

Modeling Behavioral and Epidemic Dynamics in Social Contact Networks

Kirti Jain¹, Vasudha Bhatnagar¹ and Sharanjit Kaur²

1. Department of Computer Science, University of Delhi, Delhi, India

2. Department of Computer Science, Acharya Narendra Dev College, University of Delhi, Delhi, India

Email: {kjain1,vbhatnagar}@cs.du.ac.in, sharanjitkaur@andc.du.ac.in

Abstract—Contagious diseases spread in population through the contact network and their spread is a function of the complex interplay of the biology of the contagion and behavior of the population.

In this paper, we present an *Individual-based Fear Model (IBFM)* that associates *fear-index* with the individuals to indicate the extent to which an individual follows health and hygiene protocols as a self-protective measure against disease. We also study the impact of the collective behavior of individuals in communities on the size and span of the epidemic. We propose a framework to create a wire-frame that mimics the social contact network of the population in a geography by lacing it with demographic information and improves the estimates of epidemic variables for network simulation. We test the effectiveness of the demography-laced contact network using real-life COVID data for two Indian states.

Since network simulations of epidemic spread are computationally expensive, we launch a systemic investigation into the possibility of predicting three epidemic variables, viz. *peak day*, *peak infections*, and *span* of the epidemic using the Regression Chain Model. We find that the predictions are fairly accurate for two out of three variables.

Index Terms—Epidemic dynamics, Epidemic spread model, Social contact network, Human behavior, Regression chain model

I. INTRODUCTION

The study of infectious disease dynamics is a challenging area of research. The dominant underlying reason is the influence of the patterns of human connectance and individual behavior apart from the biology of contagion, on the progression of the diseases in the population. Unlike differential equation-based modeling, which assumes homogeneity, network-based modeling is apt for understanding epidemic dynamics due to the inherent ability to capture heterogeneity of interactions [1]–[4]. The individual-based approach for epidemic spread modeling captures the heterogeneity of the contact patterns in network with *nodes* representing individuals and *edges* depicting interactions between individuals.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ASONAM '23, November 6-9, 2023, Kusadası, Turkey

© 2023 Association for Computing Machinery.

ACM ISBN 979-8-4007-0409-3/23/11...\$15.00

<https://doi.org/10.1145/3625007.3631605>

However, there are multiple extraneous factors like awareness of the disease, public health education, and risk, information dissemination by news media, religious and cultural congregations, etc. that influence the dynamics of disease spread. For example, the prevalence of the disease stimulates behavioral changes among individuals to safeguard themselves, thereby retarding the spread of the disease. The patterns of human connectance that arise due to population density variations and demography considerably influence the progression of infectious diseases. Ergo, it is prudent to develop mechanisms that factor in these aspects in network simulations of epidemic spread to improve estimates of crucial epidemic variables.

This work focuses on modeling realism and the development of models that incorporate social and behavioral aspects. Specific contributions to this Ph.D. work are listed below.

- 1) Framework for associating an index to measure individual behavior (*fear-index*) and integrating it with the spread of disease on the contact network (Sec. II).
- 2) Modeling the collective behavior of individuals in a community and understanding its influences on the course of an epidemic (Sec. III).
- 3) Framework for constructing close-to-reality social contact networks, where the interaction patterns are explicitly modeled in *family*, *social*, and *work* spaces using Census and demography data (Sec. IV).
- 4) We explore machine learning as an alternative to expensive network simulation to predict epidemic variables (Sec. V).

In the following sections, we describe the methodology for each contribution along with selected experimental results.

II. INDIVIDUAL-BASED FEAR MODEL

Since emotion of *fear* (of disease) is one of the characteristic attributes of an individual, *individual-based model* (IBM) is an appropriate instrument to integrate the spread of *fear* and *disease*. We propose an Individual-based Fear Model (IBFM) that integrates *fear-of-disease* and *spread-of-disease* [5]. The model furnishes a mathematical framework to quantify individual fear and reflect changes in behavior with time.

We use a compartmental-based SEIHRD (Susceptible-Exposed-Infected-Hospitalized-Recovered-Deceased) epidemic model [1]. Individuals in compartment (*S*) are *susceptible*, who when *exposed* to the pathogen, transit to

compartment E . *Infected* individuals move to compartment I). Some of the *infected* individuals move to *Hospitalized* compartment (H), while those who *recover* transit to compartment R . Individuals in compartment H , who *recover* from the illness move to compartment R , while those who unfortunately *die* move to compartment D . All transitions take place with a fixed probability, specific to the pathogen.

A. Methodology

1) *Archetypal model of fear*: The emotion of fear spreads in the population like a biological contagion, albeit with distinctly different dynamics. Both the “fear” and contagion are interdependent as the spread of “fear” influences the spread of “disease” and vice-versa in a feedback loop [6]. We state below the heuristics employed in the proposed fear model.

- 1) Each individual in the population possesses a *fear-index* denoted by ϕ_t^u at time t , that quantifies his/her fear level and characterizes the behavioral response to the epidemic. The high fear level experienced by an individual induces extremely cautious behavior and lowers the risk of infection. In contrast, low *fear-index* is associated with a casual attitude and reluctance to follow recommendations for safe behavior, resulting in the least precautions, and a high risk of infection.
- 2) The fear level of an individual changes over a period of time during an epidemic in accordance with the health status of the individual or his/her social contacts.
- 3) The probability of contracting the disease at a point in time depends on the *fear-index* of an individual. Thus, individual behavior overrides the basic (biological) probability of infection.
- 4) Cultural, religious, and social events in the population suppress individuals’ rational behavior, leading to modulation of their fear level.

Figure 1 shows a snapshot of a toy contact network of seven individuals in different disease states with their respective *fear-index* values as per the proposed IBFM. Susceptible and exposed individuals ($T, U, V, \text{ and } Z$) have medium fear levels, while W and X are infected and have higher fear levels. Recovered individual Y possesses a lower level of fear.

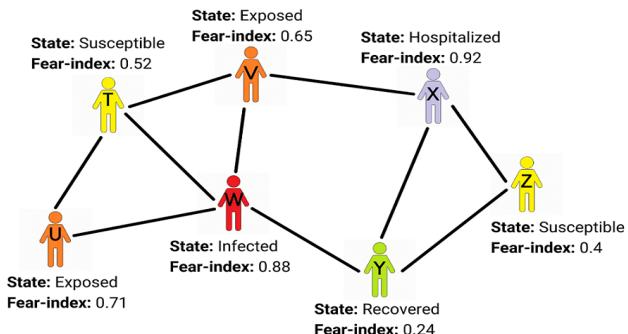


Fig. 1. Individual-based fear model of seven individuals showing their states and *fear-index* values.

2) *Initial fear in the population*: The underlying emotion of “fear”, intrinsic in the culture, explains the divergence in societal behaviors with respect to conformation to rules, regulations, and social norms in different societies. Considering the fear level of an individual as a continuous random variable (*fear-index*), we conjecture that cross-cultural variations in average fear level (μ_0) and heterogeneity in fear levels (σ_0) within a population are aptly modeled by normal distribution $N(\mu_0, \sigma_0)$. During initial phase of the epidemic innate fear level ϕ_0^u of an individual u is initialized following normal distribution ($\phi_0^u \sim N(\mu_0, \sigma_0)$).

3) *Individual transmission probability*: A pathogen spreads and infects susceptible individuals with base probability β . In IBFM, we model the probability of infection as a function of base probability and individual behavior. Let i denote the number of infected neighbors among all neighbors k of individual u . The high fraction of infections ($\frac{i}{k}$) in the neighborhood increases an individual’s fear ϕ_t^u of contracting the disease, in addition to the average fear of the population (μ_t). We model the perceived risk of infection (R_t^u) of an individual u at time t as

$$R_t^u = \exp \left[- \left(\mu_t + \phi_t^u * \frac{i}{k} \right) \right] \quad (1)$$

With an increase in the fraction of infected individuals in the neighborhood, individual fear increases, and consequently, the risk of infection decreases. An individual’s perceived risk of infection (R_t^u) alters the base transmission probability (β) to effective transmission probability $\hat{\beta}_t^u$ as

$$\hat{\beta}_t^u = 1 - [1 - \beta * R_t^u]^i \quad (2)$$

When fear is not taken into consideration, the effective transmission probability is computed as $\hat{\beta}_t^u = 1 - [1 - \beta]^i$.

4) *Fear dynamics among individuals*: The emotion of fear of an individual varies over time, conforming with the changes in the environment. Specifically, we consider the following fear dynamics among individuals.

- 1) *State-based fear update*: Change in *fear-index* due to state transition of an individual and/or her neighbor(s). Fear level increases with different degrees when an individual transits to I, H , or D compartments. When an individual recovers, the fear level decreases as he/she feels relieved. We update the *fear-index* ϕ_{t+1}^v of individual v ($v \in S, E$) having neighbor u ($u \in I, H, R, D$), by averaging their *fear-index* (Equation 3).

$$\phi_{t+1}^v = \frac{\phi_t^v + \phi_t^u}{2} \quad (3)$$

- 2) *Prevalence-based fear update*: Change in *fear-index* due to the rate of change of infection. An increase in the number of infected cases and fatalities instills fear in the population, while the decrease diminishes the emotion of fear. Let δ denote the rate of change of infections computed over a period of time (say, p). IBFM captures this phenomenon by updating *fear-index* of susceptible

and exposed individuals as follows.

$$\phi_{t+1}^u = \phi_t^u * e^\delta \quad (4)$$

Note that *fear-index* of individuals in states I , H , D , and R remain unaltered by the prevalence of the disease since they can not contract the disease again.

- 3) Event-based fear update: Change in *fear-index* due to social/political/religious events. This collective behavior in IBFM is shaped by uniformly reducing the fear of all susceptible and exposed individuals by a decay factor $\lambda(>0)$, according to Eq. 5.

$$\phi_{t+1}^u = \phi_t^u * e^{-\lambda} \quad (5)$$

B. Experimental Results

1) *Efficacy of IBFM*: We examine the impact of the initial societal fear on epidemic dynamics along with the absence of fear as a special case to demonstrate the efficacy of the proposed fear model. Figure 2 shows active cases for five distinct levels of initial fear, μ_0 as $\{0.1 (\text{very-low}), 0.3 (\text{low}), 0.5 (\text{medium}), 0.7 (\text{high}), 0.9 (\text{very-high})\}$ keeping individual heterogeneity same ($\sigma_0 = 0.15$).

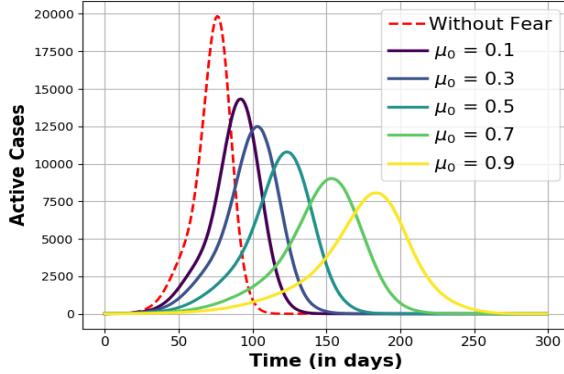


Fig. 2. Epidemic dynamics without fear and with different initial fear levels.

Our observations indicate that the epidemic curve corresponding to the “Without fear” scenario (represented by the red-dashed curve) exhibits distinct characteristics. In this scenario, disease transmission is solely influenced by the biological properties of the contagion, leading to an unimpeded spread throughout the population. Consequently, the maximum number of peak cases is observed in this scenario, and as the initial fear levels increase, the number of peak cases decreases. Furthermore, the impact of increasing initial fear levels is reflected in the shifting of the peak days to the right, indicating a delay in the timing of the epidemic peak. These results demonstrate that the proposed model is effective in distinguishing disease progression for varying initial fear levels within the population. It highlights how societal fear at the beginning of an epidemic can significantly influence the trajectory of the disease.

2) *Sensitivity Analysis*: We study the interplay between biological infectivity (i.e., base transmission probability β) and

initial fear (μ_0) in the population and plot the heatmap of peak active cases as shown in Fig. 3.

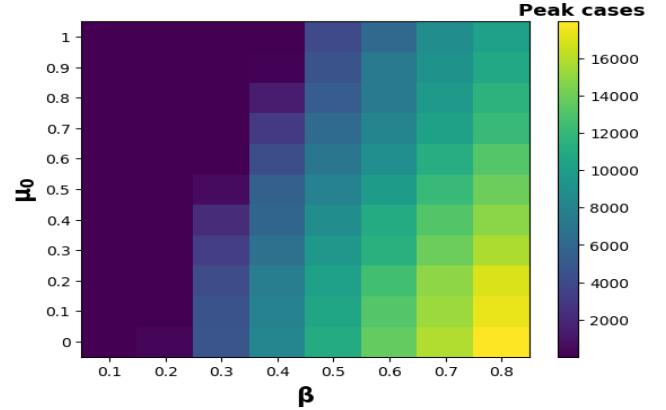


Fig. 3. Peak active cases with varying base transmission probability (β) and initial fear (μ_0).

The results of the experiment indicate that fear and the pathogen exert opposing forces on epidemic dynamics. When fear is low and the transmission probability is high, it results in an uncontrolled outbreak of the disease. Conversely, the epidemic dynamics undergo a significant shift when either fear decreases or the pathogen’s virulence weakens.

3) *Prevalence-based spread of fear*: We study how the convolution of the prevalence of disease and fear in the population develops the feedback mechanism, evincing multiple waves of the epidemic. We experiment using the periodic update function shown in Eq. 4, and vary base transmission probability β ($= 0.3, 0.4, 0.5, 0.6$) and periodicity p ($= 7, 14, 28$ days). We simulate epidemic spread for $\mu_0 = 0.5$ and plot seven days moving average of active cases in Fig. 4.

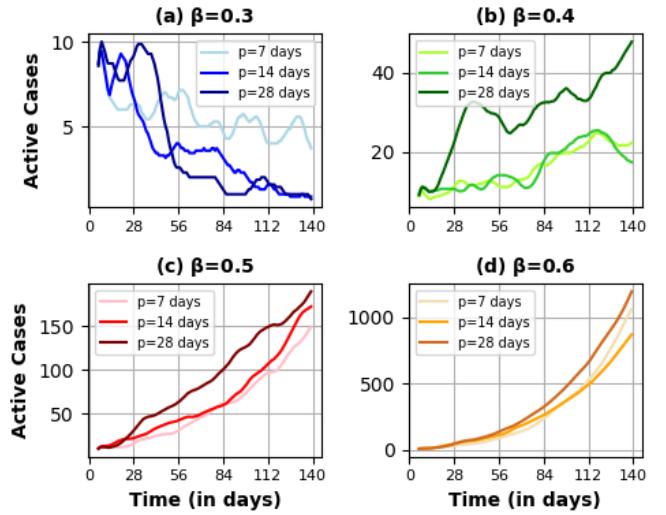


Fig. 4. Epidemic dynamics with periodic fear update for base transmission probability $\beta = \{0.3, 0.4, 0.5, 0.6\}$ and periodicity $p = \{7, 14, 28\}$ days.

Our observations reveal that when the base transmission probability of the disease is low, periodic updates in fear

levels lead to distinct and pronounced multiple waves of infection. However, as the transmission probability increases, these multiple waves become less prominent, and the epidemic dynamics are more subdued. Additionally, when the periodicity of fear updates increases, the feedback mechanism between fear and disease dynamics weakens. As a result, the epidemic waves become less frequent. This observation underscores the strong connection between population behavior driven by fear and the progression of the epidemic.

4) *Event-based spread of fear:* To demonstrate the impact of public events on the spread of the epidemic, we study the effect of mass gatherings by decaying the fear of susceptible and exposed individuals in the population. We simulate the mass gatherings on specific days, namely the 25th, 50th, 75th, 100th, and 125th day from the beginning of the epidemic. We use medium initial fear level, $\beta=0.6$, and model the effect of mass gatherings, we incorporated decay factors, denoted as λ , which were set to three different values: $\lambda = 0.1, 0.2, 0.3$.

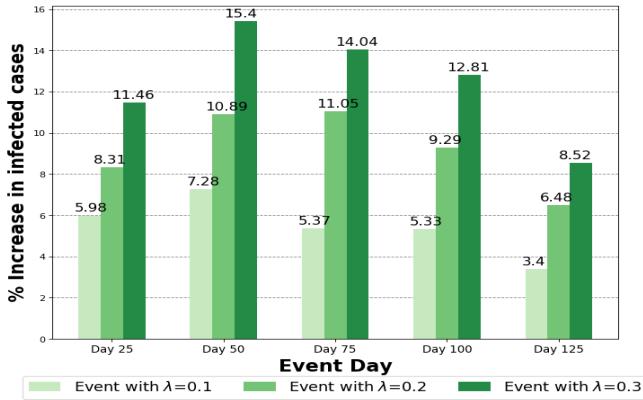


Fig. 5. Increase in number of infected cases (in %) in lieu of external events.

Figure 5 shows the percentage increase in the number of infected individuals compared to those reported in the absence of any event. The figure demonstrates that a more rapid decay in fear levels, as indicated by higher decay factors, accelerates the spread of the disease, which aligns with our expectations. This observation underscores the reliability of the IBFM in accurately representing changes in epidemic dynamics in response to sudden shifts in population behavior brought about by external events.

III. EPIDEMIC DYNAMICS IN COMMUNITY-STRUCTURED NETWORK

Disparities in societal cultures and innate psychological differences vary with communities and manifest the fear-index of an individual. The main objective is to show the importance of collective behavior in terms of prevailing fear emotion within a society that may trigger or control the disease spread and hence must be studied critically while projecting the epidemic variables [7].

A. Methodology

1) *Generation of community-structured network:* The stochastic block model (SBM) is a popular method for con-

structing graphs with clustering structure and high modularity [8]. SBM is a generative model for random graphs, which places N vertices into m communities such that pairs of nodes are connected with specified edge probabilities using a symmetric matrix P . The diagonal elements of P indicate cohesion within the community, and off-diagonal elements indicate coupling density between two communities.

2) *Initial fear in a community:* Initially, collective fear in community C_i is normally distributed ($N(\mu_0^i, \sigma)$) among individuals with mean μ_0^i and standard deviation σ .

We conjecture that the dispersion of epidemic spread is dependent both on the community structure and the mean behavior of individuals in the community. Highly aware sub-population suppresses the spread over the lackadaisical group by slowing down the transmission probability of the contagion.

B. Experimental Results

1) *Collective behavior in community and epidemic dynamics:* We consider two extreme initial fear levels (*very-low* and *very-high*) of the seeded community (C_1). Other communities (C_2-C_5) have *medium* initial fear level. The simulation begins at time $t = 0$, with the most central node in C_1 as the seed node. Figure 6 (a) and (b) show the snapshots of epidemic spread in five-community network with *very-low* and *very-high* initial fear levels respectively in C_1 .

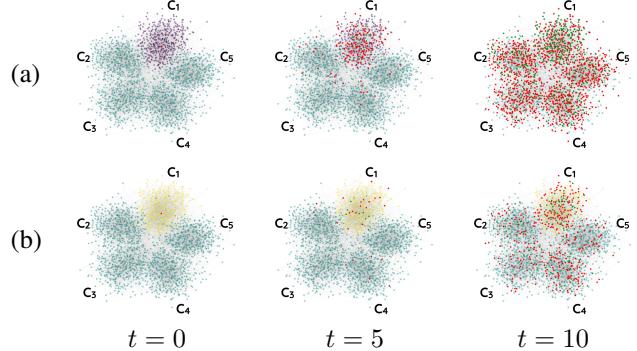


Fig. 6. Panels (a) and (b) show the snapshots of a community-structure network during the course of the epidemic for two initial fear levels, viz. *very-low* and *very-high* in community C_1 (nodes marked in violet and yellow colors respectively at $t = 0$). Other communities initially have *medium* fear (nodes marked in bluish-green color). Susceptible individuals are marked in the colors of their respective communities, whereas infected and recovered individuals are marked with red and green colors respectively.

The level of fear among individuals in community C_1 has a significant impact on the course of the epidemic, as illustrated in the two panels (a) and (b). In panel (a), where individuals in community C_1 exhibit a "very-low" level of fear (indicated by the violet color), the epidemic progresses more rapidly compared to panel (b), where the seeded community has a "very-high" fear level. We also observe that the infection is largely confined to the seeded community during the early stage of spread ($t = 5$) for both cases. During the early stage of the epidemic, we observe that the infection is primarily confined to the seeded community in both cases. A high number of infections in panel (a) is attributed to *very-low*

fear of individuals in the community, which elevates the probability of infection. As the epidemic advances to time $t = 10$, we observe that the infection begins to spread to other communities beyond the initially seeded community. In this context, the number of infected individuals (represented by the red-colored nodes) is notably higher in panel (a) compared to panel (b).

2) *Sensitivity to fear of seed node:* We conjecture that fear of the seed node plays an important role in controlling the spread. To validate this conjecture, we vary the fear of the seed node in steps of 0.1 while setting the initial fear of all the communities to *medium*. Figure 7 shows the fraction of total infected cases in the network at two-time instances $t = 3$ (represented by orange bars) and $t = 5$ (represented by blue bars). Additionally, the lines of the same color in the figure show the number of infected cases within the seeded community, with variations in the initial fear level of the infected seed node.

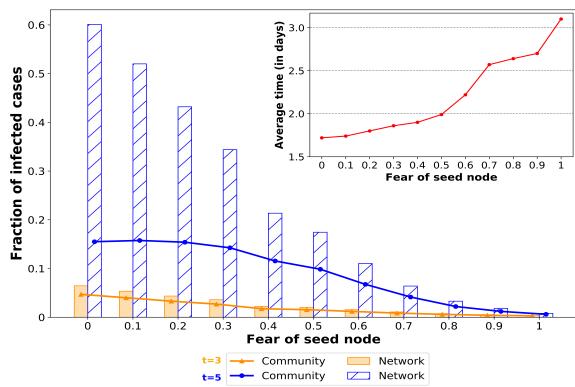


Fig. 7. Fraction of total infected cases at time $t = 3$ (orange bars) and $t = 5$ (blue bars), with varying fear of the seed node. The inset figure shows the average time (in days) for infection to reach outside the seeded community, with variations in the fear level of the seed node.

The figure illustrates that a low level of fear in the seed node leads to a higher fraction of cases in the early days of the epidemic (orange bars). As the fear level of the seed node increases, the transmission probability decreases, resulting in a reduced fraction of infected cases. At time $t = 3$, most of the infection remains within the community (orange line), with very few cases outside the initially seeded community (orange bar). However, by time $t = 5$, the infection has spread to other communities, resulting in a significantly higher fraction of infected cases (blue bar). This spread to other parts of the network diminishes considerably with an increase in the fear level of the seed node (blue line).

The inset in Fig. 7 reveals a significant trend: when the seed node has low fear, the infection spreads rapidly to other communities, taking only about two days. Conversely, when the fear level of the seed node increases, it slows down the infection's spread within the initially seeded community. As a result, the time it takes for the first infection to reach outside that community also increases.

3) *Community-wise infection spread:* We assess the effectiveness of the enforced restrictions in controlling the epidemic dynamics. We model different types of community behavior by setting different initial fear levels for the five communities $C_1 - C_5$ as *{very-low, low, medium, high and very-high}*. We show community-wise infections each day in Fig. 8.

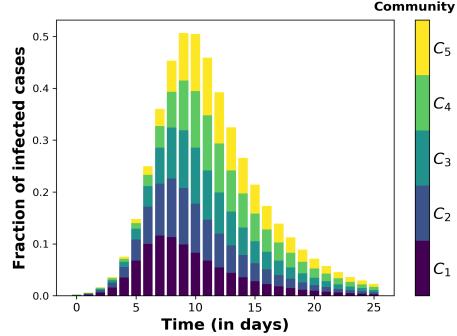


Fig. 8. Stacked bar of infections in communities $C_1 - C_5$ with different initial fear levels *{very-low, low, medium, high and very-high}*.

The findings highlight that communities with high levels of fear and compliance with restrictions may still experience a significant disease burden due to the non-compliance of other communities with low fear levels. Communities with low fear levels are ‘high-risk’ groups and should be targeted for awareness campaigns. Additionally, these groups should be prioritized for vaccination and other preventive measures to protect the overall well-being of the entire population.

IV. CENSUS-CALIBRATED MODULAR CONTACT NETWORK

We propose a framework to create a wire-frame, that proxies the social contact network of a population by considering the density and demography of the geography [9]. Since these two aspects directly impact human interactions, running simulations on a density- and demography-aware social contact network is expected to deliver more realistic estimates of the epidemic variables. The proposed framework is grounded in parsimony and uses basal demography data from the census. We estimate the interaction patterns by explicitly modeling social contacts in *family, social, and work* spaces.

A. Methodology

We represent the contact patterns of the population in a geographical unit (state/city/town) as an interconnection of multiple networks (modules), each corresponding to a sub-unit (zone) of the geography. Nodes in a module denote individuals in the zone, and the edges denote interactions (contacts) between individuals. Consider a geography of a toy city U comprising six zones as shown in Fig. 9 (a). The corresponding modular contact network of U is modeled as an interconnection of six networks (Fig. 9 (b)).

1) *Embedding Demography in Network:* Consider a geography U divided into m zones $\{z_1, \dots, z_m\}$ with compatible population vector \vec{P} , density vector $\vec{\Delta}$, household vector \vec{H} , and vector \vec{K} denoting mean household size of zones. Element

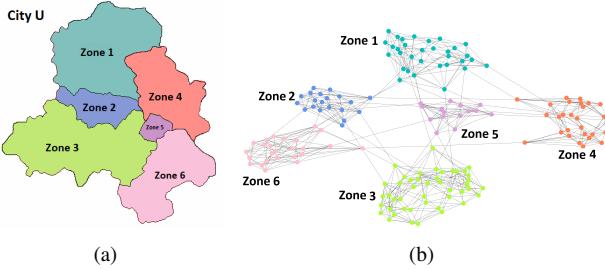


Fig. 9. a) Illustration of a geography U with six zones, b) Modular contact network of geography U modeled as an interconnected network.

p_i in \vec{P} denotes the population of zone z_i , δ_i in $\vec{\Delta}$ denotes its population density, h_i in \vec{H} denotes the number of households, and k_i in \vec{K} denotes the mean household size.

We generate small-world networks $\{G_1, \dots, G_m\}$ corresponding to zones $\{z_1, \dots, z_m\}$, with number of nodes equal to their respective populations. The degree of the regular base lattice for G_i is set to k_i (mean household size), and the dimension is based on the population density δ_i of z_i .

2) *Modeling interactions in family space:* We define a function $\mathcal{M} : \vec{\Delta} \longrightarrow \vec{D}$ to map the population density of a zone to the dimension of the base lattice of the corresponding small-world network (module) as follows.

$$d_i = \mathcal{M}(\delta_i) = \begin{cases} 1 & \text{if } \delta_i \leq \Delta_{avg} \\ 2 & \text{otherwise,} \end{cases} \quad (6)$$

where Δ_{avg} denotes the average density of m zones in the geography. Small-world networks thus created corresponding to the m zones embody variation in the spatial density of the geography U . The regular base lattice of dimension d_i (using Eq. 6) is created with p_i nodes of degree k_i , and is rewired with probability ρ to yield a small-world module G_i . In this way, we obtain m networks, each infused with the basic demographic information of the corresponding zones in U , with edges ensconcing interactions in *family* space. The network is subsequently loaded with edges representing contacts in *social* and *work* space.

3) *Modeling interactions in social space:* We consider social interactions as a function of the number of households in zone z_i . Element h_i in vector \vec{H} denotes the number of households in zone z_i . Since *social* interactions in geography are influenced by culture, demography, seasons, etc., sociability index τ provides the handle to capture these nuances in the wire-frame. Accordingly, τ is used as a multiplier to determine the number of social edges ($\tau \times h_i$) in G_i .

4) *Modeling interactions in work space:* We connect the modules (corresponding to zones) to create a realistic social contact network-of-networks for U . We use the coupling density vector \vec{Q} of size m to interconnect modules. Element q_i in \vec{Q} denotes the fraction of individuals in z_i denotes the size of the working population that mobilizes within and across zones for livelihood, thereby modeling contacts in *work* space. The inter-and intra-modular edges thus created represent interactions in the *work* space.

B. Case Study of Delhi and Goa

In this case study, we prove the validity of our conjecture that a demography-laced surrogate modular network delivers better estimates than one with random patterns of connectance. The generated wire-frames for states can be productively used as a tool for studying zonal outbreaks, mobility restrictions, vaccine administration, and other socioeconomic and behavioral scenarios to aid policy planners and administrators in general. We create surrogate social contact networks (SSCN) for the two Indian states (Delhi, and Goa). The selection of states is motivated by the diversity in the rural-urban agglomeration and population density. Delhi is the capital of India and the largest metropolitan city-state, with high urban agglomeration. Among small states, Goa has the largest urban population. Table I summarizes the basal details of two states.

TABLE I
TWO INDIAN STATES WITH DEMOGRAPHIC DETAILS AS PER CENSUS 2011.

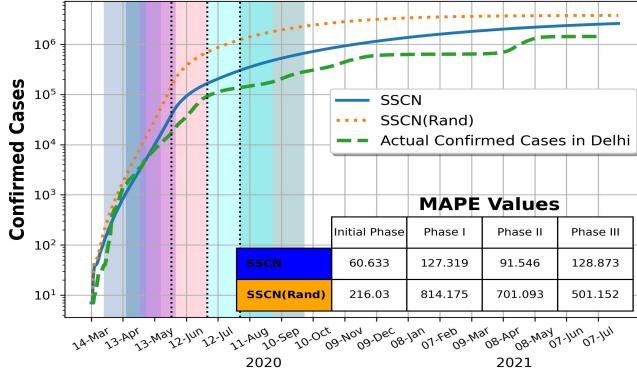
| State | Population | Area (km ²) | Population Density (people/km ²) | Urban Agglomeration (%) |
|-------|------------|-------------------------|--|-------------------------|
| Delhi | 16,787,941 | 1,483 | 11,320 | 97.50 |
| Goa | 1,458,545 | 3,702 | 394 | 62.17 |

We construct two networks specifically for *Delhi* and *Goa*, omitting demographic considerations. Within these networks, we introduce *social* and *work* contacts at random between the modules. These networks are labeled as SSCN(Rand). In Figure 10 (a and b), we present a comparison of cumulative confirmed cases predicted by the epidemic simulation with the actual confirmed cases for the two states. Notably, different events are marked with distinct colors in the figures to illustrate the non-pharmaceutical interventions (NPIs) that were implemented during the observed period.

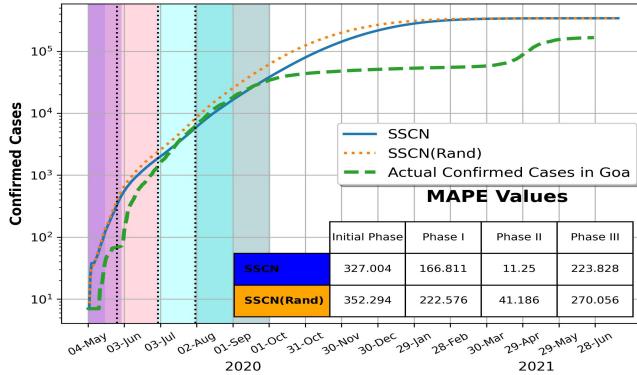
In both states, it is evident that the cumulative case curves for SSCN are more closely aligned with the actual curves compared to the SSCN (Rand) curves. This bolsters our confidence in the framework and strengthens our hypothesis that mimicking social contact patterns in the wire-frame is advantageous and gives better predictions.

We further strengthen the supposition by computing the errors for predicted variables by SSCN and SSCN(Rand) wire-frames. We compute mean absolute percentage error (MAPE) for both sets of epidemic variables for SSCN and SSCN(Rand) and display values in the inset table within Fig. 10. Our observations reveal that the error in predicting epidemic spread using the SSCN framework is lower, especially in the initial phases of the epidemic, which tend to be relatively short. In all three states, the error for Phase III, which covers a period of more than a year and represents a full-blown pandemic, is also consistently lower in the SSCN framework compared to SSCN(Rand).

Comparison of the epidemic curves on SSCN and SSCN(Rand) wire-frames testifies that a demography-laced



(a) SSCN vs SSCN(Rand) of Delhi



(b) SSCN vs SSCN(Rand) of Goa

Fig. 10. Epidemic curves for actual confirmed cases and those predicted by demography-based SSCN and SSCN(Rand) of Delhi and Goa. Inset tables show values of mean absolute percentage error (MAPE) for the phases.

surrogate social network for a geography delivers better estimates than the one with random patterns of connectance.

V. MACHINE LEARNING FOR PREDICTING EPIDEMIC VARIABLES

Real-time detection and forecasting of disease dynamics are critical for healthcare authorities during epidemics. We report a systemic investigation into the possibility of predicting three epidemic variables, viz. *peak day*, *peak infections*, and *span* using the Regression Chain Model as an alternative to computationally expensive epidemic simulations on large networks [10].

A. Methodology

1) Dataset Construction: In the absence of any real dataset for predicting epidemic variables, we curate a rich dataset called *EpiNet*, consisting of five network properties (features) and three epidemic variables (targets to be predicted) for networks with varying sizes (N) and average degrees ($\bar{k} \in [6, 40]$). We generate 15K small networks ($N \in [20K-60K]$), 10K medium networks ($N \in [60K-150K]$), and 10K large networks ($N \in [150K-300K]$), with an equal number of instances belonging to three network families- Random, Small-world, and Community-based networks [1].

We compute five topological properties, viz. average degree, normalized network density, degree variance, average clustering coefficient, and average shortest path length (approximation, as given by [11]) on each constructed network. Subsequently, we simulate SEIR epidemic spread for specific parameters and note three dependent epidemic variables, viz. peak day (in year), the fraction of peak infections, and the span of the epidemic (in year). Based on the network sizes, we split *EpiNet* into three partitions, corresponding to small networks (D-SN), medium networks (D-MN), and large networks (D-LN).

2) Regression Chain Model: A Regression Chain Model (RCM) is an ensemble built over a chain of regressors to capture dependencies between the output variables (interested readers may refer to [12] for an extensive account of RCM). We set the order for the regression chain as *peak day* (PD), *peak cases* (PC), and *span* (Span). Figure 11 shows the framework for the Regression Chain mechanism.

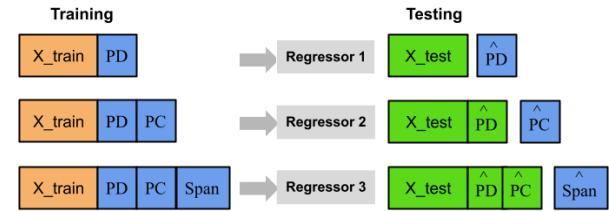


Fig. 11. RCM with the order as peak day (PD), peak cases (PC), and span (Span). Regressor 1 predicts PD, Regressor 2 predicts PC using predicted \hat{PD} and Regressor 3 predicts Span using predicted \hat{PD} and \hat{PC} .

We use Random Forest and XGBoost as representative base algorithms (estimators). To quantify the assessment of prediction quality, we use two performance metrics, viz. predicted coefficient of determination, $\langle R^2 \rangle$, and root mean squared error, $\langle RMSE \rangle$, both averaged over three target variables.

B. Experimental Results

1) Model Validation: We use ten-fold cross-validation method to assess the competence of RCM for all three training sets, followed by testing the model on unseen data. Figure 12 shows the heatmap of $\langle R^2 \rangle$ scores (lower triangle) and $\langle RMSE \rangle$ values (upper triangle) of the RCM using two base algorithms, (a) Random Forest, and (b) XG Boost.

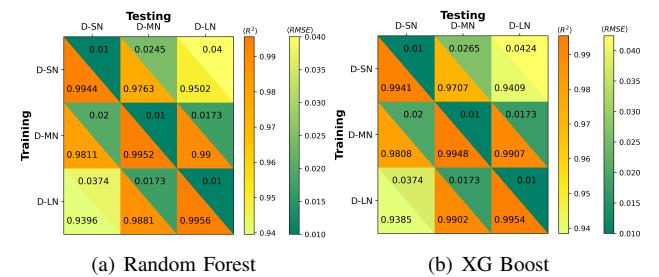


Fig. 12. Predictive performance of the RCMs trained and tested on topological properties of small (D-SN), medium (D-MN), and large (D-LN) networks.

The heatmap displays the cross-validated performance of both regressors, with the diagonal cells indicating high performance. The non-diagonal cells in the heatmap correspond to the performance of the model on unseen data. We observe that performance degrades marginally, i.e. low $\langle R^2 \rangle$ scores and high $\langle RMSE \rangle$, for the model trained on D-SN and tested D-LN, and vice versa. The overall high performance in all cases and for both base algorithms demonstrates the competence of the RCMs for predicting epidemic variables in networks.

2) *Sensitivity Analysis*: Having observed that the model performance degrades for networks of dis-similar sizes, we examine the sensitivity of the predicted variable to the size of the network. We train the model using D-SN and test the model on the pooled D-MN, and D-LN data sets. We group instances into eight batches (Figure 13) with approximately equal numbers of records per batch and report batch-wise R^2 score and RMSE value for Random Forest-based RCM.

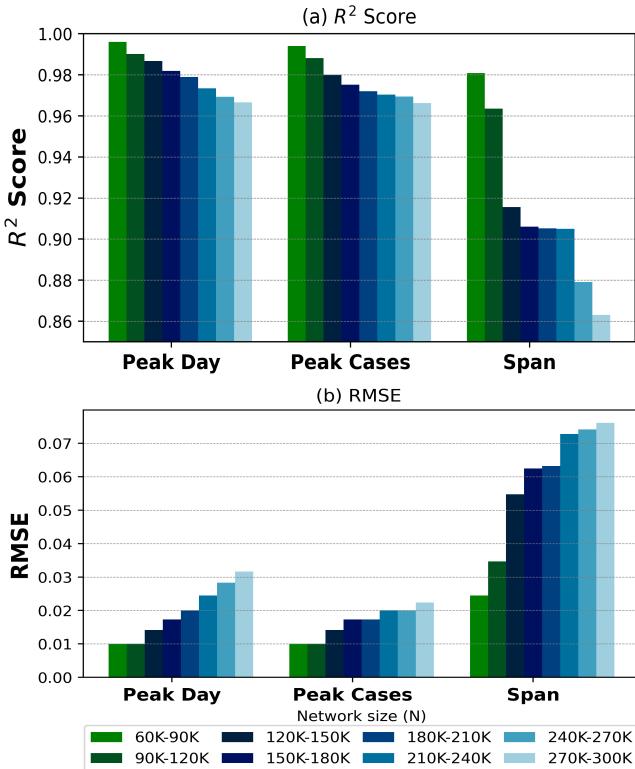


Fig. 13. Predictive accuracy for three epidemic variables for Random Forest-based regressor trained on small networks (D-SN) and tested on larger networks (D-MN and D-LN) grouped by network size.

We observe a marginal decrease in R^2 scores and a marginal increase in RMSE values for *peak days* and *peak cases* as network size increases. However, the R^2 score degrades notably for the *span* variable for larger networks ($N > 120K$). This observation ratifies our earlier conclusion that the regression chain model trained using topological properties of small networks (D-SN) is capable of reliably predicting *peak day* and *peak cases* for medium and large networks, thereby saving computational expense incurred by epidemic simulations. Prediction accuracy for the *span* variable is lower

and further study is required to understand the role of other topological properties on the *span* of the epidemic.

VI. CONCLUSION AND FUTURE WORK

This Ph.D. project focuses on modeling realism and the development of models that incorporate social and behavioral aspects. We propose an Individual-based Fear Model (IBFM) that couples the *fear-of-disease* with the *spread-of-disease* and demonstrates the impact of human behavior on epidemic spread. IBFM is potentially an effective tool for studying information diffusion applications on social media platforms and digital marketing campaigns, which are primarily governed by human behavior. We also study the complex interplay of collective behavior in community-structured networks and epidemic dynamics. For realistic estimates of epidemic spread, we propose a framework for creating a surrogate social contact network that embodies the contact patterns arising from variations in population density and demography of the constituent zones in a geography. A case study using the geography and demography of two Indian states is presented to prove the efficacy of the proposed demography-laced network for studying epidemic dynamics.

We also explore the possibility of predicting epidemic variables using a Regression Chain Model trained on the topological properties of the underlying contact networks as a substitute for costly epidemic simulations on large networks. We plan to extend the proposed *Individual-based Fear Model* on a multi-layer network to expand its applicability to incorporate more extraneous factors that influence spread.

REFERENCES

- [1] Albert-László Barabási. *Network science book*. Cambridge University Press, Cambridge, 2014.
- [2] Mark DF Shirley and Steve P Rushton. The impacts of network topology on disease spread. *Ecological Complexity*, 2(3):287–299, 2005.
- [3] Matt J. Keeling and Ken T. D. Eames. Networks and epidemic models. *Journal of The Royal Society Interface*, 2(4):295–307, 2005.
- [4] Michael Small and David Cavanagh. Modelling Strong Control Measures for Epidemic Propagation with Networks - A COVID-19 Case study. *IEEE Access*, 8, 2020.
- [5] Kirti Jain, Vasudha Bhatnagar, Sadanand Prasad, and Sharanjit Kaur. Coupling fear and contagion for modeling epidemic dynamics. *IEEE Transactions on Network Science and Engineering*, pages 1–14, 2022.
- [6] Joshua M. Epstein, Jon Parker, Derek Cummings, and Ross A. Hammond. Coupled contagion dynamics of fear and disease: Mathematical and computational explorations. *PLoS ONE*, 3(12):1–11, 12 2008.
- [7] Kirti Jain, Vasudha Bhatnagar, and Sharanjit Kaur. Collective behavior in community-structured network and epidemic dynamics. In *Proceedings of IHIC 2022*, pages 167–183. Springer, 2023.
- [8] Paul W Holland, Kathryn B Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.
- [9] Kirti Jain, Vasudha Bhatnagar, and Sharanjit Kaur. Epidemic dynamics in census-calibrated modular contact network. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 12(1):14, 2023.
- [10] Kirti Jain, Vasudha Bhatnagar, and Sharanjit Kaur. Regression chain model for predicting epidemic variables. In *Proceedings of SBP-BRIMS*, pages 285–294. Springer, 2023.
- [11] Reginald D Smith. Average path length in complex networks: Patterns and predictions. *arXiv preprint arXiv:0710.2947*, 2007.
- [12] Hanen Borchani, Gherardo Varando, Concha Bielza, and Pedro Larraaga. A survey on multi-output regression. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 5(5):216–233, 2015.