The effects of evolutionary adaptations on spreading processes in complex networks

Article *in* Proceedings of the National Academy of Sciences · March 2020

DOI: 10.1073/pnas.1918529117

CITATIONS
61

READS
145

5 authors, including:

Kathleen M Carley
Carnegie Mellon University
837 PUBLICATIONS **27,950** CITATIONS
SEE PROFILE

Osman Yagan
Carnegie Mellon University
145 PUBLICATIONS **2,440** CITATIONS
SEE PROFILE

H. Vincent Poor
Princeton University
3,003 PUBLICATIONS **132,040** CITATIONS
SEE PROFILE

The Effects of Evolutionary Adaptations on Spreading Processes in Complex Networks

Rashad Eletreby,[1] Yong Zhuang,[1] Kathleen M. Carley,[2] H. Vincent Poor,[3] and Osman Yağan[1,*]

[1] *Department of Electrical and Computer Engineering, Carnegie Mellon University Pittsburgh, PA 15213 USA*

[2] *Institute for Software Research, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213 USA*

[3] *Department of Electrical Engineering, Princeton University, Princeton, NJ 08540 USA*

(Dated: August 29, 2019)

A common theme among the proposed models for *network epidemics* is the assumption that the propagating object, i.e., a pathogen (in the context of infectious disease propagation) or a piece of information (in the context of information propagation), is transferred across the nodes without going through any modification or *evolution*. However, in real-life spreading processes, pathogens often *evolve* in response to changing environments and medical interventions and information is often modified by individuals before being forwarded. In this paper, we investigate the *evolution* of spreading processes in complex networks with the aim of i) revealing the role of evolutionary adaptations on the threshold, probability, and final size of epidemics; and ii) exploring the interplay between the structural properties of the network and the evolutionary adaptations of the process. We start by considering the case where *co-infection* with different pathogen strains (respectively, different variations of information) is not possible. In this case, we develop a mathematical theory that accurately predicts the epidemic threshold and the expected epidemic size as functions of the characteristics of the spreading process, the evolutionary pathways of the pathogen (respectively, information), and the structure of the underlying contact network. In addition to the mathematical theory, we perform extensive simulations on random and real-world contact networks to verify our theory and reveal the significant shortcomings of the classical mathematical models that do not capture evolution. Our results reveal that the classical, single-type bond-percolation models may accurately predict the threshold and final size of epidemics, but their predictions on the probability of emergence are *inaccurate* on both random and real-world networks. We then consider the case when *co-infection* is possible, i.e., a susceptible individual who receives *simultaneous* infections with multiple pathogen strains (respectively, multiple variations of information) becomes co-infected. We show by computer simulations that co-infection gives rise to a rich set of dynamics: it can amplify or inhibit the spreading dynamics, and more remarkably *lead the order of phase transition to change from second-order to first-order*.

Keywords: Evolution, Epidemics, Information Propagation, Phase Transitions, Spreading Processes.

I. INTRODUCTION

What causes an outbreak of a disease? How can we predict its emergence and control its progression? Over the past several decades, multidisciplinary research efforts were converging to tackle the above questions, aiming for providing a better understanding of the intricate dynamics of disease propagation and accurate predictions on its course [1–12]. At the heart of these research efforts is the development of mathematical models that provide insights on predicting, assessing, and controlling potential outbreaks [13–16]. The early mathematical models relied on the *homogeneous mixing* assumption, meaning that an infected individual is equally likely to infect any other individual in the population, without regard to her location, age, or the people with whom she interacts. Homogeneity allowed writing a set of differential equations that characterize the speed and scale of propagation (in the limit of large population size), providing insights on how the parameters of a disease, e.g., its *basic reproductive number*, indicate whether a disease will die out, or an epidemic will emerge [5, 16].

In real-life, however, the spread of a disease is highly dependent on the contact patterns between individuals. In particular, a person may only infect those with whom she interacts, and the number of contacts people have, varies dramatically between individuals. These basic observations render the homogeneous mixing models inaccurate, as they tend to underestimate the epidemic size in the initial stages of the outbreak and overestimate it towards the end [17]. As a result of the these shortcomings, *network epidemics* has emerged as a mathematical modeling approach that takes the underlying contact network into consideration [1, 3, 18–20]. Since then, a large body of research has looked into the delicate interplay between the structural properties of the contact network and the dynamics of propagation, leading to accurate predictions of the spatio-temporal progression of disease outbreaks. In addition to diseases, opinions and information also propagate through networks in patterns similar to those of epidemics [21]. Hence, research efforts on *information propagation* draw on the theory of infectious diseases to model the dynamics of propagation [8, 22–25]. Throughout, we use the term *spreading processes* to de-

* oyagan@ece.cmu.edu

note a general class of processes that propagate in contact networks, such as infectious diseases and information.

A common theme among the proposed models for network epidemics is the assumption that the propagating object, i.e., a virus or a piece of information, is transferred across the nodes without going through any modification or *evolution* [5, 22, 23, 26–31]. However, in real-life spreading processes, pathogens often *evolve* in response to changing environments and medical interventions [9, 32–35], and information is often modified by individuals before being forwarded [36, 37]. In fact, 60% of the (approximately) 400 emerging infectious diseases that have been identified since 1940 are zoonotic [38] [39, 40]. A zoonotic disease is initially poorly adapted, poorly replicated, and inefficiently transmitted [41], hence its ability to go from animal-to-human transmissions to human-to-human transmissions depends on the pathogen *evolving* to a strain that is well-adapted to the human host.

Similar patterns of evolution are observed in the way information propagates among individuals. Needless to say, one observes, on a daily basis, how information mutates unintentionally, or perhaps intentionally by an adversary, on social media platforms [36]. At a high-level, an individual may mutate the information by exaggeration, hoping for her variant to go viral. Mutations may also occur unintentionally. In particular, Dawkins [42] argued that ideas and information spread and evolve between individuals with patterns similar to genes, in a sense that they self-replicate, mutate, and respond to selective pressure as they interact with their host. Concluding, if we are to ignore evolution, we underestimate the severity of the epidemic and fail to understand the intricate interplay between the dynamics of propagation and evolution.

In this paper, we aim to bridge the disconnect between how spreading processes propagate *and evolve* in real-life, and the current mathematical and simulation models that do not capture evolution. In particular, we investigate the *evolution* of spreading processes with the aim of i) revealing the role of evolutionary adaptations on the threshold, probability, and final size of epidemics; and ii) understanding the interplay between the structural properties of the network and the evolutionary adaptations of the process. Throughout, we use the term *epidemics* to denote disease/information outbreaks that result in a positive fraction of infected individuals in the limit of large network size and *self-limited outbreaks* to denote small disease/information outbreaks for which the fraction of infected individuals tends to zero in the limit of large network size. We also use the term *strain* to denote a pathogen strain in the context of infectious disease propagation, or a particular variation of the information in the context of information propagation. At a high level, strains represent homogeneous groups within species [43] and they generally possess unique features such as virulence, infectivity, growth rate, etc.

In modeling the evolution of spreading processes, we adopt the multiple-strain model that was introduced by Alexander and Day in [33]. Their model can be briefly outlined as follows (more details are given in Section II). Consider a multiple-strain spreading process that starts with an individual, i.e., the seed, receiving infection (from an external reservoir) with strain-1 of a particular pathogen (respectively, information). The seed infects each of her contacts independently with probability T_1, called the *transmissibility* of strain-1. Once a susceptible individual receives the infection from the seed, the pathogen may evolve within that new host prior to any subsequent infections. In particular, the pathogen may remain as strain-1 with probability μ_{11} or mutate to strain-2 (that has transmissibility T_2) with probability $\mu_{12} = 1 - \mu_{11}$. If the pathogen remains as strain-1 (respectively, mutates to strain-2) within a newly infected host, then that host infects each of her susceptible neighbors in the subsequent stages independently with probability T_1 (respectively, T_2). As the process continues to grow, if any susceptible individual receives strain-1, the pathogen may remain as strain-1 with probability μ_{11} or mutate to strain-2 with probability $\mu_{12} = 1 - \mu_{11}$ prior to subsequent infections. Similarly, if any susceptible individual receives strain-2, the pathogen may remain as strain-2 with probability μ_{22} or mutate to strain-1 with probability $\mu_{21} = 1 - \mu_{22}$ prior to subsequent infections. The process continues to grow until no additional infections are possible. We remark that it is straightforward to extend the model to the general case, where there are m possible strains for some finite integer $m \geq 2$. More details are given in Section III.

Note that as multiple strains propagate throughout the population, a susceptible individual may simultaneously get into infectious contact with neighbors infected with strain-1 as well as neighbors infected with strain-2. This gives rise to the possibility of a susceptible individual becoming *co-infected* with multiple pathogen strains. Indeed, co-infection with multiple pathogen strains is prevalent in disease-causing protozoa, helminths, bacteria, fungi, and viruses and is known to cause significant implications [43–47]. However, from a mathematical standpoint, the possibility of co-infections creates phase discontinuities (see Section VI) that render the process mathematically intractable.

We start by considering the case when co-infection is ignored, meaning that a susceptible individual may *only* be infected with a single strain. In particular, a susceptible individual who simultaneously receives x infections of strain-1 and y infections of strain-2 becomes infected by strain-1 (respectively, strain-2) with probability $x/(x+y)$ (respectively, $y/(x+y)$). In this case, we develop a mathematical theory that draws on the tools developed for analyzing the zero-temperature random-field Ising model on Bethe lattices [48] as well as on random graphs [49, 50]. Our theory fully characterizes the process and accurately predicts the epidemic threshold, expected epidemic size and the expected fraction of individuals infected by each strain (all at steady state). These metrics are computed

as functions of the characteristics of the spreading process (i.e., T_1 and T_2), evolutionary adaptations (i.e., μ_{11} and μ_{22}), and the structure of the underlying contact network (e.g., its degree distribution).

In addition to the mathematical theory, we perform extensive simulations on random graphs with arbitrary degree distributions (generated by the configuration model [51–53]) as well as with real-world networks (obtained from SNAP dataset [54] as well as [70] and [71]) to verify our theory and reveal the significant shortcomings of the classical mathematical models that do not capture evolution. In particular, we show that the classical, single-type bond-percolation models [3, 55–57] may accurately predict the threshold and final size of epidemics, but their predictions on the probability of emergence are *significantly inaccurate* on both random and real-world networks. This inaccuracy sheds the light on a fundamental disconnect between the classical single-type, bond-percolation models and real-life spreading processes that entail evolution.

We then focus on the case where co-infection is possible. Although recent studies have shown that co-infection with multiple pathogen strains is prevalent in nature [43–47], there has been a lack of models that explain its occurrence, reveal its implications, and investigate its delicate interplay with the underlying contact network. Note that a considerable amount of literature has examined the case where co-infection with multiple *diseases* is possible [58–61], yet multiple-disease co-infection is fundamentally different from multiple-strain co-infection (see Section I in the Supplementary Material). In this paper, we use computer simulations to explore the case where multiple-strain co-infection is possible. In particular, a susceptible individual who gets infected with strain-1 and strain-2 *simultaneously* becomes co-infected, and starts to transmit the co-infection, i.e., the mixture of the two strains, with a transmissibility T_{co}.

The transmissibility T_{co} could be larger than the maximum of T_1 and T_2 (e.g., modeling a synergistic cooperation between the two resident strains) or smaller than their minimum (e.g., modeling a negative competition among the two resident strains), and it may also fall anywhere in between. We show that co-infection gives rise to a rich set of dynamics: it can amplify or inhibit the spreading dynamics, and more remarkably *lead the order of phase transition to change from second-order to first-order*. We investigate the interplay between the characteristics of co-infection, the structure of the underlying contact network, and evolutionary adaptations and reveal the cases where such interplay induces a *first-order* phase transition for the expected epidemic size.

Summary: We consider the evolution of spreading processes in complex networks. We start with the case where co-infection is ignored. In this case, we develop a mathematical theory that unravels the relationship between the characteristics of the spreading process, the structure of the underlying contact network, and the process of evolution, thereby, providing accurate predictions on the epidemic threshold, expected epidemic size, and the expected fraction of individuals infected by each strain at steady state. In addition to the mathematical theory, we perform extensive simulations on random and real-world networks to verify our theory and reveal the significant shortcomings of the classical mathematical models that do not capture evolution. Then, we use computer simulations to explore the case where co-infection is possible and show that co-infection could *lead the order of phase transition to change from second-order to first-order*. We investigate the interplay between the characteristics of co-infection, the structure of the underlying contact network, and evolutionary adaptations and explain how such interplay controls the order of phase transition for the expected epidemic size.

Structure: The rest of the paper is organized as follows. In Section II, we present the multiple-strain model for evolution and demonstrate how we model the underlying contact network. In Section III, we present and derive the main results of this work, while in Section IV, we confirm our theoretical results via computer simulations. We empirically consider the case where co-infection is possible in Section VI. In Section V, we consider evolution on real-world networks and reveal the significant shortcomings of the classical mathematical models that do not capture evolution. Finally, Section VII concludes the paper.

II. MODEL DEFINITIONS

A. A multiple-strain model for evolution

In [33], Alexander and Day proposed a *multiple-strain model* that accounts for evolution. Their model is captured by two matrices, namely, the transmissibility matrix $\boldsymbol{T}$ and the mutation matrix $\boldsymbol{\mu}$, both with dimensions $m \times m$ for a finite integer $m \geq 2$ denoting the number of possible strains. The transmissibility matrix $\boldsymbol{T}$ is a $m \times m$ diagonal matrix, with $[T_i]$ representing the transmissibility of strain-i, i.e.,

$$\boldsymbol{T} = \begin{bmatrix} T_1 & 0 & \dots & 0 \\ 0 & T_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & T_m \end{bmatrix}.$$

The mutation matrix $\boldsymbol{\mu}$ is a $m \times m$ matrix with μ_{ij} denoting the probability that strain-i mutates to strain-j. Note that $\sum_j \mu_{ij} = 1$, hence $\boldsymbol{\mu}$ is a row-stochastic matrix. One example for the transmissibility and mutation matrices was given by Antia et al. in [34], where the fitness landscape consisted of m strains, with strain-1 through $m-1$ having identical transmissibility such that $R_{0,i} < 1$ for $i = 1, \dots, m-1$, with $R_{0,i}$ denoting the basic reproductive number of strain-i. Strain-m has transmissibility T_m such that $R_{0,m} > 1$, hence the emergence of

the pathogen requires evolution from strain-1 to strain-m. Antia et al. considered the the so-called *one-step irreversible mutation* [33, 34] where the pathogen must acquire $m-1$ mutations (in order and one at a time) to evolve to strain-m , i.e.,

$$\boldsymbol{T} = \begin{bmatrix} T_1 & 0 & 0 & \dots & 0 \\ 0 & T_1 & 0 & \dots & 0 \\ 0 & 0 & T_1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & T_m \end{bmatrix}$$

and

$$\boldsymbol{\mu} = \begin{bmatrix} 1-\mu & \mu & 0 & \dots & 0 & 0 & 0 \\ 0 & 1-\mu & \mu & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1-\mu & \mu \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}$$

The multiple-strain model proposed by Alexander and Day [33] works as follows. Consider a spreading process that starts with an individual, i.e., the seed, receiving infection with strain-1 from an external reservoir. Since strain-1 has transmissibility T_1, the seed infects each of her contacts independently with probability T_1. Once a susceptible individual receives the infection from the seed, the pathogen may evolve within that new host prior to any subsequent infections. In particular, the pathogen may remain as strain-1 with probability μ_{11} or mutate to strain-i (that has transmissibility T_i) with probability μ_{1i} for $i = 2, \dots, m$. If the pathogen remains as strain-1 (respectively, mutates to strain-i), then the host infects each of her susceptible neighbors in the subsequent stages independently with probability T_1 (respectively, T_i). Observe that as the process continues to grow, multiple strains may coexist in the population as governed by the transmissibility matrix $\boldsymbol{T}$ and the mutation matrix $\boldsymbol{\mu}$. At an intermediate stage, if any susceptible individual receives strain-j, the pathogen may remain as strain-j with probability μ_{jj} or mutate to strain-ℓ with probability $\mu_{j\ell}$ for $\ell \in \{1, 2, \dots, m\} \setminus \{j\}$ prior to subsequent infections. The process terminates when no additional infections are possible. A graphical illustration for the case when $m = 2$ is given in Figure 1. In this paper, we focus on the case where $m = 2$, however, it is straightforward to extend our theory to handle the general case with m strains. More details are given in Section III.

B. Network Model: Random graphs with arbitrary degree distribution

Let $\mathbb{G}$ denote the underlying contact network, defined on the node set $\mathcal{N} = \{1, \dots, n\}$. We define the structure of $\mathbb{G}$ through its degree distribution $\{p_k\}$. In particular, $\{p_k, k = 0, 1, \dots\}$ gives the probability that an arbitrary node in $\mathbb{G}$ has degree k. We generate the network $\mathbb{G}$ according to the *configuration model* [51, 52], i.e., the degrees of nodes in $\mathbb{G}$ are all drawn independently from the distribution $\{p_k, k = 0, 1, \dots\}$. Furthermore, we assume that the degree distribution is well-behaved in the sense that all moments of arbitrary order are finite. Of particular importance in the context of the configuration model is the degree distribution of a randomly chosen neighbor of a randomly chosen vertex, denoted by $\{\hat{p}_k, k = 1, 2, \dots\}$, and given by

$$\hat{p}_k = \frac{k p_k}{\langle k \rangle}, \quad k = 1, 2, \dots$$

where $\langle k \rangle$ denotes the *mean degree*, i.e., $\langle k \rangle = \sum_k k p_k$.

III. ANALYSIS

The analysis of the probability of emergence was established by Alexander and Day in [33]. For completeness, a brief summary of their results is given in Section II in the Supplementary Material.

Our objective is to derive the expected epidemic size S and the expected fraction of individuals infected by each strain, i.e., $S_1, S_2, \dots, S_m$ for m possible strains. Note that $S = \sum_{i=1}^{m} S_i$. Below, we provide analysis for the case of two strains, but we later show how to extend our analysis to the general case with m strains, for some finite integer $m \geq 2$. We apply a *tree-based* approach that is based on the work by Gleeson [49, 50]. Their approach draws on the tools developed for analyzing the zero-temperature random-field Ising model on Bethe lattices [48]. Note that as we build our network using the configuration model, the network structure is locally tree-like with the fraction of cycles approaching zero in the limit of large network size [51–53].

Since $\mathbb{G}$ is locally tree-like, we can replace it by a tree and arrange the vertices in a hierarchical structure, such that at the top level, there is a single node (the *root*) that has degree k with probability p_k. Note that $\{p_k\}$ is a proper degree distribution with $\sum_k p_k = 1$. Each of the k neighbors of the root has degree k' with probability $k' p_{k'} / \langle k \rangle$, where $\langle k \rangle$ denotes the mean degree of the network. Furthermore, we label the levels of the tree from level $\ell = 0$ at the bottom to level $\ell = \infty$ at the top, i.e., the root.

We assume that nodes update their status starting from the bottom of the tree and proceeding towards the top. This gives rise to a delicate case, where a node at some level ℓ may be exposed to *simultaneous* infections by both strain-1 and strain-2 from her neighbors at level $\ell - 1$. In the remainder of this section, we assume that *co-infection* is not possible, hence a node that receives x infections of strain-1 and y infections of strain-2 becomes infected by strain-1 (respectively, strain-2) with probability $x/(x+y)$ (respectively, $y/(x+y)$). In Section VI, we *empirically* consider the case where co-infection is possible, i.e., a node that receives simultaneous infections by both strains becomes co-infected and starts to spread the

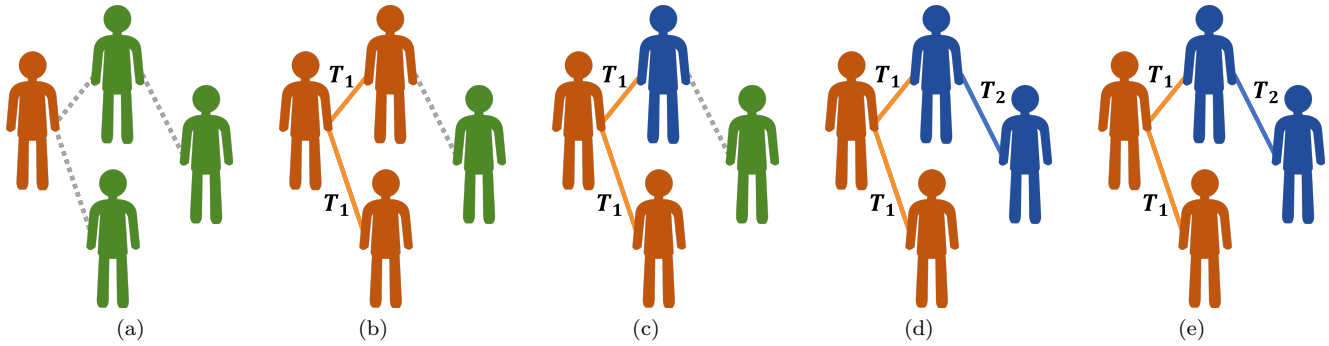

FIG. 1. **The multiple-strain model for evolution**. *(a) The process starts with a single individual, i.e., the seed, receiving infection with strain-1 (highlighted in orange) from an external reservoir. (b) The seed infects each of her susceptible neighbors (highlighted in green) independently with probability T_1. (c) The pathogen mutates independently within hosts. The pathogen remains as strain-1 with probability μ_{11} or mutates to strain-2 (highlighted in blue) with probability μ_{12}. (d) Individuals whose pathogen has mutated to strain-i infect their neighbors independently with probability T_i. (e) The pathogen mutates independently within hosts. The pathogen remains as strain-2 with probability μ_{22} or mutates to strain-1 with probability μ_{21}.*

co-infection in the subsequent rounds. In this case, co-infection may be modeled as an additional strain that has transmissibility T_{co} and never mutates back to strain-1 or strain-2.

Throughout, we say that a node is either *inactive* if it has not received any infection (i.e., still susceptible) or *active and type-i* if it has been infected and then *mutated* to strain-i, for $i = 1, 2$. With a slight abuse of notations, let $q_{\ell+1,i}$ be the probability that a node at level $\ell+1$, say node v, is active *and* type-i. Furthermore, let $q_{\ell+1} = q_{\ell+1,1} + q_{\ell+1,2}$, i.e., $q_{\ell+1}$ is the total probability that a node at level $\ell+1$ is active. We start by an arbitrary initial distribution for $\{q_{0,1}, q_{0,2}\}$ satisfying $q_{0,1} > 0, q_{0,2} > 0$. Then, we update the distribution properly until we reach the root. Note that if the degree of node v is k, then node v is using one edge to connect to her parent at level $\ell+2$, and $k-1$ edges to connect to her neighbors at level ℓ. We can condition on the *excess degree* ($\tilde{d}$) of node v to get

$$q_{\ell+1,i} = \sum_{k=1}^{\infty} \frac{kp_k}{\langle k \rangle} \mathbb{P}\left[\text{node } v \text{ becomes active and type-i} \,\middle|\, \tilde{d} = k-1\right]$$

Next, we further condition on the number of *active* neighbors of type-1 and type-2. Note that we have a Multinomial distribution for the number of active neighbors of both types. In particular, a neighbor at level ℓ may be active and type-1 with probability $q_{\ell,1}$, active and type-2 with probability $q_{\ell,2}$, or inactive with probability $1-q_\ell = 1-q_{\ell,1}-q_{\ell,2}$. Let I_i denote the number of active neighbors of type-i. Thus,

$$\begin{aligned} q_{\ell+1,i} = \sum_{k=1}^{\infty} \frac{kp_k}{\langle k \rangle} \sum_{k_1=0}^{k-1} \sum_{k_2=0}^{k-1-k_1} \binom{k-1}{k_1} \binom{k-1-k_1}{k_2} (q_{\ell,1})^{k_1} \\ \cdot (q_{\ell,2})^{k_2} (1-q_{\ell,1}-q_{\ell,2})^{k-1-k_1-k_2} \\ \cdot \mathbb{P}\left[\text{node } v \text{ becomes active and type-i} \right. \\ \left. \,\middle|\, I_1 = k_1, I_2 = k_2\right] \end{aligned}$$

Let X and Y denote the number of infections received from type-1 and type-2 neighbors, respectively. Note that conditioned on having k_1 and k_2 active neighbors of type-1 and type-2, respectively, we have

$$\begin{aligned} X &\sim \text{Binomial}(k_1, T_1) \\ Y &\sim \text{Binomial}(k_2, T_2) \end{aligned}$$

where T_i denotes the transmissibility of strain-i. Let

$$A := \mathbb{P}\left[\text{node } v \text{ becomes active and type-i} \,\middle|\, I_1 = k_1, I_2 = k_2\right]$$

Consider a particular realization (x, y) of the random variables (X, Y). Observe that if $x > 0, y = 0$, then node v becomes infected by strain-1 and eventually mutates to type-i with probability μ_{1i}. Similarly, if $x = 0, y > 0$, then node v becomes infected by strain-2 and eventually mutates to type-i with probability μ_{2i}. Finally, if $x > 0, y > 0$, then node v becomes infected by strain-1 (respectively, strain-2) with probability $x/(x+y)$ (respectively, $y/(x+y)$) and eventually mutates to type-i with probability μ_{1i} (respectively, μ_{2i}). Hence, by conditioning on X and Y, we have

$$\begin{aligned} A &= \sum_{x=0}^{k_1} \sum_{y=0}^{k_2} \binom{k_1}{x} \binom{k_2}{y} T_1^x T_2^y (1-T_1)^{k_1-x} (1-T_2)^{k_2-y} \cdot \\ &\quad \cdot \mathbb{P}\left[A \,\middle|\, X = x, Y = y\right] \\ &= \sum_{x=0}^{k_1} \sum_{y=0}^{k_2} \binom{k_1}{x} \binom{k_2}{y} T_1^x T_2^y (1-T_1)^{k_1-x} (1-T_2)^{k_2-y} \cdot \\ &\quad \cdot \Big(\mu_{1i} \mathbf{1}[x > 0, y = 0] + \mu_{2i} \mathbf{1}[x = 0, y > 0] + \end{aligned}$$

$$\left(\frac{x\mu_{1i}}{x+y}+\frac{y\mu_{2i}}{x+y}\right)\mathbf{1}[x>0,y>0]\Bigg)$$

Note that

$$\sum_{x=0}^{k_1}\sum_{y=0}^{k_2}\binom{k_1}{x}\binom{k_2}{y}T_1^xT_2^y(1-T_1)^{k_1-x}(1-T_2)^{k_2-y}\cdot$$
$$\cdot\,\mu_{1i}\mathbf{1}[x>0,y=0]$$
$$=\mu_{1i}(1-T_2)^{k_2}\left(1-\mathbb{P}(X=0)\right)$$
$$=\mu_{1i}a_2b_1$$

where $a_i=(1-T_i)^{k_i}$ and $b_i=1-a_i$. Similarly,

$$\sum_{x=0}^{k_1}\sum_{y=0}^{k_2}\binom{k_1}{x}\binom{k_2}{y}T_1^xT_2^y(1-T_1)^{k_1-x}(1-T_2)^{k_2-y}\cdot$$
$$\cdot\,\mu_{2i}\mathbf{1}[x=0,y>0]=\mu_{2i}a_1b_2$$

Thus, we have

$$q_{\ell+1,i}=\sum_{k=1}^{\infty}\frac{kp_k}{\langle k\rangle}\sum_{k_1=0}^{k-1}\sum_{k_2=0}^{k-1-k_1}\binom{k-1}{k_1}\binom{k-1-k_1}{k_2}\left(q_{\ell,1}\right)^{k_1}\left(q_{\ell,2}\right)^{k_2}\left(1-q_{\ell,1}-q_{\ell,2}\right)^{k-1-k_1-k_2}\cdot$$
$$\cdot\left(b_1a_2\mu_{1i}+a_1b_2\mu_{2i}+\sum_{x=0}^{k_1}\sum_{y=0}^{k_2}\binom{k_1}{x}\binom{k_2}{y}T_1^xT_2^y(1-T_1)^{k_1-x}(1-T_2)^{k_2-y}\left(\frac{x\mu_{1i}}{x+y}+\frac{y\mu_{2i}}{x+y}\right)\mathbf{1}[x>0,y>0]\right), \tag{1}$$

for $\ell=0,1,\ldots$ and $i=1,2$.

Observe that under the assumption that nodes do not become inactive once they turn active, the quantities $q_{\ell,i}$ appearing in (1) are non-decreasing in ℓ, and thus they converge to a limit $q_{\infty,i}$ for $i=1,2$. Finally, the final fraction of nodes that are active and type-i is equal (in expected value) to the probability that the root of the tree (at level $\ell\to\infty$) is active and type-i. Note that if the tree root has degree k, then all of these k edges will be utilized to connect with her neighbors at the lower level. Hence,

$$Q_i=\sum_{k=0}^{\infty}p_k\sum_{k_1=0}^{k}\sum_{k_2=0}^{k-k_1}\binom{k}{k_1}\binom{k-k_1}{k_2}\left(q_{\infty,1}\right)^{k_1}\left(q_{\infty,2}\right)^{k_2}\left(1-q_{\infty,1}-q_{\infty,2}\right)^{k-k_1-k_2}\cdot$$
$$\cdot\left(b_1a_2\mu_{1i}+a_1b_2\mu_{2i}+\sum_{x=0}^{k_1}\sum_{y=0}^{k_2}\binom{k_1}{x}\binom{k_2}{y}T_1^xT_2^y(1-T_1)^{k_1-x}(1-T_2)^{k_2-y}\left(\frac{x\mu_{1i}}{x+y}+\frac{y\mu_{2i}}{x+y}\right)\mathbf{1}[x>0,y>0]\right) \tag{2}$$

where Q_i for $i=1,2$ denotes the probability that the tree root is active and type-i and $q_{\infty,i}$ for $i=1,2$ is the steady-state solution of the recursive equations (1). Note that $Q=Q_1+Q_2$ is the total probability that the tree root is active.

Observe that $q_{\infty,1}=q_{\infty,2}=0$ gives a trivial fixed-point of the recursive equations (1). Indeed, this trivial solution leads to $Q=0$ by virtue of (2). Although the trivial fixed point is a valid numerical solution for the recursive equations (1), we can show that this trivial solution is *unstable*. Hence, another solution with $q_{\infty,1}>0$ and $q_{\infty,2}>0$ may exist. To test whether or not the trivial fixed point is stable, we check the spectral radius of the Jacobian matrix $\boldsymbol{J}(q_{\ell,1},q_{\ell,2})$ corresponding to the *linearization* of (1) at $q_{\ell,1}=q_{\ell,2}=0$. If the spectral radius of the $\boldsymbol{J}(q_{\ell,1},q_{\ell,2})$ at $q_{\ell,1}=q_{\ell,2}=0$ is larger than one, then the trivial fixed-point is unstable, indicating that there exists another solution with $q_{\infty,1}>0$ and $q_{\infty,2}>0$ implying the existence of a giant component. The Jacobian matrix is given by

$$\boldsymbol{J}(q_{\ell,1},q_{\ell,2})|_{q_{\ell,1}=q_{\ell,2}=0}=\begin{bmatrix}\frac{\partial q_{\ell+1,1}}{\partial q_{\ell,1}} & \frac{\partial q_{\ell+1,1}}{\partial q_{\ell,2}}\\ \frac{\partial q_{\ell+1,2}}{\partial q_{\ell,1}} & \frac{\partial q_{\ell+1,2}}{\partial q_{\ell,2}}\end{bmatrix}_{q_{\ell,1}=q_{\ell,2}=0}$$
$$=\left(\frac{\langle k^2\rangle-\langle k\rangle}{\langle k\rangle}\right)\begin{bmatrix}T_1\mu_{11} & T_2\mu_{21}\\ T_1\mu_{12} & T_2\mu_{22}\end{bmatrix}$$
$$=\left(\frac{\langle k^2\rangle-\langle k\rangle}{\langle k\rangle}\right)(\boldsymbol{T\mu})^T$$

Note that a square matrix and its transpose have the

same set of eigenvalues. It follows that a phase transition occurs when

$$\left(\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\right)\rho\left(\boldsymbol{T\mu}\right) > 1 \tag{3}$$

where $\rho\left(\boldsymbol{T\mu}\right)$ denotes the spectral radius, i.e., the largest eigenvalue (in absolute value) of the matrix product $\boldsymbol{T\mu}$.

We remark that it is straightforward to extend our analysis to the general case with m strains, for some finite integer $m \geq 2$ as long as the underlying process is *indecomposable* [33, 62, 63]. At a high level, indecomposable processes are those for which each pathogen strain i eventually gives rise to strain-j at some generation $n_{ij} \geq 1$ for $i, j = 1, 2, \ldots, m$. In other words, if an indecomposable process starts with an infection with strain-i, then as the process continues to grow, all other strains will eventually emerge. Such a property is established if, for every pair of strains (i, j), there exists a positive integer n_{ij} such that $\boldsymbol{M}^{n_{ij}}(i, j) > 0$ [33]. If the underlying process is *decomposable*, then there exist classes of strain types such that strain types belonging to the same class can eventually give rise to one another, but not to other strain types. Indeed, the existence of multiple classes leads to multiple solutions for the set of equations (2) depending on the initial distribution of $\{q_{0,1}, q_{0,2}, \ldots, q_{0,m}\}$. Hence, to guarantee the uniqueness of the solution of (2) and for mathematical tractability, we limit our formalism to the case when the underlying process is indecomposable.

IV. NUMERICAL RESULTS

A. The Structure of the Contact Network

In this section, we consider synthetic contact networks generated randomly by the configuration model, while real-world networks are considered in Section V. In particular, we consider contact networks with *Poisson* degree distribution as well as *Power-law* degree distribution.

1. Poisson degree distribution

We start by considering contact networks with Poisson degree distribution. Namely, with λ denoting the mean degree, i.e., $\lambda = \langle k\rangle$, we have

$$p_k = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, \ldots$$

In this case, condition (3) implies that phase transition occurs when

$$\lambda \times \rho\left(\boldsymbol{T\mu}\right) = 1 \tag{4}$$

where $\rho\left(\boldsymbol{T\mu}\right)$ denotes the spectral radius of the matrix multiplication $\boldsymbol{T\mu}$. Observe that condition (4) embodies the structure of the contact network (represented by λ for a contact network with Poisson degree distribution), the characteristics of propagation (represented by the matrix $\boldsymbol{T}$) and the process of evolution (represented by $\boldsymbol{\mu}$), hence it unravels how these properties interact together to yield an epidemic.

2. Power-law degree distribution

Poisson degree distribution provides a formalism for *homogeneous* networks, where the degree sequence of the graph is highly concentrated around the mean degree. However, degree sequences in real-world networks were observed to be heavily skewed to the right [1, 3, 7], meaning that the distribution is *heterogeneous*, or heavy-tailed. We consider Power-law degree distribution with exponential cutoff since they are relevant to a variety of real-world networks [3, 64]. In particular, we set

$$p_k = \begin{cases} 0 & \text{if } k = 0 \\ \left(\mathrm{Li}_\gamma\left(e^{-1/\Gamma}\right)\right)^{-1} k^{-\gamma} e^{-k/\Gamma} & \text{if } k = 1, 2, \ldots. \end{cases}$$

where γ and Γ are positive constants and $\mathrm{Li}_m(z)$ is the mth polylogarithm of z, i.e., $\mathrm{Li}_m(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^m}$. Observe that condition (3) now translates to

$$\left(\frac{\mathrm{Li}_{\gamma-2}\left(e^{-1/\Gamma}\right) - \mathrm{Li}_{\gamma-1}\left(e^{-1/\Gamma}\right)}{\mathrm{Li}_{\gamma-1}\left(e^{-1/\Gamma}\right)}\right) \times \rho\left(\boldsymbol{T\mu}\right) = 1 \tag{5}$$

Similar to (4), condition (5) indicates how the structure of the underlying network, the characteristics of propagation, and the process of evolution are intertwined together, and under what conditions their relationship would induce an epidemic.

B. Notations and Methods

Notations: In what follows, we use S, S_1 and S_2 to denote the *total* expected epidemic size, the expected fraction of nodes infected with strain-1, and the expected fraction of nodes infected with strain-2, respectively and all at the steady state, i.e., when the process terminates. We use P_1^{BP} and P_2^{BP} to denote the probability of emergence on a single-strain bond-percolated network with T_1 and the probability of emergence on a single-strain bond-percolated network with T_2, respectively.

Methods: We use the configuration model to create random random graphs with particular degree distributions. In particular, we sample a degree sequence from the corresponding distribution, then we use the configuration model to construct a random graph with that degree sequence. We use igraph [65] on both C++ and Python for simulations. Our simulation codes are available online [66]. Unless otherwise stated, we start the process by selecting a node uniformly at random and infecting it with strain-1. The node infects each neighbor

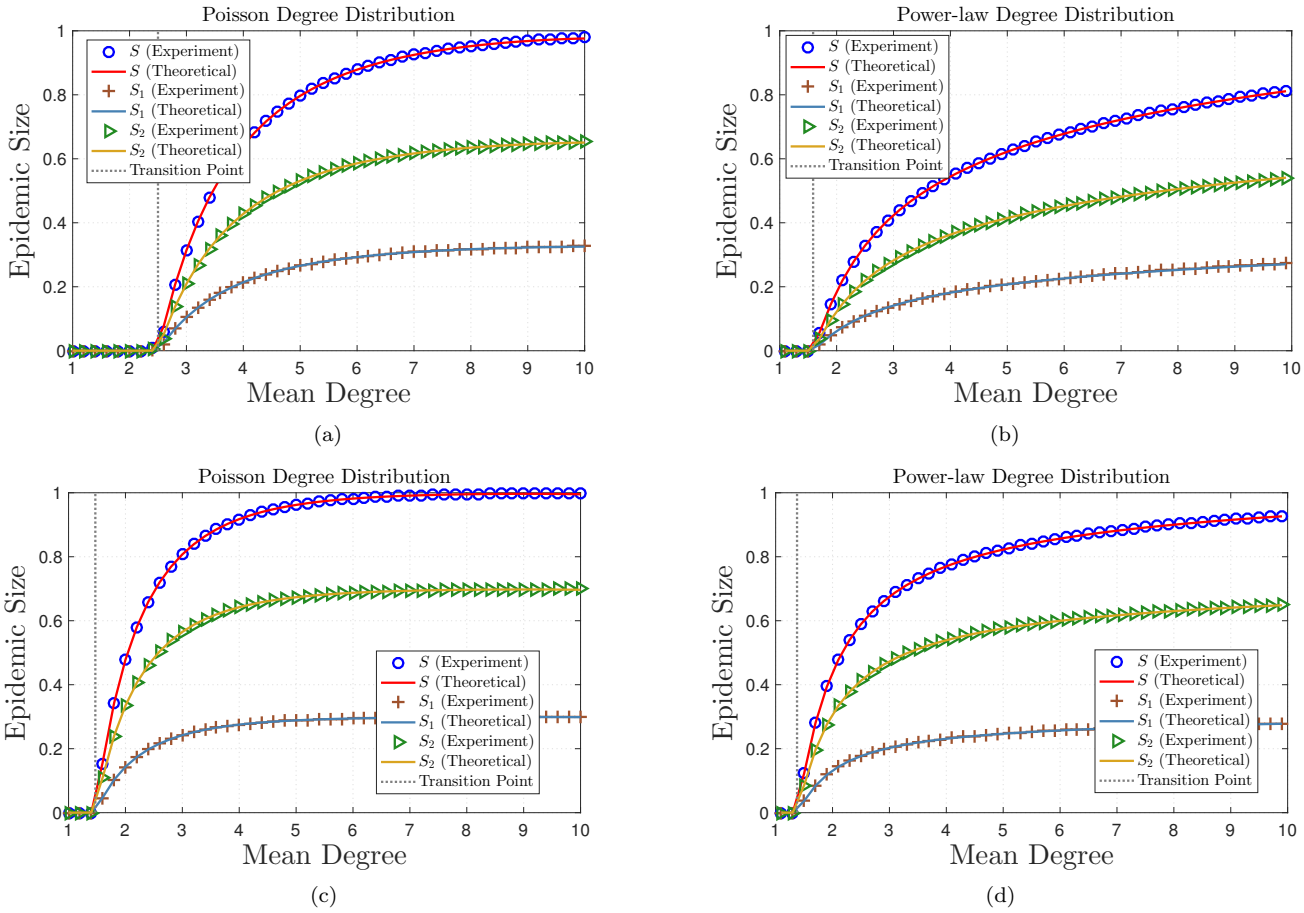

FIG. 2. **Evolution on Poisson and Power-law contact networks**. *The network size n is 2×10^5 and the number of independent experiments for each data point is* 500. *Blue circles, brown plus signs, and green triangles denote the empirical average epidemic size, average fraction of nodes infected with strain-*1*, and average fraction of nodes infected with strain-*2*, respectively. The red, blue, and yellow lines denote the theoretical average total epidemic size, average fraction of nodes infected with strain-*1*, and average fraction of nodes infected with strain-*2*, respectively. Theoretical results are obtained by solving the system of equations (2) with the corresponding parameter set. (a)-(b) We set $T_1 = 0.2$, $T_2 = 0.5$, $\mu_{11} = \mu_{22} = 0.75$. (c)-(d) We set $T_1 = 0.4$, $T_2 = 0.8$, and $\mu_{11} = 0.3$, and $\mu_{22} = 0.7$ implying that an infected node, regardless of what type of infection it has, mutates to strain-*1 *(respectively, strain-*2*) with probability* 0.3 *(respectively,* 0.7*), independently. In all cases, we observe good agreement with our theoretical results.*

.

independently with probability T_1. Each of the infected neighbors mutate independently to strain-1 with probability μ_{11}, or to strain-2 with probability μ_{12}. As the process continues to grow, both strains might exist in the population. An intermediate node that becomes infected with strain-i would mutate to strain-1 with probability μ_{i1}, or strain-2 with probability μ_{i2}, for $i = 1, 2$. When cycles start to appear, a susceptible node could be exposed to multiple infections at once. If a node is exposed to x infections of strain-1 and y infections of strain-2 simultaneously, the node becomes infected with strain-1 (respectively, strain-2) with probability $x/(x+y)$ (respectively, $y/(x+y)$) for any non-negative constants x and y. A node that receives infection at round i mutate first (by the end of round i) before it attempts to infect her neighbors at round $i+1$. The node is considered *recovered* at round $i+2$, i.e., a node is infective for only one round.

C. Epidemic Size

We start by focusing on the total epidemic size and the expected fraction of nodes that were infected with strain-1 and strain-2. The network size n is set to 2×10^5. We consider two parameter sets that emphasize the correlations between a node's eventual type (after mutation) and the type of infection it has originally received. In particular, we have

- **Parameter set 1:** $T_1 = 0.2$, $T_2 = 0.5$, $\mu_{11} = 0.75$, and $\mu_{22} = 0.75$.

- **Parameter set 2:** $T_1 = 0.4$, $T_2 = 0.8$, $\mu_{11} = 0.3$, and $\mu_{22} = 0.7$.

Observe that we have $\mu_{11} = \mu_{21}$ and $\mu_{22} = \mu_{12}$ for the second parameter set. Hence, an infected node, regardless of what type of infection it has, mutates to strain-1 (respectively, strain-2) with probability 0.3 (respectively, 0.7), independently. This is a special case that can easily be treated by our formalism given in Section III.

In Figure 2a and Figure 2b, we use the first parameter set and run 500 independent experiments for each data point. We demonstrate our results on contact networks with Poisson degree distribution (Figure 2a) and Power-law degree distribution with exponential cutoff (Figure 2b). For Figure 2b, we set $\Gamma = 15$, and vary γ with the mean degree. In particular, the mean degree λ is given by

$$\lambda = \frac{\mathrm{Li}_{\gamma-1}\left(e^{-1/\Gamma}\right)}{\mathrm{Li}_{\gamma}\left(e^{-1/\Gamma}\right)}. \tag{6}$$

Hence, we can numerically solve (6) to obtain the particular value of γ corresponding to a given value of λ.

In order to establish the validity of our analytic results given in Section III, we plot the theoretical values of S, S_1, and S_2 obtained by solving the system of equations (2) with the corresponding parameter set. We also plot a vertical line at the critical mean degree that corresponds to a phase transition (see (4) and (5)). Clearly, our experimental results are in perfect agreement with our theoretical results on both contact networks. In Figure 2c and Figure 2d, we repeat the same procedure, but with the second parameter set. Similarly, we observe perfect agreement with our theoretical results on both contact networks.

D. Probability of Emergence

In [33], Alexander and Day investigated the probability of emergence for the multiple strain model presented in Section II. However, authors did not provide a comprehensive simulation study to validate their formalism on random or real-world networks. Instead, in [33, Section 3], authors only evaluated their equations *numerically*. In this subsection, we aim to establish the validity of the results presented in [33] on random networks generated by the configuration model. For brevity, we limit our scope to contact networks with Poisson degree distribution. However, similar patterns are observed for contact networks with Power-law degree distribution.

In Figure 3, we set the network size $n = 5\times10^5$ and run a computer simulation with 10^4 independent experiment for each data point. We use the two parameter sets given in Section III.C. Namely, we set

- $T_1 = 0.2$, $T_2 = 0.5$, and $\mu_{11} = \mu_{22} = 0.75$ for Figure 3.a, and

- $T_1 = 0.4$, $T_2 = 0.8$, $\mu_{11} = 0.3$ and $\mu_{22} = 0.7$ for Figure 3.b.

Note that in Figure 3, we plot the probability of emergence conditioned on the initial node receiving infection with strain-1 [67]. We observe an agreement between our experimental results and the theoretical results given in [33]. The reasoning behind this is intuitive; the multi-type branching framework assumes that the underlying graph is tree-like, an assumption that works best for networks with vanishingly small clustering coefficient, e.g., networks which are generated by the configuration model.

E. Reduction to Single-Type Bond-Percolation

An important question to ask is whether the classical single-type bond percolation models could predict the threshold, probability, and final size of epidemics that entail *evolution*, i.e., information or diseases that propagate according to the multiple-strain model given in Section II. In pursing an answer to this question, we start by establishing a *matching condition* between single-strain models and multiple-strain models for epidemics.

In [3], Newman proposed a stochastic SIR model for the propagation of a single-strain pathogen on a contact network. Newman showed that, under some conditions, the SIR model is isomorphic to a bond-percolation model on the underlying contact network. Specifically, with the *average transmissibility* of the pathogen (denoted T_{BP}) as the bond-percolation parameter, if we are to occupy each edge of the network with probability T_{BP}, then the probability of emergence as well as the final size of the epidemic are precisely given by the fraction of nodes in the giant component of the *percolated* graph. Finally, it was shown that a phase transition occurs when

$$\left(\frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}\right) T_{\mathrm{BP}} = 1 \tag{7}$$

In other words, if the left hand side of (7) is strictly larger than 1, a giant component emerges indicating an epidemic. Otherwise, we have self-limited outbreaks.

Comparing (3) to (7) suggests the proposal of a matching that results in the same condition for phase transition. More precisely, if we are to set

$$T_{\mathrm{BP}} = \rho\left(\boldsymbol{T}\boldsymbol{\mu}\right) \tag{8}$$

then, both (3) and (7) collapse to the same condition for a given contact network. In what follows, we explore the extent to which classical, single-type bond-percolation models (under the matching condition (8)) may predict the threshold, probability, and final size of epidemics that

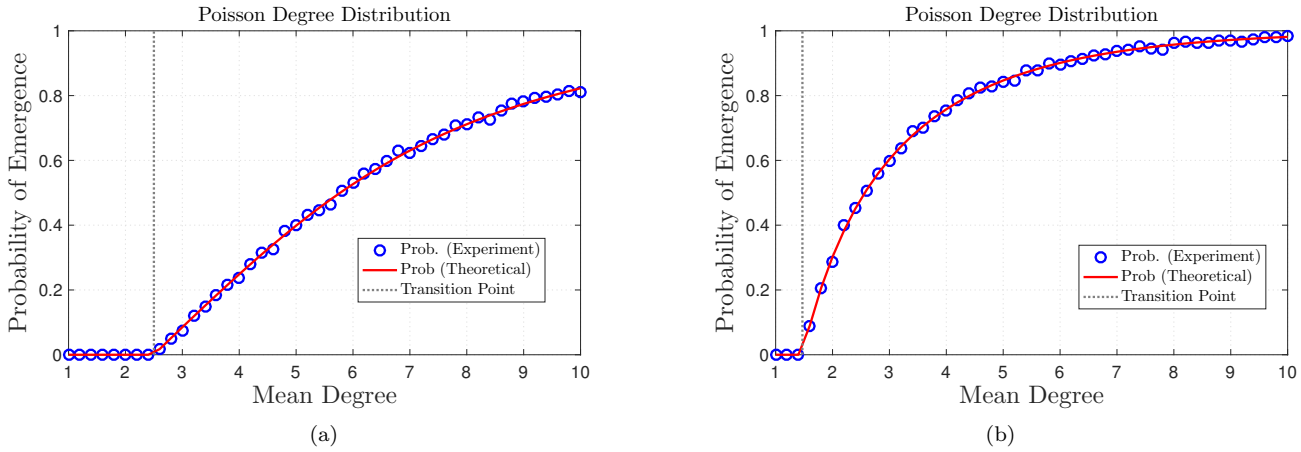

FIG. 3. **The probability of emergence on contact networks with Poisson degree distribution**. *The network size n is 5×10^5 and the number of independent experiments for data point is 10^4. Blue circles denote the empirical probability of emergence while the red line denotes the theoretical probability of emergence according to [33]. (a) We set $T_1 = 0.2$, $T_2 = 0.5$, $\mu_{11} = \mu_{22} = 0.75$. (b) We set $T_1 = 0.4$, $T_2 = 0.8$, and $\mu_{11} = 0.3$, and $\mu_{22} = 0.7$. Our experimental results prove the validity of the formalism presented by Alexander and Day in [33]*

.

entail evolution, i.e., information or diseases that propagate according to the multiple-strain model given in Section II. We focus on contact networks with Poisson degree distribution, generated by the configuration model, while we devote Section V for real-world networks.

In Figure 4, we extend Figure 3 by further adding the experimental results for the final epidemic size as well as the corresponding theoretical values for the probability of emergence on a bond-percolated network under the matching condition (8). Note that the probability of emergence is equivalent to the final epidemic size for single-type, bond-percolated networks [3]. Observe that the classical single-type bond-percolation model accurately captures the threshold and final size of epidemic but provides significantly inaccurate predictions when it comes to the probability of emergence. Similar pattern will be observed in Section V for real-world networks. This inaccuracy sheds the light on a fundamental disconnect between the classical, single-type bond-percolation models and real-life spreading processes that entail evolution. We explain the intuition behind our findings in Section IV in the Supplementary Material.

F. Effect of Mutation

When only a single evolutionary pathway is available, mutations have to occur in a particular order [68]. In [34], Antia et al. considered the case where the fitness landscape consists of m strains such that $R_{0,i} < 1$ for $i = 1, \ldots, m-1$, while $R_{0,m} > 1$. Hence, an introduced pathogen (with $R_{0,1} < 1$) must acquire $m-1$ successive mutations in order for the disease to emerge. Antia et al. derived a set of recursive equations whose solution characterizes the probability of emergence under some conditions; see [34] for more details. To gain further insights on the effect of mutation, Antia et al. proposed a theoretical approximation of the probability of emergence as a product of the probability of mutation, i.e., the probability that the introduced pathogen would eventually mutate to strain-m, and the probability of emergence of strain-m. Indeed, the probability of mutation plays a key role in the overall extinction probability. After all, if the introduced pathogen does not gain $m-1$ successive mutations, the disease would eventually die out.

Recall that the mathematical theory developed by Alexander and Day [33] defines the probability of emergence as a function of the evolutionary dynamics of the pathogen (i.e., the mutation matrix $\boldsymbol{\mu}$), the characteristics of the spreading process (i.e., the transmissibility matrix $\boldsymbol{T}$), and the structure of the underlying contact network (i.e., the degree distribution $\{p_k, \quad k = 0, 1, \ldots\}$). All of these factors are intertwined together in a way that makes it difficult to predict how the probability of mutation influences the probability of emergence. In what follows, we provide a theoretical approximation to the probability of emergence in a way that clearly distinguishes the role of mutation and shows how it strongly influences the probability of emergence.

Consider the case when the fitness landscape consists of two strains with transmissibility matrix $\boldsymbol{T}$ and mutation matrix $\boldsymbol{\mu}$ given by

$$\boldsymbol{T} = \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\mu} = \begin{bmatrix} 1-\mu & \mu \\ 0 & 1 \end{bmatrix}.$$

Assume also that $T_1 < T_2$. Note that the process starts by picking a random individual uniformly at random and infecting her with strain-1.

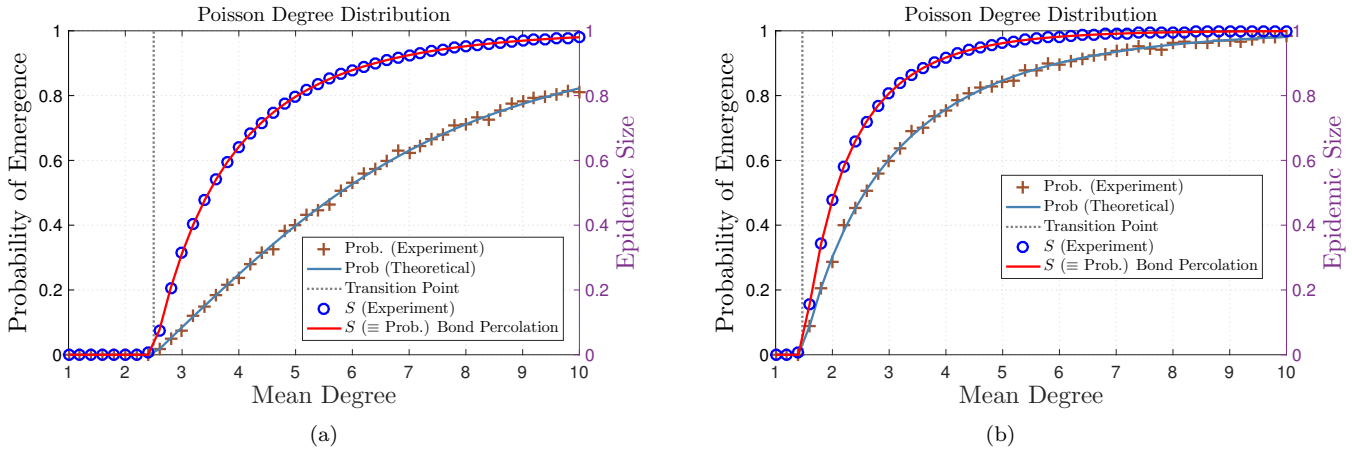

FIG. 4. **Reduction to single-type bond-percolation**. *The network size n is 5×10^5 and the number of independent experiments for each data point is 10^4. Blue circles and brown plus signs denote the empirical average epidemic size and the probability of emergence, respectively. The navy blue line denotes the theoretical probability of emergence according to [33] while the red line denotes the theoretical average epidemic size (as well as the probability of emergence) predicted by the single-type bond-percolation framework under the matching condition (8). (a) We set $T_1 = 0.2$, $T_2 = 0.5$, $\mu_{11} = \mu_{22} = 0.75$. (b) We set $T_1 = 0.4$, $T_2 = 0.8$, and $\mu_{11} = 0.3$, and $\mu_{22} = 0.7$. The classical, single-type bond percolation models may accurately predict the threshold and final size of epidemics, but their predictions on the probability of emergence are clearly inaccurate.*

Let P_μ denote the probability that at some point along the chain of infections (starting from the type-1 seed), a type-2 node would emerge. For a fixed mean degree λ of the underlying network, we may approximate the probability of emergence by (see Section III in the Supplementary Material)

$$\mathbb{P}\left[\text{emergence}\right] \geq P_\mu P_2^{\text{BP}} \tag{9}$$

where P_2^{BP} denotes the probability of emergence on a single-type bond percolated network with edge occupation probability T_2 and mean degree λ.

To confirm the validity of (9), we run a computer simulation on random networks generated by the configuration model with Poisson degree distribution. In Figure 5, we set the network size $n = 2\times10^5$ and perform 10^4 independent experiments for each data point. In Figure 5a, we set $T_1 = 0.1$, $T_2 = 1$, and $\mu = 0.01$. Observe that the bound given by (9) is *tight*, as T_2 is significantly larger than T_1. In general, we would expect a tight bound whenever $\lambda_2 \leq \lambda < \lambda_1$, where λ_1 and λ_2 denote the phase transition points (i.e., critical mean degrees) for a single-strain, bond-percolated network with T_1 and T_2, respectively, i.e., $1 \leq \lambda < 10$ for the given parameter set. As λ increases beyond λ_1, the tightness of the bound depends on the ratio between T_2 to T_1. This is illustrated in Figure 5b for the case when $T_1 = 0.2$ and $T_2 = 0.3$.

The availability of an explicit expression for the probability of mutation (see Section III in the Supplementary Material) allows for exploring the effects of mutation on the overall probability of emergence. Indeed, the way the probability of emergence behaves with respect to changes in the mean degree resembles, to a great extent, the way P_μ behaves, as illustrated in Figure 5. Hence, in what follows, we focus on the behavior of P_μ with respect to changes in the mean degree. In Figure 6, we set $T_1 = 0.1$ and plot P_μ against the mean degree for a network with Poisson degree distribution. We observe that different values for μ impacts the shape of P_μ (hence, the probability of emergence) in a remarkable way. Firstly, for all values of $\mu \in (0, 1)$, the behavior of P_μ appears to be strikingly different than the universality class of percolation models, e.g., see the shape of the probability of emergence (respectively, P_2^{B}) in Figure 3 (respectively, Figure 5). Secondly, the effect of mutation probabilities on P_μ appears to be significant as the mean degree increases from small values, reaches its peak right before the critical mean degree corresponding to P_1^{BP}, then decays as the mean degree increases further.

The reasoning behind the aforementioned observation is intuitive. Recall that the process starts with a single infection with strain-1 and note that P_μ is influenced by the structure of the underlying contact network, the transmissibility of strain-1, and the particular value of μ. As the mean degree λ increases towards λ_1, the length of the tree of infections starting from the seed [69] also increases, however, no cycles appear and the epidemic propagates on a finite, tree-like percolated network (since $\lambda < \lambda_1$). Increasing the length of the tree increases the probability that at least one intermediate node would mutate to strain-2, but the fact that the tree is finite makes the particular value of μ very crucial to P_μ. Namely, a small value of μ makes it less likely that a mutant emerges before the chain of infections is terminated, while a relatively larger value could drive the emergence of strain-2 and lead the epidemic to escape extinction. Put differently, the finiteness of the chain of infections

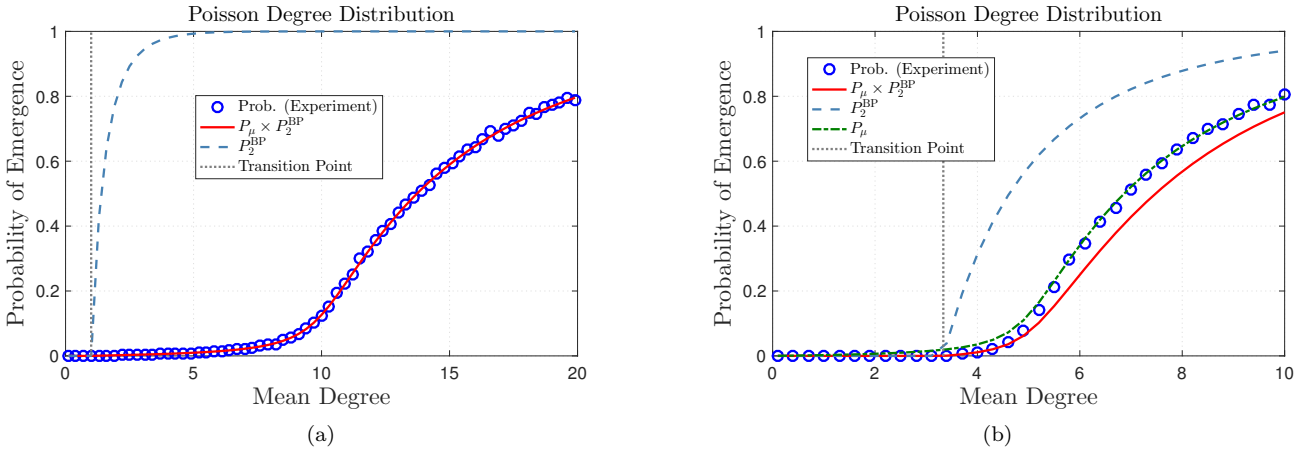

FIG. 5. **Approximating the probability of emergence:** *The network size n is 2×10^5 and the number of independent experiments for each data point is 10^4. Blue circles denote the empirical probability of emergence while the red line denotes the theoretical approximation of the probability of emergence according to (9). The light blue dashed line denotes the probability of emergence for a single-strain, bond-percolated network with T_2. (a) We set $T_1 = 0.1$, $T_2 = 1$, and $\mu = 0.01$. (b) We set $T_1 = 0.2$, $T_2 = 0.3$, and $\mu = 0.01$. We observe good agreement between the experimental results and the theoretical approximation given by (9) whenever $\lambda_2 \leq \lambda < \lambda_1$ or whenever T_2 is significantly larger than T_1.*

when $\lambda < \lambda_1$ creates a limited number of opportuniti for mutation, causing the particular value of μ to *bear t burden* of generating a mutant and driving the whole pr cess to emergence. However, as λ increases beyond λ cycles start to appear and a giant component of nod infected with strain-1 emerges. In this case, the chain infections is no longer finite, and any positive value of results in a mutation almost surely in the limit of lar network size. Put differently, when $\lambda \geq \lambda_1$, the structu of the underlying network starts to facilitate the eme gence of strain-2, hence reducing the dependence on μ

V. EVOLUTION IN REAL-WORLD NETWORKS

In Section IV.F, we explored the validity of analyzing the multiple-strain model for evolution with the available tools from the classical, single-type bond-percolation framework. We focused on *random* networks generated by the configuration model and demonstrated that a reduction to the classical, single-type bond percolation framework leads to accurate results with respect to the threshold and final size of epidemics, but significantly inaccurate results with respect to the probability of emergence. In this section, we aim to examine the universality of our findings by analyzing the probability of emergence on *real-world* contact networks obtained from SNAP data sets [54]. Our objective is twofold. Firstly, we would like to validate the multi-type branching formalism of Alexander and Day (see Section III.A) on real-world networks. Secondly, we seek to highlight and confirm the limitations of the single-type bond-percolation framework in predicting the probability of emergence on

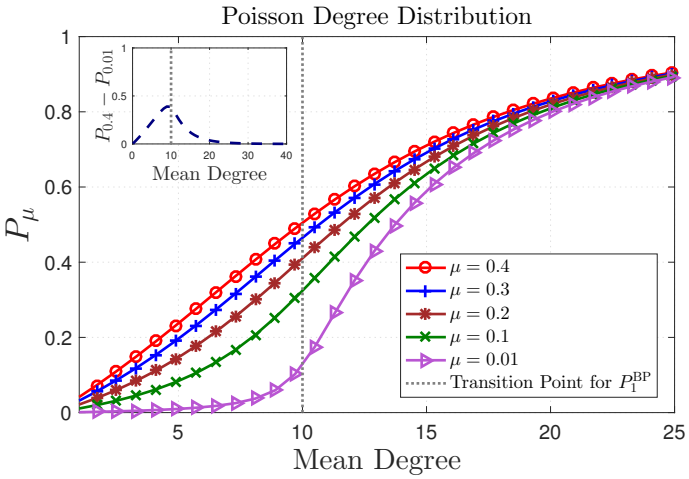

FIG. 6. **Effect of Mutation:** *We set $T_1 = 0.1$ and plot the behavior of P_μ against the mean degree for a network with Poisson degree distribution. Intuitively, different values of μ have different impact on P_μ. The impact is pronounced before the critical mean degree corresponding to a single-strain, bond-percolated network with T_1. Inset: The difference between the value of P_μ when $\mu = 0.4$ and the value of P_μ when $\mu = 0.01$ as a function of the mean degree of the underlying contact network.*

real-world networks.

Dataset: In the context of information propagation, we consider four different contact networks obtained from SNAP [54]. In particular, we consider the following contact networks:

- FACEBOOK [54, 72]: The contact network among the friends of 10 users (including those 10 users).

Network	$\lvert\mathcal{N}\rvert$	$\lvert\mathcal{E}\rvert$	$\lambda_{\text{original}}$	Φ_{original}	$\Phi_{\{\lambda=1\}}$	$\Phi_{\{\lambda=10\}}$	Φ_{random}
Facebook	4,039	88,234	43.7	0.519	0.011	0.117	0.0107
Twitter	81,306	1,342,296	33	0.170	0.005	0.051	0.0004
Slashdot	82,168	504,230	12.3	0.024	0.001	0.019	0.0001
Higgs	456,626	12,508,413	54.8	0.008	0.0001	0.001	0.0001
High school	773	6342	16.4	0.094	0.019	0.059	0.020
Hospital	73	543	14.87	0.446	0.090	0.296	0.183

FIG. 7. **Real-world contact networks**. *We consider four real-world contact networks in the context of information propagation, namely, Facebook, Twitter, Slashdot, and Higgs networks from SNAP [54] dataset. We also consider two real-world contact networks in the context of infectious disease propagation, namely, a contact network among students, teachers, and staff at a US high school [70] and a contact network among professional staff and patients in a hospital in Lyon, France [71]. For each network, we indicate the number of nodes* $|\mathcal{N}|$*, the number of edges* $|\mathcal{E}|$*, the mean degree of the original network* $\lambda_{\text{original}}$*, and the clustering coefficient of the original network* Φ_{original}*.* $\Phi_{\{\lambda=1\}}$ *(respectively,* $\Phi_{\{\lambda=10\}}$*) denotes the clustering coefficient of the original network after removing a random subset of edges such that the resulting mean degree is* 1 *(respectively,* 10*).* Φ_{random} *denotes the average clustering coefficient (over* 200 *independent realizations) of a random network generated by the configuration model with Poisson degree distribution. The random network has the same number of nodes and the same (original) mean degree of the corresponding real-world network.*

- Twitter [54, 72]: The contact network among the friends of 1000 users (including those 1000 users).

- Slashdot [54, 73]: The network contains friend/foe links between the users of Slashdot.

- Higgs [54, 74]: The Higgs data set has been collected upon monitoring the spreading processes on Twitter before, during and after the announcement of the discovery of a new particle with the features of the elusive Higgs boson on July 4, 2012. Nodes correspond to the authors of the collected tweets and edges represent the followee/follower relationships between them.

In the context of infectious disease propagation, we consider the following two contact networks:

- High school network [70]: The contact network observed at a US high school during a typical school day. The dataset covers 762,868 interactions between students, teachers, and staff. Each interaction between two individuals is characterized by their identification numbers as well as the duration of the interaction. Two individuals could have multiple interactions throughout the day, and we sum the durations of these interactions to calculate the total contact time between these two individuals over the whole day. We proceed by sampling a *static* graph out of this dataset, by assigning an edge between nodes u and v with probability $t_{uv}/t_{\max}$ where t_{uv} denotes the total contact time between nodes u and v throughout the day and $t_{\max}$ denotes the maximum total contact time observed in the dataset.

- Hospital network [71]: The contact network observed in a short stay geriatric unit of a university hospital in Lyon, France. The dataset covers five days of interactions between professional staff members and patients. Similar to the high school network, we compute the total contact time between two individuals (over the span of five days), then we sample a static graph out of the dataset, by assigning an edge between nodes u and v with probability $t_{uv}/t_{\max}$.

More details on the networks, including their clustering coefficients are given in Figure 7. We assume that all edges are unidirectional.

A. Methods

To conduct a fair comparison between the formalism given in Section III.A and the single-type bond percolation framework, we fix the parameters of the transmissibility matrix $\boldsymbol{T}$ and the mutation matrix $\boldsymbol{\mu}$, hence fixing $\rho(\boldsymbol{T}\boldsymbol{\mu})$ and T_{BP} (according to (8)). We vary the mean degree, denoted λ, for each of the contact networks between 1 and 10. For each value of λ, we remove a random subset of edges such that the resulting network is of mean degree λ (approximately). Note that the random removal of edges would indeed lower the clustering coefficient of the network, however, the resulting subgraph would remain highly clustered compared to random networks with the same mean degree (see Figure 7). In other words, the sampled networks still exhibit specific structural properties that distinguish them from synthetic contact networks generated randomly by the configuration model (with Poisson degree distribution of the same mean degree). After the mean degree is adjusted, the process proceeds similar to Section IV.B.

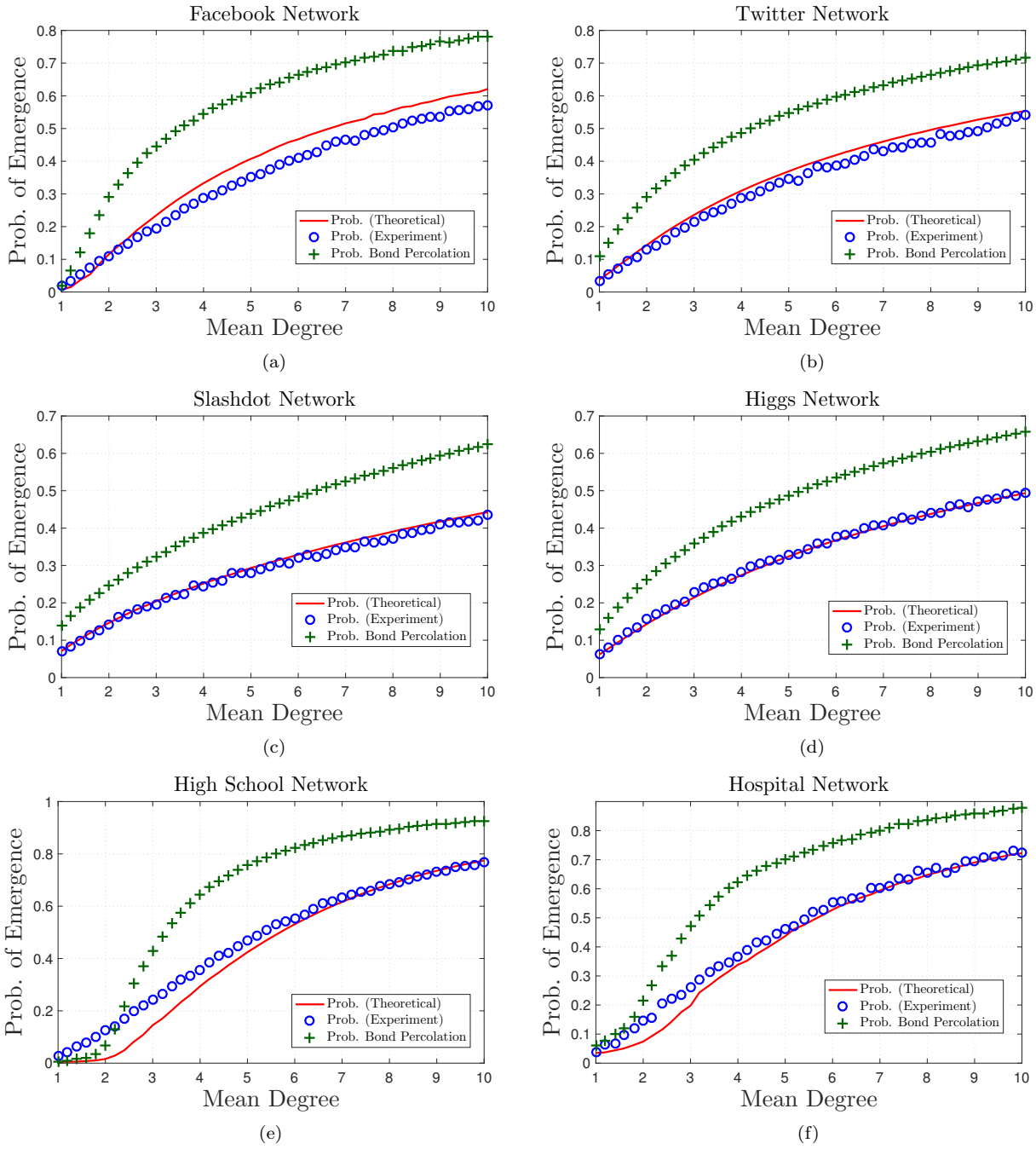

FIG. 8. **The probability of emergence on real-world contact networks.** *In the context of information propagation, we consider four contact networks sampled from SNAP data sets [54]: (a) Facebook network, (b) Twitter network, (c) Slashdot network, and (d) Higgs network. In the context of infectious disease propagation, we consider two contact networks: (e) High school contact network and (f) Hospital contact network. We set* $T_1 = 0.2$, $T_2 = 0.5$, $\mu_{11} = \mu_{22} = 0.75$ *(hence* $T_{\rm BP} = 0.4$*) and vary the mean degree, denoted* λ*, from* 1 *to* 10*. For each value of* λ*, we remove a random subset of edges such that the resulting graph is of mean degree* λ *(approximately). The sampled networks still exhibit higher clustering coefficient as compared to random networks with the same mean degree. The single-type bond-percolation framework provides inaccurate predictions on the probability of emergence, in contrast to the multiple-strain formalism given by Alexander and Day [33]. The multiple-strain formalism offers remarkably accurate predictions on a class of real-world networks with low clustering coefficient.*

B. Results

In Figure 8, we plot the probability of emergence for the four contact networks shown in Figure 7. We compare the results obtained by computer simulations with those obtained by the multiple-strain formalism (Section III.A)

and the single-type bond-percolation framework. We set $T_1 = 0.2$, $T_2 = 0.5$, and $\mu_{11} = \mu_{22} = 0.75$. It follows that $T_{\text{BP}} = 0.4$ according to (8).

Similar to our observations on random networks (Section IV.E), the single-type, bond-percolation framework provides significantly inaccurate predictions on the probability of emergence, should the underlying process entail evolution. The limitation is universal as it applies to both random and real-world networks. Section IV in the Supplementary Material explains the intuition behind our observations. In contrast, the multiple-strain formalism provides remarkably accurate predictions, especially on contact networks with low clustering coefficient. Note that the multi-type branching framework assumes that the underlying graph is tree-like; an assumption that holds for networks with small clustering coefficient. Hence, one could reasonably argue that the multiple-strain formalism would provide high prediction accuracy on such networks.

VI. CO-INFECTION CONTROLS THE ORDER OF PHASE TRANSITION

The preceding discussion considers the case when *co-infection* is not possible, hence each infected host either carries strain-1 or strain-2, but not both. However, humans, animals, plants, and other organisms may become *co-infected* with multiple pathogen strains, causing major consequences for both within- and between-host disease dynamics [43–47, 75]. For instance, in the case of human malaria, the majority of infected adults are *simultaneously* infected by more than five strains of *Plasmodium falciparum* [47, 76]. The competition and interaction patterns between the resident strains trigger significant ramifications of the disease dynamics. Also, the aggregate virulence experienced by the co-infected host could be higher than the most virulent strain, or lower than the least virulent strain, or anywhere in between [47, 77–79]. Co-infection also applies in the context of information propagation. Observe that with the growing number of news outlets, we may come across various variants of information on social media platforms. Similar to the case of infectious diseases, these variants may reinforce or weaken each other based on whether they share the same bias or not.

In this section, we seek to shed the light on the effects of co-infection on information/disease propagation. In particular, we investigate the extent to which co-infection dynamics could enhance or suppress the scale of epidemics. Of particular interest is whether co-infection could change the order of phase transition from second-order (as it is the case with most epidemic models) to first-order, leading to a phenomenon that is commonly described as *avalanche outbreaks* [58]. To that end, we extend the multiple-strain model given in Section II to account for co-infection. In particular, a susceptible individual who comes into infectious contacts with type-1 and type-2 hosts *simultaneously* becomes *co-infected* and starts to spread the *co-infection*. Henceforth, we consider the case when the co-infection has its own transmissibility T_{co} and does not mutate back to either strain-1 or strain-2. In other words, a co-infected host infects each of her neighbors independently with probability T_{co}, and infected neighbors are deemed *co-infected* with probability 1.

As with Section IV, we consider contact networks with Poisson degree distribution and Power-law degree distribution with exponential cutoff, respectively. For both cases, we set $T_1 = 0.2$, $T_2 = 0.5$, and $\mu_{11} = \mu_{22} = 0.75$. Moreover, we set the network size to 2×10^6 and the number of independent experiments for each data point to 5×10^3. To illustrate how co-infection dynamics control the order of phase transition, we simulate and compare the process for two values of T_{co}, namely $T_{co} = 0.1$ and $T_{co} = 0.8$. Finally, we plot the epidemic size, denoted by s_{co}^{BP}, for a single-strain, bond-percolated network [3].

In all cases, co-infection emerges at the phase transition point that characterizes an epidemic of strain-1 and strain-2, i.e., the mean degree for which $\rho(\boldsymbol{M}) = 1$, where $\boldsymbol{M}$ is given by

$$\boldsymbol{M} = \left(\frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right) \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix} \begin{bmatrix} \mu_{11} & \mu_{12} \\ \mu_{21} & \mu_{22} \end{bmatrix}$$

As seen in Figure 9, a *first-order* phase transition is observed on both contact networks when $T_{co} = 0.8$ due to the corresponding first order transition of S_{co}. In particular, the value of S_{co} jumps discontinuously from zero to (approximately) the corresponding value of S_{co}^{BP} for a single-strain, bond-percolated network with $T_{co} = 0.8$. Hence, a first-order phase transition is observed. In general, we conjecture that a first-order phase transition emerges whenever T_{co} is large enough such that $S_{co}^{BP} > 0$ at the critical point $\rho(\boldsymbol{M}) = 1$. If, however, T_{co} is small such that $S_{co}^{BP} = 0$ when $\rho(\boldsymbol{M}) = 1$, then a second-order phase transition is observed. This is confirmed by our simulation results for the case when $T_{co} = 0.1$.

In order to validate the order of phase transition when $T_{co} = 0.8$, we conduct an extensive simulation study around the phase transition point on both contact networks. In Figure 10, we set the number of nodes n to 15×10^6 (to alleviate finite size effects) and the number of experiments to 10^4 for each data point. We use the same parameters that were used to generate Figure 9, i.e., $T_1 = 0.2$, $T_2 = 0.5$, and $\mu_{11} = \mu_{22} = 0.75$. Our results confirm that the phase-transition is indeed first order on both contact networks. In fact, the value of S_{co} jumps discontinuously to (approximately) the corresponding value of S_{co}^{BP} with $T_{co} = 0.8$.

VII. CONCLUSION

In this paper, we have investigated the *evolution* of spreading processes on complex networks and developed

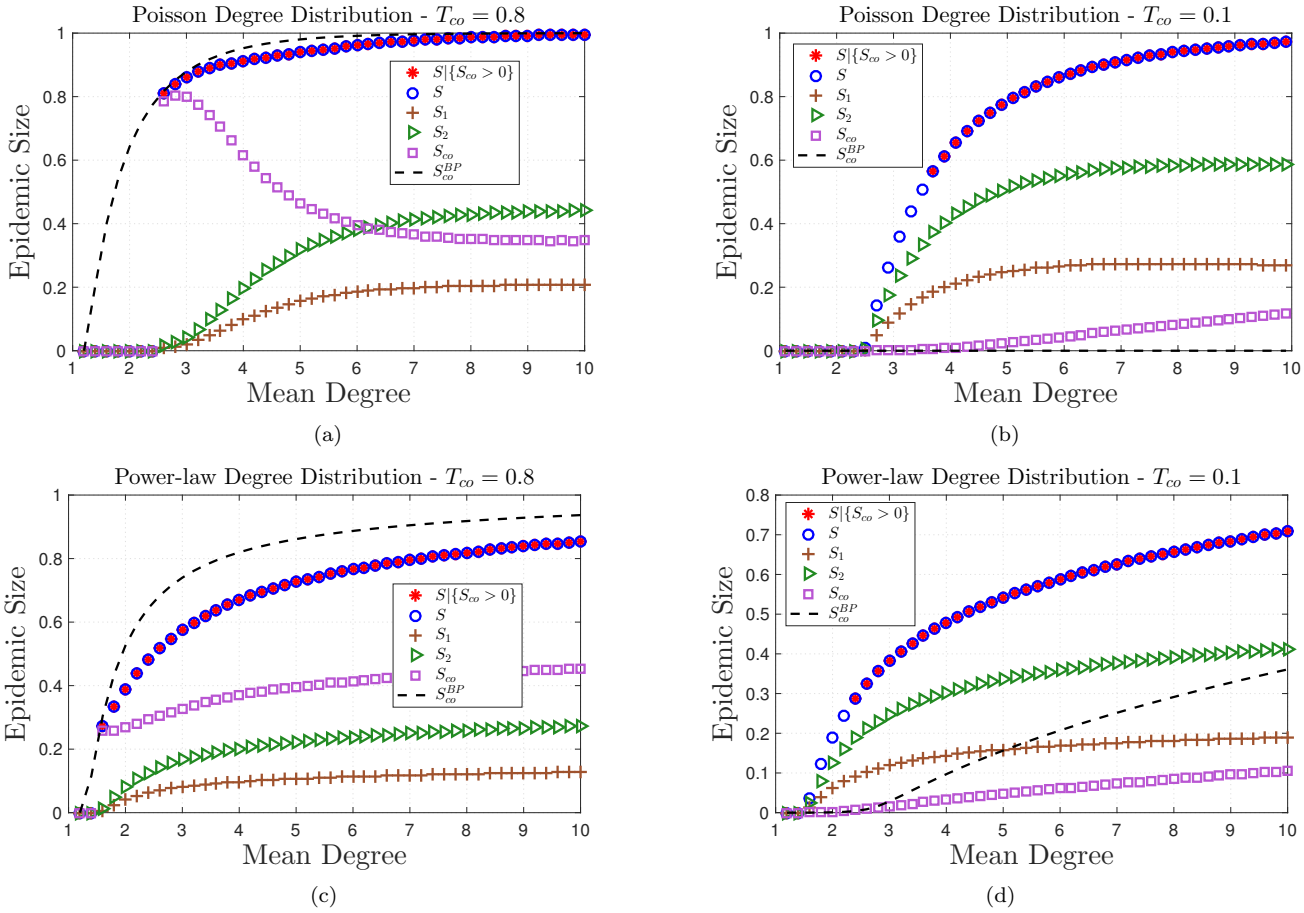

FIG. 9. **Co-infection dynamics determine the order of phase transition**. *We set* $T_1 = 0.2$, $T_2 = 0.5$, *and* $\mu_{11} = \mu_{22} = 0.75$ *for all subfigures. The network size* n *is* 2×10^6 *and the number of independent experiments for each data point is* 5×10^3. *Blue circles denote the average total epidemic size* S *and red stars denote the average total epidemic size* S *conditioned on* S_{co} *being greater than zero, i.e., conditioned on the existence of a positive fraction of co-infected nodes. Blue plus signs, orange triangles, and yellow squares denote the fraction of nodes infected with strain-1, strain-2, and co-infection, respectively. The black dashed-line denotes the epidemic size for a single-strain, bond-percolated network with* T_{co}, *i.e.,* S_{co}^{BP}. *(a) and (c): A first order phase transition is observed when* $T_{co} = 0.8$ *owing to the corresponding first order transition of* S_{co}. *Co-infection emerges at the phase transition point that characterizes an epidemic of strain-1 and strain-2. At this point, the value of* S_{co} *jumps discontinuously to (approximately) the corresponding value of* S_{co}^{BP} *with* $T_{co} = 0.8$. *Observe that* $S_{co}^{BP} > 0$ *at the transition point, hence, a first-order phase transition is observed. (b) and (d): Co-infection still emerges right at the phase transition point. However, since* T_{co} *is small,* $S_{co}^{BP} = 0$ *at the transition point. Hence, a second-order phase transition is observed.*

a mathematical theory that unravels the relationship between the characteristics of the spreading process, evolution, and the structure of the contact network on which the process spreads. Our mathematical theory was complemented by an extensive simulation study on both random and real-world contact networks. The simulation results proved the validity of our theory and revealed the significant shortcomings of the classical mathematical models that do not capture evolution. A matching condition between single- and multiple-strain models was proposed and evaluated in the context of probability of emergence, epidemic size, and epidemic threshold. Under the proposed matching condition, our results revealed that the classical bond-percolation models may accurately predict the threshold and final size of epidemics that entail evolution, but their predictions on the probability of emergence are *significantly inaccurate* on both random and real-world networks. Hence, our formalism is necessary to bridge the disconnect between how spreading processes propagate and evolve on complex networks, and the current mathematical models that do not capture evolution.

We proceeded by deriving a lower bound on the probability of emergence to gain further insights on the effects of mutation. The bound was derived for the special case of one-step irreversible mutation. Our results revealed that the probability of mutation plays a key role in determining the shape and behavior of the probability of

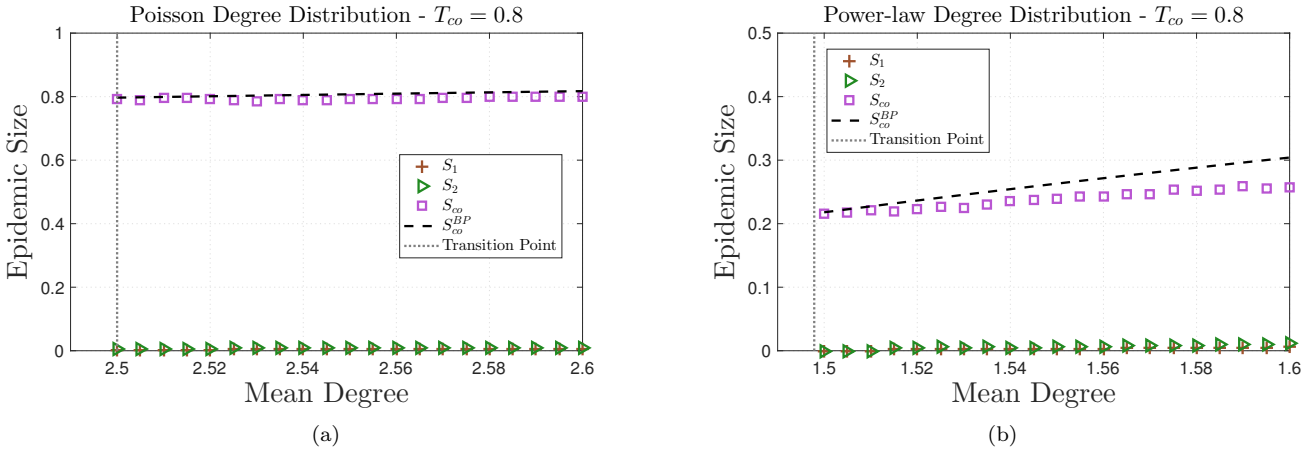

FIG. 10. **Validating the order of phase transition**. *We set the network size n to 15×10^6, the number of independent experiments for each data point to 10^4, $T_1 = 0.2$, $T_2 = 0.5$, and $\mu_{11} = \mu_{22} = 0.75$. Our results confirm that the phase-transition is indeed first order on both contact networks. The value of S_{co} jumps discontinuously to (approximately) the corresponding value of S_{co}^{BP} with $T_{co} = 0.8$.*

emergence. Moreover, the way the particular value of μ influences the probability of mutation varies according to the connectivity of the underlying contact network. Finally, we considered the case when *co-infection* is possible and showed that co-infection dynamics control the order of phase transition in an interesting way. In particular, depending on co-infection dynamics, the order of phase transition of the epidemic size could change from second-order to *first-order*, in contrast to the universality class of percolation models that are typically second-order.

ACKNOWLEDGEMENT

Research was sponsored by the Army Research Office and was accomplished under Grant Number W911NF-17-1-0587. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein. The first author was funded in part by the Dowd Fellowship from the College of Engineering at Carnegie Mellon University. The authors would like to thank Philip and Marsha Dowd for their financial support and encouragement. Research was supported in part by the Office of Naval Research (ONR) through grants N0001418SB001 and N000141512797. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research.

[1] Albert-László Barabási and Márton Pósfai, *Network science* (Cambridge university press, 2016).
[2] Christophe Fraser, Steven Riley, Roy M Anderson, and Neil M Ferguson, "Factors that make an infectious disease outbreak controllable," Proceedings of the National Academy of Sciences of the United States of America **101**, 6146–6151 (2004).
[3] Mark EJ Newman, "Spread of epidemic disease on networks," Phys. Rev. E **66**, 016128 (2002).
[4] James O Lloyd-Smith, Sebastian J Schreiber, P Ekkehard Kopp, and Wayne M Getz, "Superspreading and the effect of individual variation on disease emergence," Nature **438**, 355 (2005).
[5] Roy M Anderson, Robert M May, and B Anderson, *Infectious diseases of humans: dynamics and control*, Vol. 28 (Wiley Online Library, 1992).
[6] Romualdo Pastor-Satorras and Alessandro Vespignani, "Epidemic dynamics and endemic states in complex networks," Phys. Rev. E **63**, 066117 (2001).
[7] Yamir Moreno, Romualdo Pastor-Satorras, and Alessandro Vespignani, "Epidemic outbreaks in complex heterogeneous networks," The European Physical Journal B-Condensed Matter and Complex Systems **26**, 521–529 (2002).
[8] Clara Granell, Sergio Gómez, and Alex Arenas, "Competing spreading processes on multiplex networks: awareness and epidemics," Phys. Rev. E **90**, 012808 (2014).
[9] David M Morens, Gregory K Folkers, and Anthony S Fauci, "The challenge of emerging and re-emerging infectious diseases," Nature **430**, 242 (2004).
[10] Nathan D Wolfe, Claire Panosian Dunavan, and Jared Diamond, "Origins of major human infectious diseases,"

Nature **447**, 279 (2007).
[11] Peter Daszak, Lee Berger, Andrew A Cunningham, Alex D Hyatt, D Earl Green, and Rick Speare, "Emerging infectious diseases and amphibian population declines." Emerging infectious diseases **5**, 735 (1999).
[12] Xiping Wei, Sajal K Ghosh, Maria E Taylor, Victoria A Johnson, Emilio A Emini, Paul Deutsch, Jeffrey D Lifson, Sebastian Bonhoeffer, Martin A Nowak, Beatrice H Hahn, *et al.*, "Viral dynamics in human immunodeficiency virus type 1 infection," Nature **373**, 117 (1995).
[13] Fred Brauer, Carlos Castillo-Chavez, and Carlos Castillo-Chavez, *Mathematical models in population biology and epidemiology*, Vol. 1 (Springer, 2012).
[14] Constantinos I Siettos and Lucia Russo, "Mathematical modeling of infectious disease dynamics," Virulence **4**, 295–306 (2013).
[15] Odo Diekmann and Johan Andre Peter Heesterbeek, *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, Vol. 5 (John Wiley & Sons, 2000).
[16] Matt J Keeling and Pejman Rohani, *Modeling infectious diseases in humans and animals* (Princeton University Press, 2011).
[17] Shweta Bansal, Bryan T Grenfell, and Lauren Ancel Meyers, "When individual behaviour matters: homogeneous and network models in epidemiology," Journal of the Royal Society Interface **4**, 879–891 (2007).
[18] Matt J Keeling and Ken TD Eames, "Networks and epidemic models," Journal of the Royal Society Interface **2**, 295–307 (2005).
[19] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani, "Epidemic processes in complex networks," Reviews of modern physics **87**, 925 (2015).
[20] Joel C Miller and Istvan Z Kiss, "Epidemic spread in networks: Existing methods and current challenges," Mathematical modelling of natural phenomena **9**, 4–42 (2014).
[21] Rick Durrett, "Some features of the spread of epidemics and information on a random graph," Proceedings of the National Academy of Sciences **107**, 4491–4498 (2010).
[22] Yong Zhuang and Osman Yağan, "Information propagation in clustered multilayer networks," IEEE Transactions on Network Science and Engineering **3**, 211–224 (2016).
[23] Osman Yağan, Dajun Qian, Junshan Zhang, and Douglas Cochran, "Conjoining speeds up information diffusion in overlaying social-physical networks," IEEE Journal on Selected Areas in Communications **31**, 1038–1048 (2013).
[24] Liang Huang, Kwangho Park, and Ying-Cheng Lai, "Information propagation on modular networks," Phys. Rev. E **73**, 035103 (2006).
[25] Yamir Moreno, Maziar Nekovee, and Amalio F Pacheco, "Dynamics of rumor spreading in complex networks," Phys. Rev. E **69**, 066130 (2004).
[26] Peter Sheridan Dodds and Duncan J Watts, "Universal behavior in a generalized model of contagion," Phys. Rev. Letters **92**, 218701 (2004).
[27] Faryad Darabi Sahneh, Caterina Scoglio, and Piet Van Mieghem, "Generalized epidemic mean-field model for spreading processes over multilayer complex networks," IEEE/ACM Transactions on Networking **21**, 1609–1620 (2013).
[28] Osman Yağan and Virgil Gligor, "Analysis of complex contagions in random multiplex networks," Phys. Rev. E **86**, 036103 (2012).
[29] Mark EJ Newman, Stephanie Forrest, and Justin Balthrop, "Email networks and the spread of computer viruses," Phys. Rev. E **66**, 035101 (2002).
[30] Justin Balthrop, Stephanie Forrest, Mark EJ Newman, and Matthew M Williamson, "Technological networks and the spread of computer viruses," Science **304**, 527–529 (2004).
[31] Dajun Qian, Osman Yağan, Lei Yang, and Junshan Zhang, "Diffusion of real-time information in social-physical networks," in *IEEE GLOBECOM 2012*, pp. 2072–2077.
[32] Gabriel E Leventhal, Alison L Hill, Martin A Nowak, and Sebastian Bonhoeffer, "Evolution and emergence of infectious diseases in theoretical and real-world networks," Nature communications **6**, 6101 (2015).
[33] HK Alexander and T Day, "Risk factors for the evolutionary emergence of pathogens," Journal of The Royal Society Interface **7**, 1455–1474 (2010).
[34] Rustom Antia, Roland R Regoes, Jacob C Koella, and Carl T Bergstrom, "The role of evolution in the emergence of infectious diseases," Nature **426**, 658 (2003).
[35] Karin S Pfennig, "Evolution of pathogen virulence: the role of variation in host phenotype," Proceedings of the Royal Society of London B: Biological Sciences **268**, 755–760 (2001).
[36] Lada A. Adamic, Thomas M. Lento, Eytan Adar, and Pauline C. Ng, "Information evolution in social networks," in *ACM WSDM 2016*, pp. 473–482.
[37] Yichao Zhang, Shi Zhou, Zhongzhi Zhang, Jihong Guan, and Shuigeng Zhou, "Rumor evolution in social networks," Phys. Rev. E **87**, 032133 (2013).
[38] A zoonosis is any disease or infection that is naturally transmissible from vertebrate animals to humans [80].
[39] Stephen S Morse, Jonna AK Mazet, Mark Woolhouse, Colin R Parrish, Dennis Carroll, William B Karesh, Carlos Zambrana-Torrelio, W Ian Lipkin, and Peter Daszak, "Prediction and prevention of the next pandemic zoonosis," The Lancet **380**, 1956 – 1965 (2012).
[40] Kate E Jones, Nikkita G Patel, Marc A Levy, Adam Storeygard, Deborah Balk, John L Gittleman, and Peter Daszak, "Global trends in emerging infectious diseases," Nature **451**, 990 (2008).
[41] Colin R Parrish, Edward C Holmes, David M Morens, Eun-Chung Park, Donald S Burke, Charles H Calisher, Catherine A Laughlin, Linda J Saif, and Peter Daszak, "Cross-species virus transmission and the emergence of new epidemic diseases," Microbiology and Molecular Biology Reviews **72**, 457–470 (2008).
[42] Richard Dawkins, *The selfish gene* (Oxford university press, 2016).
[43] Oliver Balmer and Marcel Tanner, "Prevalence and implications of multiple-strain infections," The Lancet infectious diseases **11**, 868–878 (2011).
[44] Hanna Susi, Benoit Barrès, Pedro F Vale, and Anna-Liisa Laine, "Co-infection alters population dynamics of infectious disease," Nature communications **6**, 5975 (2015).
[45] Andrew F Read and Louise H Taylor, "The ecology of genetically diverse infections," Science **292**, 1099–1102 (2001).

[46] Ted Cohen, Paul D van Helden, Douglas Wilson, Caroline Colijn, Megan M McLaughlin, Ibrahim Abubakar, and Robin M Warren, "Mixed-strain mycobacterium tuberculosis infections and the implications for tuberculosis treatment and control," Clinical microbiology reviews **25**, 708–719 (2012).

[47] Samuel Alizon, Jacobus C de Roode, and Yannis Michalakis, "Multiple infections and the evolution of virulence," Ecology letters **16**, 556–567 (2013).

[48] James P. Sethna, Karin Dahmen, Sivan Kartha, James A. Krumhansl, Bruce W. Roberts, and Joel D. Shore, "Hysteresis and hierarchies: Dynamics of disorder-driven first-order phase transformations," Phys. Rev. Lett. **70**, 3347–3350 (1993).

[49] James P. Gleeson and Diarmuid J. Cahalane, "Seed size strongly affects cascades on random networks," Phys. Rev. E **75**, 056103 (2007).

[50] James P. Gleeson, "Cascades on correlated and modular random networks," Phys. Rev. E **77**, 046117 (2008).

[51] Michael Molloy and Bruce Reed, "A critical point for random graphs with a given degree sequence," Random structures & algorithms **6**, 161–180 (1995).

[52] Béla Bollobás, *Random graphs*, Vol. 73 (Cambridge university press, 2001).

[53] Mark EJ Newman, Steven H Strogatz, and Duncan J Watts, "Random graphs with arbitrary degree distributions and their applications," Phys. Rev. E **64**, 026118 (2001).

[54] Jure Leskovec and Andrej Krevl, "SNAP Datasets: Stanford large network dataset collection," `http://snap.stanford.edu/data` (2014).

[55] Linda JS Allen, Fred Brauer, Pauline Van den Driessche, and Jianhong Wu, *Mathematical epidemiology*, Vol. 1945 (Springer, 2008).

[56] Cristopher Moore and Mark EJ Newman, "Exact solution of site and bond percolation on small-world networks," Phys. Rev. E **62**, 7059 (2000).

[57] Lauren Meyers, "Contact network epidemiology: Bond percolation applied to infectious disease prediction and control," Bulletin of the American Mathematical Society **44**, 63–86 (2007).

[58] Weiran Cai, Li Chen, Fakhteh Ghanbarnejad, and Peter Grassberger, "Avalanche outbreaks emerging in cooperative contagions," Nature physics **11**, 936 (2015).

[59] N. Azimi-Tafreshi, "Cooperative epidemics on multiplex networks," Phys. Rev. E **93**, 042303 (2016).

[60] Peter Grassberger, Li Chen, Fakhteh Ghanbarnejad, and Weiran Cai, "Phase transitions in cooperative coinfections: Simulation results for networks and lattices," Phys. Rev. E **93**, 042316 (2016).

[61] Peng-Bi Cui, Francesca Colaiori, and Claudio Castellano, "Mutually cooperative epidemics on power-law networks," Phys. Rev. E **96**, 022301 (2017).

[62] Charles J Mode, *Multitype branching processes: theory and applications*, Vol. 34 (American Elsevier Pub. Co., 1971).

[63] Patsy Haccou, Patricia Haccou, Peter Jagers, Vladimir A Vatutin, and Vladimir A Vatutin, *Branching processes: variation, growth, and extinction of populations*, 5 (Cambridge university press, 2005).

[64] EA Leicht and Raissa M D'Souza, "Percolation on interacting networks," arXiv preprint arXiv:0907.0894 (2009).

[65] Gabor Csardi and Tamas Nepusz, "The igraph software package for complex network research," InterJournal **Complex Systems**, 1695 (2006).

[66] Https://github.com/reletreby/evolution.git.

[67] We remark that the formalism provided by Alexander and Day allows for computing the probability of emergence given any arbitrary initial type.

[68] Chaitanya S Gokhale, Yoh Iwasa, Martin A Nowak, and Arne Traulsen, "The pace of evolution across fitness valleys," Journal of Theoretical Biology **259**, 613–620 (2009).

[69] The length of the tree of infections can be interpreted as the size of the component (of a bond percolated network with T_1) that contains the seed.

[70] Marcel Salathé, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W. Feldman, and James H. Jones, "A high-resolution human contact network for infectious disease transmission," Proceedings of the National Academy of Sciences **107**, 22020–22025 (2010).

[71] Philippe Vanhems, Alain Barrat, Ciro Cattuto, Jean-Franois Pinton, Nagham Khanafer, Corinne Rgis, Byeula Kim, Brigitte Comte, and Nicolas Voirin, "Estimating potential infection transmission routes in hospital wards using wearable proximity sensors," PLOS ONE **8**, 1–9 (2013).

[72] Jure Leskovec and Julian J Mcauley, "Learning to discover social circles in ego networks," in *Advances in neural information processing systems* (2012) pp. 539–547.

[73] Jure Leskovec, Kevin J Lang, Anirban Dasgupta, and Michael W Mahoney, "Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters," Internet Mathematics **6**, 29–123 (2009).

[74] Manlio De Domenico, Antonio Lima, Paul Mougel, and Mirco Musolesi, "The anatomy of a scientific rumor," Scientific reports **3**, 2980 (2013).

[75] Jacobus C de Roode, Michelle EH Helinski, M Ali Anwar, and Andrew F Read, "Dynamics of multiple infection and within-host competition in genetically diverse malaria infections," The American Naturalist **166**, 531–542 (2005).

[76] CC Lord, B Barnard, K Day, JW Hargrove, JJ McNamara, REL Paul, K Trenholme, and MEJ Woolhouse, "Aggregation and distribution of strains in microparasites," Philosophical Transactions of the Royal Society B: Biological Sciences **354**, 799–807 (1999).

[77] Frank Cézilly, Marie-Jeanne Perrot-Minnot, and Thierry Rigaud, "Cooperation and conflict in host manipulation: interactions among macro-parasites and micro-organisms," Frontiers in microbiology **5**, 248 (2014).

[78] Charlotte Tollenaere, Hanna Susi, and Anna-Liisa Laine, "Evolutionary and epidemiological implications of multiple infection in plants," Trends in plant science **21**, 80–90 (2016).

[79] Sandra Lass, Peter J Hudson, Juilee Thakar, Jasmina Saric, Eric Harvill, Réka Albert, and Sarah E Perkins, "Generating super-shedders: co-infection increases bacterial load and egg production of a gastrointestinal helminth," Journal of the Royal Society Interface **10**, 20120588 (2013).

[80] World Health Organization: `http://www.who.int/topics/zoonoses/en/`.