



Applying a Probabilistic Infection Model for studying contagion processes in contact networks

William Qian^a, Sanjukta Bhowmick^{b,*}, Marty O'Neill^b, Suhasini Ramisetty-Mikler^c,
Armin R. Mikler^c

^a University of Pennsylvania

^b University of North Texas

^c Georgia State University

ARTICLE INFO

Keywords:

Computational epidemics
Outbreak simulation
SEIR model

ABSTRACT

Modeling the spread of infectious diseases is central to the field of computational epidemiology. Two prominent approaches to modeling the contagion process include (i) simulating the spread in contact networks through Monte-Carlo processes and (ii) tracking the disease dynamics using meta-population models. In both cases, the individuals are explicitly (contact networks) or implicitly (meta-population) assumed to belong to exactly one disease state (e.g., susceptible, infected, etc.).

In reality, the disease states of individuals are rarely so cleanly compartmentalized. A particular agent can exist in multiple disease states (such as infected and exposed) concurrently with varying probability. To model this stochasticity, we present a new method, that we term as the Probabilistic Infection Model (PIM). Unlike traditional models that assign exactly one state to each agent at each time step, the PIM computes the probability of each agent being in each of the infectious states.

Our proposed PIM provides a more layered understanding of the dynamics of the outbreak at individual levels, by allowing the users to (i) estimate the value of R_0 at individual vertices and (ii) instead of an all or none value, provides the probability of each infected state of an agent. Additionally, using our probabilistic approach the overall trajectories of the outbreaks can be computed in one simulation, as opposed to the numerous (order of hundreds) repeated simulations required for the Monte Carlo process.

We demonstrate the efficacy of PIM by comparing the results of the PIM simulations with those obtained by simulating stochastic SEIR models, as well as the time required for the simulations. We present results at the system and at the individual levels for three diseases; measles and two strains of influenza. We demonstrate how the PIM can be used to study the effect of varying the transmissibility of COVID-19 on its outbreak.

This paper is an extended version of a manuscript published in the proceedings of the 2020 International Conference on Computational Science (ICCS)[30]. These extensions are primarily within Sections 4 (Relationship between graph structure and probability of infection) and 5 (Effect of varying COVID-19 transmissibility on outbreak dynamics).

1. Introduction

A primary component of computational epidemics is modeling and simulating how infections spread in a population. Two main approaches to simulating the spread of disease are (i) stochastic agent-based modelling; and (ii) deterministic meta-population models [1,14].

Both models assume that the individuals are in exactly one disease

state. For example, the SEIR model, which we simulate in this paper, the states are Susceptible, Exposed, Infected, and Recovered. This framework is¹ depicted in Fig. 1. S , E , I , and R represent the number of individuals in Susceptible, Exposed, Infected, and Recovered states respectively. The total population is then given by $N = S + E + I + R$. Parameter β is the proportion of contacts between members of S and members of E that lead to disease transmission. Parameter σ is the rate at

* Corresponding author.

E-mail addresses: bqqian@sas.upenn.edu (W. Qian), sanjukta.bhowmick@unt.edu (S. Bhowmick), Marty.ONeill@unt.edu (M. O'Neill), sramisettymikler@gsu.edu (S. Ramisetty-Mikler), amikler@gsu.edu (A.R. Mikler).

¹ Births and deaths among the population are not considered in the model

which the exposed become infected. Parameter γ is the recovery rate at which the infected transition to the recovered state.

In *stochastic agent-based models* each individual (or group of individuals) in a population is represented as *agents*. Dyadic interactions between agents are governed by functions of the agents' characteristics, or their environment. These interactions can be used to form a contact network. The infection spreads through the connections in this network. *Meta-population models* use a system of differential equations to approximate the rate of change of the number of individuals in each disease state (e.g., susceptible, infected, etc.). Here the specific connections between the individuals is not modeled.

Both these models exhibit competing benefits and drawbacks. The advantage of the stochastic agent-based approach is that it can model population heterogeneity including variations in the numbers of contacts of each individual, as well as by varying the infection parameters, such as γ , σ , per individual. The disadvantage is that due to the reliance on stochastic processes, a single run of an outbreak simulation is not representative of an expected outcome. Hundreds of repeated executions per unique set of parameters are needed in order to adequately estimate trends in the data.

In an almost exact reverse, meta-population disease models are computationally efficient due to their deterministic nature. Further, closed form approximations of significant epidemiological parameters such as the basic reproduction number R_0 (i.e. the expected number of secondary cases resulting from a single infectious individual in a completely susceptible population) can be derived analytically using meta-population models. But these models do not represent the diversity of the individuals, and assume a homogeneous mixing rate within a homogeneous population. Motivated by these trade-offs, our goal is to combine the advantages of these two popular epidemiological models. We introduce the Probabilistic Infection Model (PIM), which combines the heterogeneity of the stochastic models with the computational efficiency and deterministic nature of the meta-population models. The **key idea** of PIM is to calculate the probabilities of the four SEIR states associated with that vertex for each vertex in a contact network.

To compute the probability function, we leverage the research conducted in escape probabilities by Thomas and Weber [32]. The probabilities for each state and each vertex are compounded over windows of time corresponding to the latent and infectious periods of the given disease. This allows for probabilistic values of different states over time at the individual levels and also provides the expected values of the sizes of the SEIR sub-populations corresponding to each state. As an added advantage, our proposed PIM allows us to compute an expression for $R_0(v_0)$, which yields the value of R_0 for specific single infectious individuals in an otherwise susceptible contact network. In Table 1 we provide a comparison between the stochastic model, the meta-population model and our proposed PIM.

We applied our model to a contact network created from class enrollment data from the University of North Texas, as well as on two other contact networks that are available online; (i) on a network of friendship of students in high school and (ii) a network of students living in a residential hall. We conducted our experiments by simulating the following epidemics; two varieties of influenza, measles, and Covid-19. We compared simulation results as well as the timing of the PIM with

those produced by the stochastic models. Our results demonstrate that the PIM simulations are similar to those produced by averaging trials from Monte Carlo models. This similarity is most notable when simulating diseases that are highly infectious, such as measles.

2. The Probabilistic Infection Model

In this section we describe our novel Probabilistic Infection Model (PIM). In Table 2, we provide a list of the terms that we use in our computations, along with their definitions. The input to both the stochastic model (SM) and the PIM is a contact network among individuals. In the SM model, a contact event is simulated by a vertex selecting a *single neighbor with a given probability*. Due to this inherent stochasticity of the model, the simulation must be executed multiple times to estimate how population sizes for each SEIR state change over simulated time.

In our Probabilistic Infection Model, *all neighbors of a specific vertex have a probability to make a contact*. For any given contact event, we set the contact probability per pair of vertices to be proportional to the weight of their corresponding edge. The probability that vertex v will be contacted by vertex u as a result of a single contact expended by u is $\Psi(u, v) = \frac{w(u, v)}{\sum_{x \in N(u)} w(u, x)}$; $w(u, v)$ is the weight of the edge (u, v) and $N(v)$ is the set of neighbors of vertex v . This function is not commutative. The probability of a contact from vertex u to vertex v , will differ from the probability of a contact from vertex v to vertex u , depending on each vertex's number of neighbors and weights of the adjacent edges.

Each time v is contacted by an infectious individual u , there is a transmission probability $T(u, v)$. The probability that vertex v is infected by u on day t as a result of a single contact made by u is then given by

$$\delta_t(u, v) = \Psi(u, v) \cdot I_t(u) \cdot T(u, v) \quad (1)$$

i.e. the product of the probability of contact between u and v , the probability the u is infected on day t , and the transmission probability between u and v .

Lemma 1. Given that a vertex v is in the exposed state, i.e. $E_x(v) > 0$ and $I_x(v) = 0$ on day x , v will have $I_t(v) > 0$, i.e. be in an infectious state on day t for some $t > x$, only if it was contacted by an infectious vertex within the critical infection window $[t - (\gamma_v + \sigma_v) + 1, t - \sigma_v]$.

Proof. We note that since each partial infection received by v has a latent period σ_v , the infection probability of v , for a day r prior to day t , will remain unchanged for $t - \sigma_v + 1 \leq r \leq t$. Moreover, because the infectious period is γ_v , any infections that arose from interactions made by v on or before day $t - (\gamma_v + \sigma_v)$ would have expired by day t . Thus, taking these together, the time between $t - (\gamma_v + \sigma_v) + 1$ and $t - \sigma_v$ is the **critical infection window** where an infectious contact will take v to an infectious state on day t . $\square$

Fig. 2 depicts how this critical window affects the state of the vertex. For ease of explanation, we consider the probabilities in this example to be 0 or 1. Consider the vertex v to be in an exposed state ($E_x(v) = 1$). In case 1, if an infectious contact occurs within the critical infection window, then v will be in an infected state ($I_t(v) = 1$) on day t . If, case 2, the infectious contact occurs after the critical infection window then v will

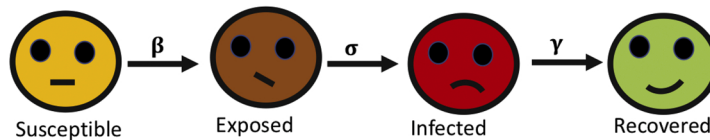


Fig. 1. A pictorial representation of the SEIR model, along with the modeling equations.

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta SI}{N} \\ \frac{dE}{dt} &= \frac{\beta SI}{N} - \sigma E \\ \frac{dI}{dt} &= \sigma E - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned}$$

Table 1

Comparison of the properties of the meta population model, the stochastic model and our proposed Probabilistic Infection Model.

Property	Meta population model	Stochastic model	Probabilistic Infection Model
Diversity of interactions	Does not model diversity of interactions	Can model diversity of interactions	Can model diversity of interactions
Execution time	Deterministic. Requires only one execution	Stochastic. Requires Multiple executions	Captures probability of infection. Requires only one execution
Measuring spread of infection	Provides cumulative number of individuals per state of infection	Provides state of infection of each individual	Provides state of infection of each individual
R_0 computation	Can be computed	Needs multiple simulations to be computed or approximated based on degree distribution	Can be computed

Table 2

Notation used in equations.

Notation	Definition
$S_t(v)$	Probability that a vertex v is susceptible on day t
$E_t(v)$	Probability that a vertex v is exposed on day t
$I_t(v)$	Probability that a vertex v is infectious on day t
$R_t(v)$	Probability that a vertex v is recovered on day t
$N(v)$	Set of neighbors of vertex v
σ_v	The incubation period, time between exposed to infected state, for vertex v
γ_v	The infectious period, time between infected to recovered state, for vertex v
$\Omega_t(v)$	The number of contacts that vertex v makes on day t
$\Psi(u, v)$	Probability that vertex u contacts vertex v as a result of a single contact expended by u
$\delta_t(u, v)$	Probability that vertex v is infected by u on day t as a result of a single contact expended by u
$T(u, v)$	Probability that an infectious vertex u infects vertex v upon contact

remain in exposed state ($E_t(v) = 1$) on day t . If, case 3, the infectious contact occurs before the critical infection window then v will be in recovered state ($R_t(v) = 1$) on day t .

2.1. Computing the probability for each state

We now derive the expressions for computing the probability of each state for a given vertex v and a day t . We assume at the beginning of the simulation, i.e. at day 0, all vertices are either completely (with 100% probability) in the susceptible state or in the infected state.

Let $\Omega_t(u)$ denote the number of contacts that u makes on day t . The probability of v not being infected due to one contact made by u on day t is $1 - \delta_t(u, v)$. Taking all neighbors of v , the probability that v is not infected by any of the neighbors is $\prod_{u \in N(v)} (1 - \delta_t(u, v))^{\Omega_t(u)}$, where we make

the approximation that each event where vertex v is not infected by some contact is independent.

Susceptible state: The probability that the vertex is in a susceptible state is the probability that v is not infected by any of the neighbors since day 0 to current day t . Thus;

$$S_t(v) = \prod_{n=0}^t \prod_{u \in N(v)} (1 - \delta_n(u, v))^{\Omega_n(u)} \quad (2)$$

Exposed state: Any susceptible vertex that was infected σ_v (the incubation period) days earlier will be exposed. Thus the probability of the exposed state is the probability of being in the susceptible state on day $\max(0, t - \sigma_v)$ minus the current probability of the susceptible state on day t .

$$E_t(v) = S_{\max(0, t - \sigma_v)}(v) - S_t(v) \quad (3)$$

Infectious state: Any susceptible vertex that was infected $\sigma_v + \gamma_v$ (the incubation period + infectious period) days earlier will be in an infectious state. The probability of the exposed state is the probability of being in the susceptible state on day $\max(0, t - \sigma_v)$ minus the current probability of the exposed state on day t .

$$I_t(v) = S_{\max(0, t - (\sigma_v + \gamma_v))}(v) - S_{\max(0, t - \sigma_v)}(v) \quad (4)$$

Recovered state: Any susceptible vertex that was infected before the critical infection window $t - (\sigma_v + \gamma_v)$ will have recovered by day t . The probability of the recovered state is 1 minus the probability that the vertex was still susceptible $\gamma_v + \sigma_v$ days prior.

$$R_t(v) = 1 - S_{\max(0, t - \gamma_v - \sigma_v)}(v) \quad (5)$$

The total number of individuals ever infected at the end of an outbreak can be computed by several methods. One method is to take the expected number of recovered individuals by summing over $R_L(v)$ for all v , where L is the last day of the simulation. Another way to approximate this quantity is to integrate the expected number of infected individuals $\sum_{v \in V(G)} I_t(v)$ over time and divide the result by the

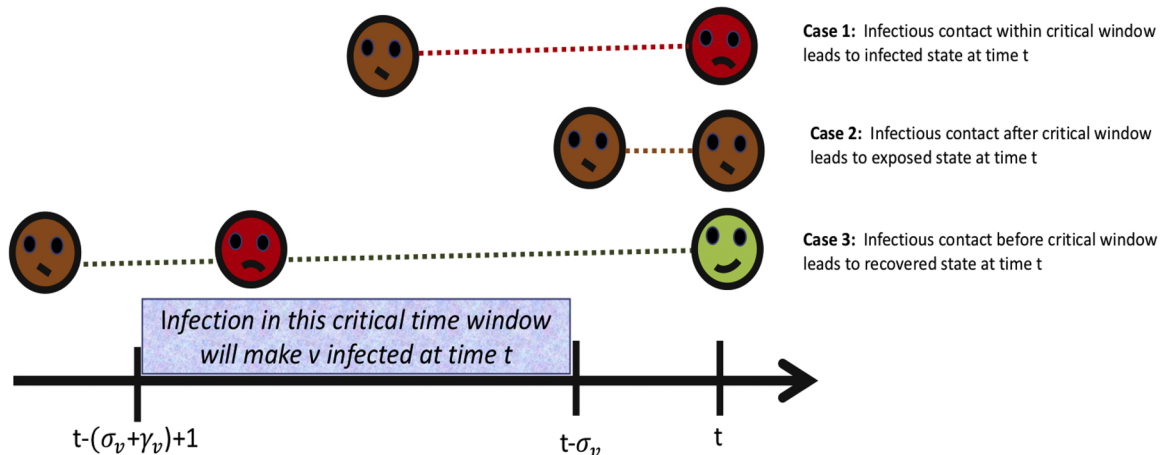


Fig. 2. A pictorial representation of the duration of infections with respect to the critical infection window.

disease's infectious period to account for over-counting. Since time is counted in discrete steps, this integral can be reduced to a sum.

Thus, given an outbreak of length L in days;

$$\sum_{v \in V(G)} R_L(v) \approx \sum_{n=0}^L \sum_{v \in V(G)} \frac{1}{\gamma_v} I_n(v) \quad (6)$$

This is satisfied in standard Monte Carlo models as well as in our PIM model.

Moreover, using PIM, we can calculate the value of the basic reproduction rate, R_0 , for a specific single infectious individual v_0 in a contact network where all other vertices are susceptible, as follows:

$$R_0(v_0) = \sum_{v \in N(v_0)} (1 - \prod_{n=0}^{\gamma_{v_0}-1} (1 - T(v_0, v) \Psi(v_0, v)^{\Omega_n(v_0)})) \quad (7)$$

Here the $\delta()$ is replaced by the product of transmission and contact probabilities, as $I_n(v_0) = 1$ for $0 \leq n < \gamma_{v_0}$.

2.2. Infection redundancy correction

One critical issue in using the PIM model is the effect of infection redundancy. This problem is illustrated in Fig. 3. Consider on day t , vertex v is exposed to the infection $\delta_t(u, v)$ through contact with vertex u . Once v reaches an infected state on day $t + \sigma_v$, it will expose vertex u to the infection $\delta_{t+\sigma_v}(v, u)$. However, note that some of the infections contributing to the value of $I_{t+\sigma_v}(v)$ have originated from u . This will result in u compounding its own probability of being infected, by incurring these redundant infections.

In order to correct this effect, we modify the infection from vertex u to vertex v by correcting each $\delta_t(u, v)$ to only factor in u 's probability of being infectious as a result of contacts from vertices other than v . This ensures that infections originating from u will not be returned to u by any of u 's direct neighbors. Making this correction will improve the accuracy provided by PIM at the expense of computation time.

To calculate this, consider

$$X = \prod_{n=\max(0, t-(\gamma_u-\sigma_u-1))}^{\max(0, t-\sigma_u)} \prod_{s \in N(u)} (1 - \delta_n(s, u))^{\Omega_n(s)}$$

and

$$Y = \prod_{n=\max(0, t-(\gamma_u-\sigma_u-1))}^{\max(0, t-\sigma_u)} (1 - \delta_n(v, u))^{\Omega_n(v)}$$

Here, X represents the probability that u was not infected in the critical infectious window by any of its neighbors (using the same logic as calculating for $S_t(v)$ earlier). Y represents the probability that u was not infected in the critical infectious window by vertex v . Since the values are given as products, the ratio of $\frac{X}{Y}$ approximates the probability that u was *not infected* in the critical infectious window by any of its neighbors *and* also discards the effect of infections from v . The probability that u is infected as a result of contacts with vertices other than v is then given by $1 - \frac{X}{Y}$. We thus modify the probability that v is infected by u on day t as a result of a single contact made by u to obtain

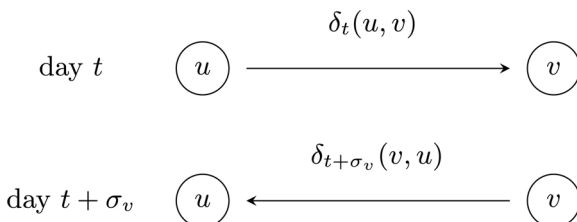


Fig. 3. An illustration of the infection redundancy problem.

$$\delta_t(u, v) = \Psi(u, v) \cdot T(u, v) \cdot \left(1 - \frac{\prod_{n=\max(0, t-\gamma_u-\sigma_u+1)}^{\max(0, t-\sigma_u)} \prod_{s \in N(u)} (1 - \delta_n(s, u))^{\Omega_n(s)}}{\prod_{n=\max(0, t-\gamma_u-\sigma_u+1)}^{\max(0, t-\sigma_u)} (1 - \delta_n(v, u))^{\Omega_n(v)}} \right) \quad (8)$$

where the factor representing the probability that u was infectious on day t has been modified to prevent infection redundancy. We note that this is an approximate correction, as it is still possible for an infection to return to its source after passing through multiple vertices. Since an infection moving down a path of vertices gets exponentially smaller in magnitude as the length of the path increases, it is expected that the effect would be increasingly negligible for higher order corrections.

Variance in results: The probability of vertices in each state, as computed by the PIM model can vary due to randomness of transmission, randomness of infectious period, etc. We provide an analytical formulation of the variance on the number of infected individuals as follows;

Let X_t be the random variable denoting the number of individuals in the infectious state on day t , and let X_t^v be an indicator random variable for whether vertex v is in the infectious state on day t .

From $X_t = \sum_{v \in V(G)} X_t^v$, $\text{Var}(X_t) = \sum_{v \in V(G)} \text{Var}(X_t^v) + \sum_{u \neq v} \text{Cov}(X_t^u, X_t^v)$. $\text{Var}(X_t^v)$ can be determined exactly as $\text{Var}(X_t^v) = I_t(v)(1 - I_t(v))$, whereas the covariance terms can be bounded by $\text{Cov}(X_t^u, X_t^v) \leq \sqrt{\text{Var}(X_t^u) \text{Var}(X_t^v)}$. From these relations, we obtain an upper bound $\widehat{\text{SD}}^2$ on the variance of the number of infectious individuals on day t given by

$$\widehat{\text{SD}}^2 = \sum_{v \in V(G)} I_t(v)(1 - I_t(v)) + \sum_{u \neq v} \sqrt{I_t(u)(1 - I_t(u)) I_t(v)(1 - I_t(v))} \quad (9)$$

3. Empirical results

In this section we present our experimental results of comparing the simulation of PIM with the stochastic Monte-Carlo simulations.

Datasets used. Creating a reliable contact network is challenging in computational epidemiology [9]. This is because such as traditional methods of determining contacts such as surveys or sensor based tracking cannot scale. Surveys are also affected by recall bias, because participants may not remember all of their contacts.

As a solution to this problem, we observe that many of the daily routines of individuals are based on scheduled activities, such as going to meetings, going to appointments, attending classes etc. Available information of scheduled activities allows us to create a reliable network of most of the frequently occurring contacts. Based on this assumption, we created a contact network of students based on the class-enrollment data for the Fall 2016 semester at the Discovery Park campus of the University of North Texas.

We created a contact network of students based on the class-enrollment data for the Fall 2016 semester at the Discovery Park campus of the University of North Texas (this network will henceforth be referred to as UNT). The dataset contained randomly generated student ids and the classes in which each student was enrolled. Online classes and classes without regular meeting times were excluded. From these data, we constructed a graph where each student was a vertex, and two vertices (students) were connected by an edge if the corresponding students shared a class. The weight of an edge was the average duration of shared class time between the students. This was an undirected weighted network with 3700 vertices and 195073 edges.

3.1. Experiment parameters

We simulated two types of influenza and measles on this contact networks with the disease-specific parameters given in Table 3. The incubation and infection rates, measles-specific parameters were adapted from [28,10], whereas influenza-specific parameters were

Table 3

The parameters used in simulations.

Disease	Incubation Period (σ) in days	Disease parameters		Contacts per hour ($\Omega_i(v)$)/(class time)
		Infectious Period (γ) in days	Transmission probability (T)	
Measles	8	5	.9	3
Influenza 1	2	5	.1	3
Influenza 2	1	3	.1	3

adapted from [3,6,8]. Two sets of parameters were chosen for influenza that varied in length of incubation and infectious periods. We used the same values of σ , γ and T for all vertices and edges.

In PIM simulations, a single vertex v_0 was selected to be infected, with $I_n(v_0) = 1$ for $0 \leq n < \gamma$, and $R_n(v_0) = 1$ for $n \geq \gamma$. The remaining vertices were initially completely susceptible. The probability values of the states of each vertex were obtained by computing the functions given in Equations (2–5) over the time period. The number of infected individuals at time t in days was determined by summing over $I_t(v)$ for all $v \in V(G)$. We terminated each simulation after day t if the outbreak activity was sufficiently small, i.e. the total number of vertices with high probability of exposed and infected states was small. We quantitatively measured this using the following conditions:

$$\sum_{v \in V(G)} E_t(v) + I_t(v) \leq 0.5$$

$$\left| \sum_{v \in V(G)} (E_t(v) + I_t(v)) - \sum_{v \in V(G)} (E_{t-1}(v) + I_{t-1}(v)) \right| \leq 0.5$$

The simulations were terminated if both these conditions were satisfied. In addition, simulations were not terminated before day 20. These bounds were selected to ensure that simulations do not end prematurely. Fig. 4 shows the state of the vertices in the UNT network as per the PIM model, on day 35. As can be seen, the measles epidemic spreads faster and takes longer time to recover (more red and less green nodes) than the influenza models.

In simulations using the stochastic model, the same graph, seed vertex of infection and parameters were used. 100 trials were run with a seeded random number generator for each of the three disease parameters. Contacts between vertices occurred randomly, with the probability of contact between vertices u and v for any given contact event

proportional to $w(u, v)$. Disease transmission occurred with probability T at the time of a successful contact between a susceptible and infectious individual.

3.2. Results

Our experiments demonstrate that PIM produces results most similar to those produced by stochastic Monte Carlo models for diseases that are more highly infectious. As seen in Table 3, the Monte Carlo model and PIM produced similar values for the total number of infected individuals in an outbreak. Additionally, while the peak number of infected individuals and day of peak infection produced by PIM tended to be within one standard deviation of the mean values produced by the Monte Carlo trials, for all disease parameters, PIM outbreaks peaked slightly earlier and higher than the average Monte Carlo trial (shown in Table 4). This becomes more apparent when the parameters for less infectious diseases are used.

We believe that earlier peaks are observed partially due to an artifact of the stochastic method. In stochastic trials with low parameters, no outbreak of the disease is likely to be observed until multiple days have passed. Outbreak trials with peaks that are lower, occur later and show greater variance in the peak day of infection are observed as a result. This contrasts with PIM, which allows the seed of infection to partially contact multiple neighbors concurrently, possibly causing slightly earlier and higher peaks of infection. In addition, the approximation that events are independent may propel the initial spread of infection at a

Table 4

A comparison of outbreak attributes between PIM and the averaged values of 100 stochastic simulations. The standard deviation is shown for each averaged value.

Disease	Total infected	Probabilistic Infection Model	
		Peak infected	Day of peak
Measles	3644.21	1059.10	38
Influenza 1	2930.08	787.61	31
Influenza 2	2077.31	454.38	22-23

Disease	Total infected	Monte Carlo model	
		Peak infected	Day of peak
Measles	3647.95 $\pm$ 0.22	1021.35 $\pm$ 132.12	38.58 $\pm$ 2.39
Influenza 1	3011.49 $\pm$ 38.04	755.90 $\pm$ 72.12	34.03 $\pm$ 4.21
Influenza 2	2094.01 $\pm$ 109.01	394.72 $\pm$ 47.80	27.01 $\pm$ 4.60

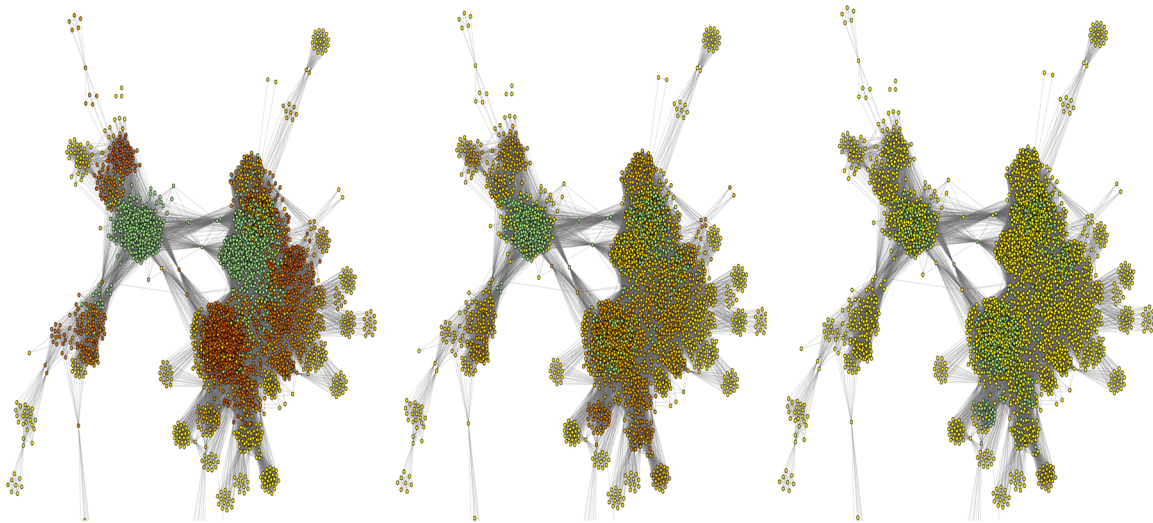


Fig. 4. States of the vertices in the contact network based on the PIM model on day 35. Yellow vertices are fully susceptible, whereas dark red vertices have a higher probability of being infected at a given time. Green vertices have a probability of 95 % or greater of being recovered. From left to right, the values are for Measles (left), Influenza 1 (middle) and Influenza 2 (right).

slightly greater rate, an effect that would be most noticeable for less infectious diseases.

Fig. 5 demonstrates that the attributes of the SEIR curves produced by PIM are similar to those of the average outbreak curves obtained from 100 stochastic trials. This similarity is most notable in simulations of highly infectious diseases, such as when using the parameters for measles; in Fig. 6 left, we show the simulation time series using measles parameters for all four states, showing that the PIM model closely follows the averaged curves of 100 trials of the stochastic model. In addition, we compare the infectious state probability curves of individual vertices produced by the PIM model: Fig. 6 right shows the $I_t(v)$ curves produced by PIM for the seed infected node as well as for 100 vertices that were randomly sampled from the set of initially susceptible vertices for the measles simulation. Most vertices reached their peak probability of being infected around day 38, which is consistent with the peak day of infection given in Fig. 5.

We also tested the performance of our model on two other contact networks obtained from the network repository Netzscheuler [26]. These networks are;

The adolescent health network (ADH) [23] which is a directed and weighted network of friendships of high school students. The connections were obtained through a social survey of the students. The network is partitioned into communities. We selected the largest community (community 50 in the repository) with 2587 vertices and 12,969 edges. Edge weights are a measure of frequency of interaction on a 1 to 5 scale.

The residence hall network of Australian National University (ANU) [12]. This is a directed weighted network, where resident a is connected to resident b if a indicated that b is a friend. Edge weights are a measure of perceived friendliness on a 1 to 5 scale. This network has 217 vertices and 2672 edges.

For both of these networks, for each pair of vertices a and b , directed edges (a, b) and (b, a) were collapsed into undirected edges $\{a, b\}$ with an edge weight that was the average of the edge weights in both directions.

Fig. 7 shows the comparison between the Monte Carlo based simulation and the PIM model. Once again, curve of the number of infected individuals using the PIM model closely follows that of the Monte Carlo simulation.

3.3. Execution time

To assess how PIM compares with Monte Carlo simulations in terms of execution time, we compare the computational efficiency of the PIM method with respect to the Monte Carlo method analytically as well as empirically. Note that both PIM and the Monte Carlo approaches we are examining are time-driven, not event-driven simulations. The number of timesteps required for a particular simulation run is not constant. Rather, a particular simulation run continues until the number of

infectious individuals falls beneath a specified threshold. As reflected in the results of experiments, given a specific contact network and set of simulation parameters, the total number of timesteps required by PIM and Monte Carlo models are comparable. Thus, to compare the complexity of PIM and Monte Carlo approaches, we have examined the computational complexity of calculations required by each approach for each time step.

Analytical time complexity. The complexity per timestep for PIM is $O(C(|V| + |E|))$, where C is the average number of contacts per vertex per timestep/day. The model requires iterating through all neighbors of all vertices, and perform an $O(1)$ calculation at each step. With backflow correction, the complexity becomes $O(C\gamma(|V| + |E|))$ (scales with infectious period). It is also worth noting that outbreaks of diseases with higher γ tend to be shorter, which may somewhat offset this factor. The computational time complexity for the Monte Carlo method is dependent on the random sampling algorithm used for selecting random contacts. It takes $O(C \log |V|(|V| + |E|))$ per timestep, where the $\log |V|$ comes from random sampling from a weighted set of choices, needed when selecting a neighbor to contact. If the selection of neighbors can be done in constant time $O(1)$, then the complexity of Monte Carlo method and the PIM method are asymptotically the same.

Empirical time complexity. In Table 5 we show the execution time of running the PIM method over the three networks and the three diseases, as well as the time to run 100 runs of the Monte Carlo simulation. As can be seen from the results, the time taken by the PIM method is comparable to the Monte Carlo method when backflow correction is not used. The time difference increases (but is within 1000 milliseconds) for larger and denser networks. For fair comparison the numbers reported are averaged over 100 trials for both the Monte Carlo method and the PIM method. Note that the Monte Carlo method requires multiple simulations, whereas the PIM requires only one simulation. As per the results, even ten runs of the Monte Carlo simulation will be slower than the PIM model. Typically about order of 100 runs are required for the Monte Carlo simulations. The difference is even larger when the backflow correction is not used.

Improving the time efficiency of PIM. While asymptotically the time to compute PIM is equivalent to the Monte Carlo method, the empirical execution time can be improved through certain approximations. The primary cost of PIM is due to contacting all neighbors of all vertices. For high degree vertices with many neighbors this becomes an expensive operation. The simulation can be made faster by processing high degree vertices at certain intervals instead of at every time step.

Effect of backflow correction. Since the backflow correction increases the simulation time, we now test by how much the correction due to redundant infection (as discussed in Section 2.2) affects the simulations. Fig. 8 shows a comparison between simulations with PIM when correction of the probability of vertex v infecting vertex u uses the modified version as in Equation (7), and one where the original Equation

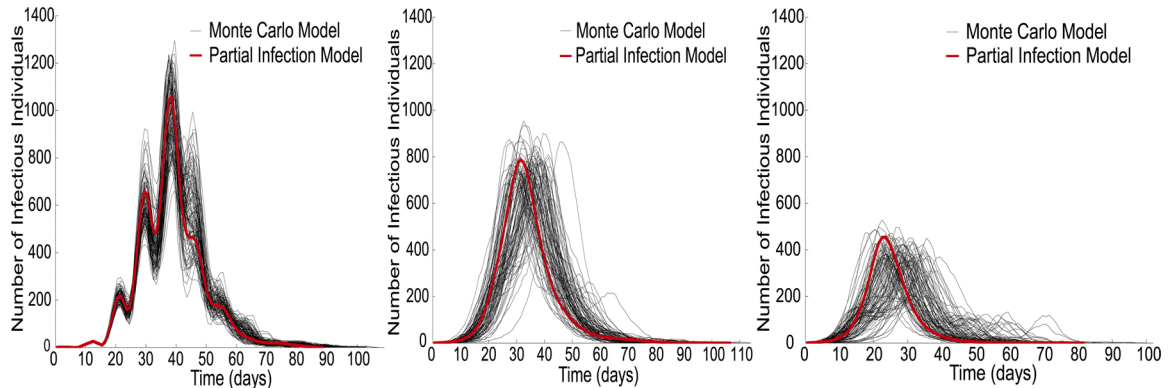


Fig. 5. A comparison between PIM and 100 simulations of the stochastic SEIR model with respect to the number of infectious individuals over the entire simulation. From left to right, the curves are for Measles (left), Influenza 1 (middle) and Influenza 2 (right).

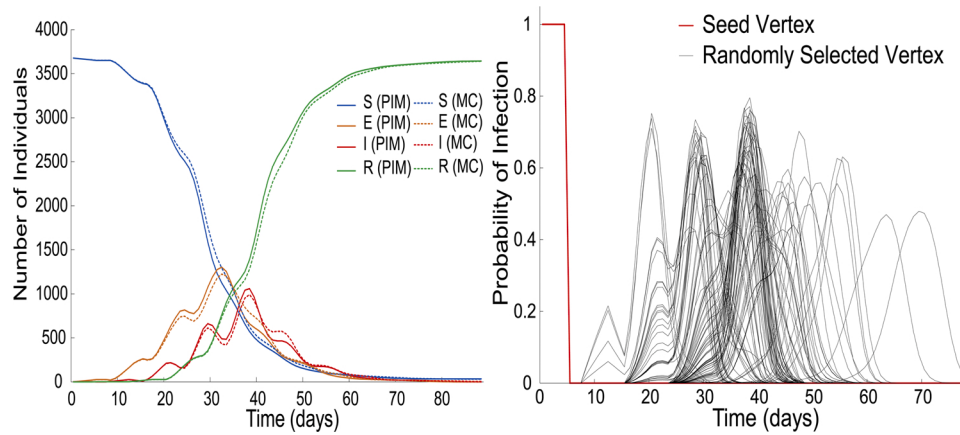


Fig. 6. State of vertices in the measles simulation. Left: Comparison between the number of vertices in each state over time for PIM and the Monte Carlo (MC) method averaged over 100 trials. Right: Probability of infection of 100 randomly selected vertices of the network. The peak occurs around days 35-45.

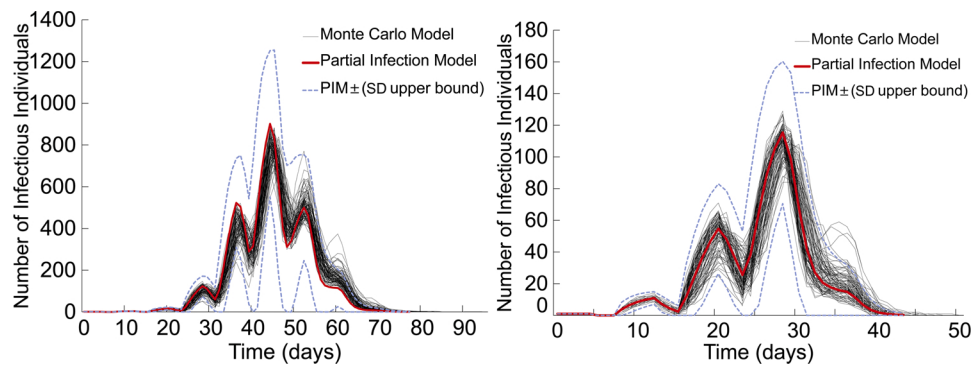


Fig. 7. Simulation of Measles, number of infected individuals on ADH network(Left) and ANY network(right). "The dotted lines show the PIM outbreak curve shifted by the upper bound on the standard deviation of the number of infectious individuals".

Table 5

Time in milliseconds for a single trial, averaged over 100 trials. These trials were run with an Intel Core i7-10510U CPU at 3 GHz.

Disease	Probabilistic Infection Model		
	Ad50 network	ANU network	UNT network
Measles	225.01	16.08	16755.81
Influenza 1	196.41	16.39	14895.05
Influenza 2	337.34	16.96	27229.71
Disease	Probabilistic Infection Model (No backflow correction)		
	Ad50 network	ANU network	UNT network
Measles	186.39	11.38	2571.78
Influenza 1	157.18	11.67	3022.39
Influenza 2	190.82	14.95	4091.78
Disease	Monte Carlo model		
	Ad50 network	ANU network	UNT network
Measles	210.47	10.12	3988.59
Influenza 1	160.69	7.5	2797.35
Influenza 2	203.69	10.55	3107.09

(1) is used. For each v_0 , the percent difference between the peak number of infected individuals produced by PIM with and without correction was less than 0.2%, suggesting that if needed the backflow correction can be eliminated for efficiency.

4. Relationship between graph structure and probability of infection

We examined the results from the Measles, Influenza 1, and Influenza

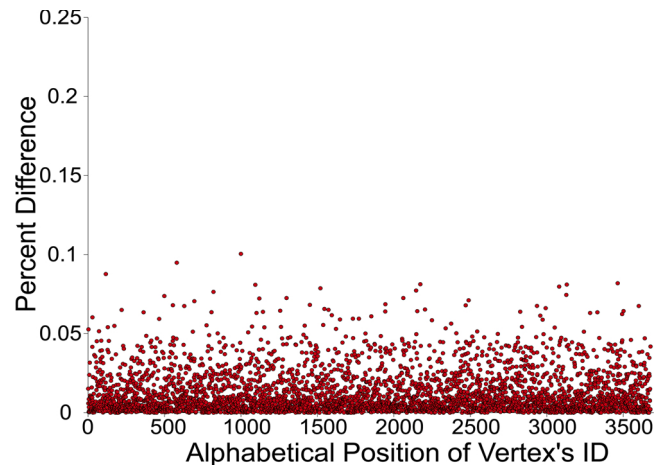


Fig. 8. The percent difference between the peak number of infected individuals is shown for simulations for measles produced by PIM with and without backflow correction for every possible initially infectious v_0 .

2 experiments to identify potential associations between graph structure and simulation outcomes. In each of these three experiments, a single vertex served as the initial source of infection. As before, edge weights in the graph were chosen to be proportional to pairwise contact probabilities and were derived from the total duration in hours of shared class time between a pair of vertices. For each vertex, the graph distance to the source of infection was calculated. For analysis, graph edge weights

were inverted before corresponding distances were calculated. In this manner, although a longer duration of shared class time between two vertices leads to a larger edge weight, a larger edge weight corresponds to a smaller distance. Thus, vertices connected to the initial source of infection via paths consisting of edges with larger weights are considered to be closer in distance to the source of infection.

First, for all three diseases, we examine the relationship between the day of peak infection probability and the distance to the source of infection for each vertex (Fig. 9). R^2 values for these analyses are 0.835, 0.784, and 0.754 for Measles, Influenza 1, and Influenza 2, respectively. Thus, the day of peak infection for each vertex appears to be heavily influenced by the distance of that vertex to the initial source of infection.

Next, to determine whether proximity to the source of infection affects a vertex's infection risk, we examine the relationship between the probability of ever becoming infected and the distance to the source of infection for each vertex (Fig. 10). R^2 values for Influenza 1 and Influenza 2 were 0.174 and 0.142, respectively. Considering these results, the probability of a vertex ever becoming infected does not appear to be heavily influenced by the vertex's distance from the infection source. Further, in the case of measles, there was no statistically significant association between distance to the source of infection and infection risk. Indeed, it is sensible that for a highly infectious disease such as measles, distance to the source of infection has little bearing on the probability of ever becoming infected.

Finally, in Fig. 11, we investigate whether a vertex's degree is predictive of the probability of it becoming infected. Note that vertex degrees are shown on a logarithmic scale. R^2 values for Influenza 1 and Influenza 2 were 0.639 and 0.642, respectively. Thus, a vertex's degree is a better predictor of whether it will ever be infected than the vertex's distance to the initial source of infection.

Based on these findings, in settings where contact networks can be derived such that individuals are represented as vertices and edges represent contact durations between pairs of individuals, outbreak mitigation activities could be informed by distance to the initial source of infection. If the index case in a disease outbreak is known, then the network distance from each vertex to this index case can be used to predict how the outbreak may unfold over time. Armed with this knowledge, responders could target testing or other mitigation resources to specific individuals at different points during the outbreak. Further, responders could leverage the structure of the social network to inform decisions regarding which individuals should be quarantined in order to temporarily modify the social network to curtail the spread of the disease.

5. Effect of varying COVID-19 transmissibility on outbreak dynamics

The PIM was used to investigate the effects of non-pharmaceutical

interventions on a simulated COVID-19 outbreak within the same university population used for Measles and Influenza simulations earlier in this manuscript. Social distancing was simulated by changing how contacts were derived from shared classtime; and mask wearing was simulated by varying the transmissibility of the disease. Due to the deterministic nature of PIM, only a single simulation run was required to generate a result corresponding to a particular set of input parameters. Multiple simulation runs were executed in order to investigate how varying parameters representative of social distancing and mask wearing impacted the outbreak. The same initial source of infection was chosen for each of the simulation runs.

Disease specific parameters used in the COVID-19 simulations were adapted from [29] and [31]. The base transmission probability T_{base} was set to a value of 0.11. The incubation period σ was set to 6 days, and the infectious period γ was set to 7 days.

Social distancing was simulated by modifying the number of contacts $\Omega_t(v)$ for each vertex v resulting from each hour of classroom time. For each simulation, the number of contacts per hour was applied uniformly across all vertices. For instance, a particular simulation run may generate 5 contacts for each hour of class time shared between a pair of vertices. Contacts per hour were restricted to values in the range (0,10).

The effect of mask wearing was simulated by implementing a transmissibility multiplier η to represent reduced potential for disease transmission resulting from wearing a mask. Thus, the transmission probability used in a given simulation was $T = \eta T_{base}$. Values for η were constrained to the range (0,1).

A total of 625 simulation runs were executed to analyze the effects of nonpharmaceutical interventions for COVID-19. For the number of contacts per hour of shared class time, simulation parameter values were started at 0.4 and were increased to 10 by increments of 0.4. These increasing values of contacts per hour were designed to simulate increasing degrees of social distancing. For the transmissibility multiplier, simulation parameter values were started at 1 and were decreased to 0.04 by increments of 0.04. The decreasing transmissibility multiplier values were designed to simulate decreasing likelihoods of transmission resulting from more effective mask wearing interventions. Results of these simulation runs are depicted in Figs. 12–14. All three of these figures indicate that either maximizing social distancing or maximizing mask wearing effectiveness will impede COVID-19 transmission by a degree sufficient to curtail the outbreak. Further, a combination of slightly relaxed social distancing and slightly less effective mask wearing are also sufficient to curtail the outbreak. However, as social distancing and/or mask wearing effectiveness decrease, the outbreak spreads to a larger population.

Plots focusing on the day of peak infectious depicted in Fig. 13 show the most dramatic results for increased contacts per hour. For each particular value of the transmissibility multiplier η , there appears to be a threshold value of the number of contacts per hour, that, if not reached,

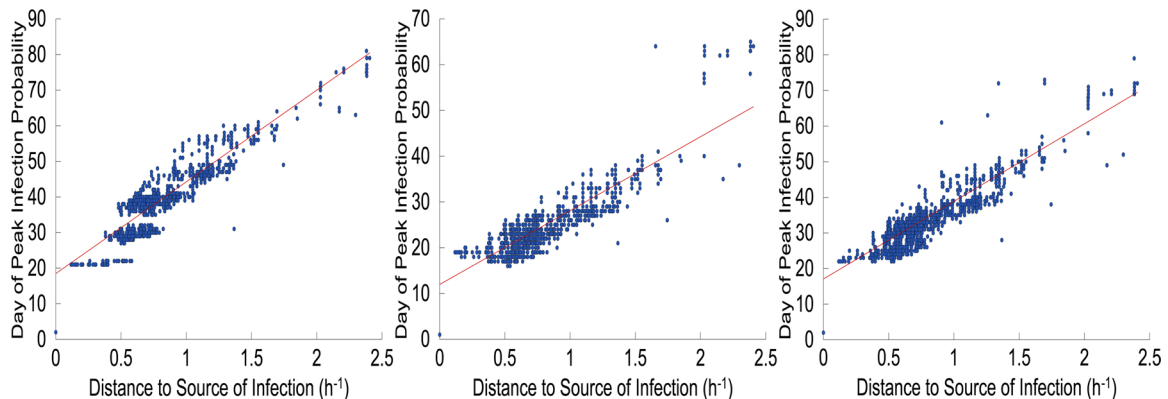


Fig. 9. Day of peak infection probability for each vertex plotted against the distance to the source of infection for Measles (left), Influenza 1 (middle) and Influenza 2 (right).

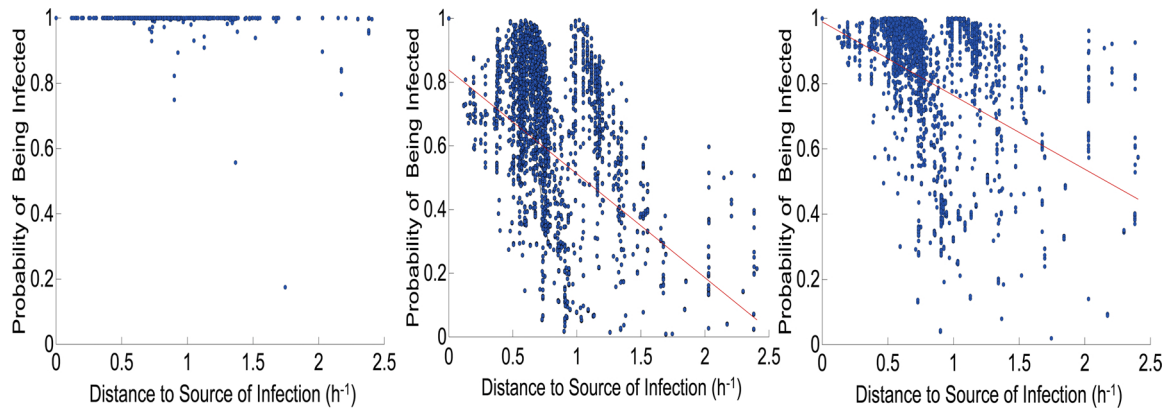


Fig. 10. Probability for each vertex to ever be infected plotted against the distance to the source of infection for Measles (left), Influenza 1 (middle) and Influenza 2 (right).

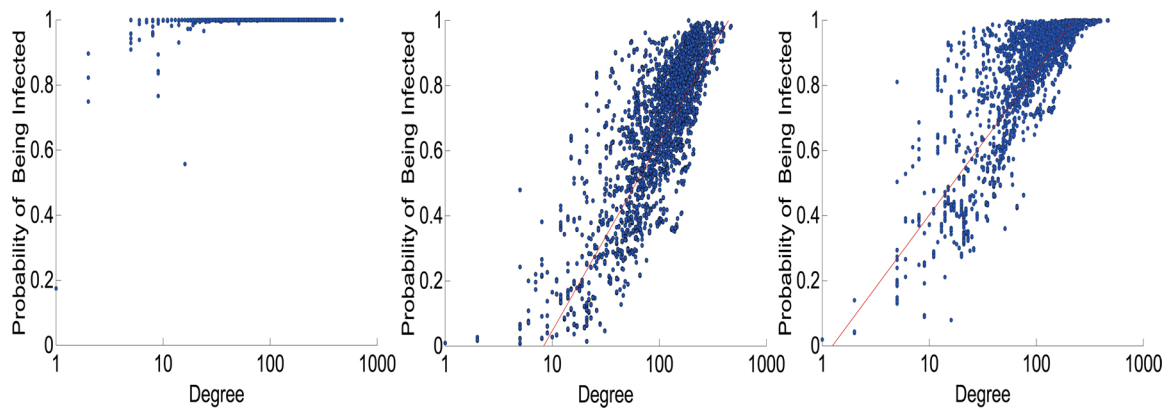


Fig. 11. Probability for each vertex to ever be infected plotted against the degree of each vertex for Measles (left), Influenza 1 (middle) and Influenza 2 (right).

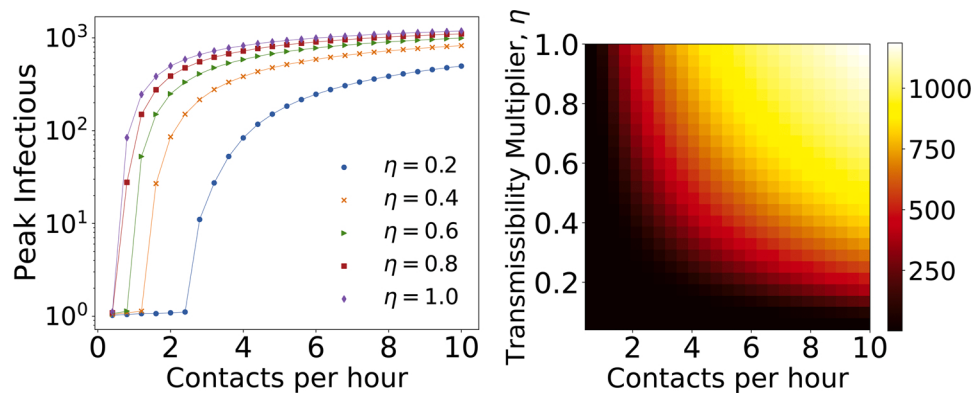


Fig. 12. Results focusing on the peak number of infectious individuals across multiple simulation runs with varying numbers of contacts per hour and transmissibility multipliers. Five different values for η are extracted in the left plot, and the right plot includes results from simulation runs using values of η in the range (0,1).

prevents an outbreak from taking hold altogether. However, once this contacts per hour threshold is reached, the day of peak infectious spikes dramatically, indicating that at this threshold, an extremely prolonged outbreak will take hold and run its course. As contacts per hour continues to increase above this threshold, the length of the outbreak decreases, indicating that more rapid outbreaks occur when contact rates are high. Nonetheless, these results need to be taken in the context of the total number of individuals ever infected. In this case, by considering the results depicted in Figs. 13 and 14 together, it can be seen that a longer outbreak does not necessarily imply a larger number of individual will (at some point) become infected.

6. Related research

Computational epidemiology is an active area of research. Despite the advances in modeling infection spread using networks several challenges exist. As discussed in [27], including developing more accurate network models from data, extending epidemic simulations to dynamic and weighted networks, understanding how the structure of the network relates to the spread of diseases and developing prevention strategies. These challenges represent on-going problems and are being addressed by several recent publications as discussed below.

The challenges of creating reliable contact networks are discussed in

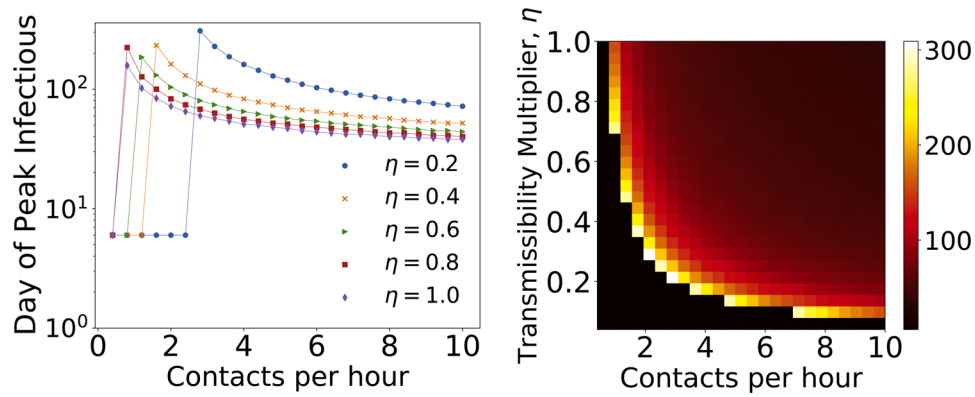


Fig. 13. Results focusing on the day the peak number of infectious individuals occurred across multiple simulation runs with varying numbers of contacts per hour and transmissibility multipliers. Five different values for η are extracted in the left plot, and the right plot includes results from simulation runs using values of η in the range (0,1).

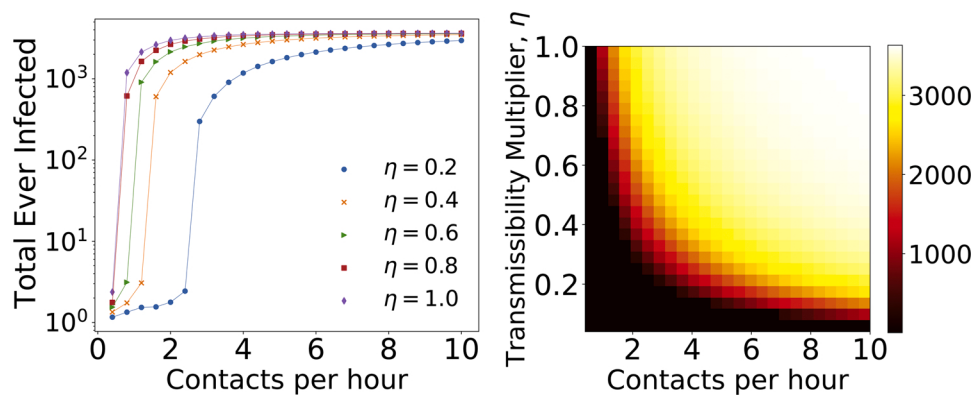


Fig. 14. Results focusing on the total number of vertices ever infected across multiple simulation runs with varying numbers of contacts per hour and transmissibility multipliers. Five different values for η are extracted in the left plot, and the right plot includes results from simulation runs using values of η in the range (0,1).

[9]. In 2008 [24], a cross-sectional survey on 7,290 participants conducted by different public health institutes or commercial companies was conducted to build a contact network. Another study [16], performed through the 2009 H1N1 flu pandemic on a population of 36 people based on communication using sensors. Recent studies have also looked into the dynamic contact networks [4] and the effect of misinformation in developing contact networks [15]. A survey on spread of epidemics on temporal networks is discussed in [34] and [20].

A large number of epidemic studies are modeled on synthetic models [19], such as the Erdos-Reyni model for random networks, the Barabasi-Albert model for scale free networks, etc. A primary focus in on leveraging the structure of the network to predict and/or to prevent the spread of epidemics. Two important methods as discussed in [5] for modeling the spread of epidemics in contact network are the Reed-Foster which models the spread of epidemic in discrete time and the Markovian model which looks at the epidemic spread in continuous time. Based on the degree distribution of the network, specifically the random network, it is possible to theoretically estimate the R_0 , basic reproductive number, and R_v , basic preventive number after v individuals have been vaccinated.

Epidemiology simulation using contact networks have gained further prominence with the onset of covid as in for studying control strategies [11], studying the spread of covid [33,21] and developing dynamic real world models of covid outbreaks [18]. There has also been studies on how asymptomatic cases can spread the diseases, which advocate contact tracing in addition to quarantine of infected individuals [22]. Different strategies for creating contact networks [2,17] has also gained prominence, including several tools for digital contact tracing.

Several software tools for simulating disease over a population have

been developed including EpiSims [13] and DiSimS [7] that use high performance computing, and Broadwick [25] which uses a sequential, but modular framework that can be modified for various disease parameters. As discussed below, our PIM method can be modified to be implemented to be parallel, and thus can be executed on large networks.

A common theme in these papers are the challenges in obtaining a contact network, both in terms of accuracy, as well as in terms of preserving the privacy of the participants. In this context our method of creating contact networks from scheduled data has an advantage in that it can be built from publicly available information (with appropriate anonymization), and can potentially scale to very large numbers.

7. Conclusion and future work

In this paper, we introduce a Probabilistic Infection Model for simulating the spread of infectious diseases on contact networks. Our model encapsulates the advantages of both deterministic meta-population models as well as stochastic models on contact networks. We further propose a method of obtaining contact networks based on the scheduled activities of individuals in specific environments (e.g., businesses, schools, etc.), and simulate our model on a contact network built from a university's class enrollment data. Comparisons of the results obtained from stochastic modelling and PIM on the contact network of university students demonstrate that our approach produces similar results to the stochastic model. In addition, our model gives a tractable framework for probabilistic analysis of outbreak dynamics at the individual level. It should be noted that the conclusions in Section 4 and 5 on the effect of the graph structure and non pharmaceutical interventions on the transmissibility of infection is on the scheduled activity data.

While we anticipate that similar results will hold on other networks, further experiments need to be conducted.

Although our proposed PIM method has a similar asymptotic computational complexity to the stochastic method, the main computational cost in practical terms comes from applying backflow correction on high degree vertices. This cost can be reduced by updating the backflow correction of these vertices at frequent intervals but not at every time step. Further, PIM removes the need for numerous repeated trials that is often required when using Monte Carlo methods.

Our PIM model can also compute $R_0(v)$, the basic reproduction number based on the vertex in the contact network from which the infection started. While an average R_0 provides the reproduction number with respect to the network, vertex based R_0 can help in identifying vertices that are more critical to the spread of diseases.

One might argue that the stochastic agent-based approach can be trivially parallelized, thereby giving it an additional computational advantage. It is however unfortunately quite difficult to determine the number of model executions necessary to obtain a reliable estimate of the outbreak characteristics. The required number of executions is generally not determined a priori. Instead, the model execution sequence is terminated once the cumulative results fall within a pre-determined confidence interval (e.g., 95%) and a specific level of precision (e.g., $\pm 1\%$), or the width of the confidence interval. Each of the individual model executions can hence be viewed as a sample of the outcome space, and depending on the sample standard deviation of the specific measurements that define the outbreak characteristics, an extensive number of executions may be required. In contrast, the PIM model requires only one execution, and the states of vertices are being computed based on the states of their neighbors only. This makes the PIM model more amenable to parallelization, since each vertex can be updated independently.

An important aspect of our PIM model is that we can model the values of the infectious period and incubation period per individual. In this paper, we are modeling individuals of about the same age group (students in university or school), and therefore set the values to be uniform.

As part of our future work, we will explore how we can factor in information about individuals such as age, health conditions and other demographic information to obtain more individualized latent periods and infectious periods as well as transmission probabilities selected from distributions rather than as static values. We also will develop approximate algorithms to improve the execution time of PIM. Moreover, we will pursue further studies of vaccine distribution and other individual-level outbreak intervention strategies by applying PIM's approximations for individual SEIR state-probabilities.

The source code for the PIM model, along with the UNT class enrollment network data, can be found at <https://github.com/wqian0/Partial-Infection-Model>.

Author statement

Qian developed the code and conducted the experiments. Mikler, Bhowmick, O'Neill and Ramisetty-Mikler provided the conceptual ideas and designed the experiments. All the authors participated in writing the paper.

Acknowledgement

Bhowmick's research was supported by the National Science Foundation under Grant No. 1916084.

Conflicts of interest: None declared.

References

- [1] M. Ajelli, B. Gonçalves, D. Balcan, V. Colizza, H. Hu, J.J. Ramasco, S. Merler, A. Vespignani, Comparing large-scale computational approaches to epidemic

- modeling: agent-based versus structured metapopulation models, *Bmc Infect. Dis.* 10 (2010).
- [2] J. Amann, J. Sleight, E. Vayena, Digital contact-tracing during the covid-19 pandemic: an analysis of newspaper coverage in germany, austria, and switzerland, *PLOS ONE* 16 (2) (2021 02) 1–16, <https://doi.org/10.1371/journal.pone.0246524>.
- [3] D. Balcan, H. Hu, B. Gonçalves, P. Bajardi, C. Poletto, J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W. Van den Broeck, V. Colizza, Seasonal transmission potential and activity peaks of the new influenza a (h1n1): a monte carlo likelihood analysis based on human mobility, *BMC Med.* 7 (1) (2009).
- [4] S. Bansal, J. Read, B. Pourbohloul, L.A. Meyers, The dynamic nature of contact networks in infectious disease epidemiology, *J. Biol. Dyn.* 4 (5) (2010) 478–489, <https://doi.org/10.1080/17513758.2010.503376>, PMID:22877143.
- [5] T. Britton, Epidemic models on social networks-with inference, *Stat. Neerl.* 74 (3) (2020) 222–241, <https://doi.org/10.1111/stan.12203>, <https://onlinelibrary.wiley.com/doi/abs/10.1111/stan.12203>.
- [6] A. Cori, A.J. Valleron, F. Carrat, G.S. Tomba, G. Thomas, P.Y. Boëlle, Estimating influenza latency and infectious period durations using viral excretion data, *Epidemics* 4 (3) (2012).
- [7] S. Deodhar, K.R. Bisset, J. Chen, Y. Ma, M.V. Marathe, An interactive, web-based high performance modeling environment for computational epidemiology, *ACM Trans. Manage. Inf. Syst.* 5 (2) (2014 Jul), <https://doi.org/10.1145/2629692>.
- [8] K. Drewniak, J. Helsing, A.R. Mikler, A method for reducing the spread of epidemics by allocating vaccines according to centrality, *ACM Conference on Bioinformatics, Computational Biology, and Health Informatics* (2014).
- [9] K. Eames, S. Bansal, S. Frost, S. Riley, Six challenges in measuring contact networks for use in modelling, *Epidemics* 10 (2015) 72–77, <https://doi.org/10.1016/j.epidem.2014.08.006>, challenges in Modelling Infectious Disease Dynamics, <http://www.sciencedirect.com/science/article/pii/S1755436514000413>.
- [10] W.T. Enanoria, F. Liu, J. Zipprich, K. Harriman, S. Ackley, S. Blumberg, L. Worden, T.C. Porco, The effect of contact investigations and public health interventions in the control and prevention of measles transmission: a simulation study, *PLoS One* 11 (12) (2016).
- [11] J. Firth, J. Hellewell, P. Klepac, S. Kissler, A. Kucharski, L. Spurgin, Using a real-world network to model localized covid-19 control strategies, *Nat. Med.* 26 (10) (2020 Oct) 1616–1622, <https://doi.org/10.1038/s41591-020-1036-8>.
- [12] L.C. Freeman, C.M. Webster, D.M. Kirke, Exploring social structure using dynamic three-dimensional color images, *Soc. Netw.* 20 (2) (1998 Apr) 109–118, [https://doi.org/10.1016/S0378-8733\(97\)00016-6](https://doi.org/10.1016/S0378-8733(97)00016-6).
- [13] M.E. Halloran, N.M. Ferguson, S. Eubank, I.M. Longini, D.A.T. Cummings, B. Lewis, S. Xu, C. Fraser, A. Vulliamy, T.C. Germann, D. Wagener, R. Beckman, K. Kadau, C. Barrett, C.A. Macken, D.S. Burke, P. Cooley, Modeling targeted layered containment of an influenza pandemic in the United States, *Proc. Natl. Acad. Sci.* 105 (12) (2008) 4639–4644, <https://doi.org/10.1073/pnas.0706849105>, <https://www.pnas.org/content/105/12/4639>.
- [14] S. Henson, F. Brauer, C. Castillo-Chavez, Mathematical models in population biology and epidemiology, *Am. Math. Mon.* 110 (3) (2003).
- [15] P. Holme, L.E.C. Rocha, Impact of misinformation in temporal network epidemiology, *Netw. Sci.* 7 (1) (2019) 52–69, <https://doi.org/10.1017/nws.2018.28>.
- [16] S. Jain, S.R. Benoit, J. Skarbinski, A.M. Bramley, L. Finelli, for the 2009 Pandemic Influenza A (H1N1) Virus Hospitalizations Investigation Team, Influenza-associated pneumonia among hospitalized patients with 2009 pandemic influenza A (H1N1) virus-United States, 2009, *Clin. Infect. Dis.* 54 (9) (2012 03) 1221–1229, <https://doi.org/10.1093/cid/cis197>.
- [17] C.E. Juneau, A.S. Briand, T. Pueyo, P. Collazzo, L. Potvin, Effective contact tracing for covid-19: a systematic review, *medRxiv* (2020), <https://doi.org/10.1101/2020.07.23.20160234>, <https://www.medrxiv.org/content/early/2020/07/25/2020.07.23.20160234>.
- [18] A. Karaivanov, A social network model of covid-19, *PLOS ONE* 15 (10) (2020 10) 1–33, <https://doi.org/10.1371/journal.pone.0240878>.
- [19] M.J. Keeling, K.T. Eames, Networks and epidemic models, *J. R. Soc. Interface* 2 (4 (Sep)) (2005) 295–307, <https://doi.org/10.1098/rsif.2005.0051>.
- [20] J. Leitch, K.A. Alexander, S. Sengupta, Toward Epidemic Thresholds on Temporal Networks: A Review and Open Questions, 2019.
- [21] F. Liu, X. Li, G. Zhu, Using the contact network model and metropolis-hastings sampling to reconstruct the covid-19 spread on the “diamond princess”, *Sci. Bull.* 65 (15) (2020) 1297–1305, <https://doi.org/10.1016/j.scib.2020.04.043>, <https://www.sciencedirect.com/science/article/pii/S2095927320302814>.
- [22] S.M. Moghadas, M.C. Fitzpatrick, P. Sah, A. Pandey, A. Shoukat, B.H. Singer, A. P. Galvani, The implications of silent transmission for the control of covid-19 outbreaks, *Proc. Natl. Acad. Sci.* 117 (30) (2020) 17513–17515, <https://doi.org/10.1073/pnas.2008373117>, <https://www.pnas.org/content/117/30/17513>.
- [23] J. Moody, Peer influence groups: identifying dense clusters in large networks, *Soc. Netw.* 23 (4) (2001 Oct) 261–283, [https://doi.org/10.1016/S0378-8733\(01\)00042-9](https://doi.org/10.1016/S0378-8733(01)00042-9).
- [24] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R.T. Mikolajczyk, M. Massari, S. Salmaso, G.S. Tomba, J. Wallinga, J.C.M. Hejblum, M. Sadkowska-Todys, M. Rosinska, W.J. Edmunds, Social contacts and mixing patterns relevant to the spread of infectious diseases, *PLoS Medicine* 5 (2008) 1083–1087.
- [25] A. O'Hare, S. Lycett, T. Doherty, L. Monteiro Salvador, R. Kao, Broadwick: a framework for computational epidemiology, *BMC Bioinform.* (2016 2), <https://doi.org/10.1186/s12859-016-0903-2>.
- [26] T.P. Peixoto, The Netzschleuder Network Catalogue and Repository, 2020. <https://networks.skewed.de/>.
- [27] L. Pellis, F. Ball, S. Bansal, K. Eames, T. House, V. Isham, P. Trapman, Eight challenges for network epidemic models, *Epidemics* 10 (2015) 58–62, <https://doi.org/10.1016/j.epidem.2015.05.001>.

org/10.1016/j.epidem.2014.07.003. , challenges in Modelling Infectious Disease Dynamics, <https://www.sciencedirect.com/science/article/pii/S1755436514000334>.

- [28] J.M. Ponciano, M.A. Capistrán, First principles modeling of nonlinear incidence rates in seasonal epidemics, *PLoS Comput. Biol.* 7 (2) (2011).
- [29] K. Prem, Y. Liu, T.W. Russell, A.J. Kucharski, R.M. Eggo, N. Davies, Centre for the mathematical modelling of infectious diseases COVID-19 working group, Jit, M., Klepac, P.: the effect of control strategies to reduce social mixing on outcomes of the covid-19 epidemic in wuhan, china: a modelling study, *The Lancet. Public Health.* 5 (5) (2020 May) e261–e270, [https://doi.org/10.1016/s2468-2667\(20\)30073-6](https://doi.org/10.1016/s2468-2667(20)30073-6). <https://europepmc.org/articles/PMC7158905>.
- [30] W. Qian, S. Bhowmick, M. O'Neill, S. Ramisetty-Mikler, A.R. Mikler, A probabilistic infection model for efficient trace-prediction of disease outbreaks in contact networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12137 LNCS (2020) 676–689.
- [31] Sha He, L.R. Sanyi Tang, A discrete stochastic model of the covid-19 outbreak: forecast and control, *Math. Biosci. Eng.* 17 (2020) 2792, <https://doi.org/10.3934/mbe.2020153> (mbe-17-04-153).
- [32] J.C. Thomas, D.J. Weber, *Epidemiologic Methods for the Study of Infectious Diseases*, Oxford University Press, 2001.
- [33] S. Thurner, P. Klimek, R. Hanel, A network-based explanation of why most covid-19 infection curves are linear, *Proc. Natl. Acad. Sci.* 117 (37) (2020) 22684–22689, <https://doi.org/10.1073/pnas.2010398117>. <https://www.pnas.org/content/117/37/22684>.
- [34] Q. Yin, T. Shi, C. Dong, Z. Yan, The impact of contact patterns on epidemic dynamics, *PLOS ONE* 12 (3) (2017 03) 1–13, <https://doi.org/10.1371/journal.pone.0173411>.



Susie Ramisetty-Mikler is a Research Associate Professor in Epidemiology and Biostatistics, Population Health Sciences, School of Public Health At Georgia State University, Atlanta, GA. Research & Activities of Interest: Survey design, sampling, psychometrics, and statistical analysis of large general population health data; Alcohol epidemiology, alcohol related problems including family, partner violence, depression; child/Adolescent risk behaviors, alcohol/drug use, sexual risk taking-etiology, consequences; current research - birth defects, pre-conception health.



Marty O'Neill is the Director of the Center for Computational Epidemiology and Response Analysis (CeCERA), and a Research Associate Professor in the Department of Biological Sciences at University of North Texas (UNT). He obtained his Ph.D. in Computer Science from UNT in 2014. His current projects include outbreak modeling, visualization of complex data, geospatial analysis, and response plan design.



William Qian is an undergraduate student in the complex systems lab at University of Pennsylvania. His research interests are in synchronization networks, dynamical systems, and temporal networks. He worked on the research published in this paper as a Texas Academy of Maths and Sciences(TAMS) student at University of North Texas.



Armin R. Mikler is a Professor and Chair of the Department of Computer Science at Georgia State University. He received his PhD in Computer Science from Iowa State University in 1995. As a professor of Computer Science and Engineering at the University of North Texas from 1997 -2020, Dr. Mikler directed the Center for Computational Epidemiology and Response Analysis (CeCERA). His research interests include Computational Epidemiology and Disaster Informatics with focus on data-driven response plan design and plan optimization.



Sanjukta Bhowmick is an Associate Professor in the Computer Science and Engineering department at the University of North Texas. She obtained her Ph.D. from the Pennsylvania State University. She had a joint postdoc at Columbia University and Argonne National Lab. Her current research is on understanding change in complex network analysis, with a focus on developing scalable algorithms for large dynamic networks and developing uncertainty quantification metrics for network analysis.