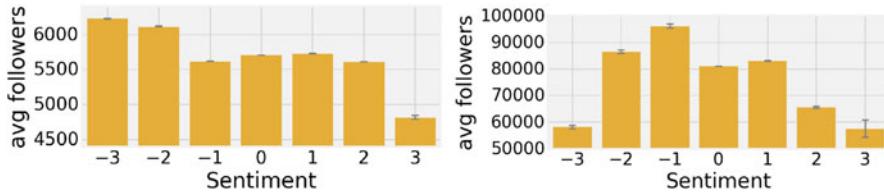
For the sake of example, let us discuss the left panel of Fig. 9 that shows the distribution of the *average number of tweets* posted by users, which were retweeted at most once, as a function of sentiment.

Let us consider sentiment  $S = 3$  (there are about 1300 such tweets in our dataset, cf. Fig. 8): the average number of tweets posted by the users who posted one such tweet with sentiment  $S = 3$  is about 92K. This is significantly higher than for every other sentiment score, denoting the fact that users who post strongly positive tweets (e.g., promotional tweets) on average posted significantly more tweets than the others. It is also worth noting that an average value of tweets nearing the hundred of thousands clearly denotes very highly-active accounts, and likely some form of automatic posting—a common feature of spam bots.

The right panel of Fig. 9 shows how this pattern is preserved even for the set of tweets that have been retweeted more than once: moreover, the distribution takes a U-like shape, suggesting that also accounts that post negative tweets exhibit much more activity than average. This suggests that some spam campaigns may not be necessarily positive. Indeed, if one compares this result with the previous case study on the manipulation of political campaigns, some interesting similarities emerge. In other words, spam at times can aim to smear some products, e.g., those from competitors.

Figure 10 shows another interesting patterns. The left panel again captures tweets that have been retweeted at most once; the right panel captures more popular tweets and exhibits a striking difference if compared to the left one: increasingly positive sentiment yields significantly more retweets. This is known as *positivity bias*, i.e., the emergence of a strong preference for retweeting positive messages; such bias was already observed in our prior Twitter analysis [34]. Strongly positive tweets obtain on average more than twice the number of retweets than negative or neutral ones. It is worth hypothesizing that, in the spam scenario, this pattern may also conceal some form of coordinated activity, i.e., bots may retweet other bots' spam in an orchestrated fashion.

Further clues supporting this hypothesis come from Fig. 11, in particular the right panel: users associated with positive tweets that are retweeted very often all exhibit a number of friends that are nearly twice as much as others. Inspecting users who follow on average over 7K accounts revealed strong reciprocity—another very common bot characteristic highlighted multiple times above.



**Fig. 12** Average number of user followers as a function of sentiment, calculated only on tweets retweeted at most once (left) and on those that have been retweeted more than once (right)

Looking at the complementary picture, i.e. the distribution of followers reported in Fig. 12, reinforces our hypothesis: left and right panels illustrate two very different scenarios, with the latter showing how users who post very positive or very negative tweets attracted significantly fewer followers than others: bots involved in spam campaigns do not commonly exhibit large followership (cf. Fig. 6).

Concluding, our diagnostics revealed characteristic patterns that may conceal clues to decode the strategies employed by spam bots to spread the content they produce, and try giving spam a legitimate appearance.

## 4 Conclusions

Social bots have become a pervasive presence in social media platforms. Applications of social bots have been documented in a variety of scenarios, including for public opinion manipulation and for social spam campaigns. The focus of this chapter was to investigate both these domains, and in particular to study the interplay between bots and information diffusion in the two scenarios.

In Sect. 2, we reviewed how social bots are created, and how they operate in social media platforms. We also briefly discussed the challenges of, and the methods to detecting them, covering techniques based on graph-centric detection, crowdsourcing, and traditional feature-based supervised learning.

Section 3.1 presented our first case study, discussing how social bots have been used during the 2016 US Presidential Election to sway the conversation around the presidential candidates. In this section we revised in detail the tools we used for social bot detection, namely *Bot Or Not*, for Sentiment Analysis, namely *SentiStrength*, and for partisanship detection.

We also summarized the results of our study on political manipulation [11], providing in particular two data-driven insights: first, we noted that social bots generate as much engagement, at least in terms of obtained retweets, than humans, suggesting the fact that humans cannot tell apart bots from other humans very easily when rebroadcasting politics-related information on Twitter. Second, we illustrated the interplay between content sentiment and social bots, highlighting a few partisanship differences (e.g., Trump bots single-handedly generated the most positive supporting content of their candidate in the entire analyzed dataset).

Finally, in Sect. 3.2 we proposed a second case study, and new results and analyses about the effects of social spam bots on the diffusion of social spam campaigns within the tobacco-related conversation on Twitter. First, we identified the presence of three types of spam campaigns: (1) relative to tobacco products; (2) relative to products unrelated to the tobacco industry, e.g., entertainment products; and, finally, (3) instances of topic hijacking, namely the use of hashtags and keywords related to the tobacco industry to attract individuals' attention on issues completely unrelated to that, e.g., social issues connected to news events in the offline world.

By means of a newly-introduced method named *Dynamical Activity-Connectivity map*, we also revealed the existence of different classes of spam accounts, including traditional spammers and social spam bots; we also discussed a statistical characterization of their most typical features. In conclusion, we provided an analysis of the interplay between sentiment and spam bots, revealing patterns that may conceal strategies of bot coordination, and the resulting effects in terms of spam diffusion.

Our findings in both case studies exemplify the potential for social media abuse: whether at stakes is the right to exercise unbiased elections and therefore democracy itself, or the exposure to illegitimate spam and propaganda, social media manipulation can have devastating societal effects. This study encourages future efforts of the research community to address the various facets of this form of abuse.

## References

1. Abokhodair N, Yoo D, McDonald DW (2015) Dissecting a social botnet: growth, content, and influence in twitter. In: Proceedings of the 18th ACM conference on computer-supported cooperative work and social computing. ACM, New York
2. Adamic LA, Glance N (2005) The political blogosphere and the 2004 us election: divided they blog. In: 3rd international workshop on link discovery. ACM, New York, pp 36–43
3. Ahn Y-Y, Han S, Kwak H, Moon S, Jeong H (2007) Analysis of topological characteristics of huge online social networking services. In: Proceedings of the 16th international conference on world wide web. ACM, New York, pp 835–844
4. Aiello LM, Deplano M, Schifanella R, Ruffo G (2012) People are strange when you're a stranger: impact and influence of bots on social networks
5. Allem J-P, Ferrara E (2016) The importance of debiasing social media data to better understand e-cigarette-related attitudes and behaviors. *J Med Internet Res* 18(8):e219
6. Alvisi L, Clement A, Epasto A, Lattanzi S, Panconesi A (2013) Sok: the evolution of sybil defense via social networks. In: 2013 IEEE symposium on security and privacy. IEEE, Piscataway, pp 382–396
7. Aral S, Walker D (2011) Creating social contagion through viral product design: a randomized trial of peer influence in networks. *Manag Sci* 57(9):1623–1639
8. Barabasi A-L (2005) The origin of bursts and heavy tails in human dynamics. *Nature* 435(7039):207–211
9. Barberá P, Wang N, Bonneau R, Jost JT, Nagler J, Tucker J, González-Bailón S (2015) The critical periphery in the growth of social protests. *PLoS One* 10(11):e0143611
10. Bekafigo MA, McBride A (2013) Who tweets about politics? Political participation of twitter users during the 2011 gubernatorial elections. *Soc Sci Comp Rev* 31(5)

11. Bessi A, Ferrara E (2016) Social bots distort the 2016 US presidential election online discussion. *First Monday* 21(11):1–14
12. Boshmaf Y, Muslukhov I, Beznosov K, Ripeanu M (2011) The socialbot network: when bots socialize for fame and money. In: Proceedings of the 27th annual computer security applications conference. ACM, New York, pp 93–102
13. Boshmaf Y, Muslukhov I, Beznosov K, Ripeanu M (2013) Design and analysis of a social botnet. *Comput Netw* 57(2):556–578
14. Boyd D, Crawford K (2012) Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Inf Commun Soc* 15(5):662–679
15. Carlisle JE, Patton RC (2013) Is social media changing how we understand political engagement? An analysis of facebook and the 2008 presidential election. *Polit Res Q* 66(4):883–895
16. Catanese SA, De Meo P, Ferrara E, Fiumara G, Provetti A (2011) Crawling facebook for social network analysis purposes. In: ACM WIMS '11: international conference on web intelligence, mining and semantics. ACM, New York, pp 52–59
17. Centola D (2011) An experimental study of homophily in the adoption of health behavior. *Science* 334(6060):1269–1272
18. Cha M, Haddadi H, Benevenuto F, Gummadi KP (2010) Measuring user influence in twitter: the million follower fallacy. In: Fourth international AAAI conference on weblogs and social media (ICWSM 2010). AAAI Press, Palo Alto, pp 10–17
19. Chu Z, Widjaja I, Wang H (2012) Detecting social spam campaigns on twiter. In: International conference on applied cryptography and network security. Springer, Berlin, Heidelberg, pp 455–472
20. Coburn Z, Marra G (2011) Realboy: believable twitter bots. <http://ca.olin.edu/2008/realboy/>
21. Conover M, Ratkiewicz J, Francisco MR, Gonçalves B, Menczer F, Flammini A (2011) Political polarization on twitter. *ICWSM* 13:89–96
22. Conover MD, Davis C, Ferrara E, McKelvey K, Menczer F, Flammini A (2013) The geospatial characteristics of a social movement communication network. *PLoS One* 8(3):e55957
23. Conover MD, Ferrara E, Menczer F, Flammini A (2013) The digital evolution of occupy wall street. *PLoS One* 8(5):e64679
24. Davis CA, Varol O, Ferrara E, Flammini A, Menczer F (2016) Botornot: a system to evaluate social bots. In: WWW '16 companion proceedings of the 25th international conference companion on world wide web. ACM, New York, pp 273–274
25. DiGrazia J, McKelvey K, Bollen J, Rojas F (2013) More tweets, more votes: social media as a quantitative indicator of political behavior. *PLoS One* 8(11):e79449
26. Effing R, Hillegersberg JV, Huibers T (2011) Social media and political participation: are facebook, twitter and youtube democratizing our political systems? In: International conference on electronic participation. Springer, Berlin, pp 25–35
27. El-Khalili S (2013) Social media as a government propaganda tool in post-revolutionary Egypt. *First Monday* 18(3)
28. Elovici Y, Fire M, Herzberg A, Shulman H (2013) Ethical considerations when employing fake identities in online social networks for research. *Sci Eng Ethics* 20:1–17
29. Elyashar A, Fire M, Kagan D, Elovici Y (2013) Homing socialbots: intrusion on a specific organization's employee using socialbots. In: Proceedings of the 2013 international conference on advances in social networks analysis and mining. ACM, New York, pp 1358–1365
30. Ferrara E (2015) Manipulation and abuse on social media. *ACM SIGWEB Newsletter* (4). ACM, New York
31. Ferrara E (2017) Contagion dynamics of extremist propaganda in social networks. *Inf Sci* 418:1–12
32. Ferrara E (2017) Disinformation and social bot operations in the run up to the 2017 French presidential election. *First Monday* 22(8)
33. Ferrara E, Yang Z (2015) Measuring emotional contagion in social media. *PLoS One* 10(11):e0142390
34. Ferrara E, Yang Z (2015) Quantifying the effect of sentiment on information diffusion in social media. *Peer J Comput Sci* 1:e26

35. Ferrara E, De Meo P, Fiumara G, Baumgartner R (2014) Web data extraction, applications and techniques: a survey. *Knowl-Based Syst* 70:301–323
36. Ferrara E, Varol O, Davis C, Menczer F, Flammini A (2016) The rise of social bots. *Commun. ACM* 59(7):96–104
37. Ferrara E, Varol O, Menczer F, Flammini A (2016) Detection of promoted social media campaigns. In: 10th international AAAI conference on web and social media, pp 563–566
38. Gao H, Hu J, Wilson C, Li Z, Chen Y, Zhao BY (2010) Detecting and characterizing social spam campaigns. In: Proceedings of the 10th ACM SIGCOMM conference on internet measurement. ACM, New York, pp 35–47
39. Gao H, Barbier G, Goolsby R (2011) Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intell Syst* 26(3):10–14
40. González-Bailón S, Borge-Holthoefer J, Rivero A, Moreno Y (2011) The dynamics of protest recruitment through an online network. *Sci Rep* 1:197
41. González-Bailón S, Borge-Holthoefer J, Moreno Y (2013) Broadcasters and hidden influentials in online protest diffusion. *Am Behav Sci* 57:943–965. <https://doi.org/10.1177/0002764213479371>
42. Hadgu AT, Garimella K, Weber I (2013) Political hashtag hijacking in the us. In: Proceedings of the 22nd international conference on world wide web. ACM, New York, pp 55–56
43. Heymann P, Koutrika G, Garcia-Molina H (2007) Fighting spam on social web sites: a survey of approaches and future challenges. *IEEE Internet Comput.* 11(6):36–45
44. Howard PN (2006) New media campaigns and the managed citizen. Cambridge University Press, Cambridge
45. Howard PN, Kollanyi B (2016) Bots, #strongerin, and #brexit: computational propaganda during the uk-eu referendum. Available at SSRN 2798311
46. Hwang T, Pearce I, Nanis M (2012) Socialbots: voices from the fronts. *Interactions* 19(2):38–45
47. Jackson SJ, Welles BF (2015) Hijacking# mynypd: social media dissent and networked counterpublics. *J Commun* 65(6):932–952
48. Jagatic TN, Johnson NA, Jakobsson M, Menczer F (2007) Social phishing. *Commun ACM* 50(10):94–100
49. Jain N, Agarwal P, Pruthi J (2015) Hashjacker-detection and analysis of hashtag hijacking on twitter. *Int J Comput Appl* 114(19):17–20
50. Jin X, Lin C, Luo J, Han J (2011) A data mining-based spam detection system for social media networks. *Proc VLDB Endowment* 4(12):1458–1461
51. Jindal N, Liu B (2007) Review spam detection. In: Proceedings of the 16th international conference on world wide web. ACM, New York, pp 1189–1190
52. Klotz RJ (2007) Internet campaigning for grassroots and astroturf support. *Soc Sci Comput Rev* 25(1):3–12
53. Kollanyi B, Howard PN, Woolley SC (2016) Bots and automation over twitter during the first us presidential debate. Technical report, COMPROP Data Memo
54. Kramer AD, Guillory JE, Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci* 111(24):8788–8790
55. Kümpel AS, Karnowski V, Keyling T (2015) News sharing in social media: a review of current research on news sharing users, content, and networks. *Social Media+ Society* 1(2):2056305115610141
56. Kwak H, Lee C, Park H, Moon S (2010) What is twitter, a social network or a news media? In: Proceedings of the 19th international conference on world wide web, pp 591–600
57. Latonero M, Shklovski I (2013) Emergency management, twitter, and social media evangelism. In: Using social and information technologies for disaster and crisis management. IGI Global, Hershey, pp 196–212
58. Lazer D, Pentland AS, Adamic L, Aral S, Barabasi AL, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M et al (2009) Life in the network: the coming age of computational social science. *Science* (New York, NY) 323(5915):721

59. Lee K, Caverlee J, Webb S (2010) The social honeypot project: protecting online communities from spammers. In: Proceedings of the 19th international conference on world wide web. ACM, New York, pp 1139–1140
60. Lee K, Caverlee J, Webb S (2010) Uncovering social spammers: social honeypots+ machine learning. In: Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval. ACM, New York, pp 435–442
61. Lutz C, Hoffmann CP, Meckel M (2014) Beyond just politics: a systematic literature review of online participation. First Monday 19(7)
62. Lyon TP, Maxwell JW (2004) Astroturf: Interest group lobbying and corporate strategy. *J Econ Manag Strateg* 13(4):561–597
63. Markines B, Cattuto C, Menczer F (2009) Social spam detection. In: Proceedings of the 5th international workshop on adversarial information retrieval on the web, pp 41–48
64. Mayzlin D, Dover Y, Chevalier J (2014) Promotional reviews: an empirical investigation of online review manipulation. *Am Econ Rev* 104(8):2421–2455
65. Messias J, Schmidt L, Oliveira R, Benevenuto F (2013) You followed my bot! transforming robots into influential users in twitter. First Monday 18(7)
66. Metaxas PT, Mustafaraj E (2012) Social media and the elections. *Science* 338(6106):472–473
67. Mønsted B, Sapiezyński P, Ferrara E, Lehmann S (2017) Evidence of complex contagion of information in social media: an experiment using twitter bots. *PLoS One* 12: e0184148
68. Morstatter F, Pfeffer J, Liu H, Carley KM (2013) Is the sample good enough? Comparing data from twitter's streaming API with twitter's firehose. In: 7th international AAAI conference on weblogs and social media
69. Mukherjee A, Liu B, Glance N (2012) Spotting fake reviewer groups in consumer reviews. In: Proceedings of the 21st international conference on world wide web, pp 191–200
70. Pang B, Lee L et al (2008) Opinion mining and sentiment analysis. *Found Trends Inf Retr* 2(1–2):1–135
71. Ratkiewicz J, Conover M, Meiss M, Gonçalves B, Flammini A, Menczer F (2011) Detecting and tracking political abuse in social media. *ICWSM* 11:297–304
72. Ratkiewicz J, Conover M, Meiss M, Gonçalves B, Patil S, Flammini A, Menczer F (2011) Truthy: mapping the spread of astroturf in microblog streams. In: Proceedings of the 20th international conference companion on world wide web. ACM, New York, pp 249–252
73. Shorey S, Howard PN (2016) Automation, algorithms, and political automation, big data and politics: a research review. *Int J Commun* 10:24
74. Song J, Lee S, Kim J (2011) Spam filtering in twitter using sender-receiver relationship. In: International workshop on recent advances in intrusion detection, pp 301–317
75. Stein T, Chen E, Mangla K (2011) Facebook immune system. In: Proceedings of the 4th workshop on social network systems, p 8. ACM, New York
76. Stringhini G, Kruegel C, Vigna G (2010) Detecting spammers on social networks. In: Proceedings of the 26th annual computer security applications conference, p 1–9. ACM, New York
77. Subrahmanian V, Azaria A, Durst S, Kagan V, Galstyan A, Lerman K, Zhu L, Ferrara E, Flammini A, Menczer F et al (2016) The DARPA Twitter bot challenge. *IEEE Comput* 49(6):38–46
78. Sutton JN, Palen L, Shklovski I (2008) Backchannels on the front lines: emergency uses of social media in the 2007 Southern California wildfires. University of Colorado, Boulder
79. Thelwall M (2013) Heart and soul: sentiment strength detection in the social web with sentistrength. In: Proceedings of the CyberEmotions, pp 1–14
80. Thelwall M, Buckley K, Paltoglou G, Cai D, Kappas A (2010) Sentiment strength detection in short informal text. *J Am Soc Inf Sci Technol* 61(12):2544–2558
81. Theocharis Y, Lowe W, van Deth JW, García-Albacete G (2015) Using twitter to mobilize protest action: online mobilization patterns and action repertoires in the occupy wall street, indignados, and aganaktismeno movements. *Inf Commun Soc* 18(2):202–220

82. Thomas K, Grier C, Song D, Paxson V (2011) Suspended accounts in retrospect: an analysis of twitter spam. In: Proceedings of the 2011 ACM SIGCOMM conference on internet measurement conference. ACM, New York, pp 243–258
83. Thomas K, McCoy D, Grier C, Kolcz A, Paxson V (2013) Trafficking fraudulent accounts: the role of the underground market in twitter spam and abuse. In: Usenix security, vol 13, pp 195–210
84. Varol O, Ferrara E, Ogan CL, Menczer F, Flammini A (2014) Evolution of online user behavior during a social upheaval. In: Proceedings 2014 ACM conference on web science, pp 81–90
85. Varol O, Ferrara E, Davis C, Menczer F, Flammini A (2017) Online human-bot interactions: detection, estimation, and characterization. In: International AAAI conference on web and social media
86. Varol O, Ferrara E, Menczer F, Flammini A (2017) Early detection of promoted campaigns on social media. EPJ Data Sci 6(1):13
87. Wagner CH (1982) Simpson's paradox in real life. Am Stat 36(1):46–48
88. Wang G, Mohanlal M, Wilson C, Wang X, Metzger M, Zheng H, Zhao BY (2013) Social turing tests: crowdsourcing sybil detection. In: NDSS. The Internet Society, Reston
89. Yang C, Harkreader R, Zhang J, Shin S, Gu G (2012) Analyzing spammers' social networks for fun and profit: a case study of cyber criminal ecosystem on twitter. In: Proceedings of the 21st international conference on world wide web. ACM, New York, pp 71–80
90. Yang X, Chen B-C, Maity M, Ferrara E (2016) Social politics: agenda setting and political communication on social media. In: International conference on social informatics. Springer, Berlin, pp 330–344
91. Yates D, Paquette S (2011) Emergency knowledge management and social media technologies: a case study of the 2010 haitian earthquake. Int J Inf Manag 31(1):6–13
92. Yin J, Lampert A, Cameron M, Robinson B, Power R (2012) Using social media to enhance emergency situation awareness. IEEE Intell Syst 27(6):52–59
93. Zangerle E, Specht G (2014) “Sorry, I was hacked” a classification of compromised twitter accounts. In: SAC: the 29th symposium on applied computing
94. Zhang X, Zhu S, Liang W (2012) Detecting spam and promoting campaigns in the twitter social network. In: IEEE 12th international conference on data mining (ICDM), 2012. IEEE, Piscataway, pp 1194–1199

# Network Happiness: How Online Social Interactions Relate to Our Well Being



Johan Bollen and Bruno Gonçalves

## 1 Introduction

In the normal course of our daily lives we naturally interact with many other individuals: the barista that prepares our daily *venti white chocolate mocha frappuccino*, the bus driver whom we ask for information, the supermarket teller that rings us out, our online acquaintances that we discuss scifi literature with, our coworkers, and our family and loved ones. However, it is clear that not all of these interactions carry the same weight or importance. We may not remember the name of the bus driver or the barista, but we would be remiss if we didn't remember the birthday of our significant other.

In an offline context it is relatively intuitive to observe and distinguish which relationships matter most to us. A small group of people with which we have close personal relations account for most of our social interactions while we dedicate less time or attention to more transactional interactions, such as those with service providers or strangers. Unfortunately for social scientists, it has proven difficult to quantitatively measure the strength and extent of real-world relationships at large scale without intrusive procedures and interventions.

However, our online activities now provide a unique opportunity to conduct such measurements as a by-product of the way in which such systems function. Every “*Like*,” “*retweet*,” or “*mention*” of billions of individuals is recorded and stored in large-scale databases that provide a unique perspective on how individuals interact socially, how they communicate with one another, and which aspects of their

---

J. Bollen (✉)

School of Informatics and Computing, Indiana University, Bloomington, IN, USA  
e-mail: [jbollen@indiana.edu](mailto:jbollen@indiana.edu)

B. Gonçalves (✉)

Center for Data Science, New York University, New York, NY, USA

social lives capture most of their attention [3]. Different systems naturally provide different features, different modes of interaction and, consequently, different views on human social behavior.

In this work we focus on Twitter which is a popular microblogging platform that as of June 30, 2016<sup>1</sup> was used world-wide by over 328 million users. Twitter was designed from the start to allow users to share their content to the world at large in the easiest way possible. As a result, all user-generated Twitter content is public by default and easily accessible through the use of an extensive API,<sup>2</sup> a fact that has long since made Twitter an invaluable resource for academic and industry researchers interested in the study of human behavior, information diffusion, and social network dynamics. Through the Twitter API one is able to easily access both the content user share and their social relations, a feature that makes it particularly suitable for the purposes of studying the relation between individual psychological states and social relationships.

## 2 Social Interactions

Modern online social network platforms provide a rich set of features and possibilities for users to interact socially. Twitter, in particular, allows users to unilaterally “Follow” another user, “Mention” another user, “Retweet” someone’s tweet, or “Like” one another user’s tweet. Each of these manners of interaction has a different meaning and, potentially, represents a different type of relationship.

Based on these types of interaction there are different possibilities to decide whether or not two users are in fact “friends.”

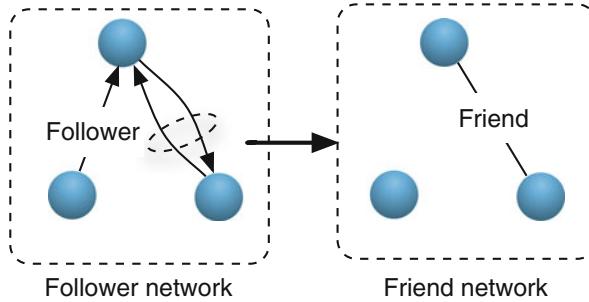
The simplest method is to define friends as users who regularly engage each other in conversation (via replies). This definition is based on the assumption that the active exchange of information between two parties indicates a social relation. This definition has been used [4] previously by us to empirically verify the well-known Dunbar’s number (a cognitive limit on the typical number of active social relationships). While likely corresponding to a “real,” offline, relationship, this definition does have the disadvantage of being rather strict; not all friendships involve the active exchange of information through Twitter replies.

Another method which we apply for the rest of this manuscript is to define friendship simply as two users who follow each other, as in Fig. 1. After all, friendship implies a reciprocated, symmetric relation. Celebrities can be followed by thousands and even hundreds of thousands of other users, and might on occasion even reply to messages, but they are not necessarily friends with their Followers, since the relation is not symmetric. While it is rather unlikely that all symmetrical Follow relations correspond to actual friendships, it does provide us with an

---

<sup>1</sup><https://about.twitter.com/company>.

<sup>2</sup><https://dev.twitter.com/rest/public>.



**Fig. 1** Friendship ties are by their very nature symmetric, but Twitter connects users by asymmetric Follower relations. This means that Twitter users may Follow other users, but the Follow relationship does not have to be reciprocated. Twitter's Follow relations are thus not sufficient to establish the existence of a Friendship tie between users. In our work, we adopt a minimal definition of a Friendship relation in the Twitter network as two users that share a *reciprocal Follow* relation. This approach does not require additional metadata such as content or frequency of information exchanges and it satisfies the minimum condition that a Friendship tie be symmetric. However, as a result, it does not account for the intensity or degree of the relationship—See [1]

operational definition of online friendship in which information (in the form of tweets) may in principle flow both ways so that each user may potentially influence the other.

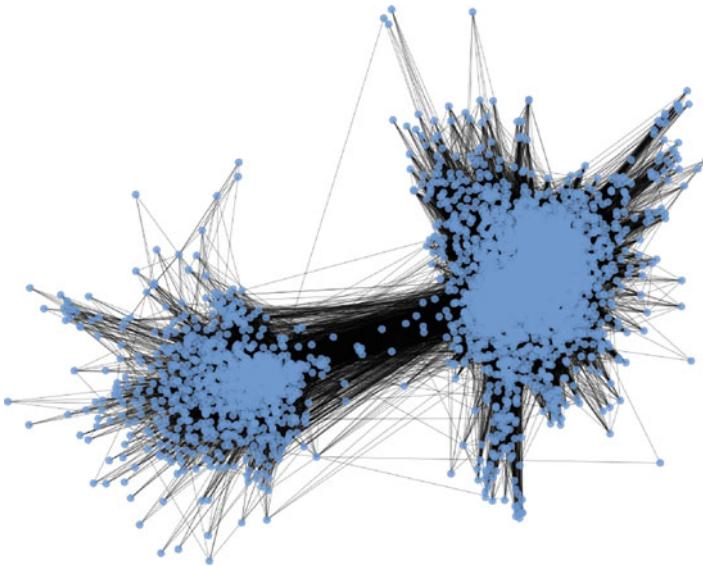
The first step of our analysis is to build an empirical friendship network from our Twitter follow data. For this purpose, we collected about 129 million tweets covering the period between November 28, 2008 to May 2009. The API provides us only with a 10% sample of all tweets produced. To avoid issues due to this sampling limitation, we expanded this dataset by retrieving the complete twitter history of all the users in our (sampled) dataset, as well as their follower network. The final mutual Twitter Follower network contains a total of 4,844,430 users (including followers of our users for which we did not collect timeline information).

From this dataset, we eliminate any user that has, on average, less than one tweet per day in the period of our study. In this way, we eliminate spurious users that are unlikely to have a significant impact on their neighbors. Finally, we remove all nonmutual connections to define the friendship network shown in Fig. 2. The giant connected component of our final network has over 102 thousand and a relatively large diameter. Further network statistics can be found in Table 1.

To each edge, we associate a weight,  $w_{ij}$ , that measures the social overlap between the two nodes. The overlap is defined as:

$$w_{i,j} = \frac{\|C_i \cap C_j\|}{\|C_i \cup C_j\|}, \quad (1)$$

where  $C_i$  is the set of friends of node  $i$ . Our goal in defining the strength of each connection in this way is twofold: first, this definition of social overlap is purely topological and insensitive to the number of actual interactions between the two



**Fig. 2** A force-directed visualization of a sample of the Friendship Network that resulted from our analysis of the Twitter Follow relations between more than 102,000 users. See [1]

users. Second, it gives us a parameter that we may threshold in order to control for shared social context as users with more mutual friends are more likely to be subjected to similar content.

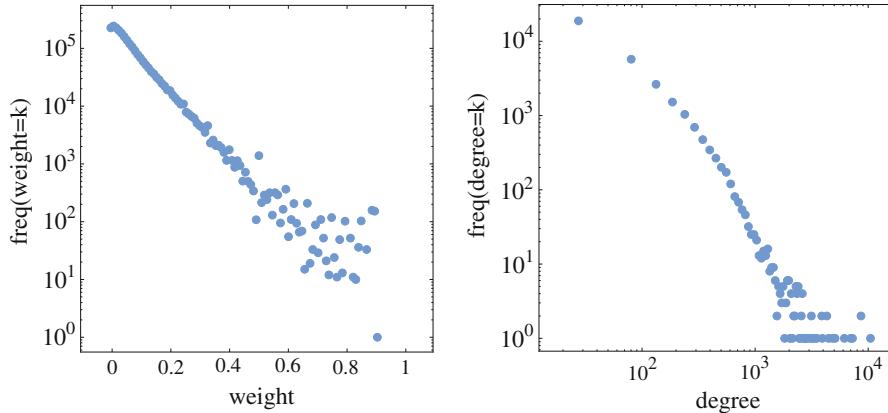
### 3 Network Structure

We now explore the structure of the friendship network we generated in the previous section. Some fundamental network statistics are listed in Table 1. Particularly significant are the relatively large average degree,  $\langle k \rangle = 46.3$  and clustering coefficient,  $\langle C \rangle = 0.262$ . The high clustering value is typical of real-world social networks [5, 6] with tightly knit groups of friends that are loosely connected through mutual acquaintances.

This type of structure is a result of the definition we used for friendship and helps to explain the relatively large diameter 14 that we observe. Above, we imposed that a link between two individuals is only created if they mutually follow each other. Inside dense friend groups, this happens naturally over the course of repeated interactions and also thanks to the fact that in many cases, these groups are to some degree topical [7]. On the other hand, our strict definition also makes it less likely for us to observe mutual follower relationships between individuals in distant groups, directly increasing the diameter of the network. The full degree distribution can be observed on the right-hand side of Fig. 3. The degree distribution we observe displays a clear broad tailed behavior. This provides further clues to the structure of

**Table 1** Network statistics for a Friendship network derived from the Twitter Follow relationships between 102,009 users

Nodes	102,009
Edges	2,361,547
Density	0.000454
Diameter	14
$\langle k \rangle$	46.300
$\langle C \rangle$	0.262



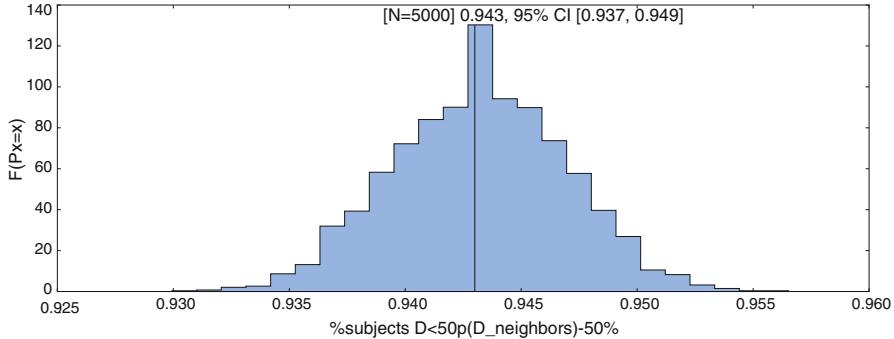
**Fig. 3** Distributions of edge weights and node degrees showed considerable skewness indicating that the large majority of connections have low connection weights whereas a few have very high connection weights, and that most users have very few Friendship relations in the network whereas a few individuals having order of magnitude higher number of friends. See [1]

the network as it shows that most nodes have relatively small degrees while a small number of them, the hubs, have collected several thousands of edges and help bind the network together as a whole.

However, not all links are created equal. We assign to each edge a weight corresponding to the number of mutual friends of the two individual at each end of the connection. Naturally, we expect that edges within groups will have higher weights while external edges should correspond to significantly smaller values resulting in a broad tailed weight distribution as shown on the left-hand side of Fig. 3.

## 4 Friendship Paradox

Hubs, by their very nature, play an important role in maintaining the connectivity of the network. The simple fact that they have such large degree implies that they *must* be connected to nodes in very different locations on the network. However, the picture is even more interesting if we take the opposite perspective, that of the ordinary node that is connected to a hub.



**Fig. 4** Histogram of the number of users with lower degree than the median of their friends over 5000 bootstrap realizations. See [2]

As most of us will remember from High School or college, there are advantages to being friends with the most popular kid in school. They know everyone and are better plugged in to the zeitgeist so they can act as brokers of information and introduce us to others we might be interested in meeting. As a result, they have a disproportionately large weight in our lives. This means that we will likely try to connect to them and others like them increasing both their global reach and impact in our lives. If we extend this way of thinking just a couple of steps further we reach a startling conclusion: everyone is trying to connect to these few hubs, resulting in a locally star-like graph. However, we have already observed strong assortativity effects. How can these two phenomena co-exist in the same system?

This observation is indeed paradoxical, but real none the less and is known as the Friendship Paradox: your friends are similar to you, but they also have more friends than you on average [8]. We investigate this paradox by measuring the fraction of users for whom the median degree of their friends is higher than their own. The result of this measurement is a single number of which we have no further information. We study its robustness through a simple bootstrapping procedure. Instead of measuring it over the full network, we evaluate it repeatedly over a small randomly selected fraction of the network. In Fig. 4 we plot the histogram of the fraction of nodes whose degree is smaller than the median of their friends' degrees taken over 5000 bootstrapping procedures. As we can see, our network displays a very strong Friendship Paradox. A large majority of users find themselves less popular than their friends on average.

## 5 Subjective Well-Being

After the data mining and preparation procedure outlined above, we have the complete Twitter history of all users in our network for a period of 6 months.

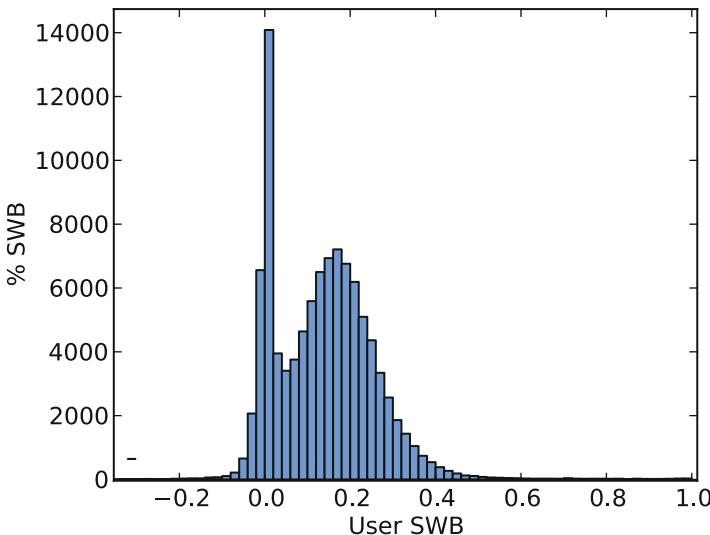
We define the Subjective Well-Being of an individual as the average valence (+ or -) of the content produced by him or her. To this end we apply the OpinionFinder (OF)<sup>3</sup> lexicon that assigns a positive (+1) or negative (-1) valence value to a set of 8630 words (2718 positive and 4912 negative words).

The subjective well-being  $S(u)$  of user  $u$  is then defined as the fractional difference between the number of tweets that contain positive OF terms and those that contain negative terms:

$$S(u) = \frac{N_+(u) - N_-(u)}{N_+(u) + N_-(u)}, \quad (2)$$

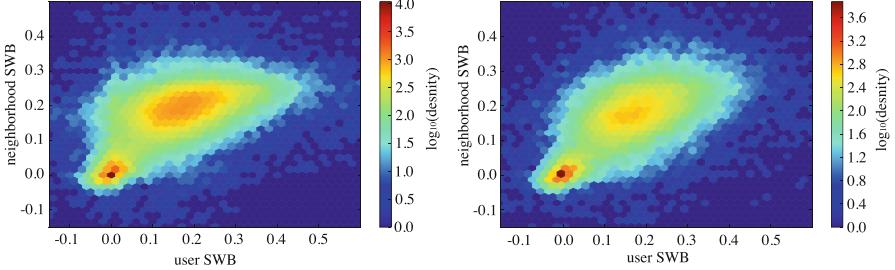
where  $N_-(u)$  and  $N_+(u)$  represent, respectively, the numbers of positive and negative tweets for user  $u$ .

After this procedure was applied, each node in our undirected, weighted network, has associated with it the average emotional polarity of the respective user, defined on a scale of  $[1, +1]$ . The empirical distribution of SWB is shown in Fig. 5. Despite the fact that the OF Corpus contains almost twice as many negative as positive words, we find a skew in the distribution towards positive SWB values with the positive values displaying an almost symmetrical distribution centered at  $\text{SWB} = 0.2$ . It is also worth to note that most negative values are close to zero, but how are these nodes connected to one another?



**Fig. 5** Distribution of Subjective Well-Being values over all users in our network reveals a strongly bi-modal distribution with two peaks: one slightly below zero and one around  $\text{SWB} = 0.2$ . See [1]

<sup>3</sup><http://www.cs.pitt.edu/mpqa/opinionfinderrelease/>.



**Fig. 6** 2D histogram of SWB values for users ( $x$ ) and their neighborhood ( $y$ ). Left: all edges included. SWB assortativity = 0.689,  $N = 102,009$  nodes. Right: histogram including only edges with  $w_{ij} \geq 0.1$ , SBW assortativity = 0.746,  $N = 59,952$ . See [1]

We start to answer this question by measuring the correlations between SWB of neighboring users. On the left side of Fig. 6 we plot the 2D histogram of the SWB values for users ( $x$ ) and the average value taken over their neighborhood ( $y$ ). Surprisingly, we find that this distribution is bi-modal with two clear clusters: one centered around zero and a larger one centered around  $\text{SWB} \approx 0.2$ . Most points in the figure are located close to the diagonal, indicating a large degree of SWB assortativity. Indeed, we measure the assortativity as the correlation between the two SWB values and find it to be a surprisingly large, namely  $R = 0.689$ .

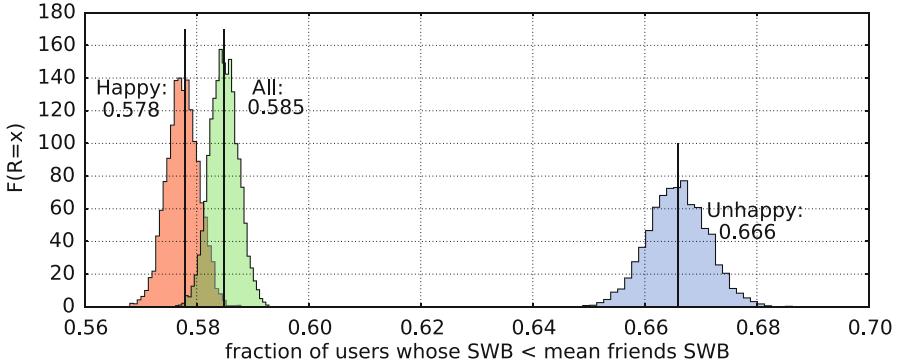
However, as we saw in Figs. 3 and 5, a large number of edges have low weight and a large number of nodes have lower values of SWB. Perhaps this is further blurring the results and might explain why we find a SWB cluster near 0? To clarify this possibility we repeat this analysis while keeping only edges with weights  $w_{ij} \geq 0.1$  (see [9] for further details). In this way, we are able to keep only the strongest (and thus likely intra-group) edges and help reduce the amount of noise in our results.

The resulting plot is shown on the right-hand side of Fig. 6. The outcome is quite striking. Not only is the second cluster near  $\text{SWB} = 0$  still present, but the assortativity has increased significantly to a whopping 0.746 providing strong evidence that these results correspond to significant features of our social system.

## 6 Happiness Paradox

We finalize our analysis by further considering the correlations between SWB values in the two clusters we found. For this, we divide our users into two groups: a "Happy" group and an "Unhappy" group. The former has high SWB values and is surrounded by friends with equally high SWB value. The latter has low SWB values and so do their friends. This way we compare SWB values only within clusters of comparable individuals.

We use a Gaussian Mixture Model (GMM) to demarcate our Happy and Unhappy groups. The location and distribution of each Gaussian component in the distribution



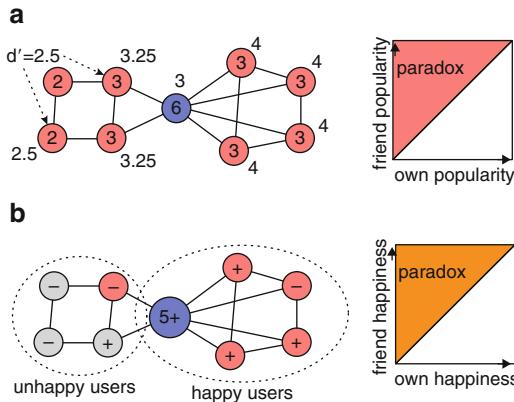
**Fig. 7** Distribution of bootstrapped estimates of the magnitude of the Happiness Paradox in our network for the Happy (red) and Unhappy (blue) group, and All (gray). See [2]

of individual vs. mean friend happiness is used to demarcate both groups by simply determining whether the SWB value of a subject and the mean SWB values of their neighbors fall within 2 standard deviations from the center of either one of the components (illustrated by the ellipses in Fig. 7).

Similarly to our Friendship Paradox analysis we also measure how the SWB value of a user is related to that of their friends through a bootstrapping procedure. In Fig. 7 we plot the histograms, taken of 5000 realizations of the bootstrapping procedure of the fraction of users whose SWB is less than the mean of their friends for the full dataset, the “Happy” and the “Unhappy” groups. As we can see, a similar behavior as the one observed for the Friendship Paradox is observed in all three cases: your friends are happier on average than you. We call this the Happiness Paradox.

One particularly interesting feature of these results is the fact that the Unhappy group, despite being the smallest, is the one for which the Happiness Paradox is strongest. This result together with Fig. 6 brings to bear the true strength of this phenomenon. Despite the fact that, in the Unhappy group, you are most likely connected with other Unhappy ( $\text{SWB} < 0$ ) users, they are **still** happier than you. In Fig. 8 we schematically represent the relation between the Friendship and the Happiness Paradoxes.

Finally, we further investigate how the Friendship Paradox manifests itself in the Happy and Unhappy groups as shown in Fig. 9. Subplot A illustrates the boundaries of each group as identified by our GMM approach, while subplots B and C illustrate the friendship paradox for each group. For clarity, we plot the log of the degrees. From this figure, it is clear that both the Friendship and Happiness paradoxes are present and statistically robust in our data set opening up new possibilities of research on the dynamical mechanisms that might help us understand these phenomena.

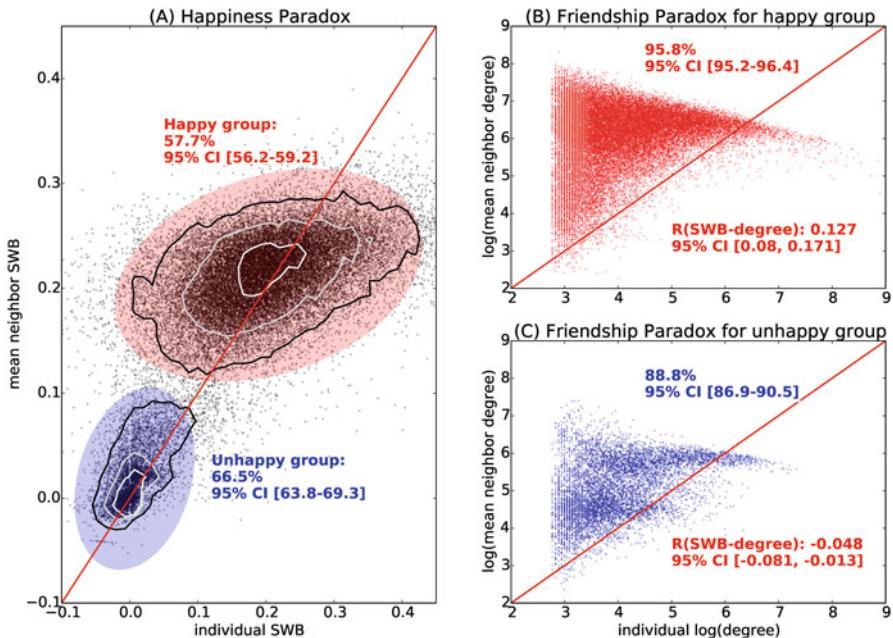


**Fig. 8** Diagram that visualizes how in many networks some user are very popular (left) and thereby inflate the average popularity of the Friendship networks they are part of, leading to a Friendship paradox where a majority of users are less popular than their friends on average (red nodes). If Popular individuals are also Happier, their presence in the network will also inflate average Happiness and lead to a Happiness Paradox (right). See [2]. (a) Friendship paradox (b) Happiness paradox

## 7 Discussion

The advent of social media has created a unique opportunity to study long-standing questions about how humans form social relations and how these relations affect their well-being, individually as well as collectively. The availability of longitudinal records of what users publish on social media allows us to assess their fluctuating mood state and overall well-being. The records of whom users talk to, about, and whom they follow, provide various perspectives on the multiplex of their social relations. In our work, we adopted an approach that combines a variety of social media data to measure otherwise difficult to quantify social constructs such as “happiness,” “well-being,” and “friendship” and establish meaningful correlations between how individuals and communities relate to each other and how it may affect their well-being over time. Social media has been in existence for almost a decade establishing records that allow us to study socio-economic phenomena as they emerge and develop over time. This allows the study of longitudinal behavioral and psychological indicators pertaining to individuals and communities over long periods of time.

We caution that many of our results may confound natural phenomena with interface and sample bias. Social media platforms are run by private enterprises that are not bound by requirements to further social science research. Researchers in this field therefore need to carefully consider the potential of self-selection, sample, interface, and social conformity bias in their work. In addition, our work pertains



**Fig. 9** This graph visualizes the magnitude of the Happiness and Friendship paradox in our sample of Twitter user. **(a)** (left): a majority of users are situated above the diagonal line in which a user's own Subjective Well-Being (SWB) is equal to the mean SWB of the user's friends. In other words, we find a significant Happiness paradox for both Happy and Unhappy groups of users. **(b and c)** (right): users in both Happy (red) and Unhappy (blue) groups experience a significant Friendship paradox, i.e. the users find themselves above the diagonal at which their own  $\log(\text{degree})$  as an indication of Popularity is equal to  $\log(\text{mean degree})$  of their Friends. See [2]

to snapshots, i.e. data that was harvested in a post-hoc manner and that pertains to specific periods of time in which we were provided access to the data. This situation does not enable controlled experiments and frequently precludes the measurement of “ground truth” with respect to social constructs that are operationalized post-ex-facto in terms of the available data.

In future research, we seek to address these shortcomings. The proliferation of social media platforms may allow the assessment of interface and sample bias as well as the correction of “opportunistic” data harvesting. In addition, we are seeing an increasing trend computational social science of using more traditional social science methods to validate computational indicators derived from social media data to establish “ground truth” and the pre-registration of trials and hypotheses. Our results can be seen as first steps towards an effort to establish a more robust understanding of socio-economic phenomena through the window of large-scale online social networking data.

## References

1. Bollen J, Gonçalves B, Ruan G, Mao H (2011) Happiness is assortative in online social networks. *Artif Life* 17(3):237–251. arxiv:1103.0784, [https://doi.org/10.1162/artl\\_a\\_00034](https://doi.org/10.1162/artl_a_00034)
2. Bollen J, Gonçalves B, van de Leemput I, Ruan G (2017) The happiness paradox: your friends are happier than you. *EPJ Data Sci* 6(4). <https://doi.org/10.1140/epjds/s13688-017-0100-1>
3. Gonçalves B, Perra N (eds) (2015) Social phenomena: from data analysis to models. Springer, Berlin
4. Gonçalves B, Perra N, Vespignani A (2011) Modeling users' activity on twitter: validation of dunbar's number. *PLoS One* 6:e22656
5. Watts DJ, Strogatz S (1998) Collective dynamics of 'small-world' networks. *Nature* 393: 440–442
6. Newman MEJ, Park J (2003) Why social networks are different from other types of networks. *Phys Rev E* 68:036122
7. Aiello LM, Barrat A, Schifanella R, Cattuto C, Markines B, Menczer F (2012) Friendship prediction and homophily in social media. In: ACM transactions on the web (TWEB), vol 6
8. Feld SL (1991) Why your friends have more friends than you do. *Am J Sociol* 96:1464–1477
9. Bollen J, Gonçalves B, Ruan G, Mao H (2011) Happiness is assortative in online social networks. *Artif Life* 17:237–251

# Information Spreading During Emergencies and Anomalous Events



James P. Bagrow

## 1 Introduction

Social networks are characterized both by their topological properties and by the dynamics they facilitate. The social spread of information is one of the most important of these dynamics [27]. Information spreading in the real world has been well studied. For example, Granovetter studied how individuals use their social ties to learn about new job opportunities [7]. Modern datasets such as social media and mobile phones have provided large-scale followups and confirmation to this seminal work [20].

However, most research on social spreading has been limited to understanding the ordinary, day-to-day dynamics. Anomalous and extreme situations, such as information spreading in the wake of an emergency or disaster, have not received as much attention [1]. Yet with appropriate data these situations provide a context by which researchers can better understand social networks and spreading phenomena. When an event occurs, a large amount of activity is generated, and this activity is all focused on that one event, leading to a strong and cohesive signal. Moreover, latent portions of the social network are likely to be activated, providing researchers a new view of the underlying social system, and there is no clearer indicator of the importance of a social tie than someone in the middle of an emergency or its aftermath choosing to reach out and communicate with that tie.

In this chapter, we discuss how natural and technological emergency and disaster events can be used to better understand social systems, human dynamics, and the spread of information and misinformation. Evidence supports a long-term increase in the frequency and severity of such events [3, 4], with driving factors including

---

J. P. Bagrow (✉)

Mathematics & Statistics, Vermont Complex Systems Center, University of Vermont,  
Burlington, VT, USA  
e-mail: [james.bagrow@uvm.edu](mailto:james.bagrow@uvm.edu)

climate change and population growth. Every effort should be made to prevent human and technological disasters, but when they inevitably occur it is important to glean as much useful information from them as possible, not only for scientific understanding but also to improve our response to future events and save lives.

The rest of this chapter is organized as follows. In Sect. 2 we describe the history of the sociology of disasters, and summarize research on using social media and telecommunications data to better understand information spreading during emergencies and disasters. In Sect. 3 we provide a case study of activity on Twitter related to the Boston Marathon Bombings. In Sect. 4 we summarize research on measuring activities in the wake of emergencies using mobile phone data taken from a country in Western Europe. In Sect. 5 we introduce an algorithm to detect unusual call activity in this country-wide mobile phone data, use it to detect 340 anomalies during a 6 month period, and define statistics to characterize the properties of these emergencies and how emergency and non-emergency events (such as music festivals) differ. We conclude with a discussion in Sect. 6.

## 2 Background

The study of the social response to emergencies, crises, and disasters has a long and fruitful history within the field of sociology, much of it due to E.L. Quarentelli, Russell Dynes, and J. Eugene Haas, who founded the Disaster Research Center at Ohio State University and pioneered the field of disaster sociology [13, 21–25]. This work, strongly influenced by the aftermath of World War II and the then-current climate of the Cold War, focused on organized behavior and emergent activity both during disasters and in their aftermaths [35]. Other pioneering work included case studies of disasters, such as panics and stampedes at large rock concerts [10]. Emergencies are complicated events, however, and the way individuals react to them is challenging to study. Indeed, even basics question, like how much panic occurs or does not occur among individuals experiencing an emergency, is a contested area of research [5].

A primary focus of disaster sociology has been understanding and improving the organizational aspects of the response to a disaster. Communication problems between competing and overlapping government agencies have hampered first responders in many large-scale emergencies, including highly unpredictable situations like the 9/11 terrorist attacks and more predictable situations such as the landfall of Hurricane Katrina [15]. Researchers have studied how organizations such as first responders use communication technology, how their use of that technology has adapted to changes and modernizations [16], and how and why such organizations either under-perform or can improve in their ability to efficiently and effectively respond to emergencies and disasters.

Since the pioneering work on disaster sociology, the rise of mobile phones, smartphones, and online social media have reshaped human communications. Individuals can now remain in constant contact with social ties if they choose,

and can quickly broadcast to a group of online followers almost anywhere. And these broadcasts can quickly go “viral,” spreading very rapidly online. Social media such as Facebook and Twitter have played key roles in recent emergency situations [16, 29, 37].

Recent work has studied how Twitter posts spread in the event of emergencies [8, 37]. Twitter is a popular microblogging platform where users can post short messages called tweets to their online followers, as well as repost or forward other tweets by “retweeting” them. Some tweets will become heavily retweeted, leading to cascades. Some tweets are geotagged, containing the geographic coordinates of the tweet poster when the tweet was made. Twitter posts are public and available through APIs to researchers, providing a wealth of text and activity data. For example, Sakaki et al. studied Twitter activity in the wake of a major earthquake, showing that tweets can be used to detect an earthquake in real-time [28]. Other work has studied how information (and misinformation) spreads during and after events including the Deepwater Horizon oil spill [30], wildfires and floods [36], Hurricane Sandy [2], the 2010 Haiti Earthquake [18, 19], and the Boston Marathon Bombing [31, 34]. Twitter is also used by government organizations such as first responders and by NGOs such as aid providers and relief organizations. Researchers have studied how these organizations use Twitter to spread information and deal with rumors and misinformation [32, 33].

Another avenue for data on emergencies and disasters is mobile phone records, specifically voice calls and text messages.<sup>1</sup> Unlike social media, these are generally not intended for broadcasting content to a group of followers, but are instead a specific, often one-on-one, communication medium. This activity is also not mixed with news media usage in the way that most journalistic organizations now rely heavily on social media. This one-on-one nature allows mobile phone data to more accurately capture individual social behavior and communication intent.

Researchers have studied mobile phone and smartphone activity in the wake of emergency events [1, 6, 26]. Kapoor et al. used mobile phone records in Africa to show that phone communications can act as early warning signals for an earthquake, and proposed an algorithm that can accurately pinpoint the epicenter of the earthquake [12]. Bagrow et al. [1] and Gao et al. [6] used mobile phone records to study a number of emergency events occurring in a country in western Europe, including a bombing and a plane crash. These records capture the temporal and spatial localization of the event from the spike in call volume immediately following the event, as well as the social propagation of information (see also Sects. 4 and 5 and Figs. 5, 6, 7, 8, 9). While mobile phone data are less readily available for researchers than public social media activity, we propose that it is an invaluable source of information parallel to social media.

---

<sup>1</sup>Although smartphone texting apps such as WhatsApp, Facebook Messenger, WeChat, Signal, SnapChat, Line, Apple Messages, etc. are now blurring the line between mobile phone SMS texts and online social media.

### 3 Twitter During and After the Boston Marathon Bombing

As an example demonstrating how emergency events provide a window into human dynamics, we performed a small case study of Twitter social media activity following the Boston Marathon Bombing.

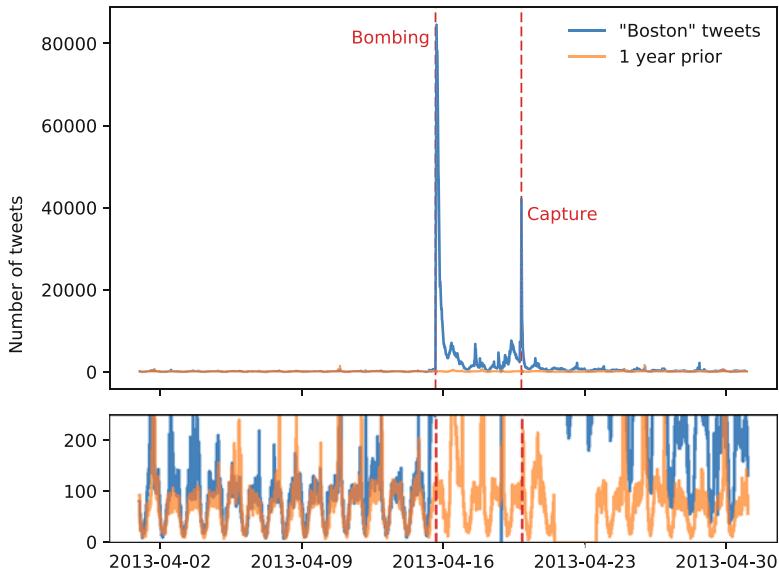
#### 3.1 *Background*

The Boston Marathon Bombing occurred on April 15, 2013 at 14:49 local time. Two improvised explosive devices exploded in a crowd near the Boston Marathon’s finish line, killing three and injuring 264 [14]. A manhunt soon unfolded, and on April 18 the FBI released photos of two suspects, Chechen-American brothers Dzhokhar Tsarnaev and Tamerlan Tsarnaev. That evening the brothers shot and killed a police officer, kidnapped a man and stole in his car, and had a shootout with police during which Tamerlan Tsarnaev was killed. The next day on April 19, Dzhokhar Tsarnaev was shot and arrested at 20:42. A police officer wounded in the April 18 shootout died the following year. Dzhokhar Tsarnaev was convicted of multiple crimes and sentenced to death in 2015 [17].

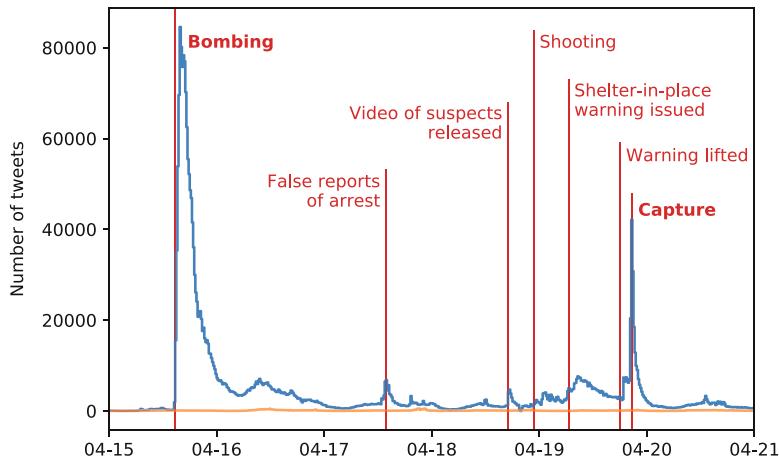
There was much related activity on Twitter in the immediate aftermath of the bombing and throughout the period of heightened uncertainty between the bombing and the capture of Dzhokhar Tsarnaev. In fact, Dzhokhar Tsarnaev himself tweeted multiple times between April 15 and April 19 [17]. Further, much misinformation and rumor propagated online, including online groups making false allegations against a missing college student [34].

#### 3.2 *Information and Rumors on Twitter*

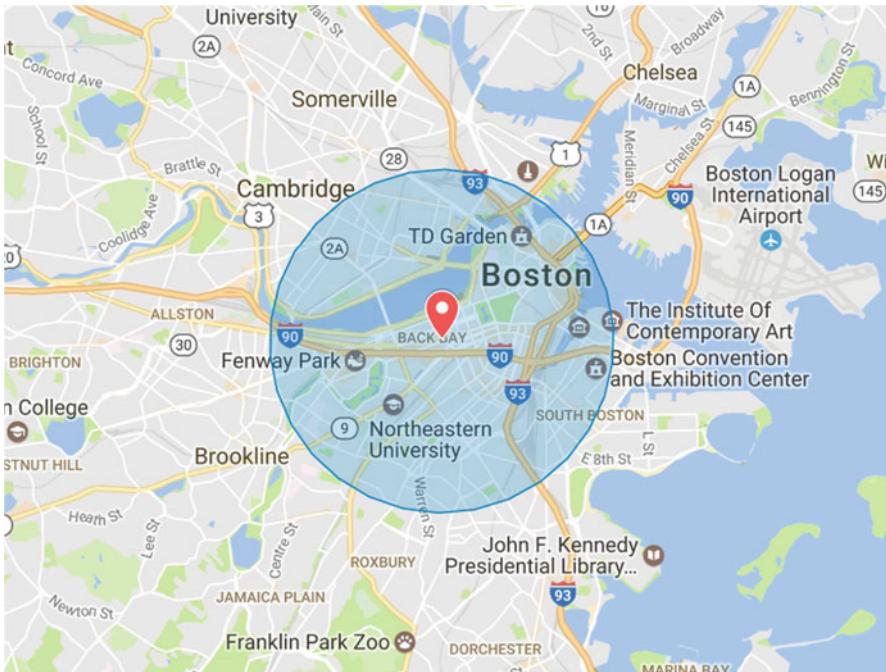
The events and rumors surrounding the Boston Marathon Bombing and how they unfolded online provide a useful case study for information and rumor spreading. Here we studied Twitter activity during and after the Boston Marathon Bombing using data captured from the “Gardenhose” feed, which captures a random 10% of all public Twitter activities. Figure 1 shows the volume of tweets (and retweets) over time containing “boston” (case-insensitive). Two strong spikes are present, coinciding with the bombing itself and Dzhokhar Tsarnaev’s capture. For comparison, we also determined the number of tweets containing “boston” 1 year prior, and superimposed the two time series. Before the bombing these time series line up very well, demonstrating much year-over-year regularity. When the bombing occurs, the volume of related tweets increases by a factor of approximately 800. The increased volume persisted for the rest of April. A closeup of the event period itself (Fig. 2) showed how well events surrounding the bombing are mirrored in the Twitter discussion.



**Fig. 1** Twitter activity surrounding the Boston Marathon Bombing. Shown are counts of tweets containing “boston” (case-insensitive) during April 2013, compared with 1 year earlier. The bombing on April 15 is clearly visible, as is the capture of Dzhokhar Tsarnaev on April 19. The year-over-year regularity of the time series before the bombing is evident in the lower plot, which shows the same time series but with a tighter range. Year-over-year deviations before the bombing are primarily due to sporting events



**Fig. 2** Closeup on the Boston Marathon Bombing time series shown in Fig. 1. The exogenous “spiking” pattern of both major events is clear, as are multiple other events occurring in the interim period between the bombing itself and the capture of the Dzhokhar Tsarnaev

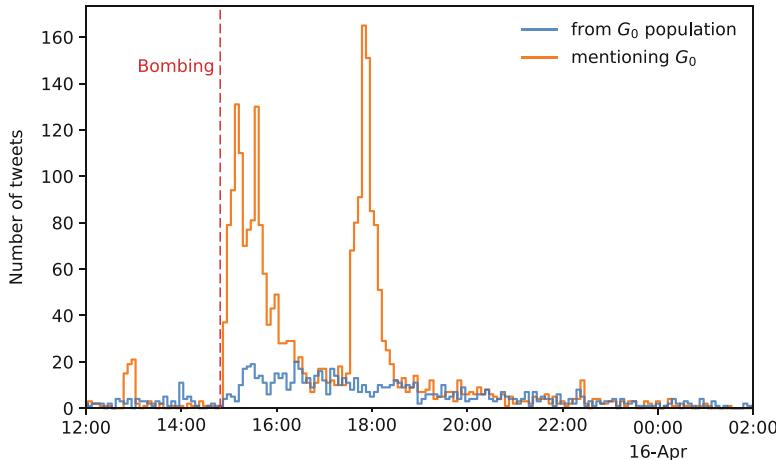


**Fig. 3** The blast site (marker) and selection area (circle) for the Boston Marathon Bombing. The circle forming the event region has radius 3 km. Users who post tweets from within the event region during the 5-h period following the blast comprise the  $G_0$  population

We next considered all geotagged tweets occurring within 3 km of the Boston Marathon Bombing blast site (Fig. 3) during the 5-h period immediately after the bombing occurred. The authors of these tweets form a population called  $G_0$ , those active “tweeters” in the vicinity of the event. We then re-scanned the Gardenhose feed, capturing all the tweets posted by individuals within  $G_0$  and all tweets which **mention** individuals within  $G_0$ . Mentions (or “at-mentions”) are a Twitter-specific term for posted tweets which contain the usernames of other Twitter users and are used to focus discussions and alert participants to online conversations; we used the mentioned usernames which Twitter extracted and provided as part of the Gardenhose feed. Time series of tweet activity for  $G_0$  individuals and mentions of  $G_0$  individuals are showed in Fig. 4.

Both time series show elevated activity levels in the aftermath of the bombing. In fact, the selection criterion for the  $G_0$  population forces the time series to display a higher activity level, as that time series is now conditioned on the fact that tweets were posted after the bombing [1]. Beyond this, we make two observations:

1. The spike in mentions of  $G_0$  individuals occurs more quickly than the spike in direct  $G_0$  activity. This implies that Twitter is not being used to get information out of the event area as much as it is being used in parallel with other media such



**Fig. 4** Volume of tweets posted by members of the Boston Marathon Bombing  $G_0$  population (Fig. 3) and tweets at-mentioning members of the  $G_0$  population. The second spike beginning at approximately 17:30 is primarily due to a highly retweeted tweet reported by a member of the  $G_0$  population about a possible suspect in custody. This event was later determined to be unrelated to the attacks

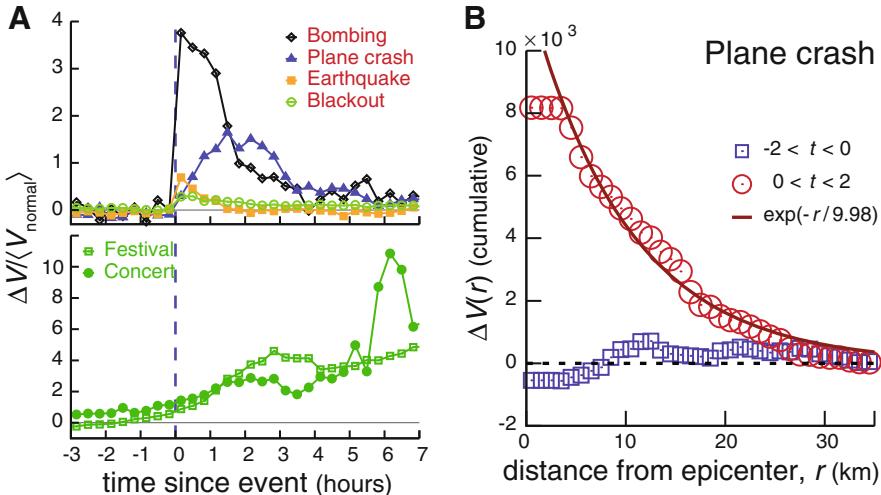
as news reports. Perhaps these tweets are people trying to reach social ties within the event region, although the nature of public Twitter activity makes it more likely that these are news reports and other media and government organizations.

2. A very strong second spike in mentions is apparent several hours after the bombing. Inspecting tweets posted at this time showed that this is due to one highly viral (heavily “retweeted”) tweet reporting the arrest of an individual as witnessed by a member of  $G_0$ . This second spike also peaked at a higher volume than the original spike, although it died out more quickly. This implies that the Twitter audience was primed to forward information during the immediate aftermath of the bombing, and the virality of any related content was much stronger. An emergency event primes the audience of social media for rumorizing and other information propagation.

Taken together, the tragic Boston Marathon Bombing provides an exemplar case study for analyzing the interplay between human dynamics, information and misinformation spread, and communication media and social media.

## 4 Mobile Phone Activity During Emergencies

Mobile phone datasets complement social media data for studying emergencies and disasters in many ways. Mobile phones are generally more established in various regions of the world, having a longer history of use and higher levels of adoption,



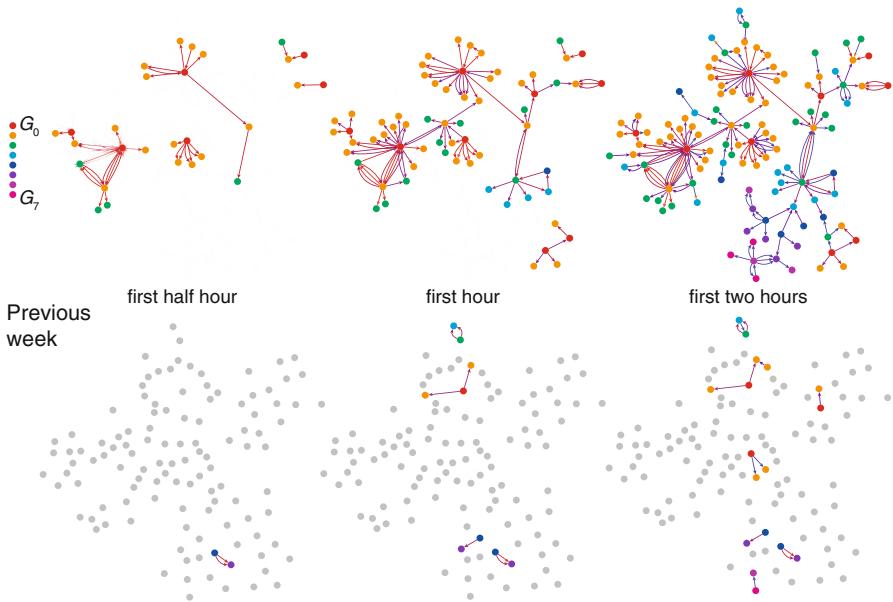
**Fig. 5** Temporal and spatial response during emergencies, as measured from the mobile phone records of a large service provider in a western European country. **(a)** The time dependence of call volume  $V(t)$  (voice and text) after four emergencies and two non-emergencies. We plot the relative change in call volume  $\Delta V/\langle V_{\text{normal}} \rangle$ , where  $\Delta V = V_{\text{event}} - \langle V_{\text{normal}} \rangle$ ,  $V_{\text{event}}$  is the call volume on the day of the event and  $\langle V_{\text{normal}} \rangle$  is the average call volume during the same period of the week. **(b)** The total change in call volume between two, 2-h periods before and after a plane crash, as a function of distance  $r$  from the epicenter of the crash. Following the event, we see an approximately exponential decay  $\Delta V \sim \exp(-r/r_c)$  characterized by decay rate  $r_c$  (figure adapted from Bagrow et al. [1]).

and providing years worth of extra historical records and large population samples. Mobile phone activity, especially voice calling, also lacks the broadcast nature of social media, acting instead as a direct communication channel. This direct communication means phone activity captures something very different than social media activity.

In an earlier work, we studied activity levels in the wake of multiple emergency events using mobile phone records from a phone provider in a western European country [1, 6]. These events included a bombing, a plane crash, and more. We found that the rapid spike in calls immediately following the emergency (Fig. 5a) was spatially localized (Fig. 5b), but rapidly propagated socially for the most serious events (Figs. 6, 7, and 8). This social propagation was measured from the time series of call activity for different populations of mobile phone users:  $G_0$ , the eyewitness group, calling from the direct vicinity of the event;  $G_1$ , those individuals who receive calls from members of the  $G_0$  group during the time period of the event;  $G_2$ , etc. As  $i$  increases, the group  $G_i$  becomes more social distant from the event itself.

The bombing in western Europe provides the most clear evidence for social information spreading based on the time series of call activity (Fig. 8a). Here we

### Bombing

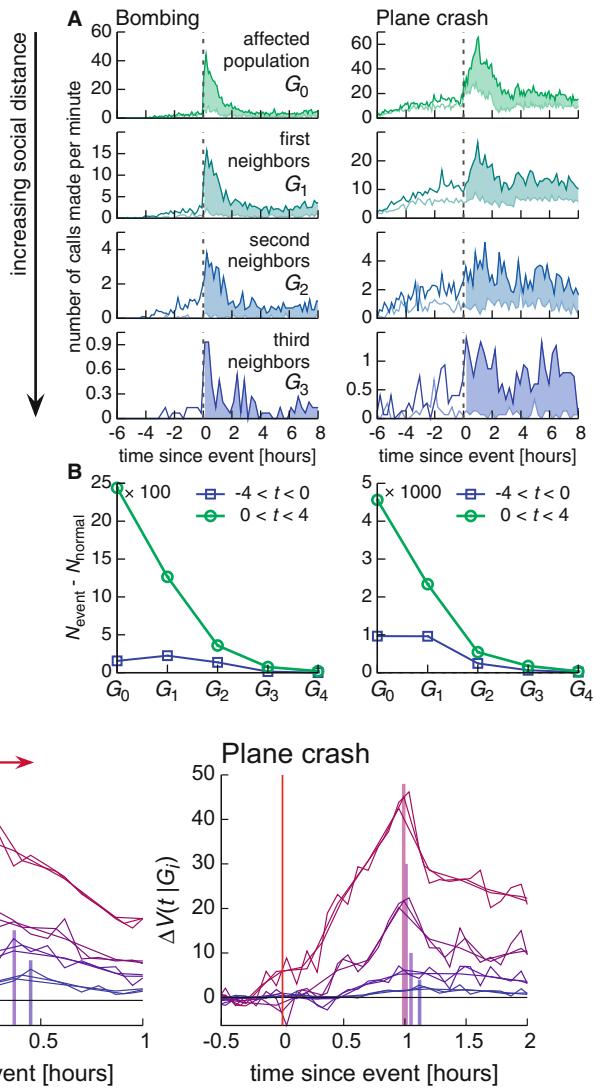


**Fig. 6** Part of the contact network formed between mobile phone users in the wake of the European bombing. Nodes are colored by group, with  $G_0$  representing phone users calling from the event region,  $G_1$  the recipients of those calls, etc. As time goes by more users are contacted as information propagates. Those same users make little contact during a corresponding time period the week before. These snapshots show the social spreading one can observe from mobile phone data (figure adapted from Bagrow et al. [1])

denote on the figure the times of the peaks of call volume for each group  $G_i$  using vertical bars. A temporal ordering is clearly evident for the bombing with the peak cascading through the populations over an approximately 20-min period (denoted by the horizontal arrow). The plane crash (Fig. 8b) does not show such clear temporal ordering of the peak. This may be due to the fact that news media were covering the crash and that social ties, particularly members of  $G_1$ , were likely to already be aware that their contacts were traveling that day. This underscores the different natures of equally unexpected emergency events.

This social spread of information outward from the wake of an emergency is intuitive, but we did not observe it in the Twitter data following the Boston Marathon Bombing. Indeed, in that case, unaffected individuals mentioning affected individuals spiked in activity before those who tweeted within the Boston Marathon Bombing event region (Fig. 4). This underscores the strong influence the communication channel has: for mobile phones, it is a direct communication channel and that limited scope requires a  $G_0$  individual to carefully choose who to contact, but for Twitter it is a secondary broadcast meant to update many followers. Those followers

**Fig. 7** Social spread of activity following emergencies, as measured from mobile phone records. The most serious events show strong propagation across the contact network. (a) Time series of call volume before and after the event, for each population  $G_0, \dots, G_3$ . The shaded regions denote the extra or anomalous call volume from that population compared with their activity the week prior. (b) The total difference in call volume compared to the prior week during time periods before and after the event, for each  $G_i$ . The two events and their  $G_i$  populations are those studied from Bagrow et al. [1]



**Fig. 8** Outward social information spread is most evident for the bombing. Different time series curves for each of  $G_0, \dots, G_3$  correspond to 5-, 10-, and 15-min time bins, intended to smooth the curves. Vertical marks denote the approximate peaks of each time series

are likely to be less socially close than contacts reached by mobile phone call, and it is probable (though definitely not certain) that a  $G_0$  individual will turn to phone calls first in the wake of an event, and then only later begin to use social media. And of course, mobile phones are not confounded by news organizations, government entities, and journalists the way social media are.

## 5 Detecting Anomalous Events

Given that emergencies and disaster events are useful for understanding and observing how information spreads in context, it is also worth understanding how rare these events are. To estimate the rate of emergencies and non-emergency events (collectively called *anomalies*) in modern datasets, as well as provide an example of the types of analysis now possible, here we introduce and apply an anomaly detection method to a country-wide mobile phone dataset, and use several basic descriptive measures on the identified anomalies to characterize their features. Such algorithms can in principle be used to detect the onset of an emergency event in real-time. This is a crucial application for first responders. However, here our focus is only on discovering anomalous events after they occur, so that they may be retroactively studied.

### 5.1 Detecting Anomalies

We implement a basic event detection algorithm and apply it to the six-month time series' of call volume taken from the mobile phone call detail records. This algorithm exploits the periodicity and recurrent nature of mobile phone activity patterns and performs well with noisy data. (A more advanced method, the Markov-modulated Poisson process [9], proved inadequate for this dataset.)

We first pre-processed the data. To help with heterogeneous tower densities, we began by dividing the country into equally-spaced squares of size  $1 \times 1$  km (one can also use  $10 \times 10$  km grids). All cell towers sharing a grid space were merged so that the total volume  $V_{\mathbf{x}}(t)$  of phone calls at grid space  $\mathbf{x}$  is the sum of the call volumes of all towers within  $\mathbf{x}$ . Grid spaces that do not contain cell towers were neglected. We now refer to each square grid space as a location. Since our goal is to find events that can yield good statistics, we ignored locations that are mostly unoccupied by only considering locations that average at least one phone call per minute over the entire six-month period. These time series are then binned into 10-min intervals so that their total length is  $6 \times 24 \times 7 \times W$  (covering  $W$  weeks).

The algorithm uses two calculations to flag *runs* of suspiciously high call volume for each time series, where a run is a time period denoted by a start time and a stop time. Runs that overlap in time (or nearly overlap) are merged.<sup>2</sup> After mergers, a run must have at least one time bin flagged as suspicious by both calculations and have a duration of at least five time bins to be considered an anomaly.

The two calculations to flag runs of suspiciously high activity use the *variance* (Sect. 5.1.1) and the *recurrence* (Sect. 5.1.2) of the  $V_{\mathbf{x}}(t)$ .

---

<sup>2</sup>Specifically, two adjacent runs of suspicious time periods are merged if they overlap in time or they are separated by less than four time bins and at least one of the two runs is longer than four time bins.

### 5.1.1 Variance Calculation

Each location's time series  $V_x(t)$  is copied  $W$  times, with each copy circularly rotated by one week from the previous copy. Now each 10-min bin  $t$  can be compared to all the other bins that occur at that same time of the week. Dropping the location index, let us denote  $V(t)$  as the original time series,  $\langle V_{\text{shifted}}(t) \rangle$  as the average of element  $t$  over the  $W$  rotated copies, and  $\sigma(V_{\text{shifted}}(t))$  as the standard deviation of the  $W$  rotated copies. Now we construct a new vector  $Z(t)$ ,

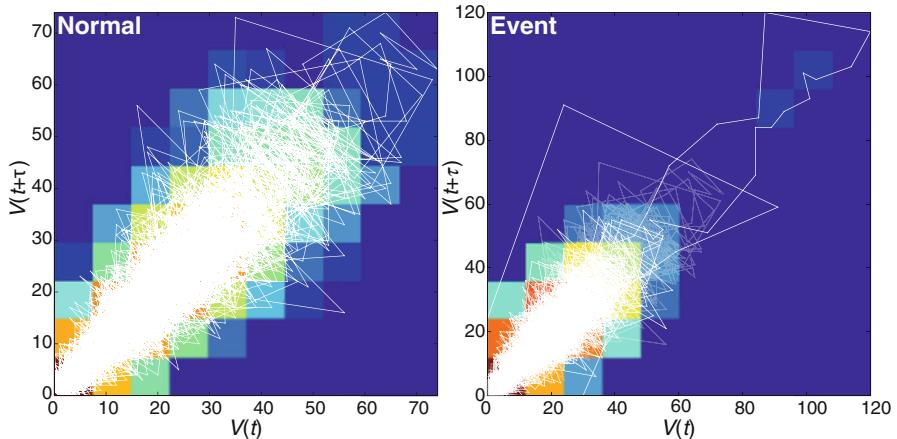
$$Z(t) = \frac{V(t) - \langle V_{\text{shifted}}(t) \rangle}{\sigma(V_{\text{shifted}}(t))}. \quad (1)$$

Finally, we flag as suspicious those contiguous times  $t_s \in [t_{\text{start}}, t_{\text{stop}}]$  where  $Z(t_s) > Z_{\text{thr}}$  for all  $t_s$ . In other words, a suspicious event's  $t_{\text{start}}$  and  $t_{\text{stop}} > t_{\text{start}}$  are determined by those times  $t$  where  $Z(t)$  crosses and remains above  $Z_{\text{thr}}$ . Events where only a single time bin was flagged ( $t_{\text{start}} = t_{\text{stop}}$ ) are ignored. For this work we use  $Z_{\text{thr}} = 2.5$ .

### 5.1.2 Recurrence Calculation

Take the original time series  $V(t)$  (suppressing location index) and rotate it by  $10\tau$  min ( $\tau$  elements). One can construct a recurrence or Poincaré plot by plotting the original time series  $V(t)$  against the rotated series  $V(t + \tau)$ .

If the time series is periodic, the plot will trace out a circular trajectory (Fig. 9). Deviations away from the normal pattern will appear as regions of the phase space



**Fig. 9** Recurrence plots of  $V(t)$  for  $\tau = 10$  min. Colored squares indicate the log of the probability for a randomly chosen point to fall in that bin. The run of points in an otherwise unoccupied region in the upper-right corner of the right plot indicates a persistent deviation from the expected recurrence and is flagged as suspicious

with relatively few points. To detect these regions we bin the phase space into squares of size  $20 \times 20$  min. The probability for a bin to contain a randomly chosen point is estimated as the fraction of points that fall within that bin. We flag points as suspicious if the probability to be in that bin is less than  $1/(24 \times 7 \times W)$ . For this work we use two rotations, one being 10 min ( $\tau = 1$ ), which primarily looks for sudden changes in activity, and the other being 1 week ( $\tau = 6 \times 24 \times 7$ ), which focuses on changes from the weekly periodicity. A point in time is suspicious if it is flagged in either recurrence plot.

### 5.1.3 Results

We applied the algorithm defined above to 6 months of mobile phone data records, and detected a total of 340 call anomalies. This corresponds to an average of 1.8 anomalies per day. Therefore, we conclude that researchers with access to years or decades of activity data may have records of hundreds or even thousands of small-, medium-, and large-scale anomalies to study. While many of the 340 events detected are not emergencies,<sup>3</sup> even non-emergency events provide a view into social activity and information spreading that is not available when one is limited to studying normal periods of activity.

## 5.2 Characterizing Detected Events

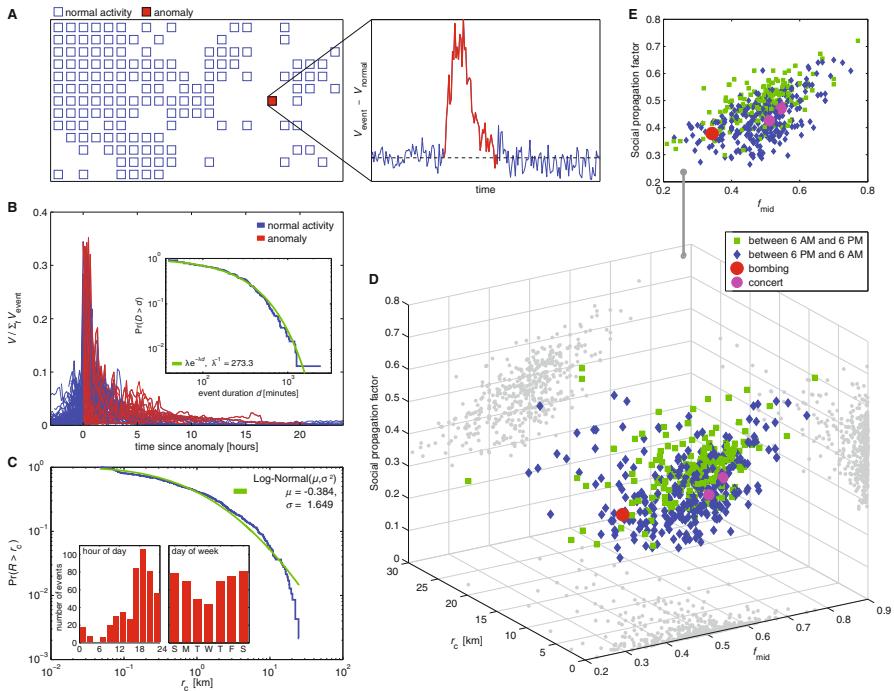
After identifying a call anomaly using the above procedure (Fig. 10a), we can characterize its temporal, spatial, and social properties:

**Temporal** The temporal nature of an event can be captured by how quickly it peaks. However, the time of the peak itself is often difficult to measure accurately from a noisy time series and may be influenced by any binning of the time series. Instead, we measure  $f_{\text{mid}}$ , the midpoint fraction, defined as the fraction of time it takes for half of the total anomalous call activity to occur. When there is a sharp spike in call volume, as shown in the red curve in Fig. 10a,  $f_{\text{mid}}$  will be low. Specifically,  $f_{\text{mid}} = (t_{\text{mid}} - t_{\text{start}})/(t_{\text{stop}} - t_{\text{start}})$ , where  $t_{\text{mid}}$  is defined such that

$$\begin{aligned} & \int_{t_{\text{start}}}^{t_{\text{mid}}} (V_{\text{event}}(t) - \langle V_{\text{normal}} \rangle(t)) dt \\ &= \frac{1}{2} \int_{t_{\text{start}}}^{t_{\text{stop}}} (V(t)_{\text{event}} - \langle V_{\text{normal}} \rangle(t)) dt. \end{aligned} \quad (2)$$

---

<sup>3</sup>We inspected the anomalies manually and determined the origins of many of the events using Google News, but cannot share this information as it will reveal the country of origin of the data, breaking our non-disclosure agreement with the mobile phone provider.



**Fig. 10** Systematic anomaly detection to estimate the rate of anomalies captured by mobile phone data records. **(a)** The full country is divided into  $1 \text{ km} \times 1 \text{ km}$  grids. Assigned to each grid space is a six-month time series corresponding to activity from mobile phone towers within that space. A composite detection algorithm, exploiting daily, weekly, and seasonal periodicities in call activity, is then used to flag anomalous call periods (highlighted). The final result is a corpus of 340 anomalies. The bombing and several known concerts occurred during this six-month period, and were successfully identified. **(b)** Time series for some anomalies, scaled so that the total activity during the anomaly is unity. **(Inset)** The distribution of anomaly durations is approximately exponential, with an average duration of 273 min. **(c)** The distribution of characteristic spatial distances  $r_c$ . The average  $\langle r_c \rangle = 2.33 \text{ km}$  corresponds well to the events studied by Bagrow et al. [1]. A log-normal distribution is shown for comparison. **(Inset)** The distribution of anomaly start times, as a function of time of day and day of week. Fifty one percent of anomalies occur between 6 PM and midnight. **(d and e)** For each anomaly, we plot: the midpoint fraction  $f_{\text{mid}}$ , the time it takes for half the anomalous call activity to occur;  $r_c$ ; and the social propagation factor, measuring how rapidly the anomaly propagates through the social network. We see that propagation rates are independent of  $r_c$  and that the bombing shows faster propagation than the concerts. Interestingly, events that occur during the day tend to show slower social propagation than events that begin during nighttime hours

**Spatial** How much an event's call anomaly is localized spatially around the detected epicenter can be captured by its characteristic spatial decay rate  $r_c$ . We measure this by integrating the anomalous call activity in concentric rings of radius  $r$  around the event epicenter, and fit an exponential function, i.e.,  $\Delta V(r) \sim \exp(-r/r_c)$ . The spatial decay rate tells us whether the event is sharply peaked at a location (small  $r_c$ ) or spreads broadly over space (large  $r_c$ ).

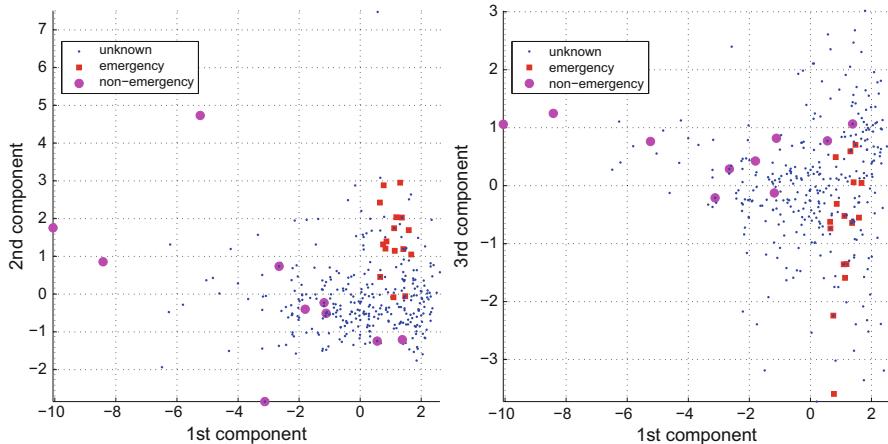
**Social** The social spread of a call anomaly can be measured by analyzing statistics of the time series of calls made by populations  $G_0, G_1, \dots$ . These populations capture those directly affected by the event ( $G_0$ ), those who receive calls from  $G_0$  but are not themselves members of  $G_0$  ( $G_1$ ), the recipients of call from  $G_1$  members not in  $G_1$  or  $G_0$  ( $G_2$ ), etc. To capture how quickly the call anomaly spreads through these populations, we define the *social propagation factor* as simply the midpoint fraction of the time series of anomalous call volume for each population  $G_i$ , averaged over  $G_i$ .

In Fig. 10b we present re-scaled time series of the call activity during the detected anomalies, compared with activity under non-anomaly circumstances. Most anomalies are short in duration, lasting under 2 h, although a few were detected lasting over 20 h. Likewise, most anomalies were spatially localized (Fig. 10c). Most anomalies occurred after 18:00 local time, although most phone activity also occurred after 18:00 so this observation may be a simple confound. Weekends were more likely to contain anomalies, with Wednesday being the weekday having the fewest detected anomalies.

The speed of the event, measured by how quickly the localized call anomaly peaks, correlates well with the social propagation factor measuring how quickly the call anomaly peaks within the social populations  $G_0, G_1, \dots$  (Fig. 10d, e). This relationship is roughly independent of spatial localization as measured by  $r_c$  (Fig. 10d). We also observe that social propagation is slower for daytime events (those occurring between 06:00 and 18:00 local time), regardless of  $f_{\text{mid}}$  itself (Fig. 10d, e). For an event to occur in the middle of the night and have a strong social propagation factor is good evidence that it is an emergency or disaster.

### 5.2.1 Principal Component Characterization

Lastly, we performed a principal component analysis (PCA) [11] on the 340 detected events (including events that were found by manual inspection to correspond to those studied by Bagrow et al. [1]). Nine measurements (or features) were determined for each event: the spatial size of the event  $r_c$ ; the speed at which the event occurs  $f_{\text{mid}}$ ; the time of day; event duration; total number of calls; affected population size  $|G_0|$ ; the “social decay rate,” the ratio of the total number of calls made by population  $G_i$  vs.  $G_{i-1}$  averaged over  $i$ ; the  $z$ -score for the total number of anomalous calls placed by population  $G_i$ , averaged over  $i$ ; and weighted social distance  $\sum_i i \times V_{\text{total}}(G_i) / \sum_i V_{\text{total}}(G_i)$ . These measures are intended to capture many different aspects of the call anomalies, and more can in principle be used, under the assumption that PCA will “net out” the most relevant linear combinations of these features. A  $340 \times 9$  data matrix is then constructed. The first three principal components are shown here (Fig. 11). We found a clustering of known emergencies, with known non-threatening events appearing mostly as outliers. The clustering of emergencies is evidence that the measures introduced here can be used to categorize events without additional information.



**Fig. 11** Principal component analysis for the anomalies detected in the mobile phone records. Anomalies were manually inspected and many were found to correspond to known emergency and non-emergency events. Emergency events showed distinct clustering, particularly in the first two principal components

## 6 Discussion

In this chapter, we discussed how to measure information or activity spreading through a social system in the wake of an emergency, disaster or other anomalous event. Such emergency events act as “found experiments,” providing researchers with new contexts and windows on the underlying social system. We presented a case study of information spread on Twitter following the Boston Marathon Bombing, and described measures of social spreading within a western European country captured from mobile phone records. Mobile phone data are limited in scope—lacking, for example, contextual details such as the text information available in social media—and are generally less freely available to researchers. But mobile phone datasets strongly complement other data such as those taken from Twitter because phone calls and text messages represent one-on-one, direct communication and are not confounded by broadcast effects and news media the way Twitter is.

Comparing the spreading dynamics on Twitter surrounding the Boston Marathon Bombing with the western European Bombing captured from mobile phone records underscores how different these communication media are, both in who uses these media and what is expected from these media. Researchers must account for these differences when studying and comparing across media. Even within a single type of media there may be great differences: a photo-oriented platform like SnapChat may present vastly different dynamics than a microblogging platform like Twitter or a chat platform like WhatsApp or Facebook Messenger. Further, spreading in a single

platform does not take place in isolation: the dynamics of Twitter users following the Boston Marathon Bombing are strongly influenced by information they (or their social ties) receive from traditional, broadcast media.

As communication services continue to evolve, and online activity continues to adapt to new services, researchers will be confronted with both technical challenges to overcome but also a wealth of new opportunities brought about by new data. Recent advances in machine learning and artificial intelligence may prove fruitful here, for example. Deep learning for computer vision may soon allow researchers to better understand and analyze video feeds and imagery created by eyewitnesses of emergency events, particularly when those feeds are generated in large volumes, keeping pace with new smartphone video streaming services such as Facebook Live and Periscope.

**Acknowledgements** We thank S. Lehmann and Y.-Y. Ahn for organizing this book and inviting us to contribute, C.M. Danforth for useful comments on the Twitter data, and we gratefully acknowledge the resources provided by the Vermont Advanced Computing Core. This material is based upon work supported by the National Science Foundation under Grant No. IIS-1447634.

## References

1. Bagrow JP, Wang D, Barabási AL (2011) Collective response of human populations to large-scale emergencies. PLoS One 6(3):e17680
2. Chatfield AT, Scholl HJ, Brajawidagda U (2014) # Sandy Tweets: citizens' co-production of time-critical information during an unfolding catastrophe. In: 2014 47th Hawaii international conference on system sciences (HICSS) IEEE, New York, pp 1947–1957
3. EM-DAT C (2010) The OFDA/CRED international disaster database. Université catholique
4. Eshghi K, Larson RC (2008) Disasters: lessons from the past 105 years. Disaster Prev Manag Int J 17(1):62–82
5. Fischer HW (1998) Response to disaster: fact versus fiction & its perpetuation: the sociology of disaster. University Press of America, Lanham
6. Gao L, Song C, Gao Z, Barabási AL, Bagrow JP, Wang D (2014) Quantifying information flow during emergencies. Sci Rep 4:3997
7. Granovetter M (1973) The strength of weak ties. Am J Sociol 78(6):1360–1380
8. Hughes AL, Palen L (2009) Twitter adoption and use in mass convergence and emergency events. Int J Emerg Manag 6(3–4):248–260
9. Ihler A, Hutchins J, Smyth P (2007) Learning to detect events with Markov-modulated Poisson processes. ACM Trans Knowl Discov Data 1(3):13
10. Johnson NR (1987) Panic at “The Who concert stampede”: an empirical assessment. Soc. Probl. 34(4):362–373
11. Jolliffe I (2002) Principal component analysis. Wiley Online Library
12. Kapoor A, Eagle N, Horvitz E (2010) People, quakes, and communications: inferences from call dynamics about a seismic event and its influences on a population. In: AAAI spring symposium: artificial intelligence for development
13. Kennedy P, Ressler E, Rodriguez H, Quarantelli EL, Dynes R et al (2009) Handbook of disaster research. Springer Science & Business Media, Berlin
14. Kotz D (2013) Injury toll from Marathon bombs reduced to 264. The Boston Globe 24

15. Lind BE, Tirado M, Butts CT, Petrescu-Prahova M (2008) Brokerage roles in disaster response: organisational mediation in the wake of Hurricane Katrina. *Int J Emerg Manag* 5(1–2):75–99
16. Merchant RM, Elmer S, Lurie N (2011) Integrating social media into emergency-preparedness efforts. *N Engl J Med* 365(4):289–291
17. Morrison S, O’Leary E (2015) Timeline of Boston marathon bombing events. Boston.com
18. Muralidharan S, Rasmussen L, Patterson D, Shin JH (2011) Hope for Haiti: an analysis of Facebook and Twitter usage during the earthquake relief efforts. *Public Relat Rev* 37(2): 175–177
19. Oh O, Kwon KH, Rao HR (2010) An exploration of social media in extreme events: rumor theory and Twitter during the Haiti earthquake 2010. In: ICIS, p 231
20. Onnela JP, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási AL (2007) Structure and tie strengths in mobile communication networks. *Proc Natl Acad Sci* 104(18):7332–7336
21. Quarantelli EL (1954) The nature and conditions of panic. *Am J Sociol* 60(3):267–275
22. Quarantelli EL (1978) Disasters: theory and research. Sage, Thousand Oaks
23. Quarantelli EL (1988) Disaster crisis management: a summary of research findings. *J Manag Stud* 25(4):373–385
24. Quarantelli EL (2005) What is a disaster?: a dozen perspectives on the question. Routledge, London
25. Quarantelli EL, Dynes RR (1977) Response to social crisis and disaster. *Annu Rev Sociol* 3(1):23–49
26. Raento M, Oulasvirta A, Eagle N (2009) Smartphones: an emerging tool for social scientists. *Sociol Methods Res* 37(3):426–454
27. Rogers EM (2010) Diffusion of innovations, 4th edn. Simon and Schuster, London
28. Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web. ACM, New York, pp 851–860
29. Schultz F, Utz S, Göritz A (2011) Is the medium the message? Perceptions of and reactions to crisis communication via Twitter, blogs and traditional media. *Public Relat Rev* 37(1):20–27
30. Spiro ES, Fitzhugh S, Sutton J, Pierski N, Greczek M, Butts CT (2012) Rumoring during extreme events: a case study of Deepwater Horizon 2010. In: Proceedings of the 4th annual ACM web science conference. ACM, New York, pp 275–283
31. Starbird K, Maddock J, Orand M, Achterman P, Mason RM (2014) Rumors, false flags, and digital vigilantes: misinformation on Twitter after the 2013 Boston Marathon Bombing. In: iConference 2014 proceedings
32. Sutton J, Spiro ES, Fitzhugh S, Johnson B, Gibson B, Butts CT (2014) Terse message amplification in the Boston bombing response. In: Proceedings of the 11th international ISCRAM conference, Pennsylvania State University, University Park, PA, pp 612–621
33. Sutton J, Spiro ES, Johnson B, Fitzhugh S, Gibson B, Butts CT (2014) Warning tweets: serial transmission of messages during the warning phase of a disaster event. *Inform Commun Soc* 17(6):765–787
34. Tapia AH, LaLone N, Kim HW (2014) Run amok: group crowd participation in identifying the bomb and bomber from the Boston Marathon Bombing. In: Proceedings of the 11th ISCRAM
35. Theirney KJ (2007) From the margins to the mainstream? Disaster research at the crossroad. *Annu Rev Sociol* 33:503–525
36. Vieweg S, Hughes AL, Starbird K, Palen L (2010) Microblogging during two natural hazards events: what Twitter may contribute to situational awareness. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM, New York, pp 1079–1088
37. Wang D, Lin YR, Bagrow JP (2014) Social networks in emergency response. In: Encyclopedia of social network analysis and mining. Springer, pp 1904–1914

## **Part IV**

# **Controlled Studies**

# Randomized Experiments to Detect and Estimate Social Influence in Networks



Sean J. Taylor and Dean Eckles

## 1 Introduction

There is a long tradition in the social sciences of examining how individual level behaviors diffuse and aggregate, including influential work by Schelling [107–109] and Granovetter [56], among many others [25, 87, 101, 119]. Stylistic models from this tradition have been used to explain some of the most important human phenomena, from which innovations are likely to gain widespread usage to who people vote for in elections. The fundamental building blocks of diffusion models are assumptions about how people change their behaviors in response to the behaviors of people they observe or interact with. These assumptions can vary in their disciplinary origins and sophistication—from epidemiological models to game-theoretic models with multiple equilibria.

Randomized experiments provide a useful tool for testing theories. The increasing digitization and connectedness of human behaviors has made digital field experiments cheaper and easier to apply to social behaviors via contemporary communication technologies. This methodological paradigm shift has created opportunities for researchers hoping to understand the underpinnings of large-scale social behaviors in order to improve theory, make predictions, and compare hypothetical policies.

---

S. J. Taylor (✉)  
Facebook, Menlo Park, CA, USA  
e-mail: [sjt@fb.com](mailto:sjt@fb.com)

D. Eckles  
Sloan School of Management and Institute for Data, Systems and Society, Massachusetts Institute of Technology, Cambridge, MA, USA  
e-mail: [eckles@mit.edu](mailto:eckles@mit.edu)

In this review, we hope to make randomized experimentation more accessible to researchers seeking to contribute to our understanding of social influence and diffusion in social systems. We first discuss how randomized experiments can rule out potential confounding factors (Sect. 1.1). Because experiments require that the researcher intervenes in the social system, we devote Sect. 1.2 to discussing the ethical consideration associated with employing digital field experiments.

Section 2 outlines the four components of a randomized experiment to detect or estimate social influence. This facilitates discussing the many design choices experimenters have, including defining the relevant network, what treatments can be employed, and how those treatments may be randomly assigned to subjects. In Sect. 3, we turn to the analysis of experiments in networks, where we focus on Fisherian randomization inference. Section 4 discusses how the analysis of experiments can be extended in various ways in order to increase the usefulness of the results.

This review complements more general references on design and analysis of randomized experiments [15, 55, 64]. Design and analysis with disjoint groups has received substantial attention in economics and epidemiology (e.g., 18, 58, 105, 121]. On the other hand, there are few other reviews of design and analysis of experiments in networks. Compared with extant reviews [5, 122], we aim to integrate all the methods reviewed into a single causal model and discuss some design choices and analysis methods in detail.

What exactly counts as social influence? Different fields distinguish among various processes by which people affect each other. For example, economists distinguish between peer effects caused by constraint, preference, and expectation interactions [78], while other fields may make different distinctions. Thus, for some prior work, “social influence” denotes something more specific. However, given our methodological focus here, we choose to remain agnostic about the mechanisms and define *social influence* to include all processes by which an individual’s behaviors affect another’s, either directly or indirectly. Thus, we could have instead referred to “peer effects,” “diffusion,” or “social contagion.” Further theory-specific distinctions may motivate additional design and analysis choices.

## 1.1 What Makes Randomized Experiments Different?

We privilege information gained through randomized experiments because they create a different kind of knowledge than observational studies: We know exactly how units are assigned to treatments. Thus, a properly implemented experiment rules out all alternative explanations for an observed correlation besides the causal one, and allows for both unbiased estimation of the effect of our intervention and statistical inference that is exact in finite samples [53].

Any observational analysis intended to answer questions about social influence must confront several potential biases that make it difficult to trust its conclusions. First, social networks are known to exhibit strong homophily [44, 73, 80, 83],

creating network correlation of attributes, opinions, and behaviors through people's preferences for whom they spend time with. For instance, people with similar political beliefs may be more likely to form friendships [23, 57, 62], and therefore homophily may readily explain cases of apparent political persuasion. Second, people who are connected in social networks are subject to similar exogenous shocks to their behavior, as when neighbors are exposed to similar marketing messages on billboards.

A substantial program of research has been devoted to proving that what economists call "identification problems" in social influence are likely to be insurmountable without randomized experiments [77, 112]. The intuitive reason is that without intervening in the social system, there are usually reasonable alternative explanations for correlations that do not involve a social influence effect.

Despite their clear advantages, the use of randomized experiments is not a panacea for social scientists. Experiments are usually more costly to design and implement than observational studies because the researcher must alter people's behaviors or interactions in a social system in some way. Interventions require substantial upfront costs for planning and implementation, including: recruitment of subjects, cost of the interventions themselves (financial or logistical), evaluation and exposure of risks of harm to subjects [65]. Because field experiments require researchers to impact the social systems they study at potentially very large scale, they can be associated with different ethical challenges from other methods, which we summarize in Sect. 1.2.

Experiments can also be problematic because, although they reduce concerns about bias, the variance of estimation becomes a first-order concern and the possibility of type II errors (commonly known as issues with experimental power) dominate due to the cost of sample size or the impossibility of the researcher creating large effects [19].

On a more positive note, we will see in Sect. 3 that some well-designed experiments can require more straightforward analysis than observational studies. In addition, there are reduced internal validity concerns with experiments, as they can provide unbiased estimates for the social influence effects they were designed to measure. The two main constraints of an experimental methodology are which estimates are possible and the precision of those estimates.

The randomization the researcher employs and structure of the network together determine what causal quantities of interest can be credibly estimated [117]. These estimands address counterfactual questions about which individual-level behaviors would obtain under alternative interventions. In one simple case, we may be able to answer the question of how much an individual's probability of a behavior is increased by having exactly one peer (rather than no peers) who engages in that behavior. A more complex causal estimand might be the distribution of that behavior in the total population after a series of targeted (e.g., marketing) interventions or a policy change (e.g., by a government). As we will see, a single experiment will generally not answer all possible causal questions and the experiment should be designed with some estimands in mind.

The second constraint from using experiments is the precision of the effect estimates. Experimental data is often more costly to collect than observational data because treatments are not free and observational data is abundant. The power to detect social influence is limited by the direct effects of the intervention (weaker ones provide less experimental power) and the available sample size [55].

## 1.2 Ethical Considerations for Digital Field Experiments

The reduced cost and increased feasibility of digital field experiments (DFEs) has led to increased experimentation over the past decade. While DFEs may help researchers answer many important questions about social influence, they can present more ethical challenges than observational research and even pre-digital lab and field experiments. To ground the discussion, we will refer to the four ethical principles proposed in the Belmont Report [96] and the subsequent Menlo Report [47] which are meant to provide guidance on human subjects research. Those principles—Respect for Persons, Beneficence, Justice, and Respect for Law and Public Interest—are briefly summarized in Table 1.

For an in-depth, thorough treatment of ethics in research in the digital age, we refer the reader to Chapter 6 of [104] and for a recent discussion of institutional review processes to mitigate risk please see [65]. Rather than review those materials exhaustively, we use this subsection to discuss five ethical considerations that we consider to be particularly salient for digital field experiments.

First, DFEs are implemented in software and therefore have very low variable costs with respect to the size of the treated population. It is no longer unusual for experiments to deploy treatments to millions of people [31], amplifying their potential harm compared to more modest sample sizes. Additionally, treatments with network effects can, by research intention or not, cause detrimental effects for people who were not in the original treated population. Researchers acting in accordance with the ethical principle of Beneficence may have a more difficult time evaluating the potential risks of DFEs in networks because their potential effects on social systems are not obvious, intuitive, or even measured.

**Table 1** Ethical principles for human subjects research

Principle	Description
Respect for persons	Treating people as autonomous and acting in accordance with their wishes
Beneficence	Recognizing the potential risks and benefits of research and striking a balance between them
Justice	Ensuring that the risks and benefits of research are fairly distributed
Respect for law and public interest	Recognizing the risks and benefits for all relevant stakeholders, not just research subjects

Second, it may be difficult to identify whether subjects in DFEs are members of a vulnerable or protected population. When designing a DFE, researchers might find it challenging to estimate risks of harm because there is uncertainty about how many subjects could be adversely affected. Researchers might also be unable to reason about whether the benefits and risks of the research are distributed equitably across the population, in accordance with the principle of Justice. On many online or mobile platforms, researchers may not know if users are a reasonable age for consent or are particularly vulnerable to risk from the planned experiment.

Third, DFEs typically use automated, large-scale collection of potentially sensitive and/or identifying information, e.g. location information or exchange of personal communication. Indeed, these data can be integral to the ability of the experiment to answer the research question of interest. For instance, a log of email communications can be used to infer a social network [73], which is a key component for social influence studies. Persistent records of sensitive or identifying information can potentially be used for unintended purposes, causing harm to experimental subjects [89, 91].

Fourth, because of their large scale and integration with existing technologies, DFEs often pose unique challenges for receiving informed consent, which is sometimes an implication of the ethical principle of Respect for Persons. Informing subjects of the experiment and receiving their consent can be disruptive to their normal experiences using various platforms and products (particularly if experiments are frequent, as is becoming more common). Furthermore, requiring informed consent can limit or bias the experimental population or prime the subjects, undermining or altering the treatment effects. Although informed consent is important component of Respect for Persons, deception may be permissible if the experiment complies with all other ethical principles and the deception does not strongly violate the norms of that setting [98]. Some experiments with potentially important benefits *require* deception in order to ensure the research question can be suitably answered. For instance, in the employment discrimination field experiments [98] discuss, one could not credibly measure discrimination after informing employers of the nature of the research.

Fifth, it may be difficult for researchers to comply with all laws, contracts, terms of service, or social norms because DFEs may involve partnerships with companies, span countries or other legal boundaries, or include subjects from many cultures. Inconsistent, overlapping, and sometimes unclear rules and norms lead to challenges for researchers hoping to understand all potential stakeholders and their associated goals and risks.

These five considerations are not meant to be exhaustive—there are certainly other ways in which DFEs can present new ethical challenges for researchers. But we hope that this subsection has made clear that while experimental research has become easier to conduct on some dimensions, it has become more fraught on others—in particular in evaluating and mitigating the risk of harm to subjects.

### 1.2.1 Recommendations for Ethical Research

Taking into account the challenges identified, more research is needed to address the ethical implications in DFEs and to develop mitigating and creative strategies. In the meantime, researchers should do the utmost to:

- Ensure that the research is ethical and beneficial for subjects; that it does not expose them to risk or harm (this may require escalation and further deliberation with other teams within the company, along with the assessment of alternative research methods that could be used).
- Carefully assess if the collection and processing of sensitive data is essential for the research being conducted.
- Determine if an experiment is strictly necessary for the objectives of the research (or if the same results can be obtained through less risky research, e.g. a smaller sample size).
- Whenever possible, ensure that such collection and processing is done with prior informed consent given by the data subjects.
- Only keep that data for the minimum necessary period of time and ensure the proper de-identification of that data according to most effective and updated industry standards.

## 2 Components of a Randomized Experiment

The randomized experiment methodology has four main components:

1. A **target population** of units (i.e., individuals, subjects, vertices, nodes) who are connected by some **interaction network** (Sect. 2.1).
2. A **treatment** which can plausibly affect behaviors or interactions (Sect. 2.2).
3. A **randomization strategy** mapping units to probabilities of treatments (Sect. 2.3).
4. An **outcome** behavior or attitude of interest and measurement strategy for capturing it (Sect. 2.4).

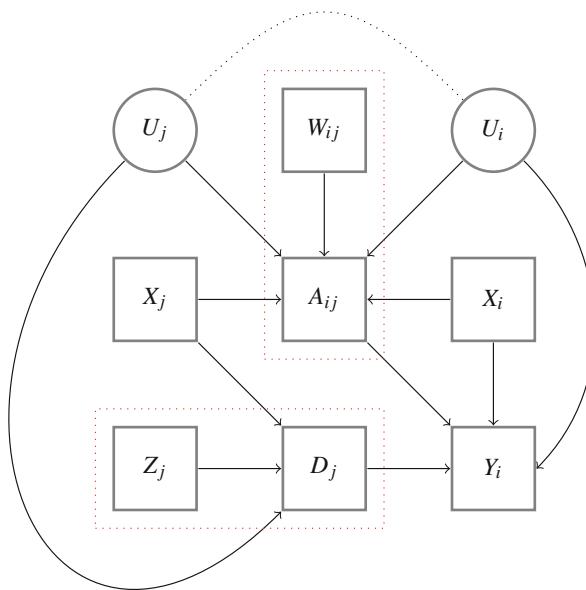
To summarize the relationship of these components, the researcher applies a treatment (2) to a target population (1) using a randomization strategy (3) and then measures the outcome behavior (4).

The following four sections describe these four components, characterize the space of possibilities for each one, and provide examples from existing research. We introduce notation along the way that we will use in Sects. 3 and 4. For convenience that notation is summarized in Table 2. Lowercase letters designate particular fixed values of interest.

**Table 2** Definitions of terminology and notation used in this review

Term	Definition
$X_i$	A vector of pre-treatment covariates about subject $i$
$U_i$	A vector of unobserved covariates about subject $i$
$D_j$	A behavior of the peer $j$ that could affect the subject $i$
$A_{ij}$	Edge between $i$ and $j$ in the interaction network that mediates social influence
$Z_j$	Researcher-determined treatment status for peer $j$
$W_{ij}$	Researcher-determined treatment status for relationship $ij$
$Y_i$	The outcome of interest for subject $i$ , measured post-treatment
Subject	A focal individual whose outcome variable $Y_i$ is studied
Peer	The person whose behavior $D_j$ could influence $Y_i$

See Fig. 1 for a graphical depiction of the relationship between these quantities



**Fig. 1** Causal diagram for the random variables in our example. Squares are observed variables and circles are unobserved. Here  $i$  is the ego or focal subject for whom we will measure outcome  $Y_i$ . We believe that her friend  $j$  can affect  $Y_i$  through her behavior  $D_j$ , which is affected by our treatment  $Z_j$ . The strength of their friendship  $A_{ij}$  can moderate this effect and is exogenously affected by treatment  $W_{ij}$ .  $U_i$  and  $U_j$  are unobserved confounders that cause  $i$  and  $j$  to become friends and may also affect  $D_j$  and  $Y_i$ . By conditioning on  $A_{ij}$ , the backdoor path indicated by the dashed line is activated and provides an alternative explanation for any association we observe between  $D_j$  and  $Y_i$ .

## 2.1 Target Population and Interaction Network

The target population is the set of people whose interactions and behaviors the researcher seeks to study. If we were studying whether peers affect which movies we choose to watch, the population of interest might be movie-goers. Before and during the experiment the target population generates some data, which we list here:

- We observe  $N$  individuals from some **target population**, indexed by  $i$ . This might be a sample or it may be the entire finite population.
- We observe **pre-treatment covariates** for the individuals:  $X_i$ . Commonly researchers collect demographic information such as gender, physical location, or age. Often it is also useful to measure pre-experimental behaviors that are similar to the outcome of interest.
- We observe an **interaction network** between people in the population  $A_{ij}$  where  $i, j \in [1, \dots, N]$ ; alternatively, this is a network  $G = (V, E)$ . This interaction network determines an exposure model—which individuals we expect to potentially influence each other and with what intensity.
- We observe when the population engages in some behavior of interest  $D_i$ .
- We observe some **outcome variable** associated with each individual,  $Y_i$ . For instance, the researcher might survey them to ask often they smoke. We will discuss outcome measurement in more depth in Sect. 2.4.

Substantively, we care about the effect of the behavior  $D_j$  on the outcome  $Y_i$  in the population. The special case where  $D_i = Y_i$  can be termed *in-kind peer effects* and is frequently studied, but it is easy to envision cases where the peer behavior of interest is different from the outcome (e.g., my friend’s studying habits, measured as  $D_j$ , affect my probability of applying to college,  $Y_i$ ).

Selecting the target population often involves tradeoffs between external validity and the ability to collect data about behaviors, outcomes of interest, and relevant social interactions—and intervene. Researchers have used the following three strategies to solve this *recruitment* problem.

First, researchers have continued to recruit convenience samples. As with classic lab experiments in social influence e.g., [13], these are often students from universities and colleges for studies. These samples can facilitate either construction of artificial networks or measuring the subjects’ networks (with, e.g., surveys, asking them to log into Facebook, measuring co-location). The latter strategy can be used to conduct “lab experiments in the field” as existing networks are combining with artificial choices and treatments e.g., [75]. The former strategy has been increasingly used in combination with online labor markets (such as Amazon’s Mechanical Turk), which has created an important new source of experimental subjects [81]. These individuals can be assigned to positions in networks by researchers or through economic games played by the subjects themselves [82, 94, 95, 113]; of course, this may limit external validity.

Second, the last decade has led to a dramatic increase in experiments that are conducted on online social networks or in collaboration with the companies that run online communication services. [6, 7] constructed a Facebook application in

order to gather social network information, introduce a treatment (presence of viral features), and measure the outcome of interest (adoption of the application). Other researchers have worked directly with Internet firms to conduct experiments. [31] and [114] conducted experiments by implementing them in partnership with Facebook (see Bond et al., this volume), while [88] partnered with a social news website to introduce an experimental change.

Third, researchers in education, development, labor economics, and ecology have conducted ambitious field experiments in samples of schools or classrooms [37, 93], villages [35, 71], and animals [4, 52] for which networks can be measured.

### 2.1.1 Measuring or Constructing the Interaction Network

We use the term “interaction network,” which is vague, because what is usually denoted by “social network” will often not be the causal network of interest.<sup>1</sup> In most settings there is some specific type of interaction we hypothesize to transmit the behavior we care about. An intuitive definition is that interaction is “ $i$  considers  $j$  to be her friend,” but, even when this can be operationalized, further consideration of a particular research question may lead to other choices:

- $i$  saw a story  $j$  posted on Facebook [21]
- $i$  is made aware that her friend  $j$  likes a product [20]
- $i$  lives with  $j$  in a dormitory for a year [103]
- $i$  lives in the same household as  $j$  [90]
- $i$  is in the same training class as  $j$  [37]

The researcher hopes that the chosen network captures salient interactions for the influence process she expects. This definition can vary depending on the outcome behavior of interest. In the case of the Sacerdote [103], who study educational outcomes, the interaction network is prolonged co-habitation, while in the case of Bakshy et al. [20], who study clicks on ads, it is merely that a Facebook friend’s name can appear next to an advertisement. A more prolonged, socially important interaction network can plausibly cause larger changes in subject behavior. In the former case, the researchers can study changes in more important and ingrained behaviors like studying habits, while in the latter the researchers must study more proximate outcomes (clicks on ads).

There are many different possibilities for measuring, eliciting, or directly constructing interaction networks. If the research setting is an articulated social network (e.g., an online social network such as Facebook, Instagram, Twitter, or Pinterest), the researcher may use that network’s definition (followers, friends, subscriptions). This approach is convenient but often not the precise interaction network of interest. Most people have online “friends” with whom they never

---

<sup>1</sup>Another related term is “exposure model”—a model that determines which subjects are exposed to which other subjects.

interact in person, as well as “real life” friends who they have not articulated ties with online. Facebook, Instagram, and Twitter use algorithmic ranking to determine which content users see, meaning that a friend or follow relationship on those platforms may not necessarily imply content visibility. If the plausible mechanism of influence is offline, then using an online network might bias estimates of causal effects. A misspecification of the interaction network can, even with randomization, bias measurement of social effects.

In digital settings, interaction networks may be constructed incrementally as people’s interactions in the social system are logged (i.e.,  $A_{ij} = 1$  if  $i$  chatted with person  $j$  during some period). For instance in Bakshy et al. [20], the interaction network is determined by Facebook users seeing advertisements during their browsing sessions. The salient interaction network is easily captured by logging which users see which ads. In addition, logging the interactions which have the potential for transmitting behaviors can improve precision by omitting interactions with no potential to transmit influence. Bakshy et al. [20] could have used other definitions of the interaction network (e.g., Facebook friendship), but these would have yielded biased and/or higher variance estimates of effects.

Like observational research [e.g., the US National Longitudinal Study of Adolescent Health (AddHealth) study [97]], much measurement of social networks for randomized experiments has involved asking subjects who their friends, kin, etc., are. The specific questions can be selected to elicit the possibly domain-specific network of interactions. For example, Cai et al. [35] asked heads of rural households to household heads to list five friends that they most frequently discuss farming and finance with, anticipating that this would be a relevant network for social influence in adoption of weather insurance and spillovers from their intervention. Such questions require being able to uniquely identify the named peers, which may be challenging in the presence of common names and/or limited literacy. Kim et al. [71] thus used a complete photographic census of the villages in which they planning to intervene. When the goal is to measure an objective fact about behavioral interactions, incentives for subjects to truthfully report their friends and tie-strength to researchers could be helpful. For example, Leider et al. [75] use a game in which individuals report how much time they spend with peers and paying them more money if this report matches the peer’s report.

Researchers can infer interaction networks from communication meta-data, especially when it covers enough time to precisely measure interaction rates and the communication medium (e.g., email) is likely to be the medium through which influence is transmitted. Influential observational research has measured networks by counting exchanges of emails [73] or instant messages [9]. Beyond allowing for constructing a binary network, directed behaviors between individuals predict self-reported tie strength [66]. In a randomized experiment, these measures can then be used to estimate how spillovers [31] or social influence [8, 20, 21, 32] varies by tie strength. Choices by researchers in inferring networks from communications data can be non-trivial and have a substantive impact on results [45].

Finally, studies can be designed to *directly construct* the interaction network for the subjects, a strategy which is enabled by running digital experiments even if they

happen to be conducted synchronously in behavioral research labs [69, 82, 94, 95]. For example, Suri and Watts [113] randomly varies the networks on which Amazon Mechanical Turkers play a public goods game. Since creating the interaction network requires the researcher to intervene in the social system, we will discuss this strategy in more depth in Sect. 2.2.

### 2.1.2 Extensions to This Framework

Thus far we have described a randomized experiment with a single time period of post-treatment observation and a single outcome of interest. The DAG in Fig. 1 does not allow for the subject's behavior to affect the peer's behavior, which in turn affects the subject's behavior. There are two simple extensions which may be useful and more realistic. First, we might study the outcome at different points in time (e.g., instead of  $D_i$  and  $Y_i$  we might observe  $D_i(t)$  and  $Y_i(t)$  where  $t$  denotes either discrete or continuous time). Time-dependent behavior is a challenging empirical setting because the researcher will often need to model how the interaction network varies across time, as well as how the individual behavior evolves over time [99].

The second extension is from a single peer behavior and outcome of interest to multiple behaviors and outcomes. We might observe a set of people make decisions about a collection of products, ads, content items, or behaviors, meaning we would measure  $D_{ik}$  and  $Y_{ik}$ , where  $k$  indexes the items. Multiple items present an important opportunity to observe social influence processes play out repeatedly in the same population of individuals across the same interaction network. Studies which measure effects across multiple behaviors might provide a more generalizable estimate of effects or allow the researcher to understand effect heterogeneity on other dimensions. As we discuss in Sect. 3.3, this may offer additional opportunities in analysis.

## 2.2 Experimental Treatments

Treatments are the means by which the researcher intervenes in the social system. The space of treatments is often very limited based on cost and practical constraints, risks to subjects, and simply what changes a researcher can possibly apply in a social system.

We will consider the researcher intervening by setting variables  $Z_j$  and  $W_{ij}$ , usually through some random assignment procedure. Note that we do not assume the researcher can directly change  $D_j$  and  $A_{ij}$ , as these variables are chosen by individuals and can often only be affected through the researcher-controlled instruments. The case where this is possible is the special case of perfect compliance, which is rare in field experiments. Instead, we posit a (potentially estimable) compliance

model that produces  $D_j$  and  $A_{ij}$  and which may also include pre-treatment variables and random noise. This section focuses on defining these treatments; we defer their random assignment to Sect. 2.3 below.

### 2.2.1 Subject-Level Treatments

A binary subject-level treatment is denoted by  $Z_j \in \{0, 1\}$ , where  $Z_j = 0$  by default, and where this treatment is expected to affect behavior such that  $D_j(z_j) = f_i(z_j, \epsilon_j)$ , with observed  $D_j = f_i(Z_j, \epsilon_j)$ . The direct effects of the treatment may sometimes be of interest (e.g., effects of a message on voter turnout), but the idea here is that  $Z_j$  functions as an encouragement or instrumental variable with respect to  $D_j$ , allowing interpretation of spillovers from treatment as social influence via  $D_j$ . Thus, researchers can create these treatments primarily for this purpose of detecting social influence. For the treatment to be effective as an instrument, we must believe that  $f_i$  is such that changing  $Z_j$  sometimes changes  $D_j$ ; for example, perhaps  $D_j(z_j) = \mathbb{1}\{\alpha + \beta z_j + \epsilon_j > 0\}$  with  $\beta \neq 0$ , which can be tested. Many interventions (e.g., providing information, advertisements) cause only small changes in the behavior, making detecting downstream social influence difficult.

The special case where  $D_j = Z_j$  is known as perfect compliance. Noncompliance may also be only one-sided, such that if  $Z_j = 1$  then  $D_j = 1$ . Say we are interested in social influence in adoption of a paid upgrade of a music streaming service. We could, as do [24], purchase the upgrade for active users at random, thus producing one-sided, rather than two-sided, noncompliance (i.e., users could still purchase the upgrade on their own if we did not).<sup>2</sup> Two-sided noncompliance seems to be more common in the social sciences, particularly among the difficult-to-change behaviors which are often most interesting to study (e.g., health behaviors, costly product purchases).

Experiments using subject-level treatments within groups (i.e., networks consisting of disconnected cliques) to detect and estimate social influence—sometimes called *partial population experiments* [86]—have been adopted in economics and political science [2, 48, 54, 84, 90]. These designs are based on the expectation that treating a fraction of subjects can induce detectable changes in the population of individuals connected to them. A smaller number of such experiments have been conducted in networks; these too have often relied on having a network multiple connected components (e.g., villages, schools) [e.g. 35, 43, 71, 93] with few exceptions [31].

Knowledge of the interaction network can be crucial for the success of subject-level treatments. If there is uncertainty about *which* peers may be affected by

---

<sup>2</sup>Of course, in such cases we may wonder whether  $D_j$  (i.e., having the upgrade) was really the behavior we were interested in. Perhaps so—if most of the effects of peers’ upgrades on subjects would be via a single indicator on the peers’ profiles that they had upgraded.

a subject's treatment, then detecting effects can become more burdensome from a statistical standpoint because omitting edges or including irrelevant ones adds additional random variation in estimation.

### 2.2.2 Interaction-Network Treatments

In an interaction-network treatment, the researcher intervenes by setting  $W_{ij}$ , which affects the interaction network of the subjects in the experiment; that is,  $A_{ij}(w_{ij}) = g_{ij}(w_{ij}, U_i, U_j, v_{ij})$ , with observed  $A_{ij} = g_i(W_{ij}, U_i, U_j, v_{ij})$ . As above, if  $A_{ij}$  is binary, we may posit that  $A_{ij}(w_{ij}) = \mathbb{1}\{\gamma + \delta w_{ij} + v_{ij} > 0\}$  with  $\delta \neq 0$ . Then particular edges may exist ( $\delta > 0$ ) or not ( $\delta < 0$ ) because of the treatment.

In the edge-formation case of  $\delta > 0$ , we have treatments such as suggesting that two people become friends or introducing them [17, 110]. Not all suggested edges will form, but we expect that some will. Researchers sometimes define the interaction network such that there is perfect compliance. There are numerous examples of randomized *group formation* with ostensibly perfect compliance. Hasan and Koning [60] used a novel group randomization to understand how the constituents of groups affect ideation. Sacerdote [103] and Carrell et al. [37] use random assignment of college roommates and squadrons in order to understand how these groups affect various learning and development outcomes. Note that the degree of “compliance” depends on how the network is defined. Although roommate assignment creates perfect compliance for the network of *roommates*, there is still two-sided noncompliance for the network of *friendships*.

Encouraging edge removal, preventing formation, or attenuating interaction ( $\delta < 0$ ) can also be possible, if challenging in practice, and would rely on the researcher discouraging at least one type of interaction between individuals in the population. As an extreme example, researchers studying smoking cessation could ask subjects to delete phone contacts for any friend they believe might encourage them to continue smoking.

In the context of online communication technologies, whether some binary treatment should be understood as encouraging or discouraging interaction is relative to an arbitrary and temporary status quo. For example, Eckles et al. [49] analyze an intervention that modifies the display of  $i$ 's posts to  $j$ , varying the salience of the user interface elements for commenting on the post.

Perfect compliance, or at least one-sided noncompliance, can also occur when there is some exhaustive channel by which interaction occurs. For example, Aral and Walker [7] randomizes along which edges notification for their Facebook application are sent, thus defining an interaction network that is a random subset of the Facebook friendship network. Similarly, Bakshy et al. [20] randomizes whether or not a friend appears as social context for an advertisement. We have elsewhere called these *mechanism* experimental designs since they randomize whether particular mechanisms for social influence are active [49].

## 2.3 Randomization Strategy

A randomization strategy  $\phi$  specifies a probability distribution over treatment assignments; here,  $\pi_\phi(Z)$  or  $\pi_\phi(W)$ , where  $Z$  is the  $N$ -vector of subject-level treatments  $Z_i$  and  $W$  is the matrix of edge-level treatments  $W_{ij}$ . The marginal distribution is thus a function that maps a subject ( $j$ ) or edge ( $ij$ ) to probability of treatment. More advanced experiments might additionally allow this function to depend on pre-treatment covariates  $X_j$  or the existing interaction network  $A_{ij}$ . The specific form of the randomization determines what causal questions the experiment is capable of, or especially suitable for, answering.

### 2.3.1 Implementing Randomization

In practice, researchers tend to implement randomization using deterministic cryptographic hash functions to generate pseudo-random variables with specified distributions [22, 72]. PlanOut is a domain-specific language for specifying randomization strategies that is used at Facebook and several other companies.<sup>3</sup> Using variable-specific cryptographic salts, PlanOut provides functionality for independent random assignment for multiple experiments, multiple variables, and multiple types of units (e.g., users, clusters, items, edges). The determinism of the hash functions ensures that a random assignment is “persistent,” without requiring the assignments be stored; that is, the assignments can be computed online and statelessly, as subjects arrive. PlanOut code implementing the i.i.d. randomization we described in the previous paragraph as well as some more advanced randomizations are shown in Listing 1.

**Listing 1** Example PlanOut code for subject-level treatment assignment.

```
# i.i.d. random assignment
smoking_program = uniformChoice(choices=[0,1],
    unit=subject_id);

# block random assignment
smoking_program = uniformChoice(choices=[0,1],
    unit=subject_group_id);

# hierarchical block random assignment
smoking_program_prob = randomFloat(min=0, max=1,
    unit=subject_group_id);
smoking_program = bernoulliTrial(p=smoking_program_prob,
    unit=subject_id);
```

---

<sup>3</sup>The design of PlanOut is described in Bakshy et al. [22] and it is available from <https://github.com/facebook/planout>.

### 2.3.2 Subject-Level Treatment Randomizations

Here we consider randomizations for subject-level treatments. Consider the simplest possible randomization for a subject-level treatment is independent and identically distributed (i.i.d) Bernoulli random variable:  $Z_j \sim \text{Bernoulli}(0.5)$ .<sup>4</sup> Say  $Z_j$  is assignment to a smoking prevention program. We hypothesize that the program will reduce how much people in the study smoke (i.e.,  $D_j$  is lower in expectation when  $Z_j = 1$ ), and are further interested in using this randomization to learn about social influence in smoking. In the context of disjoint groups (i.e., a network consisting of multiple disjoint cliques), we can think of this randomization as a *partial population experiment* [18, 86], in that some of the population is treated and we can study behavior of their peers. This design is analogous to marketing interventions which seek to exploit spillovers or network effects in demand by providing discounts or promotions to a small subset of consumers [59].

In order for our randomization to enable detecting and estimating social influence, we will generally need variation in the treatments of the peers of our subjects. While many measures of peer treatment can be used, we will illustrate the points in this section with the fraction of  $i$ 's peers who are treated:

$$T_i = \sum_{j=1}^N \bar{A}_{ij} Z_j,$$

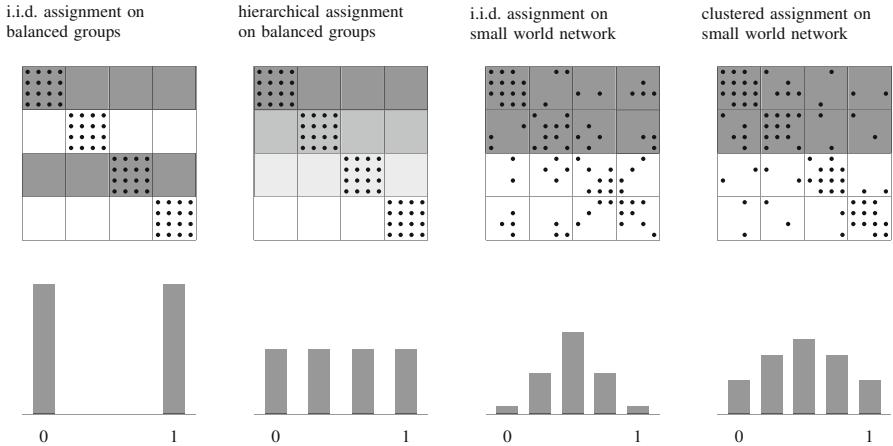
where  $\bar{A}_{ij} = A_{ij} / \sum_{j=1}^N A_{ij}$  an entry in the row-normalized adjacency matrix, with  $\bar{A}_{ij} = 0$  if  $\sum_{j=1}^N A_{ij} = 0$ .

The i.i.d. subject-level assignment described above and shown in the third panel of Fig. 2 has a very important limitation: if a subject has a substantial number of peers, then there is a vanishingly small probability that they will all be assigned to treatment; for example, if subject  $i$  has 10 peers, then  $\Pr(T_i = 1) = \Pr(\sum_{j=1}^{10} Z_j = 10) < .01$ . So we are unlikely to be able to use an experiment with this type of randomization to answer counterfactual questions about having all (or even a large percentage) of a person's friends participate in the program. For some asymptotic sequences with growing degree, this will mean the variance of sample means for units with, e.g., all treated peers diverges [118]. Thus, we will often want to consider other randomizations.

At the opposite extreme, we could assign treatment at the level of groups or clusters. For instance, if students are grouped by classrooms, we could do the smoking prevention assignment at the classroom-level. Let  $c(j)$  be the classroom for subject  $j$ . Then a group-level randomization would be to assign each group an i.i.d Bernoulli,  $P_c \sim \text{Bernoulli}(0.5)$  and assign each student her group's assignment,  $Z_j = P_{c(j)}$ . If we think the classrooms are disjoint cliques, we might

---

<sup>4</sup>Often the literature on randomized experiments e.g., [55, 64] starts with a *completely randomized design*, in which some fixed number  $N_1$  of the  $N$  subjects are assigned to treatment. However, in the case of large digital field experiments implemented as described in Sect. 2.3.1, this cannot easily be done in online (i.e., streaming) assignment without complications.



**Fig. 2** Various subject-level randomizations illustrating how they each induce different distributions of treatment status for a subject’s friends. Each of the four squares is an adjacency matrix (dots represent undirected friendships). The horizontal grey bars represent treatment probabilities, with the darkest color indicating treatment is assigned to subjects in that row with 100% probability. The stylized histograms beneath the squares indicate the fraction of friends who are treated induced by the randomization strategy above it

posit an interaction network that is a block-diagonal matrix, such that  $A_{ij} = \mathbb{1}\{c(i) = c(j)\}$ . Note that in the case of disjoint groups, such an “everyone or nobody” randomization abandons the partial population idea. This randomization can help answer questions about what will happen should we deploy the program to everyone, but it cannot answer questions about social influence and thus whether the program can be deployed more cost-effectively by treating a smaller proportion of students. We may be able to dramatically reduce smoking in a classroom by encouraging 25% of the students to not smoke. In this group randomization, we will never observe a classroom with any quantity other than 0% or 100% of the students treated; see the first panel of Fig. 2.

Intermediate designs between these two extremes use a hierarchical (or, in this case, two-stage) randomization to create additional dispersion in the quantity of students per classroom assigned to the treatment, but also make subject’s own treatment and their peers not perfectly dependent. For example, we can first draw a random uniform variable per classroom,  $P_g \sim \text{Uniform}(0, 1)$ , and then for each student, we draw a Bernoulli random variable with their group’s probability,  $Z_j \sim \text{Bernoulli}(p_{c(j)})$ .<sup>5</sup> For some randomization  $\psi$ , we call it *overdispersed* because  $\text{Var}_\psi(T_i) > \text{Var}_{\text{iid}}(T_i)$ ; that is, it has greater variability in the fraction

<sup>5</sup>With a small number of groups, we may want to use a completely randomized design, rather than independent draws of  $P_c$ . Baird et al. [18] consider optimal two-stage randomizations in the context of disjoint groups, given the goal of estimating some particular direct or indirect effects.

of peers treated than from i.i.d. subject-level randomizations. An overdispersed randomization could be useful for selecting a number of students to treat per classroom, given some budget, that will minimize smoking because it can provide an estimate smoking behavior under different many levels of treatment.

Block-diagonal networks (e.g., villages assumed to not interact) make overdispersed randomizations easy to implement. With more general networks, there are more design choices, and it can be difficult to generate arbitrary degrees of overdispersion in friend treatment assignment probabilities. We may prefer a randomization such that the distribution of  $T_i$  has certain properties; for example, one heuristic is we should have *positivity* such that  $\Pr(T_i = k) > \varepsilon$  for all feasible fractions  $k$  given  $i$ 's degree. Or we may aim to maximize  $\Pr(T_i = k)$  for  $k \in \{0, 1\}$ . One recently popular way to do so is to partition the network into clusters using existing graph partitioning algorithms, and then proceed with the cluster-randomized design (i.e., *graph cluster randomization*; [51, 106, 118, 123]). Given the structure of the network, there will still be edges between clusters (fourth panel in Fig. 2). For example, say we use state-of-the-art methods to partition the Facebook friendship network; with only 1000 clusters, already over 40% of edges will be between clusters [111]. Not only is graph partitioning challenging in large networks, but standard min-cut objectives will often just be a heuristic: we would instead prefer to optimize bias or total error in estimation of particular quantities. To facilitate such optimization, one can further treat the clusters, or some other model fit to the network (e.g., a more general stochastic block model [68]), as an approximation to the observed network. Thus, Basse and Airoldi [27] propose using optimal designs for approximations to the observed network.

A final design possibility with subject-level treatment randomizations is that treatment assignment probabilities can depend on pre-treatment covariates  $X_i$  in order to increase precision. *Blocking* or *pre-stratification* exactly balances some covariates between treatments, rather than simply balancing them in expectation, thus reducing the variance in effect estimates is attributable to the random assignment of treatments causing covariate imbalance in small samples [55, ch. 4]. For instance, in a small sample it could make a large difference in estimates if a subject who is very active or who has many friends is assigned to treatment or not. State-of-the-art blocking methods allow improving balancing on high-dimensional covariates and lead to higher-precision estimates of treatment effects [61]. While usually large samples make blocking irrelevant because post-stratification or regression adjustment can provide similar precision gains [85], use of graph cluster randomization again reduces the effective number of units being randomized, perhaps making blocking a relevant design consideration.

Pre-treatment covariates can be used to target specific subjects who may have certain network positions or be likelier to cause social influence based on some hypothesis or prior analysis. If a researcher wanted to test a seeding strategy based on network position, a reasonable design would be to select a set of influential candidate subjects [70] and treat a random fraction of them while reserving some others as a control [cf. 28, 71].

### 2.3.3 Interaction-Level Treatment Randomization

Treatments defined at the level of individual edges allow for further choices in randomization. Because this design space is so large, we consider some notable examples.

Historically, many examples of interaction-level treatments come from experiments in the formation of random groups. Here the interaction network is set in advance by the researcher or by some exogenous process. From a notational standpoint, these designs amount to setting  $W_{ij} = 1$  for blocks of subjects to induce variation in  $A_{ij}$  and, in turn, the distribution of quantities such as the fraction of adopting peers,  $\sum_{j=1}^N D_j \bar{A}_{ij}$ . An important aspect of this type of randomization is that the resulting groups must exhibit variance on  $D_j$ , the behavior of interest. For the same reason that i.i.d. assignment in subject-level treatments may not cause sufficient variation in peer exposures, large random groups are unlikely to be useful for identifying causal effects [cf. 3]. As with subject-level treatments above, it may be desirable to introduce overdispersion in group composition.

The random group assignment designs generally leverage existing group formation policies. In the case of [103], which exploits the fact that roommate assignments at Dartmouth college are conditionally randomly assigned (directly setting  $A_{ij} = 1$  for the “is roommate” relation), we may even consider this a natural experiment. On the other hand, Carrell et al. [37, 38] introduce novel group formation policies for squadrons at the United States Air Force Academy; here squadrons are groups of roughly 30 that cadets are required to spend the majority of their time with. As a further refinement, random group formation can be performed dynamically to allow for repeated measurements of the same individuals as they change social contexts. Hasan and Koning [60] uses such a randomization to measure how group interactions between entrepreneurs affect their ideation. Their approach allows them to not only measure how changing groups affects their outcome of interest, but allows for longitudinal measurements of individual outcomes as well.

Without leveraging existing group formation policies, researchers may be limited to encouraging the formation edges that involve less prolonged contact. Several experiments have randomly assigned subjects to different graph structures whether in an artificial setting e.g., [69] or in the context of an online health-related service [39]. Here the experiment is generally conceptualized at the level of entire replications of a particular graph. Thus, the outcomes and analyses may be defined and conducted in aggregate rather than at the individual level. One can think of these designs as randomizing  $A$  directly and then observing some aggregate network outcome, which is slightly more complex than the framework we propose here.

Other edge-level treatments are best understood as conditional on peer behaviors and a pre-treatment network. These include what we have called mechanism experimental designs, which work by randomizing whether a social signal is delivered via particular channel. Mechanism designs [e.g., 7, 21, 31] are equipped to answer counterfactuals about how peer behavior would be affected in the amplification or

attenuation of the influence channel of interest.<sup>6</sup> For example, in [20], the only peers eligible for the experiment are those who have already liked a page on Facebook (conditioning on  $D_j = 1$ ) and the randomization (assigning  $W_{ij}$  as a Bernoulli random variable with perfect compliance for  $A_{ij}$ ) determines whether this behavior will be displayed when the focal user sees an ad. Aral and Walker [7] uses another mechanism design in exploiting the fact that notifications in their Facebook application are delivered to a random set of the user’s friends. If we believe that these notifications are the only mechanism through which a Facebook friend might adopt the application, this amounts to randomly amplifying values of  $A_{ij}$  for the friends who received the notifications, while leaving it un-amplified for the remaining Facebook friends that were collected when the user installed the application.

Edge-level randomizations need not be i.i.d. For instance, Bakshy et al. [20] selects random subsets of edges involving the same subject. In the context of a treatment that encourages providing feedback (likes and comments on Facebook, in this case) along a specified directed edge, Eckles et al. [49] compare different possible randomizations. One sender-clustered design would randomly assign vertices to an encouragement to give all of their peers more feedback. Another recipient-clustered design would randomly assign vertices to have *all of their peers* encouraged to give them feedback. This latter design is used in Eckles et al. [49], as simulations suggest it will often have precision advantages. Finally, other designs could, like some of the designs we considered in the previous section, interpolate between i.i.d. assignment of edges and either of these clustered designs.

## 2.4 Outcome Measurement

Perhaps an underrated requirement of randomized experiments is the ability to measure an outcome appropriate to the research question at hand. Sometimes researchers invest more time and expense in intervening with their treatment than in measuring the outcome. However, precise, valid, and complete measurement plays a large role in the success of randomized experiments.

A simple example is that, if outcomes are measured with noise, the resulting estimates will be less precise. Even more problematic are cases where some outcomes are missing, either randomly or not. Coey and Bailey [42] shows that matching ad exposures to conversions via cookies—where matching is random but plausibly independent of treatment status—results in a substantial loss of experimental power. Other experiments might rely on surveys or self-reports to measure outcomes, which yields either a biased measurement (e.g., social desirability) or a treatment effect

---

<sup>6</sup>An additional refinement of the model we outline here is subjects may be connected via multiple overlapping networks, such as in-person vs online interactions, and an experiment may cause changes in some of those networks but not others.

estimate for only a biased sub-population (survey takers). Berry and Taylor [30], who studies social influence for comment quality in public discussions, can only measure comment quality improvements for the set of subjects who choose to write comments. Bond et al. [31] measures voter turnout by matching Facebook users to people in the state voter files, which is a noisy process (match rates were about 40%) that was limited to 13 states because of the expense of acquiring voter file data.

Digital field experiments present some opportunities and also limitations for experimenters. Many important outcomes are potentially observable, such as clicks on ads [20], sharing and production of user-generated content [21, 49], and adoption of apps (both free and paid) [6, 24]. However, digital platforms create comprehensive logs of *digital* behaviors, which are perhaps not the only behaviors of theoretical interest. For instance, while [74] applies a reasonable text-analysis procedure to measure people’s emotions at scale, it is debatable whether a change in emotion is adequately captured by the text they choose to share on Facebook [29]. The sheer volume of data produced on digital platforms is a signal of how trivial the actions they collect can be. Despite dramatic advances in observability of human behavior, it continues to be a central research challenge to measure important outcomes and join them to experimentally assigned treatments.

### 3 Analyzing Randomized Experiments

One frequent consequence of having a well-designed randomized experiment is that the data analysis is then straightforward. While this is true to some degree in experiments about social influence in networks, estimation and inference can both be complicated by the network. Causal and statistical inference in networks remains an active research area, with contemporary contributions to basic problems such as laws of large numbers and asymptotic inference in networks [12, 76, 116, 120].

In this section, we review methods for estimation and inference (e.g., hypothesis testing) for social influence in network experiments. The known randomization of subjects or edges to treatments provides a “reasoned basis” for inference [53, p. 14] with minimal assumptions even when we only observe a network with a single giant component. We thus focus on Fisherian randomization inference, but briefly review other methods.

As with the experimental design, the primary goal in analysis is learning about social influence. Ideally, this means learning about effects of  $D_j$  on  $Y_i$  or of  $A_{ij}$  on  $Y_i$ . It will often be more straightforward to simply detect any effects of  $Z_j$  or  $W_{ij}$  on  $Y_i$ . This is because (a) the experimenter sets these, but usually only affects  $D_j$  on  $Y_i$  indirectly and (b) in measuring  $D_j$  and  $A_{ij}$ , we may not capture all of the ways that our treatments can affect subjects. Thus, we can often take evidence about effects of  $Z_j$  or  $W_{ij}$  as evidence of social influence, without being about to denote these effects in terms of peer behaviors. We start with this simpler case.

### 3.1 Effects of Randomized Treatments

In this section, we consider how to conduct inference about effects of randomized treatments. We start by considering inference about spillovers in experiments where subjects are randomly assigned to subject-level treatments; that is, we are interested in questions about whether subjects' outcomes are affected by others' treatments. If we assume that others' treatment only affect an individual through others' behaviors (Fig. 1), then these tests are also tests of social influence.

#### 3.1.1 Testing Sharp Null Hypotheses About Spillovers

Consider the null model in which there is a direct effect of a subject's own treatment, but no effects of others' treatments, including those of peers.

**Hypothesis 1 (No Spillovers with Constant Direct Effects)** *There exists some  $\tau$  such that  $Y_i(z_i) = \tau z_i + \xi_i$  for all  $z \in \mathbb{Z}^N$  and  $i \in V$ .*

Note that under this null hypothesis  $Y_i - \tau Z_i$  does not vary under alternative treatment assignments. This null hypothesis is a composite of null hypotheses of the form:

**Hypothesis 2 (No Spillovers with Constant Direct Effects,  $\tau_0$ )**  *$Y_i(z_i) = \tau_0 z_i + \xi_i$  for all  $z \in \mathbb{Z}^N$  and  $i \in V$ .*

Hypothesis 2 is a *sharp null hypothesis*, which allows inferring all of a unit's potential outcomes from its single, observed potential outcome. We can thus use Fisherian randomization inference, in which we exploit our knowledge of the distribution of  $Z$  (which we or the experimenter chose), to test this null hypothesis. This is often implemented as a permutation test with a test statistic chosen to be sensitive to the kinds of deviations from the null that we expect. For example, consider a larger model that includes a linear effect of the fraction of treated peers:

$$Y_i = \tau Z_i + \rho \sum_{j=1}^N Z_j \bar{A}_{ij} + \xi_i \quad (1)$$

where  $\bar{A}_{ij}$  is an entry in the row normalized adjacency matrix. A non-zero  $\rho$  would correspond to a particular violation of Hypothesis 1. The score statistic for  $\rho$  can be used as a test statistic [16], as can many other test statistics.

Algorithm 1 tests Hypothesis 2 using Fisherian randomization inference. To test Hypothesis 1, researchers would generally test many particular values of  $\tau$  (e.g., in a grid, or through a search algorithm) and take the supremum.

**Algorithm 1 (Randomization Inference for Hypothesis 2)** *Inputs:* test statistic  $T(\cdot, \cdot) : \mathbb{Y}^N \times \{0, 1\}^N$  that is a function of units' residual outcomes and the treatment vector; posited direct effect  $\tau_0$ .

1. Compute residual outcomes given  $\tau_0$ ,  $\tilde{Y} := Y - \tau_0 Z$ .
2. For every  $r \in \{1, \dots, R\}$  and some  $\tau_0$ :
  - a. Draw a new treatment vector  $Z^*$  consistent with the original randomization.
  - b. Compute value of test statistic with observed outcomes and permuted treatment  $T_{\text{null},r} := T(\tilde{Y}, Z^*)$ .
3. Compare observed and null test statistics, yielding

$$\widehat{\text{p-value}}(\tau_0) = \frac{1}{R} \sum_{r=1}^R \mathbb{1}\{T(\tilde{Y}, Z) > T_{\text{null},r}\}.$$

We would then reject Hypothesis 2 for small  $p$ -values, instead concluding subjects are affected by others' treatments.

*Remark 1 (Randomization Inference and Permutation Tests)* While randomization inference frequently makes use of permutation tests, the two are not identical. Fisherian randomization inference makes use of knowledge about the exact distribution of variables that were randomized to conduct exact causal inference for a finite population of units. Often (e.g., with a single completely randomized treatment vector) this can be approximated to arbitrary precision through permutation of the treatment vector, but need not be if the distribution over treatments is more complicated. Furthermore, permutation tests of social influence are often used without the justification they are afforded by randomization; that is, they are often used when other assumptions would be needed to make them exact in finite samples or even good asymptotic approximations. For example, Anagnostopoulos et al. [1] make additional, strong assumptions about non-influence processes to justify the use of a permutation test to detect influence in observational data.

Even in the case of randomized experiments, particular permutation tests may not be readily justified by the randomization. Without explicitly considering the relevant sharp null hypothesis, it can be easy to make mistakes that make the resulting permutation test invalid. For example, Bond et al. [31] test for spillovers from a randomly assigned encouragement to vote in the 2010 U.S. elections. This was implemented as a permutation test that implicitly assumed the absence of direct effects, even though Bond et al. [31] elsewhere rejected that null hypothesis. Athey et al. [16] show that such tests can have dramatically inflated Type I error rates (i.e., they too often reject the null hypothesis when it is true).

### 3.1.2 Inference for the Magnitude of Spillovers

Say we use Algorithm 1 and reject Hypothesis 1. We may further wish to quantify the magnitude of these spillovers from treatment. These methods can also be used to construct acceptance regions for more complex positive hypotheses about the size of spillovers in the network. To do this, we can use a similar test but with Eq. (1) specifying a sharp null hypothesis given a choice of  $\tau$  and  $\rho$ . We can, for example, use a test statistic that measures model fit (e.g., sum of squared residuals) [34] and determine a region of  $\tau$  and  $\rho$  values that we do not reject (i.e., an acceptance region). With only these two parameters, grid search is often feasible, but other search algorithms can be used. For more on this topic, see Bowers et al. [33, 34].

The preceding methods require testing a sharp null hypothesis or a composite null consisting of a parametrically defined set of sharp nulls. In particular, we imposed the constant effects assumption that the direct effect of the treatment  $\tau$  was common to all units. If direct effects are heterogeneous, these tests could reject the null even when there are no spillover effects of treatment. To partially address this concern, we could expand the null model to allow effects to be heterogeneous by observed subject covariates  $X_i$ ; however, this would not allow for latent heterogeneity in direct effects. Outside the context of networks, we might be confident that, at least asymptotically, good choices of test statistics would result in tests that are not asymptotically sensitive to this heterogeneity [41]; however, we lack such asymptotic results for networks. In the next sections, we consider alternative methods that do not make use of these homogeneity assumptions. Nonetheless, the preceding methods may have some advantages in practice (e.g., greater power).

### 3.1.3 Conditional Randomization Inference in Networks

How can we use randomization inference to test for spillovers without specifying the form of direct effects? Consider a null hypothesis of no spillovers in the absence of assumptions about constant direct effects of treatment.

**Hypothesis 3 (No Spillovers)**  $Y_i(z) = Y_i(z')$  for all  $i \in V$ , and all pairs of assignment vectors  $z, z' \in \{0, 1\}^N$  such that  $z_i = z'_i$ .

This hypothesis is not sharp because it does not specify how each subject would have behaved if its treatment were different. Rather, it posits levels sets of  $Y_i(\cdot)$ . It is possible to test such non-sharp null hypotheses by using conditional randomization inference—that is, by conditioning on functions of the treatment vector  $Z$  [11, 16, 100].

Here consider the basic case of testing Hypothesis 3. In particular, we can designate a subset of subjects as *focal subjects* for which we examine their outcomes and condition on their observed treatment assignment [11, 16]. Note that, conditional on the focal subjects receiving the same treatment, Hypothesis 3 is now sharp for those subjects. We can implement this test as follows.

**Algorithm 2 (Conditional Randomization Inference for Hypothesis 2)** *Inputs:* set of focal units  $V_F \in V$ , test statistic  $T(\cdot, \cdot) : \mathbb{Y}^{|V_F|} \times \{0, 1\}^N$  that is a function of focal units' outcomes and the treatment vector.

1. Draw permuted treatment vector  $Z^*$  such that all focal units get the same treatment as observed,  $Z_i^* = Z_i$  for all  $i \in V_F$
2. Compute value of test statistic with observed outcomes and permuted treatment  $T(Y_{V_F}, Z^*)$
3. Repeat 1 and 2 for  $R$  times, storing results as the  $R$ -vector  $T_{\text{null}}$ .
4. Compare observed and null test statistics, yielding

$$\text{p-value} = \frac{1}{R} \sum_{r=1}^R \mathbb{1}\{T(Y_{V_F}, Z) > T_{\text{null},r}\}.$$

We would then reject Hypothesis 3, and thus the stronger Hypothesis 1, for small values of this  $p$ -value. This test has the correct Type I error rate without any assumptions about the model for direct effects.

How should the focal subjects be selected? Any choice is valid (i.e. results in correct Type I error rates), but this choice can affect power. First, in some cases, this choice may be obvious because of the availability of outcome data. For example, when joining treatment and network data with a second data set with outcomes, a researcher may only observe outcomes for a small fraction of subjects, which could then be designated the focal subjects e.g., [67]. Second, theory or prior observations may suggest that some subject may not respond to social influence; it may be desirable to not include them as focal units. Finally, the network itself can be used to select focal subjects to improve power [16, 26].

It is possible to apply similar approaches to testing for higher-order spillovers, testing for spillovers on a second network, and other hypotheses about spillovers. When the null hypothesis allows for, e.g., spillovers from immediate neighbors on a relatively dense network, these methods may lack sufficient power to be useful. We refer readers to Athey et al. [16] for details.

### 3.1.4 Extension to Edge-Level Treatments

We have focused on the case where subjects, rather than edges, are assigned to treatments; however, similar methods can be used when edges are assigned as long as either (a) a sharp null hypothesis can be posited or (b) a non-sharp null hypothesis implies level sets that can be conditioned on.

### 3.2 Estimating Effects of Peer Behaviors

Thus far we have described inference about effects of other subjects' randomly assigned treatments, while often the substantive questions are about effects of other subjects' behaviors (i.e. social influence). As noted above, if we assume that a subject's outcome is only affected by a peers' treatments via their behaviors, then evidence for spillovers from treatment is evidence for social influence. However, we are often interested in quantifying the size of this social influence by, e.g., estimating effects of  $D_j$  or  $A_{ij}$  on  $Y_i$ .

We can proceed as before by considering the following sharp null hypothesis, which specifies how a subject's outcomes vary with its own treatment and peers' behaviors.

**Hypothesis 4 (Constant Direct Effects and Social Influence,  $(\tau_0, \theta_0)$ )**

$$Y_i(z, d) = \tau_0 z_i + \theta_0 \sum_{j=1}^N d_j \bar{A}_{ij} + \xi_i, \quad (2)$$

for all  $z \in \{0, 1\}^N$ ,  $d \in \mathbb{D}^N$ , and  $i \in v$ .

According to Hypothesis 4, subjects are unaffected by others' treatments except as reflected in their neighbors behaviors  $D_j$ . This is a *complete mediation assumption* or *exclusion restriction* and is encoded in Fig. 1. Combined with  $Z$  having been randomized, this is sufficient for function of  $Z$  to be used as instrumental variables for social influence. Following Imbens and Rosenbaum [63], we can then test Hypothesis 4 by noting that it implies that  $Y_i(z, d) - \tau_0 z_i - \theta_0 \sum_{j=1}^N d_j \bar{A}_{ij}$  is invariant in  $z$ , and thus that Algorithm 1 can be applied with this alternative residualization of the outcomes.

### 3.3 Other Methods of Analysis

There are some other methods available for statistical inference about spillovers and social influence with randomized experiments. Under monotonicity assumptions (i.e., that treating more subjects can only increase all subjects' potential outcomes) and with bounded outcomes, it is possible to construct confidence intervals for effects attributable to the observed treatment assignment [40]. Or under local interference assumptions (i.e., subjects are only affected by immediate peers' treatments) and bounded degree, it is possible to do conservative asymptotic inference [12, 120].

In some cases the presence of replication is helpful by allowing for plausible independence assumptions. First, there may be observation of multiple plausibly independent behaviors on a single network. For example, Bakshy et al. [20] randomizes a mechanism of social influence for many different subjects and brands.

In their estimation and statistical inference, they assume that outcomes that do not have a common subject or brand are independent. They then use statistical methods that account for dependence of observations within brands and users [19, 36, 92]. Ignoring or not properly accounting for dependence in analyzing such experiments would increase the type I error rate.

Second, some networks consist of multiple sizable connected components (e.g., villages, schools), rather than a single giant component. However, often the lack of edges between components is an artifact of how the network is measured. For example, Kim et al. [71] measure kinship and friendship relationship among rural villagers in Honduras, but edges between villages are not measured. On the other hand, Cai et al. [35] measure inter-village edges, but nonetheless only allow for within-village dependence when conducting statistical inference. Thus, independence remains a potentially strong assumption.

## 4 Interpretation and Additional Analyses

The simplest possible randomized experiment with a binary treatment (i.e., an “A/B test”) could be used to estimate as little as a single causal parameter of interest—the average treatment effect. In many cases researchers have found that this is an unsatisfying conclusion to a study, especially given the costs of designing, planning, and implementing<sup>7</sup> randomized field experiments. Therefore it is common for empirical researchers to conduct more extensive analysis of experimental data or to use it as input to models or simulations. We have found that the results of field experiments, though exhibiting high internal and external validity, often motivate deeper questions about the underlying mechanism and alternative counterfactual questions that can be explored through modeling or simulation.

In Sect. 3, we made assumptions about the structure of social influence and specified models or inferential procedures to detect or estimate it. In most of these experiments we are more interested in how the effect scales with number or proportion of friends who engage in a particular behavior. But beyond estimation of that response curve, there are two other broad types of questions that can be answered by experiments.

The first is treatment effect heterogeneity—the subpopulations of products, people, or social connections where social influence is stronger or weaker. By fitting more complex models researchers can estimate heterogeneous treatment effects and these estimates can help suggest causal mechanisms or guide design of marketing efforts or public policies, analogous to finding predictors of positive response to clinical treatments in medicine.

---

<sup>7</sup>One should realistically add to this list of costs, the expected cost of failure. Researchers have not always succeeded in salvaging scientific knowledge from experiments and complexity of field experiments is associated with greater risk of unexpected problems.

The second is understanding optimal policies by simulating alternative policies. Policy simulations can be used to extrapolate the results of randomized experiments to alternative policies which were never directly tested. They are most commonly used by economists who have a rich history of using structural<sup>8</sup> models in order to measure the effects of potential policy changes.

## 4.1 *Heterogeneous Treatment Effects*

One obvious type of effect heterogeneity is what clinical researchers might call a dose-response function, which characterizes how effects tend to scale as the number of friends who are influencing the person varies [46]. Bakshy et al. [20] looks at a slightly different dose-response function: when an influential social cue from a single peer is present, how does the effect size vary with the tie-strength of that individual? This heterogeneity is important to understand because it can inform advertising strategies. For instance, knowing that close friends are far more influential than random friends, we might design a marketing campaign to encourage people to share with a small number of select friends rather than many of them.

Another common analysis is measuring how influence may be moderated by the demographic characteristics of the pair of people involved. For instance, Aral and Walker [7] observes pairwise demographic attributes of the message sender and recipient and uses this information to measure how the relative effectiveness of viral messages (Facebook notifications generated by app usage) varies as a means of identifying more influential or susceptible members of social networks.

It is completely plausible that the average treatment effect can be zero, yet obscure significant positive and negative treatment effects for many large subgroups that happen to cancel out. Taylor et al. [115] shows that the presence of some people's identity cues causes their content to receive higher and lower ratings than when their content is rendered anonymously. A distribution of effects that contains both positive and negative values is plausible in many social environments with fixed resources, such as status, reputation, or attention.

In all of the three aforementioned papers in this section, we would like to point out the effect heterogeneity does represent a “free” causal estimand. If the experiment is designed to measure the ATE (the average treatment effect over the population), the effect heterogeneity we measure is simply an association between certain subgroups of the experiment and differential effects. Researchers cannot make the claim that an intervention designed to move a subject from one subgroup to another would change their treatment effect. For instance in Taylor et al. [115], the experiment tests the effects of anonymization of a commenter's identity on

---

<sup>8</sup>Here we mean structural in the sense of imposing economic “structure,” meaning that assumptions about human behavior derived from theory are imposed in the models.

ratings, but is unable to answer questions about what rating a person's content would receive if she had some alternate identity. This type of counterfactual question is precisely the type of treatment effect heterogeneity that would drive policy decisions in the social advertising space. Such a finding can only be measured by a more complicated experiment which randomizes *which* identity is presented among some set of choices—a much more difficult experiment to design and implement.

We end this section with a note of warning about seeking results based on treatment effect heterogeneity. As researchers search dimensions by which treatment effects may exhibit differential effects, they may increase the rate of false discovery. Independently testing many heterogeneity on many possible dimensions, or for many subgroups of the experiment will invariably result in false positive results as one of the subgroups may be “lucky.”<sup>9</sup> There are reasonable methods to control this risk while still detecting interesting heterogeneity, see [14] for a detailed discussion and recent methodological development.

## 4.2 Policy Simulations

Given the obvious importance of experiments for effective policy decisions, it is natural to ask for a policy recommendation at the conclusion of a study based on a randomized experiment. Often a policy recommendation is not directly recoverable from causal quantities of interest. For instance, the average treatment effect (ATE) might tell you that the treatment has a positive effect on some outcome *on average*, but it does not necessarily follow that everyone should receive the treatment. Treatments have costs which might need to be weighed and nature of social spillovers means that treating a friend of an individual can be a substitute or a complement for treating that individual directly.

Ryan and Tucker [102] reports policy simulations, employing models containing economic structure based on assumptions about individual behavior in the presence of peer effects [59]. The key idea behind the policy simulation approach is that the experiment is used to estimate parameters of the model and then the model can be used to extrapolate the findings to more complex or interesting policies than those randomly set in the original data set. Policy simulations are often used in conjunction with natural experiment, where the researcher did not *ex ante* decide the most informative randomization and would like to answer some additional questions at the cost of imposing additional assumptions through a model.

Another example of reporting policy simulations is Aral et al. [10], which applies experimental estimates from an experiment [7] to form the basis of an optimal seeding strategy in networks. As a key feature, their experiment estimated the degree to which influence and susceptibility to influence were clustered in the network, which is an important feature for understanding diffusion processes.

---

<sup>9</sup>See for example: <https://xkcd.com/882/>.

## 5 Conclusion

We believe that credible causal inference is an important goal in the practice of social science. There is obvious utility in knowing causal structure—good policy decisions *require* that the policy-maker at least know the sign of a causal effect. But also from a purely scientific perspective, measurements which do not have causal interpretation lack usefulness and insight because they afford multiple explanations. Correlations are interesting, but they usually cannot uniquely identify an explanation for a social phenomenon.

We admit there is perhaps a bit of experimental dogma present in the social sciences and it is often possible to satisfactorily answer questions through some combination of reasonable assumptions, models, and observational data. However, the realm of social influence is one where alternative explanations are difficult to rule out without some exogenous variation which can identify causal effects [50]. Manzi [79] refers to this problematic aspect of human behavior as “high causal density.” In domains of high causal density, where there are highly dense causal graphs that can explain the observed associations in data we collect, credible causal claims often require randomization, either by the researcher or by nature.

**Acknowledgements** We would like to thank Lada Adamic, Norberto Andrade, Eytan Bakshy, and George Berry for helpful discussions in preparing this review. Disclosures: D.E. was previously an employee and contractor of Facebook; D.E. has a significant financial interest in Facebook.

## References

1. Anagnostopoulos A, Kumar R, Mahdian M (2008) Influence and correlation in social networks. In: Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 7–15
2. Angelucci M, De Giorgi G (2009) Indirect effects of an aid program: how do cash transfers affect ineligibles’ consumption? *Am Econ Rev* 99(1):486–508
3. Angrist JD (2014) The perils of peer effects. *Labour Econ* 30:98–108
4. Aplin LM, Farine DR, Morand-Ferron J, Cockburn A, Thornton A, Sheldon BC (2015) Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature* 518(7540):538–541
5. Aral S (2016) Networked experiments. In: *The oxford handbook of the economics of networks*. Oxford University Press, Oxford
6. Aral S, Walker D (2011) Creating social contagion through viral product design: a randomized trial of peer influence in networks. *Manag Sci* 57(9):1623–1639
7. Aral S, Walker D (2012) Identifying influential and susceptible members of social networks. *Science* 337(6092):337–341
8. Aral S, Walker D (2014) Tie strength, embeddedness, and social influence: a large-scale networked experiment. *Manag Sci* 60(6):1352–1370
9. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci* 106(51):21544–21549
10. Aral S, Muchnik L, Sundararajan A (2013) Engineering social contagions: optimal network seeding in the presence of homophily. *Netw Sci* 1(02):125–153

11. Aronow PM (2012) A general method for detecting interference between units in randomized experiments. *Sociol Methods Res* 41(1):3–16
12. Aronow PM, Samii C (2017) Estimating average causal effects under general interference, with application to a social network experiment. *Ann Appl Stat* 11(4):1912–1947
13. Asch SE (1955) Opinions and social pressure. Readings about the social animal. *Sci Rep* 193:17–26
14. Athey S, Imbens GW (2015) Machine learning methods for estimating heterogeneous causal effects. *Stat* 1050(5)
15. Athey S, Imbens GW (2017) The econometrics of randomized experiments. *Handb Econ Field Exp* 1:73–140
16. Athey S, Eckles D, Imbens GW (2017) Exact p-values for network interference. *J Am Stat Assoc* <https://doi.org/10.1080/01621459.2016.1241178>
17. Backstrom L, Leskovec J (2011) Supervised random walks: predicting and recommending links in social networks. In: Proceedings of the fourth ACM international conference on Web search and data mining. ACM, New York, pp 635–644
18. Baird S, Bohren JA, McIntosh C, Ozler B (2016) Optimal design of experiments in the presence of interference. working paper [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2900967](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2900967)
19. Bakshy E, Eckles D (2013) Uncertainty in online experiments with dependent data: an evaluation of bootstrap methods. In: Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, New York, pp 1303–1311
20. Bakshy E, Eckles D, Yan R, Rosenn I (2012) Social influence in social advertising: evidence from field experiments. In: Proceedings of the ACM conference on electronic commerce. ACM, New York
21. Bakshy E, Rosenn I, Marlow C, Adamic L (2012) The role of social networks in information diffusion. In: Proceedings of the 21st international conference on world wide web. ACM, WWW '12, pp 519–528
22. Bakshy E, Eckles D, Bernstein MS (2014) Designing and deploying online field experiments. In: Proceedings of the 23rd international conference on world wide web, pp 283–292
23. Bakshy E, Messing S, Adamic LA (2015) Exposure to ideologically diverse news and opinion on facebook. *Science* 348(6239):1130–1132
24. Bapna R, Umyarov A (2015) Do your online friends make you pay? A randomized field experiment on peer influence in online social networks. *Manag Sci* 61(8):1902–1920
25. Bass FM (1969) A new product growth for model consumer durables. *Manag Sci* 15(5):215–227. <https://doi.org/10.2307/2628128>
26. Basse G, Feller A (2018) Analyzing multilevel experiments in the presence of peer effects. *J Am Stat Assoc* <https://doi.org/10.1080/01621459.2017.1323641>
27. Basse GW, Airolidi EM (2015) Optimal model-assisted design of experiments for network correlated outcomes suggests new notions of network balance. arXiv preprint arXiv:150700803
28. Beaman L, BenYishay A, Magruder J, Mobarak AM (2015) Can network theory-based targeting increase technology adoption? Working paper
29. Beasley A, Mason W (2015) Emotional states vs. emotional words in social media. In: Proceedings of the ACM web science conference. ACM, New York, p 31
30. Berry G, Taylor SJ (2017) Discussion quality diffuses in the digital public square. In: Proceedings of the 26th international conference on world wide web, international world wide web conferences steering committee, pp 1371–1380
31. Bond RM, Fariss CJ, Jones JJ, Kramer AD, Marlow C, Settle JE, Fowler JH (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415):295–298
32. Bond RM, Settle JE, Fariss CJ, Jones JJ, Fowler JH (2016) Social endorsement cues and political participation. *Pol Commun* 34:1–21
33. Bowers J, Fredrickson MM, Panagopoulos C (2013) Reasoning about interference between units: a general framework. *Polit Anal* 21(1):97–124

34. Bowers J, Fredrickson MM, Aronow PM (2016) A more powerful test statistic for reasoning about interference between units. *Polit Anal* 24:mpw018
35. Cai J, De Janvry A, Sadoulet E (2015) Social networks and the decision to insure. *Am Econ J Appl Econ* 7(2):81–108
36. Cameron AC, Gelbach JB, Miller DL (2011) Robust inference with multiway clustering. *J Bus Econ Stat* 29(2):238–249
37. Carrell SE, Fullerton RL, West JE (2009) Does your cohort matter? Measuring peer effects in college achievement. *J Labor Econ* 27(3):439–464
38. Carrell SE, Sacerdote BI, West JE (2011) From natural variation to optimal policy? The lucas critique meets peer effects. National Bureau of economic research working paper series no. 16865. <http://www.nber.org/papers/w16865>
39. Centola D (2010) The spread of behavior in an online social network experiment. *Science* 329(5996):1194–1197
40. Choi DS (2017) Estimation of monotone treatment effects in network experiments. *J Am Stat Assoc* 112(519):1147–1155
41. Chung E, Romano JP (2013) Exact and asymptotically robust permutation tests. *Ann Stat* 41(2):484–507
42. Coey D, Bailey M (2016) People and cookies: imperfect treatment assignment in online experiments. In: Proceedings of the 25th international conference on world wide web, international world wide web conferences steering committee, pp 1103–1111
43. Coppock A, Guess A, Ternovski J (2015) When treatments are tweets: a network mobilization experiment over twitter. *Polit Behav* 38(1):1–24
44. Curranini S, Jackson MO, Pin P (2010) Identifying the roles of race-based choice and chance in high school friendship network formation. *Proc Natl Acad Sci* 107(11):4857–4861
45. De Choudhury M, Mason WA, Hofman JM, Watts DJ (2010) Inferring relevant social networks from interpersonal communication. In: Proceedings of the 19th international conference on world wide web. ACM, New York, pp 301–310
46. DeLean A, Munson P, Rodbard D (1978) Simultaneous analysis of families of sigmoidal curves: application to bioassay, radioligand assay, and physiological dose-response curves. *Am J Physiol Gastrointest Liver Physiol* 235(2):G97–102
47. Ditttrich D, Kenneally E et al (2012) The Menlo report: ethical principles guiding information and communication technology research. US Department of Homeland Security
48. Duflo E, Saez E (2003) The role of information and social interactions in retirement plan decisions: evidence from a randomized experiment. *Q J Econ* 118(3):815–842
49. Eckles D, Kizilcec RF, Bakshy E (2016) Estimating peer effects in networks with peer encouragement designs. *Proc Natl Acad Sci* 113(27):7316–7322
50. Eckles D, Bakshy, E (2017) Bias and high-dimensional adjustment in observational studies of peer effects, working paper. <https://arxiv.org/abs/1706.04692>
51. Eckles D, Karrer B, Ugander J (2017) Design and analysis of experiments in networks: reducing bias from interference. *J Causal Inference*. <https://doi.org/doi.org/10.1515/jci-2015-0021>
52. Firth JA, Sheldon BC, Farine DR (2016) Pathways of information transmission among wild songbirds follow experimentally imposed changes in social foraging structure. *Biol Lett* 12(6):20160,144
53. Fisher RA (1935) The design of experiments. Oliver and Boyd, Edinburgh
54. Forastiere L, Mealli F, VanderWeele TJ (2015) Identification and estimation of causal mechanisms in clustered encouragement designs: disentangling bed nets using Bayesian principal stratification. *J Am Stat Assoc* 111(514):510–525
55. Gerber AS, Green DP (2012) Field experiments: design, analysis, and interpretation. WW Norton, New York
56. Granovetter M (1978) Threshold models of collective behavior. *Am J Sociol* 83(6): 1420–1443
57. Halberstam Y, Knight B (2016) Homophily, group size, and the diffusion of political information in social networks: evidence from twitter. *J Public Econ* 143:73–88

58. Halloran ME, Hudgens MG (2016) Dependent happenings: a recent methodological review. *Curr Epidemiol Rep* 3(4):297–305
59. Hartmann WR (2010) Demand estimation with social interactions and the implications for targeted marketing. *Mark Sci* 29(4):585–601
60. Hasan S, Koning R (2017) Conversational peers and idea generation: evidence from a field experiment Stanford University Graduate School of Business Research Paper No. 17–36. <http://dx.doi.org/10.2139/ssrn.2964214>
61. Higgins MJ, Sävje F, Sekhon JS (2016) Improving massive experiments with threshold blocking. *Proc Natl Acad Sci* 113(27):7369–7376
62. Huber GA, Malhotra N (2017) Political homophily in social relationships: evidence from online dating behavior. *J Polit* 79(1):269–283
63. Imbens GW, Rosenbaum PR (2005) Robust, accurate confidence intervals with a weak instrument: quarter of birth and education. *J R Stat Soc Ser A (Stat Soc)* 168(1):109–126
64. Imbens GW, Rubin DB (2015) Causal inference in statistics, social, and biomedical sciences. Cambridge University Press, Cambridge
65. Jackman M, Kamerva L (2016) Evolving the IRB: building robust review for industry research. *Washington and Lee Law Rev Online* 72(3):442
66. Jones JJ, Settle JE, Bond RM, Fariss CJ, Marlow C, Fowler JH (2013) Inferring tie strength from online directed behavior. *PLoS one* 8(1):e52168
67. Jones JJ, Bond RM, Bakshy E, Eckles D, Fowler JH (2017) Social influence and political mobilization: further evidence from a randomized experiment in the 2012 US presidential election. *PLoS One* 12(4):e0173851
68. Karrer B, Newman ME (2011) Stochastic blockmodels and community structure in networks. *Phys Rev E* 83(1):016107
69. Kearns M, Suri S, Montfort N (2006) An experimental study of the coloring problem on human subject networks. *Science* 313(5788):824–827
70. Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 137–146
71. Kim DA, Hwong AR, Stafford D, Hughes DA, O’Malley AJ, Fowler JH, Christakis NA (2015) Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. *Lancet* 386(9989):145–153
72. Kohavi R, Deng A, Frasca B, Longbotham R, Walker T, Xu Y (2012) Trustworthy online controlled experiments: five puzzling outcomes explained. In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 786–794
73. Kossinets G, Watts DJ (2006) Empirical analysis of an evolving social network. *Science* 311(5757):88–90
74. Kramer AD, Guillory JE, Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci* 111(24):8788–8790
75. Leider S, Möbius MM, Rosenblat T, Do QA (2009) Directed altruism and enforced reciprocity in social networks. *Q J Econ* 124(4):1815–1851
76. Leung MP (2017) A weak law for moments of pairwise-stable networks, working paper. <http://dx.doi.org/10.2139/ssrn.2663685>
77. Manski CF (1993) Identification of endogenous social effects: the reflection problem. *Rev Econ Stud* 60(3):531–542
78. Manski CF (2000) Economic analysis of social interactions. Technical report, National bureau of economic research
79. Manzi J (2012) Uncontrolled: the surprising payoff of trial-and-error for business, politics, and society. Basic Books, New York
80. Marmaros D, Sacerdote B (2006) How do friendships form? *Q J Econ* 121(1):79–119
81. Mason W, Suri S (2012) Conducting behavioral research on amazon’s mechanical turk. *Behav Res Methods* 44(1):1–23

82. Mason W, Watts DJ (2012) Collaborative learning in networks. *Proc Natl Acad Sci* 109(3):764–769
83. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. *Annu Rev Sociol* 27(1):415–444
84. Miguel E, Kremer M (2004) Worms: identifying impacts on education and health in the presence of treatment externalities. *Econometrica* 72(1):159–217
85. Miratrix LW, Sekhon JS, Yu B (2013) Adjusting treatment effect estimates by post-stratification in randomized experiments. *J R Stat Soc Ser B (Stat Methodol)* 75(2):369–396
86. Moffitt RA et al (2001) Policy interventions, low-level equilibria, and social interactions. *Soc Dyn* 4:45–82
87. Morris S (2000) Contagion. *Rev Econ Stud* 67(1):57
88. Muchnik L, Aral S, Taylor SJ (2013) Social influence bias: a randomized experiment. *Science* 341(6146):647–651
89. Narayanan A, Shmatikov V (2010) Myths and fallacies of personally identifiable information. *Commun ACM* 53(6):24–26
90. Nickerson DW (2008) Is voting contagious? Evidence from two field experiments. *Am Polit Sci. Rev* 102(01):49–57
91. Ohm P (2010) Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA L Rev* 57:1701–1819
92. Owen AB, Eckles D (2012) Bootstrapping data arrays of arbitrary order. *Ann Appl Stat* 6(3):895–927
93. Paluck EL, Shepherd H, Aronow PM (2016) Changing climates of conflict: a social network experiment in 56 schools. *Proc Natl Acad Sci* 113(3):566–571
94. Rand DG, Arbesman S, Christakis NA (2011) Dynamic social networks promote cooperation in experiments with humans. *Proc Natl Acad Sci* 108(48):19193–19198
95. Rand DG, Nowak MA, Fowler JH, Christakis NA (2014) Static network structure can stabilize human cooperation. *Proc Natl Acad Sci* 111(48):17093–17098
96. Resea B, Ryan KJP (1978) The Belmont Report: ethical principles and guidelines for the protection of human subjects of research—the national commission for the protection of human subjects of biomedical and behavioral research. US Government Printing Office
97. Resnick MD, Bearman PS, Blum RW, Bauman KE, Harris KM, Jones J, Tabor J, Beuhring T, Sieving RE, Shew M et al (1997) Protecting adolescents from harm: findings from the national longitudinal study on adolescent health. *J Am Math Assoc* 278(10):823–832
98. Riach PA, Rich J (2004) Deceptive field experiments of discrimination: are they ethical? *Kyklos* 57(3):457–470
99. Rock D, Aral S, Taylor SJ (2016) Identification of peer effects in networked panel data. In: Thirty seventh international conference on information systems
100. Rosenbaum PR (1984) Conditional permutation tests and the propensity score in observational studies. *J Am Stat Assoc* 79(387):565–574
101. Ryan B, Gross NC (1943) The diffusion of hybrid seed corn in two Iowa communities. *Rural Soc* 8(1):15
102. Ryan SP, Tucker C (2012) Heterogeneity and the dynamics of technology adoption. *Quant Mark Econ* 10(1):63–109
103. Sacerdote B (2001) Peer effects with random assignment: results for dartmouth roommates. *Q J Econ* 116(2):681–704
104. Salganik MJ (2017) Bit by bit: social research in the digital age. Princeton University Press, Princeton
105. Saul BC, Hudgens MG, Halloran ME (2017) Causal inference in the study of infectious disease. In: Handbook of statistics. Elsevier, Amsterdam
106. Saveski M, Pouget-Abadie J, Saint-Jacques G, Duan W, Ghosh S, Xu Y, Airolidi EM (2017) Detecting network effects: Randomizing over randomized experiments. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp 1027–1035
107. Schelling T (1969) Models of segregation. *Am Econ Rev* 59(2):488–493

108. Schelling T (1971) Dynamic models of segregation. *J Math Sociol* 1(2):143–186
109. Schelling T (1973) Hockey helmets, concealed weapons, and daylight saving: a study of binary choices with externalities. *J Confl Resolut* 17(3):381–428
110. Schultz AP, Piepgrass B, Weng CC, Ferrante D, Verma D, Martinazzi P, Alison T, Mao Z (2012) Methods and systems for determining use and content of PYMK based on value model. US Patent App. 13/659,695
111. Shalita A, Karrer B, Kabiljo I, Sharma A, Presta A, Adcock A, Klap H, Stumm M (2016) Social hash: an assignment framework for optimizing distributed systems operations on social networks. In: NSDI, pp 455–468
112. Shalizi CR, Thomas AC (2011) Homophily and contagion are generically confounded in observational social network studies. *Sociol Methods Res* 40(2):211–239
113. Suri S, Watts DJ (2011) Cooperation and contagion in web-based, networked public goods experiments. *PloS one* 6(3):e16836
114. Taylor SJ, Bakshy E, Aral S (2013) Selection effects in online sharing: consequences for peer adoption. In: Proceedings of the fourteenth ACM conference on electronic commerce. ACM, New York, pp 821–836
115. Taylor SJ, Muchnik L, Aral S (2014) Identity and opinion: a randomized experiment, working paper. <http://dx.doi.org/10.2139/ssrn.2538130>
116. Tchetgen EJT, Fulcher I, Shpitser I (2017) Auto-G-computation of causal effects on a network. arXiv preprint arXiv:170901577
117. Toulis P, Kao E (2013) Estimation of causal peer influence effects. In: Proceedings of the 30th international conference on machine learning, pp 1489–1497
118. Ugander J, Karrer B, Backstrom L, Kleinberg JM (2013) Graph cluster randomization: network exposure to multiple universes. In: Proceedings of KDD, ACM, New York
119. Valente TW (1996) Social network thresholds in the diffusion of innovations. *Soc Netw* 18(1):69–89
120. van der Laan MJ (2014) Causal inference for a population of causally connected units. *J Causal Inference* 2:1–6
121. VanderWeele TJ, Tchetgen EJT, Halloran ME (2012) Components of the indirect effect in vaccine trials: identification of contagion and infectiousness effects. *Epidemiology* 23(5):751
122. Walker D, Muchnik L (2014) Design of randomized experiments in networks. *Proc IEEE* 102(12):1940–1951
123. Xu Y, Chen N, Fernandez A, Sinno O, Bhasin A (2015) From infrastructure to culture: A/B testing challenges in large scale social networks. In: Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 2227–2236

# The Rippling Effect of Social Influence via Phone Communication Network



Yan Leng, Xiaowen Dong, Esteban Moro, and Alex ‘Sandy’ Pentland

## 1 Introduction

We live in a connected world and are increasingly closer to each other thanks to the emerging information technologies. While the “small-world” phenomenon and the “six degrees of separation” have been traditionally studied by Milgram [14] and Watts [24], a recent research suggests that the average degree of separation between two members of the online social network Facebook is reduced to around 4.74 [5]. Furthermore, individuals are not merely connected; as a series of experiments in various domains such as obesity, happiness, cooperation, and political opinions has demonstrated, connectivity also indicates behavioral similarities of up to three degrees of separation [7, 25].

The recent availability of large-scale communication and networked data, such as emails, mobile phone records, and online social media activities, enables the studies of information diffusion and correlations of adoption behaviors as well as social contagion processes at an unprecedented scale [8, 15, 17]. In particular, the understanding of the phenomenon of and the mechanism that drives the social contagion process help promote behavioral change in domains such as commerce, public health, politics, and social mobilization at both local and global scales [3, 6, 9, 23]. As examples, Aral et al. [4] focused on the diffusion of the adoptions of mobile service application using a social network connected by instant message

---

Y. Leng · X. Dong · A. ‘Sandy’ Pentland (✉)  
MIT Media Lab, Cambridge, MA, USA  
e-mail: [yleng@mit.edu](mailto:yleng@mit.edu); [xdong@mit.edu](mailto:xdong@mit.edu); [pentland@mit.edu](mailto:pentland@mit.edu)

E. Moro  
MIT Media Lab, Cambridge, MA, USA

Departamento de Matematicas & GISC, Universidad Carlos III de Madrid, Leganes, Spain  
e-mail: [emoro@mit.edu](mailto:emoro@mit.edu)

traffic [2]. Ugander et al. [23] found that the decision to join Facebook varies with the number of distinct social groups their friends occupy. Bond et al. [6] conducted a 61-million-person experiment on Facebook and found that strong ties are instrumental for spreading political behavior through online social network. Often, the connectivity and structure of the social network play a role in the effectiveness of social contagion. For instance, both Onnela et al. [17] and Ugander et al. [23] emphasized the importance of network structure in information spreading and product adoption [26].

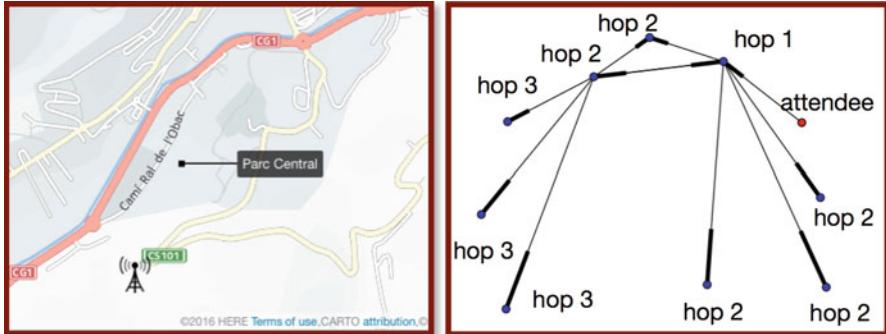
However, most of the previous works focus on online social networks, and measure influence between direct contacts concerning either long-term habits or low-cost decision-making in virtual space (such as online product adoption). In this study, we are interested in investigating how social influence propagates over a large-scale offline communication network, and how it manifests in short-time decision-making and social mobilization that are more costly than merely information diffusion or online production adoption.

We use a data set of mobile phone records with high resolution in Andorra for our analysis. We construct a large-scale communication network and mirror the contagion process of social influence, whose effect is measured by the change in the likelihood of attending a large-scale international cultural event in the capital city. In order to control for the selection bias caused by homophily and identify the causal effect of social influence, we utilize a matching method to mimic the procedure of random assignment of treatments [3, 11]. One novel aspect of our study is to condition matchings on revealed preferences, i.e., historical visitation patterns, instead of the traditionally considered demographics. Rather surprisingly, our results show that influence decays across social distance from initial attendees, but persists up to six degrees of separation, similarly to the physical phenomenon of ripples expanding across the water. Meanwhile, the patterns of communication, such as intensity and the timeliness of communication, also impact the strength of social influence, but to a lesser degree. Finally, we analyze the heterogeneous effects of social influence on the population, and observe that the effect is stronger on the geographically explorative subgroup of population.

## 2 Data and Method

Mobile phone logs have been used in various studies as a proxy for human mobility and social interactions at a societal scale [8, 21]. We leverage the detailed tracking and wide coverage of mobile phone logs in the country of Andorra to study how the likelihood of an individual attending a local Cirque Du Soleil performance, which was held repetitively in July, 2016, is affected if someone in his social circle receives phone calls directly or indirectly from past attendees of the event.

We introduce three key definitions in our study. First, we assume that people who were connected to a cell tower nearby the performance venue, as shown in the left panel of Fig. 1, during the performance hours ( $\pm 30$  min as buffer time)

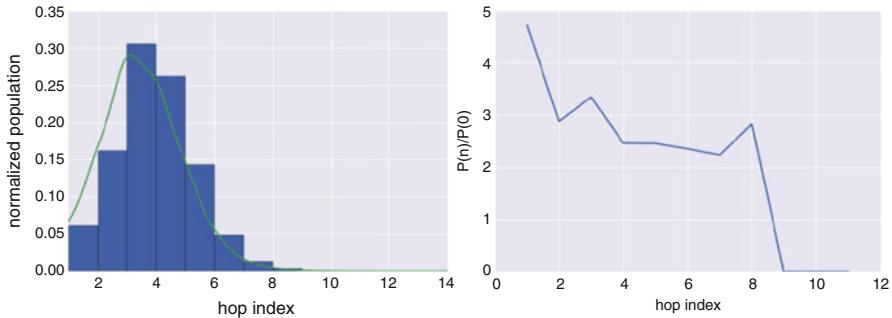


**Fig. 1** Illustration of attendee, influence cascade and hop

attended the events and are labeled as **attendees**. Next, we construct **influence cascade**, as shown in the right panel of Fig. 1, by adding links between the caller and receiver if: (1) at least one of them is linked directly or indirectly with the attendees by the time the call was initiated; (2) the calls took place within 24 h after the performance started. Finally, we use **hop** to capture the shortest social distance to any attendee via the influence cascade. Overall, we observed 16,043 attendees across the one-month observational period. Among others, the influence cascade covers 161,857 individuals. And another 71,337 population are disconnected to the influence cascade.

In order to quantify the effect of social influence in people's decision-making, the key challenge is to control for the upward estimation bias caused by homophily. We use matched sample estimation to mimic the assignment of treatment as in a randomized experiment, rather than regression analysis which only establishes correlations [3, 10, 11]. More specifically, for the influence cascade constructed for each day, we consider a treatment group in which individuals are of certain social distance from the attendees (we use treatment group on hop  $h$  to represent people that are  $h$ -degree of separation from the closest attendee), and a control group in which individuals are not connected to any attendee on that day. Individuals in treatment and control groups are matched to control group on a one-to-one basis based on their mobility patterns, which we will further explain in more detail in later section.

Before establishing causal studies, we first analyze the distribution of social distances of individuals to the attendees. As shown in Fig. 2, a large mass of population are three and four degrees of separation from the attendees. Moreover, we analyze the predictive power of the degree of separation from the attendees for attendance rate. We compare the attendance ratio between treatment group on hop  $h$  and control group. The larger-than-one ratio comes from a mixed effects of homophily, social influence and other confounding variables. The right panel of Fig. 2 shows the ratio between the likelihood of attending the social event of people on hop  $h$  and those who receive no treatment. As we see, direct contacts of the



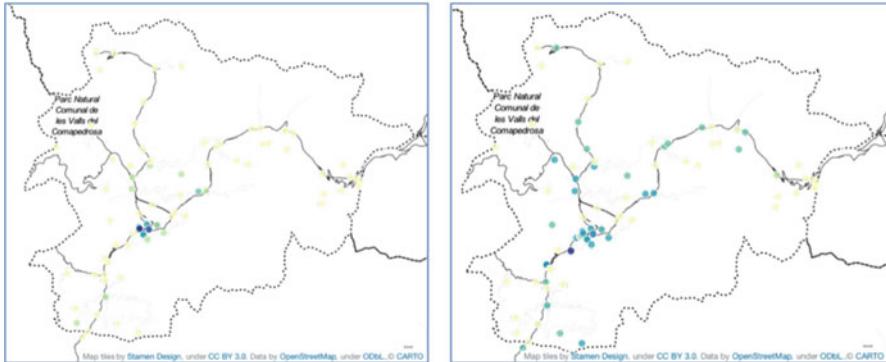
**Fig. 2** Distribution of social distances of individuals to the attendees (left) and the attendance ratio between treatment group on hop  $h$  and control group (right)

attendees are five times more likely to attend the performance than individuals who do not receive treatment. Meanwhile, individuals on hop six are 2.5 times more likely to attend the events than individuals receiving no treatment. The average decreasing trend of the likelihood indicates that the degree of separation from the attendees is an important factor in studying the likelihood of attending the performance. However, this correlation does not indicate causality, the latter of which is the main focus of our study.

## 2.1 Controlling for Homophily

It is widely argued that the adoption behavior in the social network (the decision of attending the event in our case) is a mixture of similarities over friends and contagion driven by social influence [3, 6, 22]. Similarities among peers may cause the over-estimation of social influence [3]. Therefore, we need to balance the distribution of similarities across individuals in the influence cascades and isolate the causal effect of word-of-mouth influence through phone calls in our observational study.

Empowered by the longitudinal and detailed mobility tracking via Call Detail Records, we use behavioral patterns to characterize individuals instead of the widely applied method of demographic characterization [3, 6]. The power of behavioral characterization as a control for homophily is that behavior reveals preferences regarding the same type activities that we are observing and treating [13], which is exactly what we want to control for. Specifically in our case, activities performed during their leisure time, the revealed visitation preferences, are captured via cell tower visitation frequencies over the weekend for the past 6 months [12]. As shown in Fig. 3, the left panel represents an individual who spends most of the weekends in the crowded shopping districts while the right panel stands for an individual with a diversified activity patterns.



**Fig. 3** Two examples of historical mobility patterns during the weekends for the past 6 months

## 2.2 Matching

As stated before, we use matched sample estimation to yield the estimates of social influence by conditioning matches on mobility frequency vectors. The matching results establish an upper bound to which extent social influence, rather than homophily, explains the attendance behavior<sup>1</sup> [3].

We segment individuals into two groups, the treatment group and the control group, based on whether they receive influence related to the event or not. Treatment groups are further split into eight subgroups according to the hop index. The control group consists of individuals who are disconnected to the influence cascades. Each individual in the treatment group is paired with another individual in the control group that is most similar in terms of preferences approximated by mobility patterns. By such a matching, we ensure that the main difference between the two individuals paired together is whether or not one receives the treatment of social influence [20]. The matchings depend on nearest Mahalanobis distance calculated as:

$$\text{md}(X_j, X_k) = [(X_j - X_k)^T S^{-1} (X_j - X_k)]^{1/2}, \quad (1)$$

where  $X_j$  and  $X_k$  are the covariate vectors (mobility frequency vectors) for individual  $j$  and individual  $k$ , and  $S$  is the sample covariance matrix for the mobility frequency matrix  $X$ .

We perform Principal Component Analysis on  $X$  to reduce the correlations of the visitation patterns among nearby cell towers and to reduce the number of variables used in matching. Dimension reduction is important in Mahalanobis Distance Matching, which works better in balancing fewer covariates [11].

<sup>1</sup>Unobserved confounding variables are difficult to control for by using matching-based methods. To partly address the issue that tourists may travel together and social links may not pass social influence, we remove individual pairs who are potentially on the same trip to Andorra. This can be inferred based on whether individuals stay at the same hotel at the same night.

In our setting, the difference in the attendance rate of the two groups is the average treatment effect of social influence:

$$\text{ATE}_h = E(Y_{ih} - Y_{ic}), \quad (2)$$

where  $\text{ATE}_h$  is the average treatment effect of treatment group on hop  $h$ ,  $Y_{ih}$  is the outcome for matched pair  $i$  in treatment group  $h$ , and  $Y_{ic}$  is the outcome for matched pair  $i$  in control group.

### 3 Results

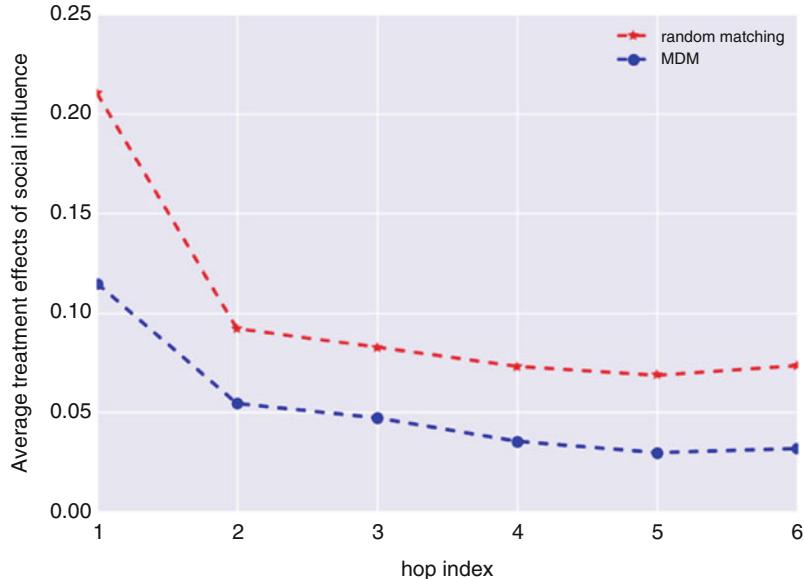
In this section, we first investigate the effect of social influence after distinguishing it from homophily using Mahalanobis Distance Matching. To evaluate the inflation bias caused by homophily, we compare our results with random matching, where we do not control for homophilous behavior and pair individuals randomly. Furthermore, we quantify both external and internal factors that affect the strength of social influence, namely, the patterns of the communications and the characteristics of the individuals.

#### 3.1 *The Decay of Influence over Social Distance from Attendee*

After distinguishing homophily and social influence, we are able to estimate the treatment effect of social influence on the likelihood of attendance. In Fig. 4, the blue-dashed line shows the average treatment effect of social influence (as in  $y$ -axis) across hops (as in  $x$ -axis). The positive treatment effects—the increasing likelihood of attending a future performance—indicate that social influence promotes the likelihood of attending the performance. More importantly, we discover a “ripple effect” of social influence over communication network: originating from the attendees and expanding across information cascade. In particular, this effect decays across social distances from the attendees and persists up to six degrees of separation. The average treatment effect of social influence is 11% on the first hop and drops dramatically to a half at the second hop. Starting from the third hop, the treatment effects decay slowly and persist until the sixth hops.

The difference between the red-dashed line and the blue-dashed line in Fig. 4 shows the overestimation of social influence without controlling for homophily. In particular, with random matching, we overestimate the effect of social influence by around 100%, which is similar to the findings in a previous study by Aral (2009) on the adoption of an online application [4].

Furthermore, we use “random shuffling” proposed by Anagnostopoulos [1] to exclude the concern that other mechanical reasons might cause the decay pattern in



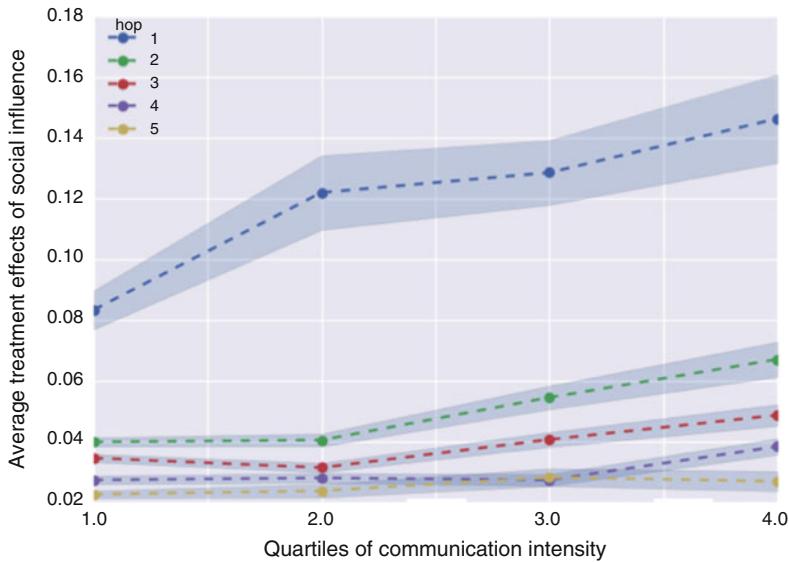
**Fig. 4** Average treatment effect of social influence via communication network

social network. We first randomly assign people to control and treatment group, as well as the hop index if assigned to treatment group, and then measure the average treatment effect with Mahalanobis Distance Matching. The average treatment effect as well as the decay pattern disappear.

In order for the estimation of treatment effects from matching results to be robust, the assignment of treatment, conditional on the Mahalanobis distance, need to be as good as randomly assigned. In other words, the covariates are required to be balanced between matched pairs in treatment and control groups. Therefore, we use standardized mean differences (SMD) to evaluate whether the covariates in the two groups demonstrate sufficient overlap [16]. SMD is calculated as the difference of means in units of pooled standard deviation as follows:

$$\text{SMD} = \frac{\bar{x}_{l,h} - \bar{x}_{l,c}}{\sqrt{(s_{l,h}^2 + s_{l,c}^2)/2}}, \quad (3)$$

where  $\bar{x}_{l,h}$  and  $\bar{x}_{l,c}$  are the means of covariate  $x_l$  for treatment group  $h$  and control group, respectively, and  $s_{l,h}$  and  $s_{l,c}$  are the standard deviation of covariate  $x_l$  for treatment group  $h$  and control group, respectively. We run the covariates balanced test and show that all of the SMDs are far below 0.1, which rejects the hypothesis that they have insufficient overlap.



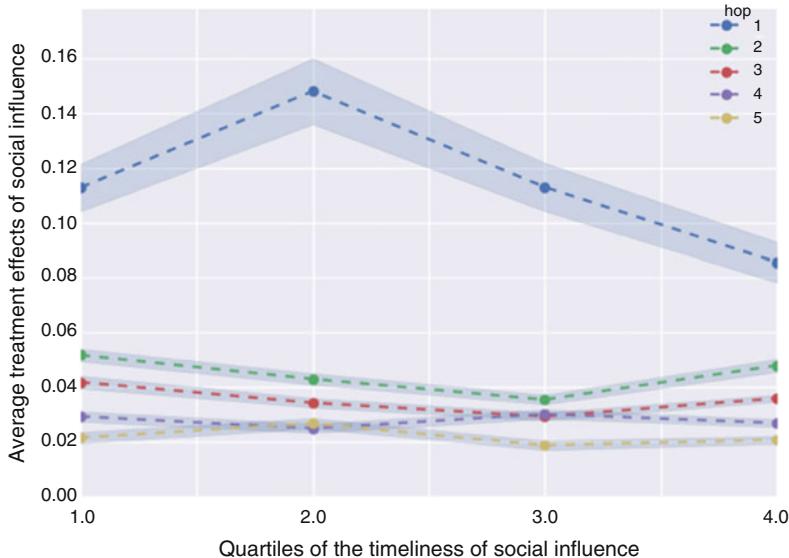
**Fig. 5** Average treatment effect of social influence with respect to intensity of communications. Different colors shown in the legend represent different hop indexes as labeled

### 3.2 Communication Patterns

In this section, we test the hypothesis that social influence and contagion process on the social network may vary according to the communication patterns. As shown in Fig. 5, more intense communications between two individuals indicates a larger treatment effect for the first three hops and stay constant afterwards. In terms of the timeliness of communication, we show in Fig. 6 that the treatment effects are significantly stronger if the calls are made immediately after the event. Similarly to intensity, this only holds up to hop three. These two empirical exercises indicate that communication patterns exert quantifiable and discernible effects on the strength of social influence up to three degrees of separation.

## 4 Discussion

In this study, we illustrate the application of a matching strategy in a large-population study to identify the effect of social influence. A novel aspect of our study is the use of matched samples as determined by previously observed behavior instead of those obtained by Randomized Control Trials (RCTs), which seems potentially quite useful in many large-scale studies. By analyzing the pattern of attendance of an international cultural event in Andorra using large-scale mobile



**Fig. 6** Average treatment effect of social influence with respect to timeliness of communication. Different colors shown in the legend stand for different hop indexes as labeled in the legend

phone data, we quantify how our decision-makings are influenced by, and how the social network propagates our influence to, people that are several degrees away from us in the communication network with matched and balanced samples.

Our results reveal the subtle and often invisible effect of social influence on decision-making via phone communication network, which, surprisingly, persists up to six degrees of separation. This is analogous to the physical phenomenon of ripples expanding across the water, which highlights the hidden relationship and connections among people in the society. More interestingly, we show that such effect is significantly larger when phone communication took place immediately after the event and lasted longer, and when those receiving calls are more explorative geographically as indicated by a more diverse mobility pattern.

The ripple effect via phone communications demonstrated through our study has far-reaching implications in domains such as viral marketing, public health, and social mobilization. Recent works have demonstrated the success of social mobilization via Internet-based services [18], but also shown that such mechanisms are not without limitations [19]. Our findings suggest that an alternative would be to exploit the hidden and often overlooked influence between people that are caused by chains of offline communication. The same strategy may also be applied into marketing or political campaigns. Our results on the impact of communication pattern and mobility pattern of individuals on the strength of influence can also help design more effective strategies to maximize social influence.

Our work also opens new possibilities in understanding social influence and contagion, in terms of both mathematical modeling and experiment-based studies.

In the context of networks, threshold-based contagion models and epidemic models have largely explored the direct interaction between neighboring nodes in the network, where the behavior of a given node is dependent on its interactions with neighboring nodes. Hidden interactions across several degrees of separation could be naturally incorporated into such models. For example, we could systematically model the treatment effect and the adoption behavior of a given node as a function of degrees of separation, as well as other network characteristics. With better models on contagion processes, we could perform counter-factual simulations over different intervention strategies to incentivize key individuals and maximize social influence for behavioral change.

It is worth noting that our study also has certain limitations. First, given that we do not have the actual records for attendance of the event, we consider people who had phone activities at cell towers close to the venue as attendees. This strategy might, therefore, have included people who just passed by the venue without actually attending the event. Second, due to the lack of demographic information, we approximate homophily in a social network by looking at the mobility history of individuals. While it is reasonable to assume that mobility patterns reflect to some extent characteristics and interests of different people, it may also make people with different demographics much more similar. Third, in the current framework, we define social distance as the length of the shortest path between an individual and the attendees, thus effectively considering only this “strongest treatment” in estimating the treatment effect. There might be a multiplicative effect in the case of more than one communication path (hence the possibility of multiple treatments), which may require slightly more complex modeling of influence. We leave such analysis for future work.

## References

1. Anagnostopoulos A, Kumar R, Mahdian M (2008) Influence and correlation in social networks. In: 14th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 7–15
2. Aral S (2011) Commentary-identifying social influence: a comment on opinion leadership and social contagion in new product diffusion. *Mark Sci* 30(2):217–223
3. Aral S (2012) Social science: poked to vote. *Nature* 489(7415):212–214
4. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci* 106(51):21544–21549
5. Backstrom L, Boldi P, Rosa M, Ugander J, Vigna S (2012) Four degrees of separation in Proceedings of the 4th annual ACM web science conference. ACM, New York, pp 33–42
6. Bond RM et al (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415):295–298.
7. Christakis N, Fowler J (2009) Connected: the surprising power of our social networks and how they shape our lives. Little Brown and Company, New York
8. Deville P et al (2016) Scaling identity connects human mobility and social interactions. *Proc Natl Acad Sci* 113(26):201525443
9. Gleeson JP, Cellai D, Onnela JP, Porter MA, Reed-Tsochas F (2014) A simple generative model of collective online behavior. *Proc Natl Acad Sci* 111(29):10411–10415

10. Hill S, Provost F, Volinsky C (2006) Network-based marketing: Identifying likely adopters via consumer networks. *Stat Sci* 21(2):256–276
11. King G, Nielsen R (2016) Why propensity scores should not be used for matching
12. Leng Y, Rudolph L, Pentland A, Zhao J, Koutsopoulos HN (2016) Managing travel demand: location recommendation for system efficiency based on mobile phone data. *CoRR* abs/1610.06825
13. Lobel I, Sadler E (2015) Preferences, homophily, and social learning. *Oper Res* 64(3):564–584
14. Milgram S (1967) The small world problem. *Psychol Today* 1(1):61–67
15. Miritello G, Moro E, Lara R (2011) Dynamical strength of social ties in information spreading. *Phys Rev E* 83(4):045102
16. Normand SLT et al (2001) Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. *J Clin Epidemiol* 54(4):387–398
17. Onnela JP et al (2007) Structure and tie strengths in mobile communication networks. *Proc Natl Acad Sci* 104(18):7332–7336
18. Pickard G et al (2011) Time-critical social mobilization. *Science* 334:509–512
19. Rutherford A et al (2013) Limits of social mobilization. *Proc Natl Acad Sci* 110(16):6281–6286
20. Stuart EA (2010) Matching methods for causal inference: a review and a look forward. *Stat Sci Rev J Inst Math Stat* 25(1):1
21. Toole JL, Herrera-Yaque C, Schneider CM, González MC (2015) Coupling human mobility and social ties. *J R Soc Interface* 12(105):20141128
22. Toulis P, Kao EK (2013) Estimation of causal peer influence effects. *ICML* 28(3):1489–1497
23. Ugander J, Backstrom L, Marlow C, Kleinberg J (2012) Structural diversity in social contagion. *Proc Natl Acad Sci* 109(16):5962–5966
24. Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393(6684):440–442
25. VanderWeele TJ (2013) Inference for influence over multiple degrees of separation on a social network. *Stat Med* 32(4):591–596
26. Zhao J, Wu J, Xu K (2010) Weak ties: Subtle role of information diffusion in online social networks. *Phys Rev E* 82(1):016105

# Network Experiments Through Academic-Industry Collaboration



Robert M. Bond, Christopher J. Fariss, Jason J. Jones, and Jaime E. Settle

## 1 Introduction

The study of social contagion and the spread of behaviors across social networks is a vibrant and dynamic field of scientific inquiry.<sup>1</sup> Massive scale datasets that are collected unobtrusively (i.e., without users noticing they are being observed) and in real time are being compiled by a large number of private companies and open up new insights into social contagion and the spread of different behaviors (e.g., [7, 15, 16, 55]). The use of services that collect such data, such as social media websites, is now so commonplace that there is very little debate about the generalizability from studies conducted on samples drawn from Facebook or other social networking platforms.

---

<sup>1</sup>The examples from this literature utilize both observational and experimental research designs but are too numerous to cite completely. Prominent examples from this literature include [2, 8, 9, 12, 16, 21–29, 36–38, 42, 44, 45, 52, 54, 55, 58–60, 62, 63], among many others.

R. M. Bond (✉)  
The Ohio State University, Columbus, OH, USA  
e-mail: [bond.136@osu.edu](mailto:bond.136@osu.edu)

C. J. Fariss  
University of Michigan, Ann Arbor, MI, USA  
e-mail: [cjfariss@umich.edu](mailto:cjfariss@umich.edu)

J. J. Jones  
State University of New York, Stony Brook, Stony Brook, NY, USA  
e-mail: [jason.j.jones@stonybrook.edu](mailto:jason.j.jones@stonybrook.edu)

J. E. Settle  
William & Mary, Williamsburg, VA, USA  
e-mail: [jsettle@wm.edu](mailto:jsettle@wm.edu)

Unfortunately, the advance of scientific knowledge has had difficulty keeping pace with the rapid development of new forms of behavior on these platforms. Social networking site companies like Facebook and Twitter strive to optimize their interfaces for user interaction and the flow of information, resulting in a near constant evolution of the particular affordances of a given site. Social scientists struggle to keep up: as the features on the site are developed, new forms of social behavior arise. What we might infer about users' offline behavior and attitudes is in flux as the form of the online data and users' online behavior co-evolve. The challenge is not only to theorize the etiology and consequences of these behaviors, but also to derive measures of theoretically important behaviors and attitudes through them while also keeping in mind the processes that give rise to the behaviors.

Herein lies the chief obstacle in this vein of research: the difficulty in developing a robust and rigorous science of social contagion in online environments, as well as the relationship between online and offline social behaviors, is the accessibility of the massive social network data itself. Why? Because massive online social media and behavioral data are not publicly available and only a handful of researchers have access to it. Over the last few years, the private companies that collect and store these data are placing increasingly stringent limitations over access and use. Moreover, the types of research questions that these gatekeepers are comfortable publicly addressing has changed as these for-profit companies face public criticism for their participation in some of the more controversial online experimental studies (e.g., [43]) that have been published to date (for additional commentary about this controversy, see [40, 48]). Perhaps due to the negative backlash against these studies, experimental research conducted internally is frequently no longer being publicly discussed by the research teams working with these companies.

This "behind closed door" mentality is not only a loss for the scientific community, but it also raises ethical concerns about the production of knowledge itself. The research at these firms goes on, but it often does not enter the public domain, adhere to the standards of university institutional review boards, or participate in a transparent and replicable peer review process.

We argue that instead, companies should embrace the opportunities for innovative massive scale behavioral research and that researchers should work together to make inroads with the owners and curators of existing massive scale databases. These types of collaborations offer opportunities that are mutually beneficial. Scholars gain access to large samples with collections of data that otherwise would be impossible to efficiently collect. Companies gain access to the expertise of scholars who are at the forefront of cutting-edge methodological and theoretical research. By working together, companies and academics have the potential to advance our scientific understanding of our complex world and answer questions that are pertinent to the company's goals. This is the promise of big data.

In this essay, we describe some experimental research that we successfully developed, implemented, and publicly disseminated with the Data Science Team at Facebook. Our relationship with this team was highly productive and positive over a 6-year period beginning in the summer of 2010, just a few months prior

to the 2010 U.S. midterm elections. Since that time, we along with many other social scientists working at the University of California, San Diego, have published a number of important observational and experimental studies based on this academic-industry collaboration (e.g., [7–9, 15, 26, 27, 36–38, 55]). We hope that what was a positive experience for us might, in the future, again serve as a model for future collaborations with for-profit companies that house massive scale behavioral data like Facebook. These data are important for uncovering countless patterns of social contagion and the spread of behaviors. They help us understand the world and, hopefully, contribute to making it better.

We make three contributions with this essay. First, we describe several experimental designs that we successfully implemented in partnership with the Data Science Team at Facebook. We hope these designs spur innovative new ideas for other experimental research designs and serve as a focal point for members of data science and analytics teams at other firms collecting massive scale data, of which the number is growing quickly. Second, we describe several design considerations that future research should take into account. These considerations should help improve on existing research by expanding the ability to put forth testable extensions based on the current state of knowledge. Third, we offer concrete suggestions based on our experience collaborating with for-profit companies. Academic-industry collaborations are essential for fulfilling the promise of understanding social contagion and the spread of behaviors across social networks.

## 2 The Experiments

### 2.1 *The 2010 Get-Out-the-Vote Experiment*

In the summer of 2010, we were fortunate to start what was to become a long-term and productive research collaboration with the Data Science Team at Facebook. The first major study to emerge from this collaboration was an experiment that was conducted during the 2010 U.S. midterm elections. Our collaborative team based at UCSD and Facebook designed and implemented a get-out-the-vote (GOTV) experiment on the Facebook website [8]. The experiment was intended to test whether a message delivered through Facebook could increase voter participation in the election. Further, as we describe below, the design of the experiment enabled us to test whether such a message would spillover from friend to friend to friend. This experiment permitted tests of a number of hypotheses about how social networks function and how interactions in online media impact our offline political behaviors.

Previous work in political science had shown that while many types of GOTV messages are effective at increasing turnout [3, 32, 33, 50], electronic messages had been shown to be much less effective [49] though text messaging was effective in one study [18]. Understanding if and how electronic messages may be effective for increasing turnout is important both for practitioners seeking to influence election

outcomes and social scientists interested in understanding social contagion. For practitioners, understanding if and when electronic messaging may be effective at increasing voter turnout is particularly important because electronic messaging is much cheaper than in-person or on-the-phone messaging. Facebook had recognized the potential of its platform for mobilization. In 2008, they had delivered a “banner” to the top of every adult American user’s News Feed with a standard message reminding people to vote and linking them to information about their polling location. Voter mobilization efforts via social media were thus “in the wild” and established as a possible way to increase voter turnout.

The challenge for researchers was thus to figure out what kind of messages could most effectively mobilize people to vote. For researchers interested in social contagion, electronic messaging, particularly through social media, enables researchers to quickly and relatively easily couple information about treatment status to information about social ties. More and more, our behaviors, attitudes, and social connections are recorded online [44]. This fact presents social scientists with unprecedented new opportunities to understand how social contagion manifests outside of small sample laboratory settings (e.g., [23]) or specific contexts which are difficult to generalize from (e.g., [12]). In our case, we were able to leverage the popularity of social media to conduct a field experiment at the scale of millions of participants and several hundred million relationships.

The design of the experiment we implemented had many advantages over other research designs in terms of both internal and external validity.<sup>2</sup> First, by placing information about voting at the top of users’ News Feeds we were confident that when users logged in to the site, they were very likely to receive the treatment. With other electronic messaging intended to increase turnout, such as through email, it is likely that at least some recipients are not exposed to the message, because they never open the email or the email goes to a spam folder and the user never even knows of its existence. Researchers typically focus on the “intent to treat” effect in those designs. However, in such cases the actual exposure to treatment may be quite low, which in turn makes the detection of treatment effects on those who actually

---

<sup>2</sup>Shadish [56], who builds on the research design tradition from [14], defines internal validity, as “[t]he validity of inferences about whether observed covariation between A (the presumed treatment) and B (the presumed outcome) reflects a causal relationship from A to B, as those variables were manipulated or measured,” and external validity as “The validity of inferences about whether the cause–effect relationship holds over variation in persons, settings, treatment variables, and measurement variables.” (4). Within the potential outcomes framework developed by Rubin [53], the focus is oriented primarily towards internal validity. However, some authors have related the SUTVA assumption from the potential outcomes framework to issues of external validity and construct validity [51, 56]. There are many sources available for additional and more detailed discussions that link together different validity types. For classic discussions of the relationships between these concepts, see [1, 11, 14, 57, 64]. For more recent treatments, see [17, 19, 20]. A focus on internal validity for massive social interventions forces the analyst to intentionally design the study to avoid violations to the Stable Unit-Treatment Value Assumption (SUTVA) [53]. Recognizing this assumption and designing the study to address are critical steps, which are necessary for exploring social contagions (e.g., [13, 51, 61]).

were exposed more difficult. With the prominent placement of the GOTV message in our experiment, it was very likely that users were exposed to the information we intended. This type of placement on a popular social media website would only be possible through collaboration with the company. Researchers are able to purchase ads appearing on sites or post their own content, but the prominence with which the treatment was presented to users could not be achieved in other ways. These design choices limit the external validity of the results.

Second, the message included an “I voted” button. While this was a useful online dependent variable—with which we were able to identify differences in the likelihood of clicking the button—it also functioned as one of the potential mechanisms through which network sharing took place. When users clicked the button a story about their voting action was automatically created and shared with their friends. Through this mechanism, the names and faces of friends who reported that they had voted became highly salient to the targeted users on Election Day.

Finally, by studying voter turnout, we were able to link an online treatment to a validated offline behavior. In the U.S., whether or not an individual has voted is a matter of public record, but the process through which a researcher may collect public voting records varies considerably from state to state. In some states, such records are easily downloadable for free. In others, there may be a cumbersome application process, a substantial fee (up to \$30,000), or requirements that the requests for turnout data be made by residents of the state. Additionally, states vary in what data they collect and make available in such records. Some states collect information about voters’ demographic characteristics, but most do not.

Because of this extensive variation, we identified thirteen states<sup>3</sup> that made available the data necessary for matching to Facebook records (first name, last name, and date of birth) at a reasonable cost per record. We then used a group-level matching process [38] that allowed for individual level inferences but avoided any direct one-to-one matching between the data from Facebook and the data from the state voter lists. The data was never linked together.

We developed the group-level matching procedure in order to preserve the privacy of individual Facebook users. By using the group-level matching procedure we were able to know probabilistically whether or not an individual had voted. This type of procedure may be helpful for instances in which two data frames need to be matched with one another, but preserving some uncertainty about individual attributes or measures is desirable. An important, if often overlooked feature of this method, is that the two datasets never need to reside on the same system. Only the repeated instances of the group-level information needs to be generated and transferred from one dataset to the other. The only way the identities of users on one dataset can be confirmed within the other dataset is if the set of users in both are completely overlapping. Otherwise, knowledge of the identity and the individual level attributes of each individual user is preserved.

---

<sup>3</sup>The states we collected data from were Arkansas, California, Connecticut, Florida, Kansas, Kentucky, Missouri, Nevada, New Jersey, New York, Oklahoma, Pennsylvania, and Rhode Island.

The GOTV experiment included three conditions. First was the “social message” condition, to which 98% of users were assigned. In this condition, users saw a message that encouraged them to vote, saw a link to a website that enabled them to search for their polling location, were offered a button to click to self-report voting, and were shown a set of up to six profile pictures of friends who had previously reported voting on the site. While we were unable to test the mechanism directly, we believe that seeing the faces of friends likely encouraged users to think of voting as a social act. As we note in the paper, however, it is also possible that the faces of friends simply made the message larger and more interesting, thereby drawing additional attention to it. The second experimental condition was the “message” condition, to which 1% of users were randomly assigned. The message looked just like the social message condition, but did not include the faces of friends. Because of this, this message was much more similar to a traditional GOTV message that encouraged voting and provided some of the information necessary to vote. Finally, the remaining 1% of users were randomly assigned to a control condition in which there was no message at the top of the users News Feeds—in essence Facebook appeared similar to how it would on any other day.

In the first stage of our analysis, we investigated the direct effects of treatment on vote reporting (clicking the “I voted” button), searching for information related to voting (clicking on the link to find a polling place), and validated voting. For each of these conditions, we found that users in the social message condition engaged in the behavior at a significantly higher rate than users in the message condition. For validated voting, we also compared the social message group to the control group and again found that being exposed to the social message led to higher rates of turnout. While we were able to use the full sample for comparisons of the online behavior, to compare rates of offline behavior (validated voting) we were restricted to the set of matched users. This meant that our sample size was about one-tenth the original number of users in the experiment, greatly reducing our statistical power in these comparisons compared to the comparisons we could make with the full sample.

In fact, one of the important lessons we learned from conducting the experiment was about the issue of statistical power. Of course, with such a large sample size, statistical power is greatly increased. However, with online experiments such as the one described here, and particularly when the dependent variable is a behavior that occurs offline, frequently the effects that one is likely to observe are quite small as well. The other issue with our design relates to the difference in the relative proportions of the three conditions: 98%, 1%, and 1%, respectively. We made this choice for two main reasons. First, Facebook wanted to maintain a consistent user experience for most of its users. Second, Facebook did not want to reduce its mobilization efforts compared to what they had done in the 2008 election. Because we hypothesized that the social message treatment would be most effective, the company prioritized that message. This choice however greatly reduced statistical power, which is an important trade-off that must be considered in any industry-academic collaboration.

As previous research had shown, email messages show little evidence of mobilization [49]. With this in mind, we knew that if the Facebook message was to be successful in mobilizing voters, the effect of the message was likely to be small in percentage terms and we would need a large sample in order to identify it statistically.

In our case, with a sample of approximately 6.3 million users who had been matched to voting records, we were able to identify a small effect that would have been undetectable with smaller samples. Although such small effect sizes may seem relatively unimportant, when treatments that have small effects are given to millions of people, their cumulative effects can add up to large changes in overall behavior. For our GOTV intervention, we estimated that the direct effect of the experiment was approximately 60,000 increased votes. Some critics might respond that with such large samples of users, that statistical significance is almost guaranteed. But this is not the case because of the sample size issue we described above. Moreover, we tested for this using a variety of auxiliary tests that we describe in detail in the supplementary appendix that accompanies the main article. These types of auxiliary tests are important for ruling out false positives in such massive scale studies.

In many ways, what we have described above is akin to a typical field experiment implemented at a very large scale. It was important to provide evidence supporting the idea that electronic GOTV efforts could yield small effects that most likely are statistically insignificant when conducted on the size sample typically available to academic researchers.

However, the more important contribution of the study was the ability to examine whether or not the effects of the experiment spilled over to other users as well. Although the literature on get-out-the-vote messages was by this point robust, few scholars investigated whether such messages had effects beyond those on directly contacted individuals. Because politics is such a social process, and the message encouraging voting was to be delivered through a social medium, we felt confident that this was an area in which the spread of the message from one individual to another was likely.

In our view, one of the biggest advantages of working with Facebook was the detailed information that the site dynamically collects about social relationships. A challenging aspect of conducting experiments in a social context is the possible presence of spillovers, intentional or otherwise (e.g., [61]). However, identifying the network ties that govern how such spillover occurs is a critical design challenge. Facebook's existing data on friendships enabled us to not only identify such network ties, but also to differentiate between them in any number of theoretically-important ways [37].

What types of relationships should be most likely to show evidence of social influence, based on findings from previous research? A long literature in the social sciences has emphasized tie strength, and its importance for understanding if and how things in networks may be transferred between individuals [10, 26, 27, 31, 35, 47]. While we knew we wanted to examine spillovers, we also knew that using the set of all Facebook friends was unlikely to provide us with the best opportunity to understand the social pathways through which information about voting spread.

Previous researchers had differentiated between friendships based on single criteria, such as photo-tagging behavior [46]. Knowing that an even more fine-grained measure of tie strength would enable us to better understand which ties mattered the most for the spread of behavior, we devised a method to assign friendship weights across a continuum [37].

To differentiate between network ties of various strengths, we coupled survey data with data from Facebook on how frequently people interact offline. We first surveyed a convenience sample of users about who their closest real-world friends were. We then matched their free-responses to this question to the list of their Facebook friends and used data on the interactions between an individual and his or her friends to predict these user-specified close friendships. Facebook collects many distinct types of digital traces that represent interactions between individuals, such as posting on one another's timeline, commenting on one another's posts, tagging each other in photos, liking a post of another, and so on. Our analysis revealed that simply counting these interactions (and accounting for the base rate of interaction with friends overall) was a good predictor for the self reported closeness of a relationship.

Once we had measures of tie strength, we investigated whether the treatment of one individual had an effect on the behavior of another. To do so, we started with the full network of Facebook friendships (among users included in the study). We assigned each individual his or her experimental condition, observed the outcome behavior (validated voting, vote reporting, searching for polling place information), and for each friendship pair we assigned a measure of friendship strength. We then observed the relationship between one individual's treatment status and the behavior of his or her friend.

While this gave us a good estimate of the relationship between an individual's treatment status and a friend's turnout behavior, we did not yet have a measure of how likely the relationship was to be observed simply by chance. That is, we knew that individuals whose friends had seen the social message were more likely to have voted than individuals whose friends had been in the control condition, but we did not yet know if that relationship was different from what we would expect by chance. Traditional statistical tools we use to test for the likelihood of such differences when treatment and behavior are measured within the same individual, such as a *t*-test or regression, would not account for the network. In particular, because individuals are tied together in the network, the structure of those ties may impact the likelihood of observing a relationship between treatment and behavior simply due to chance. So, we had to use other statistical tools, as described below, to assess whether the relationship we observed was different from chance.

We used a permutation method [30] to estimate a null distribution for the relationship between treatment and behavior. To do so, we kept the network topology and the behavior of each individual fixed and randomly shuffled the assignment to treatment. We then observed the relationship between treatment status and the turnout in this new random network, which was one example of what would have happened if treatment was not related to behavior. We repeated this process 1000 times, creating a full null distribution for the potential outcomes that may have

occurred simply due to chance. Conducting the analysis this way gave us confidence that if the true, observed relationship was outside of the null distribution, or at least in the top 2.5% of the distribution, that the relationship we observed was not due to chance.<sup>4</sup>

We repeated the above permutation analysis for the relationship between treatment of one individual and the behavior of his or her friends for increasingly close friendship relationships. Across all relationships we found that friend treatment was linked to increased probability of vote reporting. However, we found that for validated voting, only in the closest 20% of friendships was a person's friend's treatment related to his or her own behavior. It is notable that we found very small effect sizes here. For instance, for each close friend who was assigned to the social message, a user was 0.22% more likely to vote than had a friend been in the control condition. Similar to the results on direct effects, though the per-friend effect size is quite small, because the number of friends is so large, the cumulative effects of effect sizes like these can be substantial. For example, we estimated that the spillover of the message to close friends increased voting by approximately 559,000 votes.

Finally, we were interested in understanding whether the effects of the GOTV message spread even further than a single link in the social network. In other words, could we detect friend-of-a-friend spillover? We knew that for validated voting spillovers were likely to occur only for close friendships. Therefore, we assumed that the first place to look for further spillovers was through the close friends. We constructed a network of the close friends of close friends (who are not also friends). That is, if Joe and Amy are close friends and Amy and Jill are close friends, but Joe and Jill are not friends at all, we used Joe's experimental condition to predict Jill's behavior. In doing so, we were able to investigate whether or not the spillover occurred at more than one step removed from the focal individual. We found that the per-close-friend-of-close-friend effect was very small and only different from chance for self-reported vote—an increase of 0.01%.

## 2.2 *The 2012 Get-Out-the-Vote Experiment*

During the 2012 U.S. presidential election, we followed up our 2010 study with another GOTV experiment [39]. Our goals in doing so were to both replicate the 2010 study and to further examine the mechanisms that are likely to drive the social contagion we observed in the 2010 study. Both the election context and the use of social media changed from 2010 to 2012. For one thing, as opposed to the 2010 midterm election, the 2012 election involved a presidential race, and GOTV

---

<sup>4</sup>Later methodological work pointed out this procedure for simulating the null distribution rests on the unnecessary assumption of no direct effects [5]. Thus, our method of naively permuting treatments over the network could elevate the rate of false alarms. Focal unit analysis [4, 5] allows the researcher to more explicitly specify the null hypothesis and test for the presence of spillovers. The analysis performed on the 2012 election experiment data utilizes focal unit analysis.

messages are known to be less effective during high-stakes elections [34]. Secondly, between 2010 and 2012, millions more Americans had joined Facebook. A report from pew showed that in 2010, 60% of those who had access to the Internet used at least one social networking site. By 2012, that percent had climbed to 67%. The rate of growth was fastest among older Americans, the very kind of users we had found to be most responsive to the treatment in 2010. Because of the shifting context, we felt it was particularly important to replicate the findings from our previous research. As we describe below, we also changed the design of the experiment to try to better understand the mechanisms that were likely to drive changes in behavior.

The 2012 study functioned in a largely similar fashion to the 2010 study with some key differences. Again, users were exposed to a message encouraging turnout and offering a button to report voting through the site. However, in this experiment we implemented a  $2 \times 2$  design. The first factor varied whether or not individuals saw a post at the top of their News Feeds that encouraged turnout (the “banner” condition). This was very similar to the social message vs. control comparison from the 2010 study. The other factor varied whether or not users saw individual posts within their News Feed regarding friends’ voting (the “feed” condition). In the 2010 experiment, all users saw these messages within their feeds. By treating this as an experimental variable in 2012, we hoped to better understand if the banner message or the feed messages were more likely to induce behavior change, as well as if one or the other was more likely to cause contagion. Similar to the 2010 experiment, most users (96%) were in the condition that included both the banner GOTV message and the messages from friends in the newsfeed, and the remaining 4% were in control conditions for one or both of the GOTV message conditions.

In this experiment, we found largely similar results as in the 2010 experiment. In particular, we found that those in both the banner and feed conditions were significantly more likely to have voted than those in the control conditions. We also attempted to differentiate between the banner and feed, but we found no statistical differences between them, which suggests that the combination of the banner and the feed was responsible for the increase in voting.

We again investigated spillover effects. In the 2012 experiment, we found that the banner treatment caused significant spillovers, but that the feed treatment did not. So, while the combination of treatments appears to be most effective for causing direct change in behavior, the banner treatment appears to be more effective for causing spillovers. A critical component of this research was replicating the previous finding. As we noted earlier, both the affordances of social media and the ways in which people use sites change rapidly. As such, it is important to replicate findings from such studies across different sites and across time points. The fact that the 2010 and 2012 experiments had similar results gives us confidence that the processes of social influence that we observed were not dependent on aspects of the site or the ways in which the site was used at a particular time.

### 3 Design Considerations

Academic-industry collaborations provide both opportunities and challenges for the design of social science experiments. On the one hand, the wealth of data available can create unparalleled opportunities for testing effects that would be difficult to do using other kinds of data. At the same time, the imperative for experimental design must be balanced against a company's preferences and requirements for the user experience. Under certain circumstances, this may necessitate limits on optimal experimental design.

#### 3.1 *Heterogeneous Treatment Effects*

Previous work has investigated how influence may be maximized [41] in addition to how influence and susceptibility may be balanced [2, 6]. A significant advantage of working with a very large sample is the significant increase in statistical power that such sample sizes afford. Following this work, we investigated differences in treatment effect size in the 2010 turnout experiment across the pre-treatment covariates of users [9]. The reduction in sample size resulting from matching to voter records prohibited investigating treatment effect heterogeneity on validated voting. However, with the larger sample for which we had information on vote reporting and polling place search we were able to examine how treatment varied across individuals.

We found substantial differences in the likelihood that individuals responded to the treatment. For example, we found that older users were much more likely to respond to the treatment than younger users. For the number of friends that an individual has, we found an inverted u-shaped relationship: those with a moderate number of close friends were most responsive to treatment. This finding in particular should be of interest to researchers designing future experiments. If users with a moderate number of friends are most responsive to treatment, does the boost in responsiveness make up for the moderate number of friends to whom they may pass on the effects of the treatment? That is, is it better to treat nodes with many network ties, knowing that they may not be as responsive as other nodes, but that if they do respond they will pass along the effect to many others? Or is it better to have a larger direct effect on nodes that have a more limited capacity to create contagion?

In the future, researchers should examine how individual attributes, and edge attributes, are not only related to treatment effects, but also to the likelihood that contagion occurs. For example, while we found that older users were more likely than younger users to respond to the treatment themselves, we don't yet know if older users are more or less influential. By understanding this, practitioners may be better able to design and implement programs intended to effect change by accounting for not only the types of people they should contact, but also the likelihood that the people they contact will spread their message further.

### 3.2 *The Importance of Close Friends*

One of the key findings from our 2010 experiment suggested the importance of differentiating between friends online based on the strength of the relationship between the individuals offline. By estimating tie strength, we were better able to understand which types of friends are influential. For both online and offline behaviors, we found that the closest friends have particularly strong effects. In both cases, it appears that the relationship between friendship strength and likelihood of influence is non-linear—that is, close friends seem to matter much more than otherwise similar, but slightly more distant friends. Further, our analyses suggest that the closest friends are where most of the contagion is likely to take place, particularly for offline behaviors.

There is a more general lesson for network researchers in these results. If our findings are representative of how other network phenomena are likely to spread, then our work underscores the importance of measuring edge attributes in addition to node attributes.

### 3.3 *Uneven Assignment of Users to Conditions*

The final design consideration worthy of discussion is the uneven assignment of users to conditions. In both the 2010 and 2012 experiments, assignment to treatment was much more likely than to control. The proportion of people assigned to each condition was weighted in this way in order to make sure that the user experience was similar for most users. This was an important aspect of the design from Facebook’s point of view, as maintaining consistency in user experience was very important. This has important implications for the analysis and interpretation of the results, particularly the contagion results. For measuring direct effects, this impacts statistical power as it would for other, similar experiments. In particular, while we have relatively precise estimates for the condition to which most individuals are assigned, the estimates for the control conditions are noisier. For the online behaviors, where the sample size is very large, this is largely a non-issue. For the offline behavior, where the sample size is relatively small (even if large in absolute terms) this makes the identification of treatment effects more difficult.

Perhaps more important, however, is how the uneven assignment to treatment groups affects the inferences drawn from the spillover analyses. Because treatment is assigned in uneven ways, and network topology is not accounted for in treatment assignment, for most people the vast majority of their friends are assigned to treatment. That is, we observe many egos for whom all or nearly all of their friends were assigned to treatment groups, but we observe very few egos for whom few of their friends were assigned to treatment. Because of this, our findings about contagion leave many questions about how the distribution of the number or proportion of friends who were treated unanswered. For some contagions, a single friend who is treated may be enough to effect behavioral change, while for others

many friends may need to be treated. In the future, studies should implement designs that enable researchers to better understand how the distribution of treatment affects the likelihood of spillover effects.

## 4 Academic-Industry Collaboration

We do not claim to be experts at academic-industry collaboration. Indeed, we believe there are likely many researchers who are in a better position than we are to discuss how to best enable academic researchers and their industry partners to have successful collaborative relationships. However, we do believe that our experience places us in a unique position to discuss why collaboration is particularly important for the study of complex social systems. In particular, we hope that our experience might act as a guide or reference point for other scholars working towards establishing relationships with industry partners.

We began collaborating with Facebook as graduate students. We had been working to build our own app that we hoped to recruit students to use so we could collect our own data on online networks. However, the development of the app was slow-going and we were unsure whether such a research design was likely to bear fruit. Thus, when the opportunity arose to work directly with the Facebook Data Science Team, we moved forward as quickly as university bureaucracy would allow.

More generally, we recognize that social media companies possess data that is both deep (many millions of users) and wide (many datapoints on demographics and behavior), and no academic researcher has the resources to create something on the same scale. For better or worse, the best social science datasets in the world live on the servers of for-profit companies. By collaborating with industry partners, many of the limiting factors related to understanding network phenomena are greatly alleviated.

One of the advantages to such collaborations is the ability to quickly and at a relatively low cost implement experiments on existing networks. This is particularly important for studying phenomena that are unpredictable. For example, the Ebola outbreak of 2015 and the information and messaging surrounding it constituted a quickly developing and uncertain circumstance for people globally in which information and misinformation were being spread widely, often through social media. In such circumstances, social media data would enable researchers to understand how information is being spread and how correct information may be spread to greater effect. Importantly, in such circumstances data could be collected about people's beliefs and behavior in real time. Even with a large amount of resources, collecting and analyzing social network information from scratch in such an environment would take a long time—perhaps too long to implement policy changes before significant societal changes have taken place. However, if researchers were able to collaborate with social media platforms that already have information about networks and over which the transmission of information is likely to take place, such research may take place at a truly rapid pace.

## 5 Conclusion

We hope that our experience working with Facebook does not represent a unique moment in time when industry-held collections of massive scale behavioral data were open for analysis and publicly facing social science. Rather, we hope that this type of collaboration might happen again at places like Facebook and the many other new firms that are collecting and analyzing data. Social scientists can help understand and disseminate many important social patterns that could reveal new insights. Let's not black box the social data that we are all participating in producing.

## References

1. Adcock R, Collier D (2001) Measurement validity: a shared standard for qualitative and quantitative research. *Am Polit Sci Rev* 95(3):529–546
2. Aral S, Walker D (2012) Identifying influential and susceptible members of social networks. *Science* 337(6092):337–341
3. Arceneaux K, Nickerson DW (2009) Who is mobilized to vote? A re-analysis of 11 field experiments. *Am J Polit Sci* 53(1):1–16
4. Aronow PM (2012) A general method for detecting interference between units in randomized experiments. *Sociol Methods Res* 41(1):3–16
5. Athey S, Eckles D, Imbens GW (forthcoming) Exact P-values for network interference. *J Am Stat Assoc.* <https://www.tandfonline.com/doi/abs/10.1080/01621459.2016.1241178>
6. Bakshy E, Hofman JM, Mason WA, Watts DJ (2011) Everyone's an influencer: quantifying influence on twitter. In: Proceedings of the fourth ACM international conference on web search and data mining, pp 65–74
7. Bond RM, Messing S (2015) Quantifying social media's political space: estimating ideology from publicly revealed preferences on Facebook. *Am Polit Sci Rev* 109(1):62–78
8. Bond RM, Fariss CJ, Jones JJ, Kramer ADI, Marlow C, Settle JE, Fowler JH (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415):295–298
9. Bond RM, Settle JE, Fariss CJ, Jones JJ, Fowler JH (2017) Social endorsement cues and political participation. *Polit Commun* 34(2):261–281
10. Burt RS (1997) A note on social capital and network capital. *Social Netw* 19(4):355–373
11. Campbell DT (1960) Recommendation for APA test standards regarding construct, trait, or discriminant validity. *Am Psychol* 15:546–553
12. Christakis NA, Fowler JH (2008) The spread of obesity in a large social network over 32 years. *N Engl J Med* 357(4):370–379
13. Christakis NA, Fowler JH (2013) Social contagion theory: examining dynamic social networks and human behavior. *Stat Med* 32(4):556–577
14. Cook TD, Campbell DT (1979) Quasi-experimentation: design and analysis for field settings. Houghton Mifflin, Boston
15. Covello L, Sohn Y, Kramer ADI, Marlow C, Franceschetti M, Christakis NA, Fowler, JA (2014) Detecting emotional contagion in massive social networks. *PLoS One* 9(3):e90315
16. Covello L, Fowler JH, Franceschetti M (2014) Words on the web: noninvasive detection of emotional contagion in online social networks. *Proc IEEE* 102(12):1911–1921
17. Crabtree CD, Fariss CJ (2016) Stylized facts and experimentation. *Sociol Sci* 3:910–914
18. Dale A, Strauss A (2016) Don't forget to vote: text message reminders as a mobilization tool. *Am J Polit Sci* 53(4):787–804

19. Dunning T (2012) Natural experiments in the social sciences: a design-based approach. Cambridge University Press, Cambridge
20. Fariss CJ, Jones ZM (2017) Enhancing validity in observational settings when replication is not possible. *Polit Sci Res Methods*. <https://doi.org/10.1017/psrm.2017.5>
21. Farrell, H (2012) The consequences of the internet for politics. *Ann Rev Polit Sci* 15:35–52
22. Fowler JH (2005) Turnout in a small world. In: Zuckerman A (ed) *The social logic of politics: personal networks as contexts for political behavior*. Temple University Press, Philadelphia
23. Fowler JH, Christakis NA (2010) Cooperative behavior cascades in human social networks. *Proc Natl Acad Sci* 107(12):5334–5338
24. Fowler JH, Heaney MT, Nickerson DW, Padgett JF, Sinclair B (2011) Causality in political networks. *Am Polit Res* 39(2):437–480
25. Garcia-Herranz M, Moro E, Cebrian M, Christakis NA, Fowler JH (2014) Using friends as sensors to detect global-scale contagious outbreaks. *PLoS One* 9(4):e92413
26. Gee LK, Jones JJ, Fariss CJ, Burke M, Fowler JH (2017) The paradox of weak ties in 55 countries. *J Econ Behav Organ* 133(January):362–372
27. Gee LK, Jones JJ, Burke M (2017) Social networks and labor markets: how strong ties relate to job finding on facebook's social network. *J Labor Econ*. <https://doi.org/10.1086/686225>
28. Godino J, Merchant G, Norman GJ, Donohue MC, Marshall SJ, Fowler JH, Calfas KJ, Huang JS, Rock CL, Griswold WG, Gupta A, Raab F, Fogg BJ, Robinson TN, Patrick K (2016) Using social and mobile tools for weight loss in overweight and obese young adults (Project SMART): a 2 year, parallel-group, randomised, controlled trial. *Lancet Diabetes Endocrinol* 4(9):747–755
29. González MC, Hidalgo CA, Barabási A-L (2008) Understanding individual human mobility patterns. *Nature* 453(7196):779–782
30. Gordon PI (2005) Permutation, parametric, and bootstrap tests of hypotheses. Springer, Berlin
31. Granovetter MS (1973) The strength of weak ties. *Am J Sociol* 78(6):1360–1380
32. Green DP, Gerber AS (2002) Reclaiming the experimental tradition in political science. In: Katzenbach I, Milner HV (eds) *Political science: state of the discipline*. W. W. Norton, New York
33. Green D, Gerber AS (2004) *Get out the vote!: a guide for candidates and campaigns*. Brookings Institution Press, Washington
34. Green DP, Gerber AS (2008) *Get out the vote: how to increase voter turnout*. Brookings Press, Washington, DC
35. Hampton KN, Sessions LF, Her EJ (2011) Core networks, social isolation, and new media: how internet and mobile phone use is related to network size and diversity. *Inform Commun Soc* 14(1):130–155
36. Hobbs WR, Burke M, Christakis NA, Fowler JH (2016) Online social integration is associated with reduced mortality risk. *Proc Natl Acad Sci* 113(46):12980–12984
37. Jones JJ, Settle JE, Bond RM, Fariss CJ, Marlow C, Fowler JH (2013) Inferring tie strength from online directed behavior. *PLoS One* 8(1):e52168
38. Jones JJ, Bond RM, Fariss CJ, Settle JE, Kramer ADI, Marlow C, Fowler JH (2013) Yahtzee: an anonymized group level matching procedure. *PLoS One* 8(2):e55760
39. Jones JJ, Bond RM, Bakshy E, Eckles D, Fowler JH (2017) Social influence and political mobilization: further evidence from a randomized experiment in the 2012 U.S. presidential election. *PLoS One* 12(4):e0173851
40. Kahn JP, Vayena E, Mastroianni AC (2014) Opinion: learning as we go: lessons from the publication of Facebook's social-computing research. *Proc Natl Acad Sci* 111(38): 13677–13679
41. Kempe D, Kleinberg J, Tardos É (2003) 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
42. Kim DA, Hwong AR, Stafford D, Hughes DA, O'Malley AJ, Fowler JH, Christakis, NA (2015) Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. *Lancet* 386(9989):145–153

43. Kramer ADI, Guillory JE, Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci* 111(24):8788–8790
44. Lazer D, Pentland A, Adamic L, Aral S, Barabási A-L, Brewer D, Christakis NA, Contractor N, Fowler JH, Gutmann M, Jebara T, King G, Macy M, Roy D, Van Alstyne M (2009) *Comput Soci Sci Science* 323:721–723
45. Leas EC, Althouse BM, Dredze M, Obradovich N, Fowler JH, Noar SM, Allem J-P, Ayers JW (2016) Big data sensors of organic advocacy: the case of Leonardo DiCaprio and climate change. *PLoS ONE* 11(8):e0159885
46. Lewis K, Kaufman J, Gonzalez M, Wimmer A, Christakis N (2008) Tastes, ties, and time: a new social network dataset using Facebook.com. *Soc Netw* 30(4):330–342
47. Marsden PV (1987) Core discussions networks of Americans. *Am Sociol Rev* 52(1):122–131
48. Meyer M (2014) Misjudgements will drive social trials underground. *Nature* 511(7509):265
49. Nickerson DW (2007) Does email boost turnout? *Q J Polit Sci* 2(4):369
50. Nickerson DW (2008) Is voting contagious? Evidence from two field experiments. *Am Polit Sci Rev* 102(1):49–57
51. Oakes JM (2004) The (mis)estimation of neighborhood effects: causal inference for a practicable social epidemiology. *Soc Sci Med* 58(10):1929–1952
52. Onnelaa J-P, Reed-Tsochas F (2010) Spontaneous emergence of social influence in online systems. *Proc Natl Acad Sci* 107(43):18375–18380
53. Rubin DB (2008) For objective causal inference design trumps analysis. *Ann Appl Stat* 2(3):808–840
54. Settle JE, Bond RM, Levitt J (2011) The social origins of adult political behavior. *Am Polit Res* 39(2):239–263
55. Settle JE, Bond RM, Coviello L, Fariss CJ, Fowler JH, Jones JJ (2016) From posting to voting: the effects of political competition on online political engagement. *Polit Sci Res Methods* 4(2):361–378
56. Shadish WR (2010) Campbell and Rubin: a primer and comparison of their approaches to causal inference in field settings. *Psychol Methods* 12(1):3–17
57. Shadish WR, Cook TD, Campbell DT (2001) Experimental and quasi-experimental designs for generalized causal inference. Wadsworth Publishing, Belmont
58. Shakya HB, Hughes DA, Stafford D, Christakis NA, Fowler JH, Silverman JG (2016) Intimate partner violence norms cluster within households: an observational social network study in rural Honduras. *BMC Public Health* 16:233
59. Shakya HB, Stafford D, Hughes DA, Keegan T, Negron R, Broome J, McKnight M, Nicoll L, Nelson J, Iriarte E, Ordonez M, Airolidi E, Fowler JH, Christakis NA (2017) Exploiting social influence to magnify population-level behaviour change in maternal and child health: study protocol for a randomised controlled trial of network targeting algorithms in rural Honduras. *BMJ Open* 7(3):e012996
60. Shakya HB, Fariss CJ, Ojeda C, Raj A, Reed E (2017) Social network clustering of sexual violence experienced by adolescent girls. *Am J Epidemiol* 186:796–804
61. Sinclair B, McConnell M, Green DP (2012) Detecting spillover effects: design and analysis of multilevel experiments. *Am J Polit Sci* 56(4):1055–1069
62. Steinert-Threlkeld, ZC (2017) Spontaneous collective action: peripheral mobilization during the Arab spring. *Am Polit Sci Rev* 111:379–403
63. Steinert-Threlkeld ZC, Mocanu D, Vespignani A, Fowler JH (2015) Online social networks and offline protest. *EPJ Data Sci* 4(19):1–9
64. Zeller RA, Carmines EG (1980) Measurement in the social sciences: the link between theory and data. Cambridge University Press, Cambridge

# Spreading in Social Systems: Reflections



Sune Lehmann and Yong-Yeol Ahn

## 1 Introduction

As a starting point, we believe that social contagion will play a key role in shaping how society and democracy develops in the coming decades. As our world has become increasingly connected through the networks of social media, the role of social contagion has grown. Social media services, such as Twitter, Facebook, or Reddit, are becoming the main channels through which people communicate and consume news. Because of these platforms' global connectedness, a piece of news—fake or not—can spread to millions of people around the world at near instantaneous speed. Moreover, the increasing social media use, combined with sophisticated machine learning algorithms for content recommendation, means that we increasingly find ourselves within comfortable ideological bubbles. Inside each bubble, content that reinforces our beliefs and biases will spread more easily among people with shared ideologies and potentially entrench people. Such entrenchment may grow in the future. Thus, humanity's major challenges are beginning to revolve less around building the right technologies, but more around puncturing bubbles in order to reduce societal polarization.

The power to manipulate and control people's beliefs through social contagion is a double-edged sword. Such power can be used for public good—to effectively spread informed opinions on public health matters: safe sex, smoking, or vaccination to name a few examples. At the same time, however, this power can be, and has been, misused for manipulating public opinions or influencing the outcome of elections.

---

S. Lehmann (✉)  
Technical University of Denmark, Lyngby, Denmark  
e-mail: [sljo@dtu.dk](mailto:sljo@dtu.dk)

Y.-Y. Ahn  
Indiana University, Bloomington, IN, USA  
e-mail: [yyahn@iu.edu](mailto:yyahn@iu.edu)

We expect that the impact of social contagion on our society—particularly on the foundation of democracy—will keep increasing. The social responsibility of research into social contagion processes should not be overlooked.

• • •

As is clear from the fantastic contributions in this book, our understanding of social spreading processes has advanced significantly over the past decade. Still, there are of course outstanding challenges. Among many, here we discuss the following:

- How can we improve the quality, quantity, extent, and accessibility of datasets?
- How can we extract more information from limited datasets?
- How can we take individual cognition and decision-making processes into account?
- How can we incorporate other complexities from the real contagion processes?
- How can we translate research into positive real-world impact?

## 2 Please Sir, I Want Some More Data

History tells us that the availability of high-quality data is a key driving force in science. The science of social contagion is no exception. In the past decade, datasets from long-term longitudinal studies, such as the Framingham Heart Study, as well as other massive online social media datasets have been the main fuel source that propels the study of social contagion. So, a natural question is “how can we get our hands on better datasets?”

We should probably begin by asking what “better” data would even mean. “Better” may mean simply more details and larger volume. For instance, high-resolution data can reveal insights that are completely hidden when that same dataset is aggregated. Larger datasets imply increased statistical power and the ability to identify minute effects. Having more attributes can lead to the discovery of new associations or more precise control of confounding factors.

Going beyond size and detail, better data may also imply a shift from *found* data to more *designed* data [1]. Instead of re-purposing observational datasets, one can specifically design a (controlled) experiment and collect data. To do so, one should either create one’s own data collection environment (e.g., Sensible DTU [2]) or leverage existing services (e.g., controlled experiments conducted by Facebook [3, 4]). The former is more constrained by resources and difficult to scale, while the latter is more constrained by the economic incentives of the company and details of the services. Collaboration between academia and industry is a nice hybrid approach and has produced many successful insights (see, for example, Part IV, chapter “Network Experiments Through Academic-Industry Collaboration”).

Finally, we stress the importance of open access to data. Even if one collects an ideal dataset to study social contagion, the dataset may make little impact on the field

if the data cannot be shared with others. Even the validity of a study that uses this ideal data may not be ensured if no one can use the data to replicate the results. The benefits of data sharing are clear; making a dataset public can maximize the impact of the dataset and makes the resulting research more transparent and reproducible.

However, many social datasets are not easy to share in a raw form (or even collect) due to privacy concerns. There have been several efforts from industry to share anonymized datasets but many have unfortunately failed. For instance, Netflix shared a large dataset for a highly profiled recommendation engine challenge, only to find that the dataset could be easily de-anonymized [5]. Later, an anonymized Flickr social network was de-anonymized using Twitter's social network [6]. After several incidents of this kind and more theoretical developments, it has become clear that it is very difficult to properly anonymize data, in particular data that involves social networks.

There can also be other kinds of backlash related to sharing data, or simply just sharing results of studies that are conducted in industry. For instance, the emotional contagion study published by Facebook in collaboration with academic researchers [3] upset many users and put Facebook in a difficult position. The adverse reaction to this study may have suppressed the in-house research efforts across industry and reduced incentives to publish academic articles, not to mention datasets. Thus, understandably, most companies are cautious about sharing raw datasets and even results of their internal experiments.

At the same time, the push from publishers, scientists, and other advocates for open data has begun to produce practical solutions. These practical solutions aim for a compromise between level of detail within the data and privacy concerns. A common approach is to share data that is sufficiently aggregated so that the re-identification or extraction of any individual data is impossible. Another solution is to maintain a special internal repository for replication data as well as mechanisms for external researchers to access the data upon request. Such solutions may address the issue of replicability, but fall short with respect to replicating the full benefits of open datasets. It will be interesting to see whether it will become easier to access raw datasets from industry through improved privacy-conserving algorithms or whether we will see aligned efforts resulting academia-industry collaboration in the future.

### 3 Homophily or Contagion?

Although we now have unprecedented amounts of data related to social contagion—and describing social behavior in general, most available datasets are still observational. This fact imposes serious limitations. A central issue is that, because of homophily (and latent homophily) in networks, it is difficult to perform causal inference. As the heated debate regarding the series of papers using the Framingham Heart Study—an observational dataset—has demonstrated [7–14], causal inference based on observational data is a major challenge, and the effort to extract as much as information from observational datasets will continue. A number of methods

have been developed to more clearly understand the limitations and extract more information from observational datasets (see chapter “Challenges to Estimating Contagion Effects from Observational Data”). The results from observational data will remain as an important part of the social contagion research.

## 4 Micro-Contagionomics

Most existing studies assume fairly simple contagion models that do not take into account complex individual decision-making and variations across individuals. Given what is known about cognition and social psychology, another interesting avenue of research will be to incorporate cognitive and psychological models of decision-making and behavioral changes into the study of social contagion; both in theoretical and empirical studies.

Although there are many theoretical models, rich models that can capture more nuanced cognitive limitations and biases—such as complex interactions between beliefs [15] or limited attention [16]—as well as the nature of contagion [17] will be needed to fully understand and better model social contagions.

On the empirical front, we need more precisely controlled, high-resolution experiments. In spite of all the progress there has been made studying empirical patterns of information diffusion (Part III), we are still limited to examining overall patterns and the results of spreading. In fact, outside of purely theoretical models (Part II), we have little idea how to incorporate knowledge and insights from psychology and cognitive science in order to measure the microscopic mechanisms that govern the adoption of a new idea.

Within the empirical work, we mostly study *proxies* for the information that is truly spreading, whereas the work on random control trials (Part IV) focuses on observing *behaviors* resulting from a spreading process on an underlying network. Thus, a possible way forward could be through new experimental paradigms, where we study both the spreading agent on its journey through the network, along with well-defined behavioral changes on the individual-level.

To make this concrete, let us outline some thought experiments. An extreme one will be similar to a reality show, where every single conversation and related behavior is recorded [18], with added potential interventions and controls. The data then could then be analyzed to identify how exactly the information spread through the participants.

Another possible experimental design would begin with designing specific, well-defined pieces of information designed to illicit a reaction (or lack thereof) that can be measured (e.g., going to collect free beer at a certain location, pressing a certain button). Further, study participants must only be able to access these pieces of information in a way that reveals the identity of the person in question (e.g., by displaying this information on a personalized web-page or via a mobile-phone app). Finally, of course, information on how to access these pieces of information must travel in a well-defined way on the social graph independently of the communication

platform (email, online social network, face-to-face). While accessing the piece of information, we could also provide information about actions (or information state) of the network neighbors. Starting from randomized control trials and with access to both detailed spreading paths on the network and behavioral outcomes, such an experimental paradigm would allow us to begin collect reliable statistics and answer questions on the microscopic mechanisms that shape spreading and adoption, such as how the probability of spreading depends on the local network structure.

By running multiple experiments we would also be able to empirically examine the role of “stickyness” or “sexyness” of ideas in spreading, acknowledging that intrinsic properties of the spreading agent might interact with the network in a non-trivial way.

## 5 It’s Complicated: Multi-Layered, Dynamic, Co-evolving Networks

Over the past 20 years or so, we have made substantial progress in our ability to describe and analyze static complex networks. But real networks exhibit many complex features. For instance, networks change dramatically over time. The connection patterns of social networks are constantly reconfigured as we connect with friends, co-workers, and family—as we move through our daily lives, as we adopt new platforms for communication. Our theoretical foundation for analyzing and understanding temporal networks is solidifying, but we are still learning how to treat the interplay between temporal networks and the dynamics of network spreading on those networks.

Network structure is not just changing in isolation. Often the dynamic evolution of a network is due to the social contagion in the network. In other words, the structure of social network and the dynamics of social contagion co-evolve [19, 20]. We can re-examine the issue of homophily versus contagion in this context. It is not just that these two concepts are confounded (a difficult problem in its own right). It is also that reality is often a mixture of the two (an even more difficult problem in its own right). In the wild, we are likely to see a dynamic bidirectional interplay of influence and homophily on each dyad—and more generally within each network neighborhood—shaping the evolution of the network itself, as well as the dynamics of information flowing through it.

While the advent of online social media and other communication channels have opened up new ways to study society and our communication patterns, online social media have also had the less publicized effect of increasingly fragmenting social communication across multiple channels. Most people use multiple social media services, often each for different purposes. As the main communication channel for their friends, some may use Facebook, some may use Twitter, some may use Snapchat, and some others may not even use any social media services at all, using only “traditional” channels such as in-person conversations. Thus, even when

some company dominates in many markets across the globe, a single service only captures a small, biased fraction of threads in this fabric of social communication. For instance, a single instance of social contagion may manifest itself as numerous disjoint spreading events if observed through the lens of a single service. And thanks to homophily and network effects, users of a service, and the ways that they use the service, tend not to be random samples from the full population.

An important implication in terms of the study of social contagion is that even the largest studies, if they were conducted on a single platform, might be lacking significant spreading events that occur via other channels and making conclusions based on biased behavioral patterns. Thus, it is important to ask: how can we know that the observed results are not artifacts of such a fragmentation? How can we study spreading phenomena that occur across many communication channels?

We believe that progress on this topic will occur by working simultaneously on both the theoretical and empirical side, with each side complementing the other. Empirical observations describe how people juggle multiple types of social media and how information spreads across the different layers of social networks will provide good insights on how to model use of the fragmented networks. Theoretical studies on multi-layer information spreading processes will then inform hypotheses and suggest general patterns to be tested through additional empirical studies. Collaborations across multiple social media platforms are also needed to obtain proper datasets to study multi-layer diffusion. Smaller-scale, but higher-resolution studies also have great potential to deepen our understanding of how people use multi-layered social fabric. Finally, because it will be practically impossible to capture every possible social interaction, statistical inference techniques and theoretical studies to understand the effect of missing data—or missing layers—and to infer the missing data will be necessary.

## 6 Translating into Real-World Applications

The final frontier will consist in translating social contagion research to real-world social problems beyond applications to product adoption and advertisement. Such studies could focus on inducing social contagion that intends a positive impact on society. Topics that citizens of a society can democratically agree are to the benefit of everyone. For instance, researchers in Facebook have already demonstrated that it is possible to significantly increase the participation to the election by engineering the social contagion on Facebook alone [4]. Similar campaigns may be designed and implemented for the behaviors that are relevant to public health, such as hand washing, safe sex, or vaccination.

At the same time, it is essential to remain vigilant with respect to the other side of the coin: increasing our collective ability to detect and mitigate malicious manipulative campaigns or public shaming events. It has been shown that there exist ongoing efforts to manipulate public opinions through human workers, social bots, and fake news [21, 22]. This type of manipulation potentially threatens the

foundation of democracy in many countries across world and even the very concept of “truth.” Thus it will be important for researchers to ask how the study of social contagion can help us understand and improve the “post-truth” world.

• • •

And to the reader who has made it this far. We thank you!

## References

1. Salganik MJ (2017) Bit by bit: social research in the digital age. Princeton University Press, Princeton
2. Stopczynski A, Sekara V, Sapiezynski P, Cuttone A, Larsen JE, Lehmann S (2014) Measuring large-scale social networks with high resolution. *PLoS One* 9(4):e95978
3. Kramer AD, Guillory JE, Hancock JT (2014) Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci* 111(24):8788–8790
4. Bond RM, Fariss CJ, Jones JJ, Kramer ADI, Marlow C, Settle JE, Fowler JH (2012) A 61-million-person experiment in social influence and political mobilization. *Nature* 489(7415):295–298
5. Narayanan A, Shmatikov V (2008) Robust de-anonymization of large sparse datasets. In: IEEE symposium on security and privacy, 2008. SP 2008. IEEE, New York, pp 111–125
6. Narayanan A, Shi E, Rubinstei BIP (2011) Link prediction by de-anonymization: how we won the kaggle social network challenge. In: The 2011 international joint conference on neural networks (IJCNN). IEEE, New York, pp 1825–1834
7. Christakis NA, Fowler JH (2007) The spread of obesity in a large social network over 32 years. *New Engl J Med* 357(357):370–379
8. Fowler JH, Christakis NA (2008) Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the framingham heart study. *Br Med J* 337:a2338
9. Christakis NA, Fowler JH (2008) The collective dynamics of smoking in a large social network. *New Engl J Med* 358(21):2249–2258
10. Cacioppo JT, Fowler JH, Christakis NA (2009) Alone in the crowd: the structure and spread of loneliness in a large social network. *J Pers Soc Psychol* 97(6):977
11. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc Natl Acad Sci* 106(51):21544–21549
12. Shalizi CR, Thomas AC (2011) Homophily and contagion are generically confounded in observational social network studies. *Sociol Methods Res* 40(2):211–239
13. Lyons R (2011) The spread of evidence-poor medicine via flawed social-network analysis. *Stat Polit Policy* 2(1). <https://doi.org/10.2202/2151-7509.1024>
14. Christakis NA, Fowler JH (2013) Social contagion theory: examining dynamic social networks and human behavior. *Stat Med* 32(4):556–577
15. Rodriguez N, Bollen J, Ahn Y-Y (2016) Collective dynamics of belief evolution under cognitive coherence and social conformity. *PLoS one* 11(11):e0165910
16. Weng L, Flammini A, Vespignani A, Menczer F (2012) Competition among memes in a world with limited attention. *Sci Rep* 2:335
17. Weng L, Menczer F, Ahn Y-Y (2013) Virality prediction and community structure in social networks. *Sci Rep* 3:2522
18. Roy DK, Pentland AP (2002) Learning words from sights and sounds: a computational model. *Cogn Sci* 26(1):113–146

19. Snijders TAB, Steglich CEG, Schweinberger M (2007) Modeling the co-evolution of networks and behavior. In: van Montfort K, Oud H, Satorra A (eds) *Longitudinal models in the behavioral and related sciences*, Lawrence Erlbaum, p 41–71
20. Gross T, Blasius B (2008) Adaptive coevolutionary networks: a review. *J R Soc. Interface* 5(20):259–271
21. Howard P, Bradshaw P (2017) Troops, trolls and troublemakers: a global inventory of organized social media manipulation. Oxford internet institute working paper
22. King G, Pan J, Roberts ME (2017) How the chinese government fabricates social media posts for strategic distraction, not engaged argument. *Am Polit Sci Rev* 111(3):484–501

# Index

## A

- Adoption thresholds, 14, 15, 18, 20, 101, 153, 156, 157, 161, 164, 170
- Anomaly detection algorithm, 282
- Assortativity, 262, 264
- Attention, 12, 51, 53, 82, 111, 142, 153, 182–185, 197, 199, 200, 208, 213–27, 230, 232, 234, 241, 242, 246, 251, 257, 258, 269, 290, 315, 340, 354

## B

- Biological contagion, 67–70, 82, 152
- Bipartite network, 38, 40
- Blocking, 305
- Boston marathon bombing, 270, 272–275, 285

## C

- Cascade condition, 86–87, 93, 163, 166
- Cascading behaviour, 164
- Causal analysis, 48
- Causal inference, 50, 61, 310, 317, 353
- Cell phone network, 218, 221
- Cluster randomized designs, 305
- Collaboration, 296, 335–348, 352, 353, 356
- Collective influence, 125, 126, 128, 130, 131, 133, 136, 137, 139–143
- Communication networks, 12, 112, 143, 170, 213–227, 323–332
- Community structure, 88, 97, 102, 200, 204, 206, 208
- Complex contagion, 3–21, 81–93, 97–106, 115, 152–155, 162, 200, 204
- Computational social science (CSS), 179

- Confounding, 48, 49, 51–54, 57, 59–62, 290, 325, 327, 352
- Contagion, 3–21, 27–44, 47–63, 67–79, 81–93, 97–106, 115, 152–155, 159, 162–164, 166, 170, 178, 179, 200, 204, 215, 290, 323, 324, 326, 330–332, 335–338, 343–346, 351–357
- Contagion condition, 28–44
- Correlated network, 88
- Critical mass, 12, 15, 28, 29, 41–43, 72, 76–79, 200

## D

- Data availability, 179
- Data-driven modelling, 170
- Data quality, 181
- Degree-degree correlations, 35, 82, 87–89, 93
- Disaster management, 269, 270
- Dose response, 315

## E

- Echo chambers, 7, 19, 179, 180, 182–186, 188–191, 193
- Effects, 5, 6, 8, 10, 12, 15, 19, 21, 47–62, 72, 78, 89, 103, 104, 110, 116, 117, 120, 121, 126, 136, 143, 152, 153, 155, 158, 159, 161, 162, 164, 165, 167–170, 177, 180, 190, 193, 199, 200, 205, 208, 215, 222, 223, 225, 226, 229–251, 284, 290–293, 295, 296, 298–301, 303–305, 307, 309, 310, 314–317, 323–332, 338, 340–343, 345–347, 354–56
- Email network, 217, 218, 220, 222–226

Emergency management, 269, 270, 272  
 Epidemic, 7, 69, 70, 76–79, 88, 114, 115,  
   118–120, 125, 126, 204, 205, 322  
 Ethical considerations, 290, 292–294  
 Experiment, 8, 9, 11, 47, 59, 104, 106, 111,  
   142, 153, 159, 178, 179, 200, 202–204,  
   206–208, 215, 232, 267, 284, 289–317,  
   323–325, 331, 335–348, 352  
 Experimental methods, 11, 291  
 Experimental paradigms, 354, 355

**F**  
 Feature-based classification, 197, 198  
 Field experiments, 290–294, 297, 299, 303,  
   308, 314, 338, 341  
 Fisherian randomization inference, 290, 308,  
   309  
 Friendship paradox, 32, 261–262, 265–267

**G**  
 Get-out-the-vote (GOTV), 337–344  
 Granovetter’s hypotheses, 214  
 Graph partitioning, 305  
 Group formation, 301, 306

**H**  
 Happiness paradox, 264–267  
 Heterogeneous treatment, 314–316, 345  
 Heterogeneous treatment effects, 314–316,  
   345  
 Homophily, 6, 8, 12, 15, 17, 19–20, 51–53,  
   57–62, 152, 153, 159–161, 186, 193,  
   204, 290, 291, 324–328, 332, 353–356  
 Human dynamics behavioral traces, 272

**I**  
 Induction, 47  
 Industry, 5, 53, 232, 251, 258, 294, 347, 353  
 Influence maximization, 125, 126, 130,  
   134–136  
 Influence network, 324–332  
 Informational links, 224–226  
 Information diffusion, 97, 125, 137, 179, 186,  
   193, 197, 198, 226, 229–251, 258, 323,  
   324, 354  
 Information flow, 128, 140, 200, 202, 205, 355

**L**  
 Limited information, 199

**M**  
 Matching, 167–170, 307, 308, 324, 327–330,  
   339, 345  
 Mean-field approximation, 82–84  
 Meme, 10, 18, 180, 197–209  
 Memory, 15, 18, 68, 72–76, 78, 79  
 Message passing, 81–93, 97–99, 131, 134–136  
 Misinformation, 119, 177–194, 269, 271, 272,  
   275, 347

Mobile phone records, 271, 276, 278, 284, 323,  
   324  
 Modeling cognition, 354  
 Modeling decision making, 331, 354  
 Monotonic dynamics, 93

**N**  
 Network  
   dynamics, 3, 4, 6, 122, 258  
   effects, 72, 292, 303, 356  
   epidemiology, 115

**O**  
 Observational records, 271  
 Online social media, 197, 200, 214, 231, 232,  
   270, 271, 323, 336, 352, 355  
 Online social networks, 151–172, 179, 201,  
   233, 258, 267, 296, 297, 323, 324, 355  
 Optimal modularity, 97–106  
 Optimal percolation, 128, 130–134, 141, 142

**P**  
 Peer effects, 47, 60, 61, 290, 296, 316  
 Peer influence, 9, 10, 13, 16, 47  
 Permutation test, 309, 310  
 Political communication, 229  
 Popularity, 71, 151, 155, 199, 200, 204, 206,  
   229, 234, 245, 266, 267, 338  
 Prediction, 92, 115, 144, 154, 169, 197, 198,  
   201, 203, 206–209, 233, 289

**R**  
 Randomization, 11, 59, 110, 116, 117, 120,  
   121, 290, 291, 294, 298, 301–312, 316,  
   317

- Real-world impact, 352  
Rumor spreading, 128, 178
- S**  
Sentiment, 191–193, 230, 231, 234–240, 246–251  
Sharp null hypothesis, 59, 309–313  
Simple contagion, 4, 6, 7, 10, 16, 17, 28–29, 33–35, 38–40, 97, 115, 152, 153, 204, 354  
Simulations, 17, 28, 75, 78, 79, 82, 84, 86, 101, 104, 105, 109–122, 126, 127, 130, 135, 164, 167–169, 198, 307, 314–316, 332  
Social bots, 229–236, 240, 244, 246, 250  
Social contagion, 3, 4, 6, 17, 34, 40, 41, 57, 68, 70–152, 772–154, 159, 162–166, 170, 178, 179, 204, 290, 323, 324, 335–338, 343, 351–356  
Social diffusion, 11, 18, 21  
Social influence, 7–10, 59, 68, 152, 153, 155, 158–162, 199, 289–317, 323–332, 344  
Social interactions, 152, 193, 197, 215, 223, 226, 257–267, 296, 324, 356  
Social links, 215, 224–226, 327  
Social media, 5, 6, 9–11, 13–15, 18, 20, 21, 23, 109, 112, 115, 117, 137, 177, 179, 180, 197, 200–202, 207, 208, 213–215, 229–251, 266, 267, 269–272, 275, 276, 278, 284, 323, 335, 338, 339, 343, 344, 347, 351, 352, 355, 356  
Social networks, 3, 4, 7, 9, 11, 15, 16, 47–52, 54–57, 59, 61, 62, 85, 97, 122, 125, 129, 134, 137–139, 143, 144, 151–172, 179, 180, 193, 198–201, 204, 208, 215, 217, 219, 222, 224, 226, 231–233, 243, 245, 246, 258, 267, 269, 282, 290, 291, 293, 296–298, 315, 323, 324, 329–332, 335–337, 343, 344, 347, 353, 355, 356  
Social spam campaigns, 230, 240–250  
Social spreading phenomena, 153
- Spillovers, 298, 300, 303, 309–313, 316, 337, 341, 343–345, 347  
Spreading condition, 76, 79  
Spreading dynamics, 109–122, 128, 129, 230, 284  
Statistical dependence, 58  
Strong ties, 3, 214, 215, 221–226, 324  
Subjective well-being, 263, 267  
Superspreaders, 75, 77
- T**  
Temporal networks, 110–114, 116, 117, 121, 215, 355  
Threshold, 5–7, 12–20, 34, 40–43, 69–79, 81, 83, 84, 86, 87, 90, 92, 97, 98, 101–106, 115, 130, 131, 134–136, 141, 142, 152–154, 156, 157, 159, 161–165, 167, 170–172, 236, 259  
Threshold contagion, 40–43  
Tie strength, 59, 214, 215, 217–223, 225, 226, 298, 341, 342, 346  
Traffic, 201, 214, 215, 219, 221–222, 224, 226, 324  
Tumblr, 197, 198, 200–203, 207, 208, 224, 232  
Twitter, 8–11, 14, 16, 132, 134, 137, 138, 153, 197, 198, 201–203, 205, 207, 208, 216–226, 230–233, 235, 236, 240–243, 248–251, 258–261, 270–275, 277, 284, 285, 297, 298, 336, 351, 353, 355
- V**  
Virality, 14, 179, 197–202, 205, 206, 208, 275
- W**  
Watts threshold model (WTM), 81, 82, 84, 92, 97, 154  
Weak ties, 3, 4, 12, 200, 213–227