Spreading processes with population heterogeneity over multi-layer networks

Yurun Tian
Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, USA
yurunt@andrew.cmu.edu

Osman Yağan
Electrical and Computer Engineering
Carnegie Mellon Univeristy
Pittsburgh, USA
oyagan@andrew.cmu.edu

Abstract—Modeling spreading processes over complex networks has been receiving increasing attention. For example, bond percolation models considering population heterogeneity have been used to derive insights into disease spread and misinformation control. However, most works on spreading processes with population heterogeneity only concentrate on single-layer contact networks. To study how the course of a spreading process changes due to *multiple* layers of contact networks (e.g., neighborhood vs. schools or Twitter vs. Facebook) while considering population heterogeneity from a principled, mathematical lens, we propose the *Multi-layer Mask model* based on SIR dynamics. We derive analytical expressions for three fundamental epidemiological quantities: the probability of emergence, the epidemic threshold, and the expected epidemic size. Analytical results are shown to be in near-perfect agreement with the numerical results obtained through extensive simulations. These results reveal the impact of the structure of the multi-layer contact network, viral transmission dynamics, and population heterogeneity on the final state of the spreading process. Thus, they might help develop mitigation and control strategies for disease spread and information diffusion.

Index Terms—Network Epidemics, Population Heterogeneity, Multi-layer Networks, Bond Percolation, Branching process

I. Introduction

Studies on spreading processes over complex networks have captured increasing attention over the years due to recent pandemics (e.g., COVID-19, SARS) and increasing concerns on *misinformation* spread [1]. Mathematical models over complex networks have been widely studied to understand the spreading dynamics of a pathogen or a piece of information [2]–[4]. Especially, susceptible-infectious-recovered (SIR) compartmental model received significant interest due to its ability to represent the propagation of both pathogens and information [5]–[7], and its steady-state analysis being closely tied with the *bond-percolation* over networks [7], [8].

More recently, there has been interest on studying the SIR spreading process with increasing complexity of the underlying contact network (e.g., clustered networks [9]–[11] and multi-layer networks [6], [12]) and the *heterogeneity* of the population. For example, Tian et al. [13] investigated a SIR model with population heterogeneity that manifest from

different types of masks that the individuals in the population might be wearing. More broadly, population heterogeneity can be relevant in the spread of a pathogen due to differences of age, gender, socio-economic status, and access to healthcare and other resources [14]–[16] in the population. Similarly, population heterogeneity can play an important role in information diffusion, where individuals might have different habits of accepting and transmitting information based on their personalities or fact-checking behaviors [1], [17]. Allard et al. [5] also studied the SIR model with population heterogeneity and showed that their steady-state can be analyzed through a *semi-directed* bond percolation model.

This paper is motivated by the fact that most studies on spreading processes with population heterogeneity consider single-layer networks, including [14] - [13]. However, most real-world spreading processes take place over *multi-layer* networks. In viral spreading, different layers might represent viral spreading paths in different environments, e.g., community, school, and workplace, etc, each with different rate of viral transmissibility [18]. Similarly, (mis)information tends to spread over multiple social media platforms, each with different rates and dynamics of propagation. To our best knowledge, there has only been two prior efforts [5], [19] on studying the SIR model while incorporating both population heterogeneity and the multi-layer nature of the contact network. Bongiorno and Zino [19] proposed a model that incorporates both population heterogeneity and a multi-layer contact network, but they do not provide mathematical analysis for the three epidemic quantities and instead rely on simulation results. The work by Allard et al. [5] consider multi-type networks with arbitrary joint degree distribution. However, their work does not provide a detailed analysis on the impact of multi-layer network structures and the associated multi-layer transmission dynamics on the final spreading results.

Inspired by these, we propose the *multi-layer mask model* that jointly considers population heterogeneity and multi-layer contact networks. For illustrative purposes, we present the work under the *community-school viral transmission* context, where the first layer represents the community contact network, the second layer represents the school contact network. Population heterogeneity results from different types of masks (with different efficiencies) that individuals are wearing. We

also reserve one *mask-type* to represent individuals who do not wear any mask.

Our main contribution is the analytical solution of the multi-layer mask model for three key epidemiological quantities, namely the probability of emergence (PE), epidemic threshold, and expected epidemic size (ES). The *emergence* of *epidemics* represents situations where the spreading process leads to a positive fraction of the population being infected in the limit of the number of nodes going to infinity. Put differently, *epidemics* refer to *large-scale* spreading events such as viral pandemics or information *memes*. Our analytical solutions disentangle the impact of multiple factors, including the degree distribution of different layers of the contact network, inward and outward efficiencies of the masks involved, and the proportion of the population wearing different types of masks, on these three quantities of interest. Extensive simulations validate our analytical results with near-perfect match. Utilizing our results, we also identify epidemic boundaries that reveal the conditions on the parameters involved under which PE and ES are zero, meaning that epidemics will *not* take place with probability one. We believe that these results can help better understand the interplay between different factors affecting the spreading dynamics, including mask-wearing behavior and contact rate across different layers.

II. MODEL

We generate our contact network model, consisting of two layers, as follows. Consider a population of size n with individuals labeled as $\mathcal{N} = \{1, \ldots, n\}$. Within the multi-layer network, each node corresponds to an individual in $\mathcal{N}$, and an edge is drawn between two nodes if they have a chance to transmit the virus/information to each other. Let $\mathbb{C}$ stand for the first contact layer, e.g., representing the *community* contacts of individuals, defined on the node set $\mathcal{N}$. Let $\mathbb{S}$ denote the second layer, e.g., representing the school contact network, with the assumption that each node in $\mathcal{N}$ is a member of $\mathbb{S}$ with probability $\alpha \in (0, 1]$, independently from other nodes. Formally, we let

$$\mathbb{P}[i \in \mathcal{N}_S] = \alpha, \quad i = 1, \ldots, n \qquad (1)$$

where $\mathcal{N}_S$ denotes the set of individuals who also participate the school layer. Edges belonging to network $\mathbb{C}$ (resp., $\mathbb{S}$) are noted as type-c (resp. type-s) edges, representing viral transmission paths via community (resp., school) contacts.

In line with prior literature on stochastic epidemic models, we generate the contact networks $\mathbb{C}$ and $\mathbb{S}$ by the *configuration model* [20], [21]. In other words, the topology of the networks $\mathbb{C}$ and $\mathbb{S}$ are generated *randomly* from their degree distributions $\{p_k^c\}$ and $\{p_k^s\}$, respectively, where $k = 0, 1, \ldots$. Here, p_k^c (resp. p_k^s) gives the probability that an arbitrary node on network $\mathbb{C}$ (resp. $\mathbb{S}$) has degree k, i.e., it is connected to k other nodes via an *undirected* type-c (resp. type-s) edge. To model viral transmission among individuals through both layers, we introduce the multi-layer network $\mathbb{H}$ formed by taking the *disjoint* union of $\mathbb{C}$ and $\mathbb{S}$, i.e., $\mathbb{H} = \mathbb{C} \coprod \mathbb{S}$. In this setting, the *colored* degree of a node i is represented

by an integer vector $\boldsymbol{d^i} = \begin{bmatrix} k_c^i, k_s^i \end{bmatrix}$, where k_c^i (resp., k_s^i) stands for its number of community edges (resp. school edges). Assuming the independence of $\{p_k^c\}$ and $\{p_k^s\}$, the colored degree distribution $p_{\boldsymbol{d}}$ is given by

$$p_{\boldsymbol{d}} = (\alpha p_{k_s}^s + (1 - \alpha)\mathbf{1}[k_s = 0]) \cdot p_{k_c}^c, \quad \boldsymbol{d} = (k_c, k_s), \quad (2)$$

where the term $(1 - \alpha)\mathbf{1}[k_s = 0]$ accounts for the case where the node is not a member of the school layer $\mathbb{S}$, and its number of type-s edges is automatically zero.

In their seminal work [7], Newman studied the SIR (susceptible-infectious-recovered) model over a contact network generated by the configuration model. Newman's model captures complex viral transmission and recovery mechanisms via the *transmissibility* parameter T which gives the probability that an *infected* node will transmit the virus/information to each of their neighbors (independently from each other). Many authors have incorporated various node-level heterogeneity based on Newman's model [5], [6], [13], [15], [22]–[25], among which *mask model* [13], [15] presents direct modeling of population heterogeneity via the inward and outward efficiency of different types of masks that individuals might be wearing. In the *mask model*, the transmission probability from a type-i individual to a type-j individual is given by

$$\mathbf{T}[i, j] = (1 - \epsilon_{\mathrm{out},i})(1 - \epsilon_{\mathrm{in},j})T; \quad 1 \le i, j \le M$$

where $\epsilon_{out,i}$ is the outward efficiency of a type-i mask, $\epsilon_{in,j}$ is the inward efficiency of a type-j mask, M is the total number of mask types, and T is the baseline transmissibility of the virus without any masks. We have $\epsilon_{\mathrm{out},i}, \epsilon_{in,j} \in [0, 1], \forall i, j \in [1, M]$, and $\mathbf{T}$ is a $M \times M$ transmissibility matrix. It was also assumed that the mask distribution is given by $\{m_1, \ldots, m_M\}$ where m_i represents the fraction of individuals who wear mask of type-i. For notational convenience, we shall say that an individual is of type-i if they wear a type-i mask ($1 \le i \le M$).

In this work, we introduce the *multi-layer mask model* to extend the *mask model* to multi-layer networks. To this end, we assume that the *transmissibility*, i.e., the probability that an infected individual passes on the infection to their direct contacts, depends on *both* the type of masks they are wearing and the type of the link connecting them. We let T_c and T_s denote the baseline transmissibility of the virus/information on type-c and type-s links, respectively. For example, the difference in the baseline transmissibility across two layers can result from nodes having a different rate of being in close-enough contact in the school layer versus the community layer. Compared to the *mask model*, instead of a single $M \times M$ transmissibility matrix, we now have two transmissibility matrices $\mathbf{T_c}$ and $\mathbf{T_s}$, each of size $M \times M$, for layer $\mathbb{C}$ and $\mathbb{S}$, respectively. More specifically, we have

$$\mathbf{T_c}[i, j] = (1 - \epsilon_{\mathrm{out},i})(1 - \epsilon_{\mathrm{in},j})T_c, \ 1 \le i, j \le M$$

each edge in $\mathbb{S}$ can transmit the pathogen with probability

$$\mathbf{T_s}[i, j] = (1 - \epsilon_{\mathrm{out},i})(1 - \epsilon_{\mathrm{in},j})T_s, \ 1 \le i, j \le M$$

where $\mathbf{T_c}[i, j]$ gives the probability that, if infectious, a node of type-i transmits the virus/information to a node of type-j given that they are connected by a type-c link.

III. ANALYTICAL RESULTS

A. *Probability of Emergence and Epidemic Threshold*

Consider random graphs $\mathbb{C}(n, \{p_k^c\})$ and $\mathbb{S}(n; \alpha, \{p_k^s\})$ as introduced in Section II. In order to study the viral transmission in the multi-layer network $\mathbb{H} = \mathbb{C} \coprod \mathbb{S}$, we consider a branching process that starts by giving the pathogen to an arbitrary node and then recursively discovers the set of nodes that are reached and *infected* by exploring its neighbors. As already mentioned, we assume that a type-i infected node transmits the pathogen to a type-j susceptible neighbor with probability $\mathbf{T_c}[i,j] = T_c(1-\epsilon_{i,out})(1-\epsilon_{j,in})$ if the link connecting them is type-c (or, with probability $\mathbf{T_s}[i,j] = T_s(1-\epsilon_{i,out})(1-\epsilon_{j,in})$ if the link between them is type-s), independently from all the other neighbors.

We derive the *survival probability* of the aforementioned branching process is through a *mean-field* analysis utilizing the method of generating functions [7], [21]. Namely, for $1 \le i \le M$, let $h_{c,i}(x)$ (resp. $h_{s,i}(x)$) denote the generating function for "the *finite* number of nodes reached and infected by following a *randomly* selected type-c (resp. type-s) edge coming from a type-i infected node." Put differently, we have $h_{c,i}(x) = \sum_{m=0}^{\infty} v_m x^m$, where v_m denotes the "probability that an arbitrary type-c edge coming from a type-i infected node leads to a component of size m." Similarly, let $H_i(x)$ denote the generating function for "the *finite* number of nodes reached and infected by following a randomly selected type-i infected node."

We now derive $h_{c,i}(x)$ and $h_{s,i}(x)$, for each $i = 1, \ldots, M$ recursively. We find that for each $i = 1, \ldots, M$, the following self-consistency equations hold:

$$h_{c,i} = \sum_{j=1}^{M} m_j \left(1 - \mathbf{T_c}[i,j] \right. \tag{3}$$

$$\left. + \mathbf{T_c}[i,j]x \sum_{\boldsymbol{d}} \frac{p_{\boldsymbol{d}} k_c}{\langle k_c \rangle} h_{c,j}(x)^{k_c-1} h_{s,j}(x)^{k_s} \right)$$

$$h_{s,i} = \sum_{j=1}^{M} m_j \left(1 - \mathbf{T_s}[i,j] \right. \tag{4}$$

$$\left. + \mathbf{T_s}[i,j]x \sum_{\boldsymbol{d}} \frac{p_{\boldsymbol{d}} k_s}{\langle k_s \rangle} h_{c,j}(x)^{k_c} h_{s,j}(x)^{k_s-1} \right)$$

We explain each term appearing in (3) in turn. Consider a type-i node, say node v, that is infected and consider a type-c edge incident on it. We first condition on the *type* of the node on the other end of this edge, i.e., we condition on the type of the mask that the neighbor node, say node u, is wearing. Since the mask assignment is done before the spreading process and independently for all the nodes, the neighbor node u is type-j with probability m_j. Conditioned on u being type-j, it will turn infected (through its edge with v) with probability $\mathbf{T_c}[i,j]$. If the transmission fails, which happens with probability $1 - \mathbf{T_c}[i,j]$, then node v will have no offspring through its edge to u which explains the first part of (3). If the transmission is successful, which has probability $\mathbf{T_c}[i,j]$, the number of nodes reached and infected by node v

will increase by one, and this is captured by the multiplicative term x in the second half of (3). In addition to this, the total size of this *branch* will also include all subsequent nodes that might be infected by u, which gives rise to the term

$$\sum_{\boldsymbol{d}} \frac{p_{\boldsymbol{d}} k_c}{\langle k_c \rangle} h_{c,j}(x)^{k_c-1} h_{s,j}(x)^{k_s}.$$

This term is explained as follows. We first condition on the *colored* degree of node u, i.e., its number of edges in both network layers. The term $p_{\boldsymbol{d}} k_c/\langle k_c \rangle$ gives the *normalized* probability that an edge of type-c is attached to a node at the other end with colored degree $\boldsymbol{d} = (k_c, k_s)$ [7]. Given its colored degree $\boldsymbol{d} = (k_c, k_s)$ and the fact that node u is reached via a type-c edge, it can infect other nodes via its remaining $k_c - 1$ links of type-c and k_s links of type-s. Given that the number of nodes reached and infected by each of its type-c (resp. type-s) links will in turn by generated by $h_{c,j}(x)$ (resp. $h_{s,j}(x)$), we obtain the term $h_{c,j}(x)^{k_c-1} h_{s,j}(x)^{k_s}$ to describe the total number of nodes reached and infected by node u by using the *powers property* of generating functions [21]. This completes the derivation of (3). The validity of (4) can be seen in a very similar way and is omitted here for brevity.

Utilizing equations (3) and (4), we now derive the the generating function $H_i(x)$ for the entire size of the branching process. For each $i = 1, \ldots, M$, we have

$$H_i(x) = x \sum_{\boldsymbol{d}} p_{\boldsymbol{d}} h_{c,i}(x)^{k_c} h_{s,i}(x)^{k_s} \tag{5}$$

Here, the factor x corresponds to the initial node selected arbitrarily and infected. The selected node has colored degree $\boldsymbol{d} = (k_c, k_s)$ with probability $p_{\boldsymbol{d}}$. The number of nodes it reaches and infects by each of its k_c (resp. k_s) links of type-c (resp. type-s) is generated through $h_{c,i}(x)$ (resp. $h_{s,i}(x)$). Summing over all the possible colored degrees, we obtain equation (5).

With equations (3)-(5) in hand, the generating function $H_i(x)$ can be computed in the following manner. Given any x, we can solve for the recursive relations (3)-(4) to obtain $h_{c,1}(x), \ldots, h_{c,M}(x)$ and $h_{s,1}(x), \ldots, h_{s,M}(x)$, which in turn will yield $H_1(x), \ldots, H_M(x)$ in light of (5). We are interested in cases where the number of nodes reached and infected by the initial node is *infinite*, which represents the cases where a randomly chosen infected node triggers an *epidemic*. The *conservation of probability* property of generating functions indicates that there exists a trivial fixed point $h_{c,i}(1) = h_{s,i}(1) = 1$ (yielding $H_i(1) = 1$) when the number of nodes reached and infected is always *finite*. In other words, the underlying branching process is in the *sub-critical* regime, and *all* infected components have finite size. However, the fixed point $h_{c,i}(1) = h_{s,i}(1) = 1$ may not be a stable solution to the recursion (3) to (5). We can check the stability of this fixed point by the linearization of recursion (3) to (5) around $h_{c,i}(1) = h_{s,i}(1) = 1$. This yields the Jacobian matrix $\boldsymbol{J}$ with the form

$$\boldsymbol{J} = \left[\begin{array}{cc} \boldsymbol{J}_{cc} & \boldsymbol{J}_{cs} \\ \boldsymbol{J}_{sc} & \boldsymbol{J}_{ss} \end{array} \right]_{2M \times 2M} \tag{6}$$

where

$$\boldsymbol{J}_{cc}(i,j) = \frac{\partial h_{c,i}(1)}{\partial h_{c,j}(1)}; \qquad \boldsymbol{J}_{cs}(i,j) = \frac{\partial h_{c,i}(1)}{\partial h_{s,j}(1)}$$
$$\boldsymbol{J}_{ss}(i,j) = \frac{\partial h_{s,i}(1)}{\partial h_{s,j}(1)}; \qquad \boldsymbol{J}_{sc}(i,j) = \frac{\partial h_{s,i}(1)}{\partial h_{c,j}(1)} \tag{7}$$

If all eigenvalues of $\boldsymbol{J}$ are less than one in absolute value, i.e., if the spectral radius $\rho(\boldsymbol{J})$ of $\boldsymbol{J}$ satisfies $\rho(\boldsymbol{J}) \leq 1$, then the solution $h_{c,i}(1) = h_{s,i}(1) = 1$ is stable and $H_i(1) = 1$ becomes the physical solution. In this case, the fraction of nodes that are infected will tend to zero as as the number of nodes n goes to infinity. In contrast, if $\rho(\boldsymbol{J}) > 1$, the trivial fixed point is not stable, which indicates that the branching process is in the *supercritical* regime; i.e., there is a positive probability that the branching process will lead to an *infinite* component. In this case, the fraction nodes that are infected will be strictly greater than zero as as the number of nodes n goes to infinity.

In fact, when $\rho(\boldsymbol{J}) > 1$, a nontrivial fixed point exists and becomes the attractor of the recursions (3) to (5), leading to a solution with $h_{c,i}(1), h_{s,i}(1) < 1$ which in turn yields $H_i(1) < 1$. In that case, $1 - H_i(1)$ gives the probability that the spreading process initiated by a seed node of type-i yields an *epidemic*. Namely, with R denoting the *final* fraction of nodes that are infected in the limit of n going to infinity, the probability of epidemic emergence PE (with a random initiator) is given by PE $= \mathbb{P}[R > 0] = \sum_{i=1}^{M} m_i(1 - H_i(1))$. Finally, we conclude that the epidemic threshold, i.e., the boundary that separates the parameter regions where $\mathbb{P}[R > 0] = 0$ from those that yield $\mathbb{P}[R > 0] > 0$ is given by $\rho(\boldsymbol{J}) = 1$.

B. Epidemic Size

In this section, we compute the expected size of epidemics when they take place, i.e., $\mathbb{E}[R \mid R > 0]$. We will also compute the fraction of infected nodes in each type. Our approach is similar that used in [2], [3], [15], [23]. Since the multi-layer network $\mathbb{H}$ is locally tree-like as the network size approaches infinity [26], we can consider it as a tree-structure, where there is a single node of type-i at the top level (referred to as the *root*). We label the levels of the tree from $\ell = 0$ at the bottom to the top $\ell = \infty$. Without loss of generality, we assume that the spreading event starts at the bottom of the tree and proceeds toward the top. In other words, we assume that a node at level ℓ can only be infected by one of its neighbors in level $\ell - 1$. Let $q_{c,\ell}^i$ (respectively, $q_{s,\ell}^i$) denote the probability of a type-i node at level ℓ who is connected to its parent at level $\ell + 1$ through a type-c (respectively, type-s) edge is *not* infected. Our ultimate goal is to compute q_∞^i which represents the probability that the *root* node, which is of type-i, is *not* infected. Given that the root node is arbitrary, q_∞^i also gives the expected fraction of type-i nodes that will be infected during the spreading process. Put differently, we have $\mathbb{E}[R_i \mid R > 0] = 1 - q_\infty^i$ with R_i denoting the final fraction of nodes of type-i that are infected in the spreading process; we also have $R = \sum_{i=1}^{M} m_i R_i$.

With these definition in mind, we now derive $q_{c,\ell}^i$ and $q_{s,\ell}^i$ in a recursive manner. For each $i = 1, \ldots, M$, we find that

$$q_{c,\ell+1}^i = \sum_{\boldsymbol{d}=(k_c,k_s)} \frac{p_{\boldsymbol{d}} k_c}{\langle k_c \rangle} f_i(\boldsymbol{q}_{c,\ell}, \boldsymbol{q}_{s,\ell}, k_c - 1, k_s) \tag{8}$$

$$q_{s,\ell+1}^i = \sum_{\boldsymbol{d}=(k_c,k_s)} \frac{p_{\boldsymbol{d}} k_s}{\langle k_s \rangle} f_i(\boldsymbol{q}_{c,\ell}, \boldsymbol{q}_{s,\ell}, k_c, k_s - 1) \tag{9}$$

where $\boldsymbol{q}_{c,\ell} = [q_{c,\ell}^1, q_{c,\ell}^2, ..., q_{c,\ell}^M]$, $\boldsymbol{q}_{s,\ell} = [q_{s,\ell}^1, q_{s,\ell}^2, ..., q_{s,\ell}^M]$, and $f_i(\boldsymbol{q}_{c,\ell}, \boldsymbol{q}_{s,\ell}, k_c, k_s)$ is given by

$$f_i(\boldsymbol{q}_{c,\ell}, \boldsymbol{q}_{s,\ell}, k_c, k_s) = \left(\sum_{j=1}^{M} m_j(1 - \mathbf{T_c}[j,i] + q_{c,\ell}^j \mathbf{T_c}[j,i]) \right)^{k_c}$$
$$\cdot \left(\sum_{j=1}^{M} m_j(1 - \mathbf{T_s}[j,i] + q_{s,\ell}^j \mathbf{T_s}[j,i]) \right)^{k_s} \tag{10}$$

In order to see why (8) holds, let u be a type-i node at level $\ell + 1$ who is connected to its unique parent at level $\ell + 2$ with an edge of type-c. As already mentioned, $q_{c,\ell+1}^i$ gives the probability that u is *not* infected. As before, we first condition on the colored degree of u being $\boldsymbol{d} = (k_c, k_s)$ which has probability $\frac{p_{\boldsymbol{d}} k_c}{\langle k_c \rangle}$. Under the assumption that nodes can only be infected by neighbors in the layers below, node u can be infected through either one of $k_c - 1$ edges of type-c and k_s edges of type-s in layer ℓ (given that one of its type-c edges is used to connect it to the parent node in layer $\ell + 2$). We establish (8) by noting that $f_i(\boldsymbol{q}_{c,\ell}, \boldsymbol{q}_{s,\ell}, k_c, k_s)$ represents the probability that a type-i node with k_c edges of type-c and k_s edges of type-s with nodes in layer ℓ is *not* infected. Equation (9) can be seen to hold in a similar way.

We now explain why (10) holds. First, for a node with k_c edges of type-c and k_s edges of type-s in layer ℓ to be not infected, it should not receive the pathogen from any of these neighbors. Given the independence of infection events, we see that $f_i(\boldsymbol{q}_{c,\ell}, \boldsymbol{q}_{s,\ell}, k_c, k_s)$ should be of the form $f_i(\boldsymbol{q}_{c,\ell})^{k_c} f_i(\boldsymbol{q}_{s,\ell})^{k_s}$ with $f_i(\boldsymbol{q}_{c,\ell})$ (respectively, $f_i(\boldsymbol{q}_{s,\ell})$) defined as the probability that a type-i node with only one edge of type-c (respectively, type-s) with nodes in layer ℓ is *not* infected. In order to compute $f_i(\boldsymbol{q}_{c,\ell})$, we condition on the type of the node that is connected in layer ℓ, which is type-j with probability m_j. Then, we note that for a type-i node in layer $\ell + 1$ to be infected by a type-j neighbor in layer ℓ that it is connected via a type-c link if both of the following events hold: the node in layer ℓ is infected, which happens with probability $(1 - q_{c,\ell}^j)$, *and* the pathogen is transmitted from the node in layer ℓ to its parent in layer $\ell + 1$, which happens with probability $\mathbf{T_c}[j,i]$. Collecting, we see that the probability of a type-i node in layer $\ell + 1$ to be *not* infected by a type-j neighbor in layer ℓ that it is connected via a type-c is given by

$$1 - (1 - q_{c,\ell}^j)\mathbf{T_c}[j,i] = 1 - \mathbf{T_c}[j,i] + q_{c,\ell}^j \mathbf{T_c}[j,i]$$

Proceeding similarly for $f_i(\boldsymbol{q}_{s,\ell})$, we establish (10).

We are now in a position to compute q_∞^i for each $i = 1, \ldots, M$. First, solving (8)-(9) in the limit of $\ell \to \infty$ we compute $\boldsymbol{q}_{c,\infty}$ and $\boldsymbol{q}_{s,\infty}$. Using these, we then get

$$q_\infty^i = \sum_{\boldsymbol{d}=(k_c,k_s)} p_{\boldsymbol{d}} f_i(\boldsymbol{q}_{c,\infty}, \boldsymbol{q}_{s,\infty}, k_c, k_s) \qquad (11)$$

by conditioning on the colored degree of the root node. Finally, we have $\mathbb{E}[R_i \mid R > 0] = 1 - q_\infty^i$ and the expected epidemic size is given by $\text{ES} = \mathbb{E}[R \mid R > 0] = \sum_{i=1}^M m_i (1 - q_\infty^i)$.

IV. Numerical Results

We next present simulation results with an eye towards validating our analytical results (which are exact in the limit of number of nodes n goes to infinity) in the finite node regime. In doing so, we also aim to shed light on how various parameters used in the model affect the spreading process, e.g., in terms of probability and expected size of epidemics. Throughout, we fix the number of nodes as $n = 10^6$. For each parameter setting, we run 10,000 independent experiments and report the average of these independent runs. The networks $\mathbb{C}$ and $\mathbb{S}$ are generated according to the configuration model with Poisson degree distributions with mean degrees $\langle k_c \rangle$ and $\langle k_s \rangle$, respectively. The fraction of nodes that participate in network layer $\mathbb{S}$ is denoted by α. The baseline transmissibilities are denoted by T_c and T_s for networks $\mathbb{C}$ and $\mathbb{S}$, respectively. We assume there are two types of nodes in the population, type-1 and type-2; e.g., these can represent individuals wearing a surgical mask (type-1) and a cloth mask (type-2). The vector $\boldsymbol{m} = [m_1, m_2]$ represents the node type distribution in the population, with m_1 and m_2 denoting fractions of type-1 and type-2 individuals, respectively. The inward efficiencies of the two types of masks are represented by the vector $\boldsymbol{\epsilon}_{in} = [\epsilon_{in,1}, \epsilon_{in,2}]$, and outward efficiencies are given by $\boldsymbol{\epsilon}_{out} = [\epsilon_{out,1}, \epsilon_{out,2}]$.
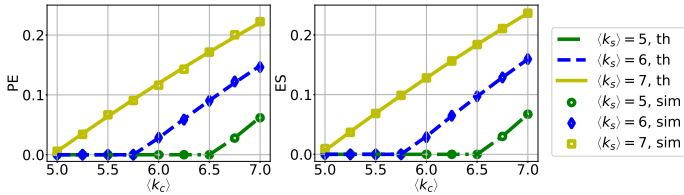

Fig. 1. *Probability of emergence (left) and the expected epidemic size (right) with varying mean degrees $\langle k_c \rangle$ and $\langle k_s \rangle$, when $\boldsymbol{m} = [0.4, 0.6]$, $\alpha = 0.6$, $\boldsymbol{\epsilon}_{out} = [0.8, 0.5]$, $\boldsymbol{\epsilon}_{in} = [0.7, 0.5]$, $T_c = 0.6$ and $T_s = 0.5$. Simulation results (marked as sim in the legend) show near-perfect agreement with our theoretical results (marked as th in the legend). For data points near the epidemic threshold (i.e., points where PE and ES transition from being zero to positive) we increased the network size to $n = 10^7$.*

In the first set of experiments, we investigate how the probability of emergence (PE) and expected epidemic size given emergence (ES) change as we vary the mean degrees $\langle k_c \rangle$ and $\langle k_s \rangle$. The results are reported in Figure 1 and show that the simulation results match the analytical solutions with near-perfect accuracy. This confirms the usefulness of our results in the finite node regime. These results can also help understand the impact of strategies that reduce the mean

degrees of contact networks, such as social distancing and quarantines, in mitigating the spread of a virus.

In the second set of experiments, we explore how PE and ES change with varying baseline transmissibilities T_c and T_s. The results are reported in Figure 2 and as before our analytical results are seen to match the simulations very well. These results can also be helpful in understanding the increased risk of epidemics in cases where high-transmissibility virus variants may emerge over time, e.g., the Delta variant for COVID-19.
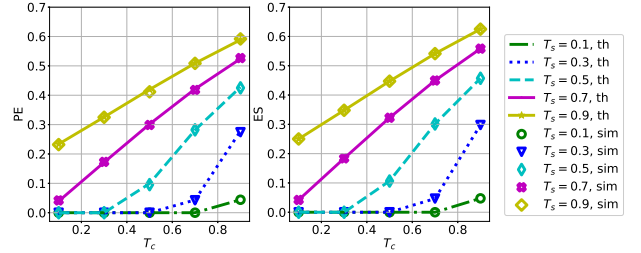

Fig. 2. *Probability of emergence (left) and the expected epidemic size (right) with varying T_c and T_s, when $\langle k_c \rangle = 6$, $\langle k_s \rangle = 8$, $\alpha = 0.6$, $\boldsymbol{m} = [0.4, 0.6]$, $\boldsymbol{\epsilon}_{out} = [0.8, 0.5]$, and $\boldsymbol{\epsilon}_{in} = [0.7, 0.5]$.*

Next, we present in Figure 3 the boundary of $T_c - T_s$ plane (Figure 3(a)) and $T_c - \alpha$ plane (Figure 3(b)) that identify the epidemic threshold $\rho(\boldsymbol{J}) = 1$. In this experiment, we assume two types of nodes: mask and no mask. m_1 is the fraction of mask-wearers and $m_2 = 1 - m_1$ is the fraction of no-mask nodes. In Figure 3(a), for each α, the curves separate the areas where epidemics can take place (north-east of the curves) from the areas where they can not (south-west of the curves). It is observed that with the same T_c, increasing α decreases the minimum T_s that is needed for epidemics to happen. Figure 3(b) presents the boundary of $T_c - \alpha$ plane separating $\rho(\boldsymbol{J}) = 1$. We set $T_s = 0.9 T_c$. Similarly, in Figure 3(b) for each m_1, the curves separate the areas where epidemics can take place (north-east of the curves) from the areas where they can not (south-west of the curves). We can see that with the same m_1, increasing T_c (correspondingly T_s) decreases the minimum α that is needed for epidemics to happen. This shows that α and the baseline transmissibilities both contribute to facilitating the transmission, and they compensate for each other. On the other hand, if locate same T_c, an increase of m_1 lifts up the epidemic boundary, i.e., having more masks increases the minimum α required for the epidemic.

We now focus on analyzing a scenario where different regions have different school participation rates and population mask distributions. We assume surgical and cloth mask wearers are in the population, with mask distribution $\boldsymbol{m} = [m_1, m_2]$. Figure 4 shows the PE and ES varying α and m_1 ($m_2 = 1 - m_1$ correspondingly) from 0.1 to 0.9 with intervals 0.1. Figure 4 shows the epidemic boundary defined by $m - \alpha$ and the trend of PE and ES when exceeding the boundary. This shows how these quantities depend on the corresponding variables and allows analyzing cases where the epidemic already exists. We can see that α monotonically
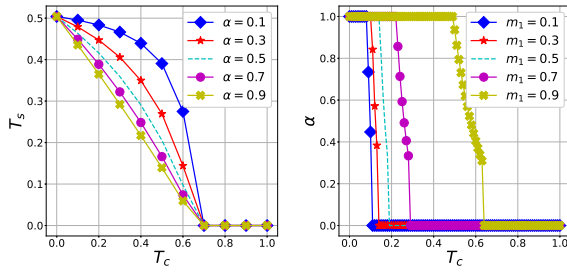
(a) $T_c - T_s$ epidemic boundary (b) $T_c - \alpha$ epidemic boundary

Fig. 3. *Epidemic boundary defined by T_c and T_s (a), and defined by T_c and α (b). $\langle k_c \rangle = 6$ and $\langle k_s \rangle = 8$, $\boldsymbol{\epsilon}_{in} = [0.7, 0]$ and $\boldsymbol{\epsilon}_{out} = [0.8, 0]$. For (a) $T_c = 0.6$, $T_s = 0.5$, $\boldsymbol{m} = [0.8, 0.2]$, and for (b) $T_s = 0.9T_c$. The north and east of each curve specify the region for which epidemics are possible, while the south and west parts of each curve stand for the region where epidemics can not occur.*

increases PE and ES. The separation by the $m_1 - \alpha$ boundary indicates that if a proper mask-wearing distribution is in place, it might be possible to find a *safe* condition (i.e., PE and ES are zeros) where the population bears no risk of an epidemic with a corresponding school participation rate.
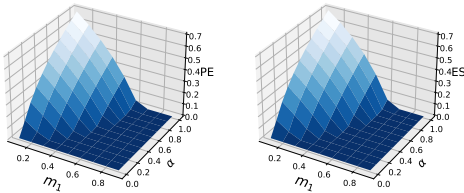

Fig. 4. *Probability of the emergence (left), the expected epidemic size given emergence (right). $\langle k_c \rangle = 6$ and $\langle k_s \rangle = 8$, $\boldsymbol{\epsilon}_{out} = [0.8, 0.5]$, $\boldsymbol{\epsilon}_{in} = [0.7, 0.5]$, $T_c = 0.6$ and $T_s = 0.5$.*

V. CONCLUSION

In this paper, we have studied spreading processes on *multi-layer* networks with *population heterogeneity*. In the model, the heterogeneous viral transmission probability between nodes depends both on the node types and the link type that connects both nodes. We provide theoretical solutions of three fundamental quantities describing such spreading processes: the probability of emergence, the epidemic threshold, and the expected epidemic size. We validate our analytical results by comparing them against agent-based simulations, showing near-perfect matches. Analytical results demonstrate how the spreading processes are impacted by the network structure, viral transmission dynamics, and mask distribution. Epidemic boundaries define safe conditions where the population bears no risk of an epidemic when trading off different factors. For future work, we can further consider layer-dependent population heterogeneity, networks with different structural properties (e.g., clustering and assortativity), and more complicated transmission dynamics such as mutation.

REFERENCES

[1] M. Tambuscio, G. Ruffo, A. Flammini, and F. Menczer, "Fact-checking Effect on Viral Hoaxes: A Model of Misinformation Spread in Social Networks," in *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, 2015.

[2] J. P. Gleeson and D. J. Cahalane, "Seed size strongly affects cascades on random networks," *Physical Review E*, vol. 75, no. 5, 2007.

[3] M. Newman, "Epidemics on networks," in *Networks* (M. Newman, ed.), 2018.

[4] M. Li, X. Wang, K. Gao, and S. Zhang, "A Survey on Information Diffusion in Online Social Networks: Models and Methods," *Information*, vol. 8, no. 4, 2017.

[5] A. Allard, P.-A. Noël, L. Dubé, and B. Pourbohloul, "Heterogeneous Bond Percolation on Multitype Networks with an Application to Epidemic Dynamics," *Physical review. E, Statistical, nonlinear, and soft matter physics*, vol. 79, 2009.

[6] O. Yağan, D. Qian, J. Zhang, and D. Cochran, "Conjoining Speeds up Information Diffusion in Overlaying Social-Physical Networks," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 6, 2013.

[7] M. E. J. Newman, "Spread of epidemic disease on networks," *Physical Review E*, vol. 66, no. 1, 2002.

[8] P. Grassberger, "On the critical behavior of the general epidemic process and dynamical percolation," *Mathematical Biosciences*, vol. 63, no. 2, 1983.

[9] Y. Zhuang and O. Yağan, "Information propagation in clustered multi-layer networks," *IEEE Transactions on Network Science and Engineering*, vol. 3, no. 4, pp. 211–224, 2016.

[10] M. E. Newman, "Random graphs with clustering," *Physical review letters*, vol. 103, no. 5, p. 058701, 2009.

[11] J. C. Miller, "Percolation and epidemics in random clustered networks," *Physical Review E*, vol. 80, no. 2, p. 020901, 2009.

[12] C. Buono, L. G. Alvarez-Zuzek, P. A. Macri, and L. A. Braunstein, "Epidemics in Partially Overlapped Multiplex Networks," *PLOS ONE*, vol. 9, no. 3, 2014.

[13] Y. Tian, A. Sridhar, C. W. Wu, S. A. Levin, K. M. Carley, H. V. Poor, and O. Yağan, "The Role of Masks in Mitigating Viral Spread on Networks," *Arxiv preprint http://arxiv.org/abs/2110.04398*, 2022.

[14] X. Chen, G. Zhu, L. Zhang, Y. Fang, L. Guo, and X. Chen, "Age-Stratified COVID-19 Spread Analysis and Vaccination: A Multitype Random Network Approach," *IEEE transactions on network science and engineering*, vol. 8, no. 2, 2021.

[15] Y. Tian, A. Sridhar, O. Yağan, and H. V. Poor, "Analysis of the Impact of Mask-wearing in Viral Spread: Implications for COVID-19," in *2021 American Control Conference (ACC)*, 2021.

[16] D.-S. Lee and M. Zhu, "Epidemic Spreading in a Social Network with Facial Masks wearing Individuals," *IEEE Transactions on Computational Social Systems*, vol. 8, no. 6, 2021.

[17] G. Bonifazi, B. Breve, S. Cirillo, E. Corradini, and L. Virgili, "Investigating the COVID-19 vaccine discussions on Twitter through a multi-layer network-based approach," *Information Processing & Management*, vol. 59, no. 6, 2022.

[18] G. J. Milne, J. K. Kelso, H. A. Kelly, S. T. Huband, and J. McVernon, "A Small Community Model for the Transmission of Infectious Diseases," *Plos One*, vol. 3, no. 12, 2008.

[19] C. Bongiorno and L. Zino, "A multi-layer network model to assess school opening policies during a vaccination campaign," *Applied Network Science*, vol. 7, no. 1, 2022.

[20] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random structures & algorithms*, vol. 6, no. 2-3, pp. 161–180, 1995.

[21] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Physical Review E*, vol. 64, no. 2, 2001.

[22] H. K. Alexander and T. Day, "Risk factors for the evolutionary emergence of pathogens," *Journal of The Royal Society Interface*, vol. 7, no. 51, 2010.

[23] R. Eletreby, Y. Zhuang, K. M. Carley, O. Yağan, and H. V. Poor, "The effects of evolutionary adaptations on spreading processes in complex networks," *Proceedings of the National Academy of Sciences*, vol. 117, no. 11, 2020.

[24] A. Sridhar, O. Yağan, R. Eletreby, S. A. Levin, J. B. Plotkin, and H. V. Poor, "Leveraging A Multiple-Strain Model with Mutations in Analyzing the Spread of Covid-19," in *ICASSP*, 2021.

[25] O. Yağan, A. Sridhar, R. Eletreby, S. Levin, J. B. Plotkin, and H. V. Poor, "Modeling and analysis of the spread of covid-19 under a multiple-strain model with mutations," *Harvard Data Science Review*, vol. 4, 2021.

[26] B. Söderberg, "Properties of random graphs with hidden color," *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 68, no. 2 Pt 2, 2003.