Symbiotic and antagonistic disease dynamics on networks using bond percolation

Peter Mann,* V. Anne Smith, John B. O. Mitchell, and Simon Dobson
School of Computer Science, University of St Andrews, St Andrews, Fife KY16 9SX, United Kingdom;
School of Chemistry, University of St Andrews, St Andrews, Fife KY16 9ST, United Kingdom;
and School of Biology, University of St Andrews, St Andrews, Fife KY16 9TH, United Kingdom

(Received 7 May 2021; accepted 13 July 2021; published 3 August 2021)

In this paper we introduce a description of the equilibrium state of a bond percolation process on random graphs using the exact method of generating functions. This allows us to find the expected size of the giant connected component (GCC) of two sequential bond percolation processes in which the bond occupancy probability of the second process is modulated (increased or decreased) by a node being inside or outside of the GCC created by the first process. In the context of epidemic spreading this amounts to both an antagonistic partial immunity and a synergistic partial coinfection interaction between the two sequential diseases. We examine configuration model networks with tunable clustering. We find that the emergent evolutionary behavior of the second strain is highly dependent on the details of the coupling between the strains. Contact clustering generally reduces the outbreak size of the second strain relative to unclustered topologies; however, positive assortativity induced by clustered contacts inverts this conclusion for highly transmissible disease dynamics.

DOI: 10.1103/PhysRevE.104.024303

I. INTRODUCTION

A network is a collection of nodes that are connected by edges. Bond percolation on complex networks is a widely studied binary-state stochastic process where the edges of a network are said to be *occupied* with probability T_1 or *unoccupied* with probability $1 - T_1$. When T_1 is small, many edges are unoccupied, and the network is fractured into small components. At some critical value $T_{1,c}$, the small components connect together to form a macroscopic giant connected component (GCC); the expected size of the GCC exhibits a second-order phase transition. As $T_1 \to 1$, the GCC occupies an increasing fraction of the network. Nodes not contained within the GCC are said to be in the residual graph (RG) of the percolation process.

The size of the GCC following bond percolation is equivalent to the fraction of nodes that are in the removed state of the SIR process. In the SIR model, nodes are either susceptible, infected, or removed. Infection occurs along edges that connect infected nodes to their susceptible neighbors. Once infected, a node remains infected for a period τ before recovering to the R state. The equivalence between SIR dynamics and bond percolation occurs when the infection period is a single-valued distribution [1]. Hence, the absorbing equilibrium of the model is binary state and is composed of only susceptible and recovered nodes.

Once a network has been percolated, it can be percolated a second time. Under disease mapping, this corresponds to a second disease, temporally separated from the first, spreading over the network. This could be a mutant strain of strain 1 invading a population or a different pathogen altogether; we will simply refer to "strain 2." If strain 2 does not interact with strain 1 in any way, then the expected dynamics will follow the traditional SIR model, and ordinary bond percolation is sufficient to describe the outbreak fraction. However, strain 1 could grant perfect cross-immunity to recovered nodes such that they do not become infected further [2–8]. Alternatively, infection by strain 1 could be a prerequisite to contracting strain 2 [9–11]. In both cases, the *infection history* of a node in the equilibrium of strain 1 is important in whether or not it contracts the second strain. The more likely scenario, and the most general disease interaction model, is that the spreading of strain 2 is only *modulated* by the presence of strain 1. The modulation could facilitate strain 2, a *partial coinfection* model, or it could hinder its spreading, a *partial immunity model*. With these definitions in place, the above two scenarios of perfect cross-immunity and perfect coinfection are the limiting logic of a smooth spectrum of interactions in which only partial interaction is observed, and it is this case that we model in this paper. We illustrate the concepts of partial coinfection and partial immunity in Fig. 1.

We choose to investigate our model for clustered networks, that is, networks containing a nonzero density of triangles. We do this because, typically, human contact networks are highly clustered: any two contacts often have a third, mutual friend, and clustering has important consequences for real epidemics [12–15]. Clustering has been well studied in the literature and is known to greatly influence the properties of the emergence and overall size of the GCC of strain 1. It is not only the number of triangles in a network that determines the percolation properties but also the nature of how those triangles connect together and to other edge topologies, degree assortativity being a leading factor. Therefore, different clustered degree distributions often lead to dichotomous conclusions regarding the effect of clustering. It was shown by Newman [16]

*pm78@st-andrews.ac.uk

 ©2021 American Physical Society

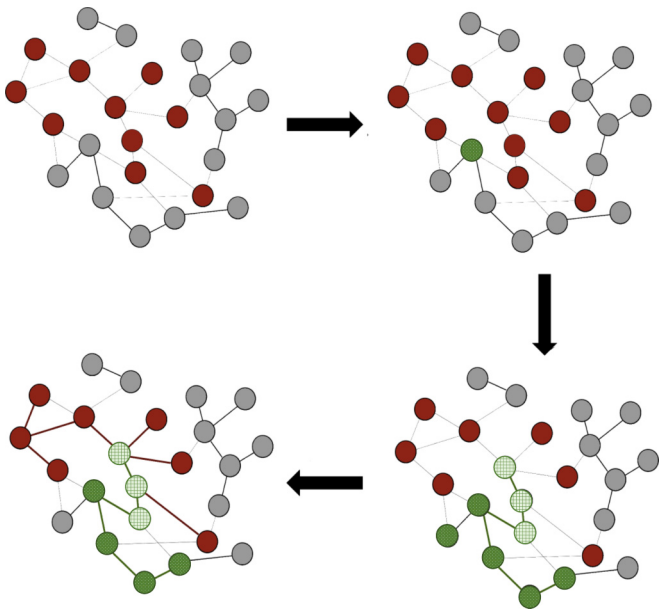

FIG. 1. From top left: A primary disease has spread over a network of susceptible nodes (gray) to create a giant component of infected hosts (red) at its equilibrium. A seed node (dotted dark green) of degree $k = 4$ has degree 2 in the RG (uninfected neighbors) and $k - l = 2$ strain 1 infected neighbors. In the general case, the second strain then spreads *both* on the RG with transmissibility T_2 and over the GCC of strain 1 (hatched light green) with modulated transmissibility T_2'. The light green nodes are *coinfected* with both strain 1 and strain 2, while the red (dark green) nodes only have strain 1 (2).

that Poisson networks with clustering reduce the epidemic threshold of the first strain, while Miller [17,18] and Gleeson and coworkers [19–21] showed that clustering can increase the epidemic threshold when assortativity is controlled, which was also reported by Hasegawa and colleagues [22,23].

Recent results [8,11] have investigated the role of clustering for a second strain in the limiting cases of perfect cross-immunity (spreading solely on the RG) and perfect coinfection (spreading solely on the GCC) using bond percolation. It was found that clustering can again exhibit polarized results depending on the details of the contact network. In this paper, we introduce a model of partial immunity using the generating function formulation and find excellent agreement with Monte Carlo simulations of bond percolation. This model generalizes the results of [8,11] by relaxing the strict conditions imposed on how the second strain interacts with the first one (spreading on both the RG and the GCC). We then study the role of clustering in shaping the evolutionary behavior of strain 2 for power law networks with tunable clustering that exhibit exponential degree cutoffs.

II. CONFIGURATION MODEL AND GENERATING FUNCTIONS

The configuration model is a prescription to create random graphs whose degrees are distributed according to some predefined distribution [24–28]. The central object of the model is the degree distribution $p(k)$, which is the probability that a node chosen at random from the network has degree k. By extending $p(k)$ to a joint distribution $p(s,t)$, where s is the number of ordinary tree-like edges and t is the number of triangles that the node is involved in (such that $k = s + 2t$), the configuration model can be used to create random graphs with clustering [16,17]. Similarly, by extending the joint distribution to include other cycles, such as squares, four-cliques, and so on, graphs with complicated clustering can be created.

To create a realization of a clustered graph of size N according to the configuration model, each node is assigned half degrees or *stubs*, in other words, an (s_i, t_i) pair drawn from the joint distribution $p(s,t)$ where the following conditions hold:

$$\sum_{i=1}^{N} s_i = 2\mathcal{S} \tag{1}$$

and

$$2\sum_{i=1}^{N} t_i = 3\mathcal{T}, \tag{2}$$

where $\mathcal{S}$ is the number of tree-like edges in the network and $\mathcal{T}$ is the number of triangles, both of which are integers to ensure that the drawn sequence is graphic. At each step of the construction, pairs of tree-like stubs from two different randomly selected nodes i and j are connected together to create a full edge; similarly, three triangle stubs are selected at random from three different nodes i, j, k to create a triangle among them. This process is repeated until all stubs have been matched among the nodes. Each constructed graph is a member of an ensemble of graphs with equivalent degree distributions and is absent of degree correlations [28].

Next, we turn our attention to the generating function formulation [16,17,24,25]. Generating functions are infinite series that can be used to encapsulate probabilities associated with certain network properties. For instance, the joint probability distribution is generated by a bivariate generating function as

$$G_0(z_\perp, z_\Delta) = \sum_{s=0}^{\infty}\sum_{t=0}^{\infty} p(s,t) {z_\perp}^s {z_\Delta}^t. \tag{3}$$

This expression is understood as follows: we choose a node at random from the network and let $z_\perp$ and z_Δ be the probabilities that a single tree-like edge and a single triangle are in states $Z_\perp$ and Z_Δ, respectively. Assuming that all tree-like edges and all triangles are independent of one another (such that $Z_\perp$ and Z_Δ are independent and identically distributed probabilities), then the probability that precisely s edges and t triangles surrounding a node are in this state is $z_\perp^s$ and z_Δ^t. We then multiply this expression by the probability that the node we chose had joint degree (s,t), which is simply $p(s,t)$, and finally, we sum over all combinations of s and t in the network. This yields the probability of choosing a node at random, or, equivalently, the fraction of nodes in the network whose tree-like edges are in states $Z_\perp$ and whose triangles are in state Z_Δ.

If $z_\perp = z_\Delta = 1$, then these states occur with certainty, and we have a normalizing condition

$$\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} p(s,t) = 1. \tag{4}$$

The probability of choosing a node at random from the network and the probability of choosing an *edge* at random and then picking one of the adjacent nodes at random do not yield equivalent probabilities in general. The reason is high degree nodes are more likely to be chosen, and so the properties of the nodes are different from one another. Further, the properties of the nodes reached by following a random tree-like edge and a random triangle are also nonequivalent in general. Thus, we must write an expression for the probability of reaching a node of joint degree (s,t) by traversing a random tree-like edge back to a node,

$$G_{1,\perp}(z_\perp, z_\Delta) = \frac{1}{\langle s\rangle}\frac{\partial G_0}{\partial z_\perp}, \tag{5}$$

and by following a random triangle back to a node,

$$G_{1,\Delta}(z_\perp, z_\Delta) = \frac{1}{\langle t\rangle}\frac{\partial G_0}{\partial z_\Delta}. \tag{6}$$

In each case, $\langle s\rangle$ is the average number of tree-like edges a node is a member of and is given by $\partial_{z_\perp} G_0(1,1)$ with an analogous expression for $\langle t\rangle$.

The bond percolation equivalence can now be understood in detail. If we set $z_\perp$ as the probability that a single tree-like edge does not connect the chosen node to the GCC and z_Δ as the probability that the chosen node's involvement in a single triangle fails to connect it to the GCC, then we can generate the probability that a randomly chosen node fails to be part of the GCC. Thus, we can find emergent macroscopic properties of the entire network from the description of the local environment of a node, which we then average over all permissible node joint degrees.

III. PARTIAL INTERACTION

We begin this section with a review of traditional generating function theory for tree-like edges [26]. To obtain the faction of the network that did not become infected, we must examine the probability that none of the edges of a degree k node transmit their infection. The failure to pass on infection occurs either because the neighbor was itself uninfected with probability $u_\perp$ or it was infected but did not transmit in this instance with probability $(1-u_\perp)(1-T_1)$. Since there are k such edges and they are all independent of one another, the probability that a node fails to become infected at all is simply the sum of these two probabilities raised to the power k,

$$[1 - T + u_\perp T_1]^k.$$

The degree is then averaged over the degree distribution of the network to obtain the total probability that the average node remains uninfected as

$$\sum_{k=0}^{\infty} p(k)[1 - T_1 + u_\perp T_1]^k.$$

At the binary-state equilibrium point of bond percolation, nodes are either infected or uninfected. Familiar generating function theory utilizes this mutually exclusive relationship to calculate the outbreak fraction of strain 1 as 1 minus the fraction of uninfected nodes.

This picture was expanded upon by Newman and colleagues [2,9] to describe the equilibrated network from the perspective of both an uninfected node and an infected node *given* they belong to the RG and the GCC, respectively. It is within this picture that we can create a partial immunity model by an extension and adaptation of Newman's work. In a partial immunity model, we cannot simply select a single node type to describe the final state of the network and use the mutually exclusive property as before. This is because subsequent strains can spread on both neighbor types, and so neglecting one of the descriptions leads to an undercounting of the full spectrum of transmission routes to the average node in the network. Further, each distinct transmission route from neighbors that have different infection histories from one another occurs with a different probability. Thus, we must describe the equilibrium of strain 1 by considering the local structure of both possible node states as

$$1 = p_{\text{uninfected}} + p_{\text{infected}}. \tag{7}$$

In the following two sections, we describe the local environment of a node in the RG and a node in the GCC following the first percolation process. During the discussion, we refer to the graph motifs in Fig. 2, where each possible neighbor state following percolation of a clustered network is displayed.

Across both focal nodes, we observe 14 different motifs that could surround a pair of nodes chosen at random. Among these, there are 18 different neighbor states with unique infection histories composed from three basis states of the neighbor node: uninfected (green), infected externally (gray), and infected directly (blue). For example, consider the infected focal node (red) in Fig. 2. There are nine different motifs (F–N) that could potentially surround the focal node. Counting each tree-like neighbor and each node within a triangle (excluding the focal node itself) that is not related by symmetry to its neighbor, there are 12 different neighboring sites; each site is occupied by one of three infection states: uninfected (green), externally infected (gray), and directly infected (blue).

A. Uninfected node description

The local environment of a node in the RG created by the first strain is considered here. It accounts for all motifs that have the yellow focal node in Fig. 2. This result was first derived in [8], and so we present and discuss the result rather than rederive it. The generating function for the probability of choosing an uninfected focal node (yellow) from the network $p_{\text{uninfected}} = F_0(\vec{x})$ is

$$\begin{aligned} F_0(\vec{x}) = G_0\big(&u_\perp x_1 + (1-u_\perp)(1-T_1)x_2, (u_\Delta x_3)^2 \\ &+ [(1-u_\Delta)(1-T_1)x_4]^2 \\ &+ 2u_\Delta(1-u_\Delta)(1-T_1)\big(1-T_1^2\big)x_5\big), \end{aligned} \tag{8}$$

where u_Δ is the probability that a neighbor node in a triangle is uninfected. The vector $\vec{x} = \{x_1, \ldots, x_5\}$ has five dimensions, one for each of the five motifs that could surround the

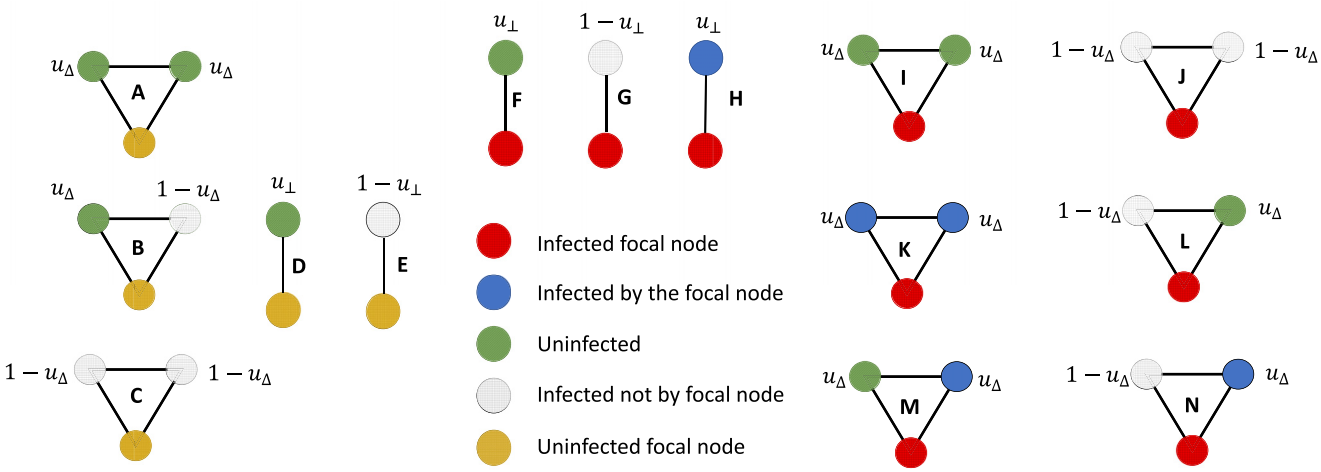

FIG. 2. The 14 motifs that surround both focal nodes in the percolation model. In each case the lowest node in the motif is the focal node, for which there are two considerations, uninfected (yellow) and infected (red). There are three states a neighboring node could be in: uninfected (green), infected externally (gray), and infected directly (blue), although the latter occurs only when the focal node is infected. Motifs A–C are the triangle motifs surrounding an uninfected focal node; D and E are the two types of tree-like edges. Motifs F–H are the tree-like edges that could surround an infected focal node, and finally, motifs I–N are the triangles that an infected focal node can belong to. Among these motifs, there are 18 unique node sites in total. Symmetric triangles (about a vertical axis bisecting the focal node) contribute only one site type, while mixed triangles contribute two site types and tree-like edges contribute one site type each. The numbering convention for mixed triangles always proceeds from left to right; for instance, in mixed state triangle B the uninfected neighbor is B1, while the infected neighbor is B2. Also, we do not have to include the mirror image of mixed triangles since they occur with equal probability.

uninfected focal node. We define two additional generating functions, $F_{1,\perp}(\vec{x})$ and $F_{1,\Delta}(\vec{x})$, by replacing $G_0(\vec{x})$ in Eq. (8) with $G_{1,\perp}(\vec{x})$ and $G_{1,\Delta}(\vec{x})$, respectively. It is well known that Eq. (8) can be used to generate the size of the GCC of strain 1 according to the following self-consistent set of equations:

$$u_\perp = F_{1,\perp}(\vec{1}), \tag{9}$$

$$u_\Delta = F_{1,\Delta}(\vec{1}), \tag{10}$$

followed by $S_1 = 1 - F_0(\vec{1})$.

B. Infected node description

The local environment of a node in the GCC created by the first strain is presented here. This result was first derived in [11], and so for brevity we do not repeat that work. The generating function $p_{\text{infected}} = H_0(\vec{y})$ for picking an (externally) infected focal node (red) from the network that is part of the GCC is given by

$$\begin{aligned} H_0(\vec{y}) = {} & G_0\big(u_\perp(1-T_1)y_1 + (1-u_\perp)y_2 + u_\perp T_1 y_3, [u_\Delta(1-T_1)]^2 y_4 + (1-u_\Delta)^2 y_5 + (u_\Delta T_1)^2 y_6 \\ & + 2u_\Delta(1-T_1)(1-u_\Delta)y_7 + 2u_\Delta(1-T_1)u_\Delta T_1 y_8 + 2(1-u_\Delta)u_\Delta T_1 y_9\big) \\ & - G_0(u_\perp(1-T_1)y_1 + (1-u_\perp)(1-T_1)y_2 + u_\perp T_1 y_3, [u_\Delta(1-T_1)]^2 y_4 + [(1-u_\Delta)(1-T_1)]^2 y_5 + (u_\Delta T_1)^2 y_6 \\ & + 2u_\Delta(1-T_1)^2(1-u_\Delta)\big(1-T_1^2\big)y_7 + 2u_\Delta(1-T_1)u_\Delta T_1 y_8 + 2(1-u_\Delta)u_\Delta T_1(1-T_1)\big(1-T_1^2\big)y_9\big). \end{aligned} \tag{11}$$

We will also define $H_{1,\perp}(\vec{y})$ and $H_{1,\Delta}(\vec{y})$ by replacing $G_0(\vec{y})$ by $G_{1,\perp}(\vec{y})$ and $G_{1,\Delta}(\vec{y})$, respectively. Additionally, we generate a description of the directly infected neighbor state (blue) as

$$\begin{aligned} J_{1,\tau}(\vec{y}) = {} & G_{1,\tau}\big(u_\perp(1-T_1)y_1 + (1-u_\perp)(1-T_1)y_2 + u_\perp T_1 y_3, [u_\Delta(1-T_1)]^2 y_4 + [(1-u_\Delta)(1-T_1)]^2 y_5 + (u_\Delta T_1)^2 y_6 \\ & + 2u_\Delta(1-T_1)^2(1-u_\Delta)\big(1-T_1^2\big)y_7 + 2u_\Delta(1-T_1)u_\Delta T_1 y_8 + 2(1-u_\Delta)u_\Delta T_1(1-T_1)\big(1-T_1^2\big)y_9\big). \end{aligned} \tag{12}$$

The size of the GCC of strain 1 can be found by solving

$$u_\perp = J_{1,\perp}(\vec{1}), \tag{13a}$$

$$u_\Delta = J_{1,\Delta}(\vec{1}), \tag{13b}$$

and then $S_1 = H_0(\vec{1})$. In relation to the uninfected node description we have that $F_{1,\tau}(\vec{1}) = J_{1,\tau}(\vec{1})$ and that $H_0(\vec{1}) = 1 - F_0(\vec{1})$. Thus, the full description of the binary-state equilibrium following bond percolation is given by the relation

$$1 = F_0(\vec{1}) + G_0(\vec{1}). \tag{14}$$

This expression constitutes a way of using the generating function formulation, and it is this key equation that allows us to create the partial immunity model.

IV. STRAIN 2

We have seen above how the GCC of the first strain can be obtained from either description of members of the percolation equilibrium: an uninfected node in the RG or an infected node in the GCC. Both methods utilize the state normalzation in Eq. (7) and the mutually exclusive property of the binary state.

To calculate the outbreak size of strain 2, we proceed as follows. For each of the 18 possible neighboring node states, we must introduce a probability that infection with strain 2 does not occur through this channel by some means. Therefore, we introduce 18 distinct probabilities that a neighbor of a given state fails to infect a given focal node with strain 2. Although arbitrary, we choose different symbols for these probabilities depending on whether the neighbor state surrounds an uninfected node or an infected node. We will see in a moment that subsets of the 18 sites are generated by the same expressions, and as such, the dimensionality of the model can be significantly reduced. However, we proceed in full for the moment.

There are six unique states that surround an uninfected focal node, and thus, we define a set of six probabilities $\{w\}$ that each holds the value of not becoming infected by strain 2 from one of these states. Specifically, there are four triangle neighbors and two tree-like neighbors, so

$$\{w\} = \left\{w_\Delta^A, w_\Delta^{B1}, w_\Delta^{B2}, w_\Delta^C, w_\perp^D, w_\perp^E\right\}. \tag{15}$$

Similarly, there are 12 states surrounding the node in the GCC, and so we introduce a set $\{v\}$ that holds the values of the probabilities of not becoming infected with strain 2 from these states. Specifically, there are three states reached by tree-like edges and nine states within the triangle motifs. Hence,

$$\begin{aligned}\{v\} = \{&v_\perp^F, v_\perp^G, v_\perp^H, v_\Delta^I, v_\Delta^J, v_\Delta^K, v_\Delta^{L1}, v_\Delta^{L2}, v_\Delta^{M1},\\ &\times v_\Delta^{M2}, v_\Delta^{N1}, v_\Delta^{N2}\}.\end{aligned} \tag{16}$$

We next need to write self-consistent expressions for each of the values in $\{w\}$ and $\{v\}$. Before we do this, we define two functions that express the probability of transmission failing through a tree-like edge, $g(v, T) = v + (1 - v)(1 - T)$, and a triangle motif

$$\begin{aligned}h(v_\mu, v_\nu, T_\mu, T_\nu) = {}& g(v_\mu, T_\mu)g(v_\nu, T_\nu)\\ &- v_\mu(1 - v_\nu)(1 - T_\nu)T_\nu T_\mu\\ &- v_\nu(1 - v_\mu)(1 - T_\mu)T_\mu T_\nu,\end{aligned} \tag{17}$$

with the convention that $h(v_\mu, v_\mu, T_\mu, T_\mu) = h(v_\mu, T_\mu)$. We will insert these functions into the $\vec{x}$ and $\vec{y}$ vectors in the arguments of the generating functions; each insertion describes the probability that strain 2 is not contracted from a particular motif. The probability of not getting infected by strain 2 from the uninfected neighbor at the end of a tree-like edge is

$$w_\perp^D = F_{1,\perp}/u_\perp. \tag{18a}$$

The probability of not contracting strain 2 from the infected neighbor at the end of a tree-like edge is

$$w_\perp^E = H_{1,\perp}/(1 - u_\perp). \tag{18b}$$

We now turn to the triangle probabilities $\{w_\Delta^A, w_\Delta^{B1}, w_\Delta^{B2}, w_\Delta^C\}$. The probability that the uninfected focal node does not get strain 2 from the symmetric susceptible site is

$$w_\Delta^A = F_{1,\Delta}/u_\Delta. \tag{18c}$$

The probability that the symmetric infected site does not transmit to the uninfected focal node is

$$w_\Delta^C = H_{1,\Delta}/(1 - u_\Delta). \tag{18d}$$

The mixed triangle follows. For the uninfected focal node, we have the probability of not becoming infected with strain 2 from an uninfected neighbor as

$$w_\Delta^{B1} = F_{1,\Delta}/u_\Delta, \tag{18e}$$

while for the infected site we have

$$w_\Delta^{B2} = H_{1,\Delta}/(1 - u_\Delta). \tag{18f}$$

We now have all of the probabilities that we require to describe the local environment of the uninfected node. We now turn to the description of the infected node in the GCC of strain 1. The three tree-like sites, $\{v_\perp^F, v_\perp^G, v_\perp^H\}$, are generated as follows: the uninfected neighbor

$$v_\perp^F = F_{1,\perp}/u_\perp, \tag{18g}$$

the externally infected neighbor

$$v_\perp^G = H_{1,\perp}/(1 - u_\perp), \tag{18h}$$

and the directly infected neighbor

$$v_\perp^H = J_{1,\perp}/u_\perp. \tag{18i}$$

We now require the nine triangle values $\{v_\Delta^I, v_\Delta^J, v_\Delta^K, v_\Delta^{L1}, v_\Delta^{L2}, v_\Delta^{M1}, v_\Delta^{M2}, v_\Delta^{N1}, v_\Delta^{N2}\}$. The probability that an uninfected neighbor fails to transmit strain 2 through a symmetric uninfected triangle I is

$$v_\Delta^I = F_{1,\Delta}/u_\Delta. \tag{18j}$$

The probabilities that the infected focal nodes in triangles J and K do not contract strain 2 are

$$v_\Delta^J = H_{1,\Delta}/(1 - u_\Delta) \tag{18k}$$

and

$$v_\Delta^K = J_{1,\Delta}/u_\Delta. \tag{18l}$$

The mixed triangle L is given by

$$v_\Delta^{L_1} = H_{1,\Delta}/(1 - u_\Delta) \tag{18m}$$

and

$$v_\Delta^{L2} = F_{1,\Delta}/u_\Delta. \tag{18n}$$

Triangle M follows as

$$v_\Delta^{M1} = F_{1,\Delta}/u_\Delta \tag{18o}$$

and

$$v_\Delta^{M_2} = J_{1,\Delta}/u_\Delta. \tag{18p}$$

Finally, triangle N is given by

$$v_\Delta^{N_1} = H_{1,\Delta}/(1 - u_\Delta) \tag{18q}$$

and

$$v_\Delta^{N_2} = J_{1,\Delta}/u_\Delta. \tag{18r}$$

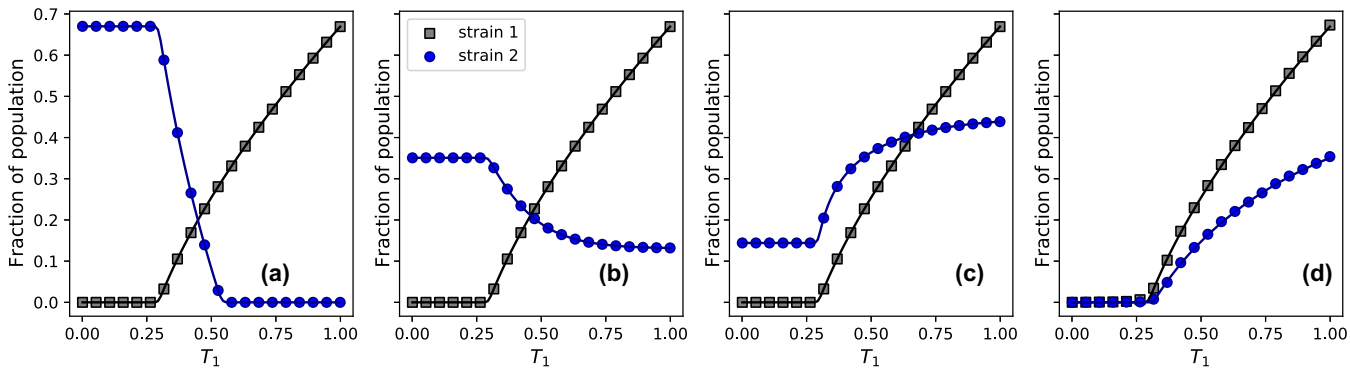

FIG. 3. The outbreak fractions for several (T_2, T_2') combinations of the model all in the absence of clustering. (a) Complete cross-immunity $(T_2, T_2') = (1, 0)$, (b) partial cross-immunity $(T_2, T_2') = (0.6, 0.39)$, (c) partial coinfection $(T_2, T_2') = (0.4, 0.7)$, and (d) perfect coinfection $(T_2, T_2') = (0, 0.6)$. Markers are the average of 50 repeats of bond percolation over CCM networks of size $N = 65\,000$, $\alpha = 2.0$, and $\theta = 0$; square markers are strain 1, while circles are strain 2. Solid lines are the theoretical results of Eq. (21) for strain 2.

At this point, we have not yet written the arguments of each generating function, $\vec{x}$ and $\vec{y}$. It happens that there are several equivalent expressions among the variables, allowing us to reduce the dimension of the problem considerably. Specifically, we notice the following redundancies among the relations: $v_\Delta^{M1} = v_\Delta^{L2} = v_\Delta^{I} = w_\Delta^{B1} = w_\Delta^{A}$, $w_\Delta^{B2} = w_\Delta^{C} = v_\Delta^{L1} = v_\Delta^{N1} = v_\Delta^{J}$, $w_\perp^{E} = v_\perp^{G}$, $w_\perp^{D} = v_\perp^{F}$, and $v_\Delta^{K} = v_\Delta^{N2} = v_\Delta^{M2}$. This overprescription affords a reduction in the number of system variables to only six independent variables, one for each of the possible neighbor nodes: uninfected, externally infected, and directly infected for tree-like edges and triangle motifs, respectively. Therefore, if we write the argument of each generating function $F_{1,\perp}, H_{1,\perp}, J_{1,\perp}, F_{1,\Delta}, H_{1,\Delta}$, and $J_{1,\Delta}$ once, it is known for all occurrences of that function in the model. Further, we observe that the only difference between $F_{1,\tau}, H_{1,\tau}$, and $J_{1,\tau}$ for $\tau = \perp, \Delta$ is the underlying $G_{1,\tau}$ function, not the argument. In other words, the arguments of $F_{1,\perp}$ and $F_{1,\Delta}$, for instance, are equivalent; we do not distinguish based on their topology. A final simplification can be achieved by noting that the arguments of $J_{1,\tau}$ and $H_{1,\tau}$ are also equivalent. Therefore, there are only two arguments to write: one for $F_{1,\tau}$ and another for $H_{1,\tau}$. These are given by $\vec{x} = \vec{\zeta}$ and $\vec{y} = \vec{\xi}$, where

$$\vec{\zeta} = \{g(w_\perp^D, T_2), g(w_\perp^E, T_2'), h(w_\Delta^A, T_2), h(w_\Delta^C, T_2'), \times h(w_\Delta^{B1}, w_\Delta^{B2}, T_2, T_2')\} \tag{19}$$

and

$$\vec{\xi} = \{g(v_\perp^F, T_2), g(v_\perp^G, T_2'), g(v_\perp^H, T_2'), h(v_\Delta^I, T_2), h(v_\Delta^J, T_2'), \times h(v_\Delta^K, T_2'), h(v_\Delta^{L1}, v_\Delta^{L_2}, T_2', T_2), \times h(v_\Delta^{M1}, v_\Delta^{M2}, T_2, T_2'), h(v_\Delta^{N_1}, v_\Delta^{N_2}, T_2', T_2')\}, \tag{20}$$

which constitute vectors of probabilities that each neighbor site fails to transmit infection to the focal node (or connect it to the GCC). With this in place, we now have an expression for all of the required probabilities $\{w\}$ and $\{v\}$. The size of the second outbreak over the network is then found by solving

$$S_2 = [F_0(\vec{1}) + H_0(\vec{1})] - [F_0(\vec{\zeta}) + H_0(\vec{\xi})], \tag{21}$$

where $[F_0(\vec{1}) + H_0(\vec{1})] = 1$. Qualitatively, this expression is 1 minus the probability that a node obtains strain 2 from either uninfected or infected neighbors.

V. NUMERICAL RESULTS

The results of the model under four different strain interactions for tree-like networks in the absence of clustering are shown in Fig. 3 as T_1 is varied. Across the simulations, the networks are built according to the clustered contact model (CCM), defined in [11], which is an example of a configuration model degree distribution with clustering. Power law contact distributions are typical of those found in real-world social networks [29]. The underlying degree distribution is given by a power law model with exponential degree cutoff (PLC) defined as

$$p^{\mathrm{PLC}}(k) = \frac{k^{-\alpha} e^{-k/\kappa}}{\mathrm{Li}_\alpha(e^{-1/\kappa})} \tag{22}$$

where κ is the degree cutoff, $\alpha \in [2, 3]$ is a power law exponent, and $\mathrm{Li}_n(z)$ is the nth polylogarithm of z [26]. Each k is then decomposed into tree degrees s and triangle degrees t according to

$$p^{\mathrm{CCM}}(k) = p^{\mathrm{PLC}}(k) \sum_{t=0}^{\lfloor k/2 \rfloor} \binom{\lfloor k/2 \rfloor}{t} \theta^t (1-\theta)^{\lfloor k/2 \rfloor - t}, \tag{23}$$

where $\lfloor \cdot \rfloor$ is the floor function and $\theta \in [0, 1]$ is the probability of a pair of edges belonging to a triangle.

We simulate bond percolation for both strains numerically using Monte Carlo simulations. Following strain 1, infected nodes are labeled, and subsequent infection with strain 2 occurs with probability T_2 for nodes in the RG or T_2' for GCC nodes.

In Fig. 3(a) we have $T_2 = 1$ and $T_2' = 0$, a perfect cross-immunity model [2] in which infection with strain 1 prevents infection with strain 2. In Fig. 3(b) we relax this hard limit, with $T_2 = 0.6$ and $T_2' = 0.39$, to obtain a partially cross-immune interaction whereby the transmission of strain 2 is reduced for strain 1 infected nodes. For $T_1 < T_{1,c}$ we observe the steady-state of strain 2 without competition from strain 1. At $T_1 = T_{1,c}$ a GCC in strain 1 emerges, and the number of

cases of strain 2 drops but does not vanish; in the limit $T_1 = 1$ strain 2 reaches its lowest incidence rate as competition is maximized. In Fig. 3(c) we observe a partial coinfection model, with $T_2 = 0.4$ and $T_2' = 0.7$. In this case, strain 2 is facilitated by the presence of strain 1 in the network, the symbiotic interaction leading to an increase in the incidence of strain 2 infected nodes. Figure 3(d) shows the hard limit of a perfect coinfection model [9] with $T_2 = 0$ and $T_2' = 0.6$; strain 2 cannot survive without a GCC of strain 1 present in the network.

With an understanding of the model without clustering, we now examine the case where $\theta \neq 0$ for both partial interaction models with $\kappa = 20$, $\alpha = 2$, and $\theta = 0.5$ (see Fig. 4). The epidemic threshold of strain 1 is reduced with clustering, so too is the overall outbreak size of strain 1 at large T_1 compared to unclustered networks. The incidence of strain 2 exhibits dual behavior over the range of T_1. For the partial immunity scenario [Fig. 4(a)], with $T_2 = 0.6$ and $T_2' = 0.4$, clustering reduces the incidence of strain 2 at low T_1; however, it increases it as $T_1 \to 1$. Conversely, for partial coinfection [Fig. 4(b)], with $T_2 = 0.4$ and $T_2' = 0.7$, having lowered the epidemic threshold of strain 1, clustering causes an increase in the incidence of strain 2 at lower T_1 values compared to the unclustered analog.

In Fig. 5 we perform a second experiment using the degree-δ model [11,17,19]. We define a distribution in which the degree of nodes involved in triangles is fixed to $k = 3$ and thus $(s, t) = (1, 1)$, while all other degrees are given by Eq. (23) for $(s, t) = (k, 0)$. With the degree correlations among triangles fixed, the epidemic threshold of the first strain increases with clustering. The partial cross-immune coupling [Fig. 5(a)], with $T_2 = 0.8$ and $T_2' = 0.65$, no longer exhibits a crossover in expected size of strain 1 and strain 2; clustering reduces the incidence of strain 2 for all values of T_1. Similarly, the partial coinfection model [Fig. 5(b)], with $T_2 = 0.6$ and $T_2' = 0.75$, exhibits a reduced incidence of strain 2 compared to the unclustered analog. As $T_1 \to 1$, however, the coinfection is reduced in the clustered graph compared to the unclustered one.

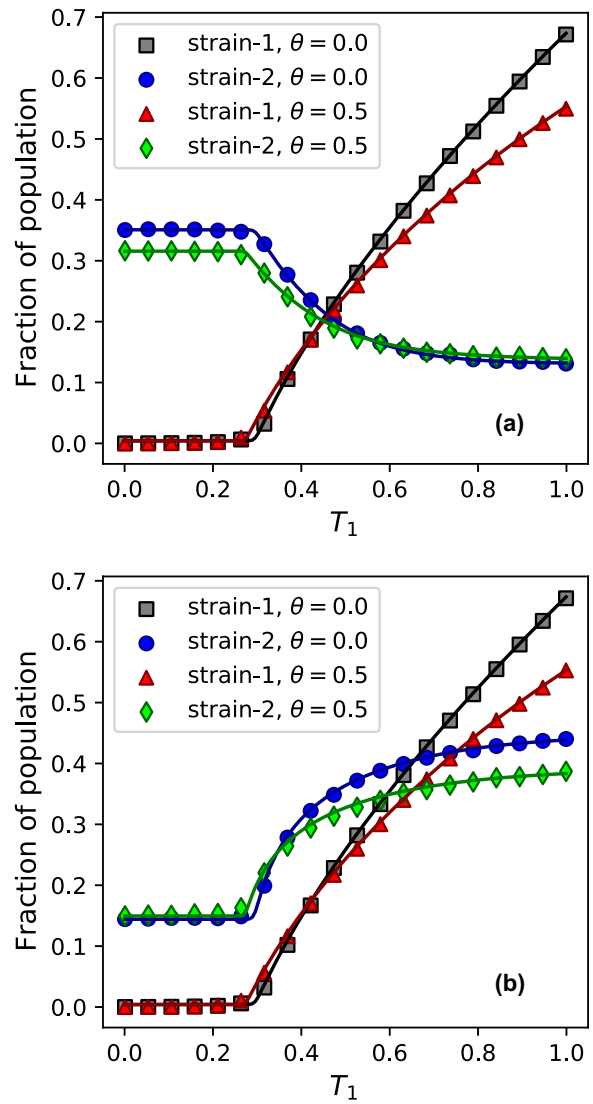

FIG. 4. The outbreak fractions for both strains for clustered [Eq. (23)] and unclustered networks as T_1 is varied under two disease couplings: (a) partial cross-immunity $(T_2, T_2') = (0.6, 0.4)$ and (b) partial coinfection $(T_2, T_2') = (0.4, 0.7)$. Simulations are the average of 50 repeats of bond percolation on networks with $N = 35\,000$ and $\theta = 0.0$, with $\alpha = 2.0$ and 0.5 for the unclustered and clustered networks, respectively. Solid lines are the theoretical results of Eqs. (11) and (21). In general, clustering reduces the extent of plural infections in the network; however, degree assortativity within the contact topology causes a reversal of this at high (low) values of T_1 in (a) [(b)].

VI. CONCLUSION

In this paper we have presented a general model of two pathogens spreading sequentially over a clustered network. The two strains can interact in a very broad manner ranging from perfect coinfection, facilitating the spreading of the subsequent strain, to perfect antagonism, competing for hosts, and all combinations in between. Our model uses generating functions to provide a theoretical understanding of Monte Carlo simulations of bond percolation, showing excellent agreement. Our model, which considers the local descriptions of both uninfected and infected nodes simultaneously, introduces a way to use generating functions to describe bond percolation. The paradigm of describing the entire absorbing state of the percolation equilibrium rather than the typical approach of only the uninfected nodes will certainly prove illuminating to many areas of epidemic spreading and network dynamics.

Our simulations on clustered networks examined how the dynamics of strain 2, under a partial disease interaction, were influenced by contact topology. We found that clustering generally reduces the prevalence of strain 2 in the network. The magnitude of this effect is dependent on the details of the clustering and the value of the transmissibility of strain 1. However, networks with finite degree correlations present (due to contact clustering) were found to have greater incidences of coinfection.

It is clear that this model can be reduced to purely tree-like networks by the removal of references to triangle motifs in the generating functions. The work herein can be generalized in a number of ways: first, we have formulated a clear recipe to

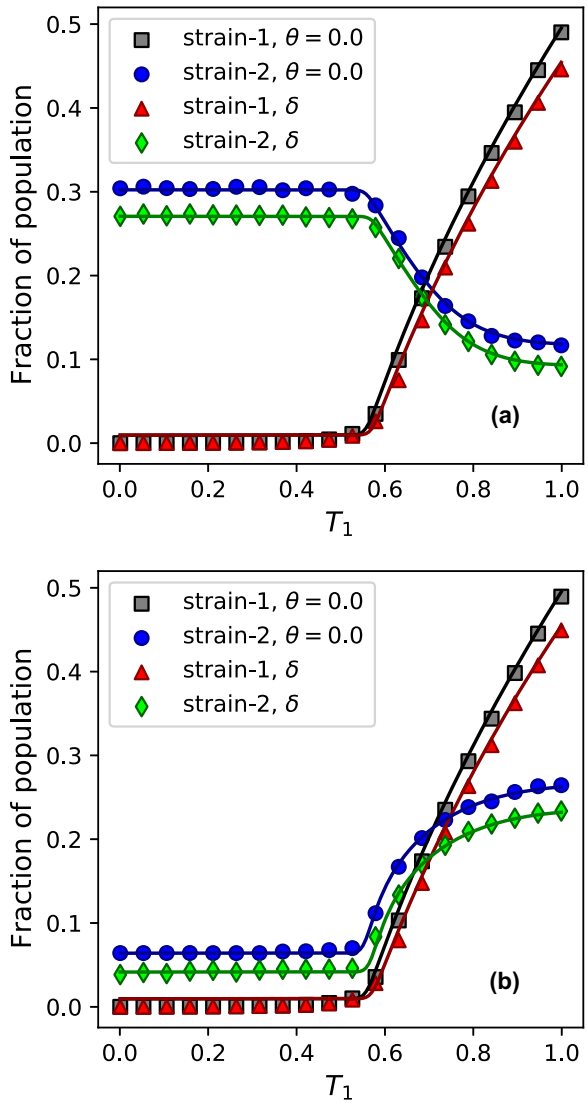

FIG. 5. The outbreak fractions of the degree-δ model with clustering constrained to low degree assortativity for (a) partial immunity $(T_2, T_2') = (0.8, 0.65)$ and (b) partial coinfection $(T_2, T_2') = (0.6, 0.75)$ disease interactions. Simulations are the average of 50 repeats of bond percolation on networks with $N = 35\,000$, $\alpha = 2.0$; solid lines are the theoretical results of Eqs. (11) and (21). Clustering reduces the fraction of the network that becomes infected by both strains for all values of T_1.

follow for the inclusion of different types of clustering, such as higher-order cliques and cycles [30], or, indeed, custom motifs such as those in [31]. In many cases, these subgraphs may represent social networks more accurately than the tree-triangle model we have considered here, and thus, this extension is important for the rationalization of disease spreading among human populations.

As presented, the model considers sequential strains that are temporally separated. Often, diseases spread simultaneously among a population. It is well known from analytical methods based on differential equations [32,33] and simulations of SIR through, for instance, Gillespie simulation that the initial conditions of each strain or disease greatly influence the dominant pathogen for two otherwise symmetric diseases [3]. Bond percolation and the method of generating functions inherently cannot capture the stochastic effects of concurrent spreading due to its equilibrium-based nature. However, previous work was conducted [3,4] to model a mutual pathogen interaction using generating functions. Extending our model to the study of concurrent strains would be a significant step forward for understanding disease spreading.

The model could be applied to multilayer clustered networks with some adaptation to increase the influence of topology on the spreading and enrich the dynamics further. Or perhaps the effects of drugs or vaccines that target particular topologies or node sites could also be investigated.

However, perhaps the most significant future generalization of this model is the extension to additional strains, in other words, subsequent percolations. As we do this, the immunological landscape becomes increasingly rich, and the node's infection histories become more diverse. We believe that the best approach would be to develop an automation, perhaps through recursion, that can write the required expressions, in a vein similar to [31]. This would allow the exact study of an N-strain seasonal influenza model and could also be used to discover once-a-century events (such as the 1918–1919 Spanish influenza and the COVID-19 pandemic), along with generational evolutionary pressures and genetic drift. We also believe that the parameters controlling disease interactions could be matched to real-world data to understand the relationship, symbiotic or antagonistic, between multiple diseases as they evolve and spread.

[1] E. Kenah and J. M. Robins, Second look at the spread of epidemics on networks, Phys. Rev. E **76**, 036113 (2007).
[2] M. E. J. Newman, Threshold Effects for Two Pathogens Spreading on a Network, Phys. Rev. Lett. **95**, 108701 (2005).
[3] B. Karrer and M. E. J. Newman, Competing epidemics on complex networks, Phys. Rev. E **84**, 036106 (2011).
[4] S. Funk and V. A. A. Jansen, Interacting epidemics on overlay networks, Phys. Rev. E **81**, 036118 (2010).
[5] V. Marceau, P.-A. Noël, L. Hébert-Dufresne, A. Allard, and L. J. Dubé, Modeling the dynamical interaction between epidemics on overlay networks, Phys. Rev. E **84**, 026105 (2011).
[6] J. C. Miller, Cocirculation of infectious diseases on networks, Phys. Rev. E **87**, 060801(R) (2013).
[7] L. Hébert-Dufresne, O. Patterson-Lomba, G. M. Goerg, and B. M. Althouse, Pathogen Mutation Modeled by Competition Between Site and Bond Percolation, Phys. Rev. Lett. **110**, 108103 (2013).
[8] P. Mann, V. A. Smith, J. B. O. Mitchell, and S. Dobson, Two-pathogen model with competition on clustered networks, Phys. Rev. E **103**, 062308 (2021).
[9] M. E. J. Newman and C. R. Ferrario, Interacting epidemics and coinfection on contact networks, PLoS One **8**, e71321 (2013).

[10] P.-B. Cui, F. Colaiori, and C. Castellano, Mutually cooperative epidemics on power-law networks, Phys. Rev. E **96**, 022301 (2017).

[11] P. Mann, V. A. Smith, J. B. O. Mitchell, and S. Dobson, Cooperative coinfection dynamics on clustered networks, Phys. Rev. E **103**, 042307 (2021).

[12] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature (London) **393**, 440 (1998).

[13] R. B. Rothenberg, J. J. Potterat, D. E. Woodhouse, S. Q. Muth, W. W. Darrow, and A. S. Klovdahl, Social network dynamics and HIV transmission, AIDS **12**, 1529 (1998).

[14] B. Szendroi and G. Csanyi, Polynomial epidemics and clustering in contact networks, Proc. R. Soc. London, Ser. B **271** (2004).

[15] J. M. Read and M. J. Keeling, Disease evolution on networks: The role of contact structure, Proc. R. Soc. London, Ser. B **270**, 699 (2003).

[16] M. E. J. Newman, Random Graphs with Clustering, Phys. Rev. Lett. **103**, 058701 (2009).

[17] J. C. Miller, Percolation and epidemics in random clustered networks, Phys. Rev. E **80**, 020901(R) (2009).

[18] J. C. Miller, Spread of infectious disease through clustered populations, J. R. Soc. Interface **6**, 1121 (2009).

[19] J. P. Gleeson, S. Melnik, and A. Hackett, How clustering affects the bond percolation threshold in complex networks, Phys. Rev. E **81**, 066114 (2010).

[20] A. Hackett, S. Melnik, and J. P. Gleeson, Cascades on a class of clustered random networks, Phys. Rev. E **83**, 056107 (2011).

[21] A. Faqeeh, S. Melnik, and J. P. Gleeson, Network cloning unfolds the effect of clustering on dynamical processes, Phys. Rev. E **91**, 052807 (2015).

[22] T. Hasegawa and S. Mizutaka, Structure of percolating clusters in random clustered networks, Phys. Rev. E **101**, 062310 (2020).

[23] T. Hasegawa and Y. Iwase, Observability transitions in clustered networks, Phys. A (Amsterdam, Neth.) **573**, 125970 (2021).

[24] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Network Robustness and Fragility: Percolation on Random Graphs, Phys. Rev. Lett. **85**, 5468 (2000).

[25] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Random graphs with arbitrary degree distributions and their applications, Phys. Rev. E **64**, 026118 (2001).

[26] M. E. J. Newman, Spread of epidemic disease on networks, Phys. Rev. E **66**, 016128 (2002).

[27] B. K. Fosdick, D. B. Larremore, J. Nishimura, and J. Ugander, Configuring random graph models with fixed degree sequences, SIAM Rev. **60**, 315 (2018).

[28] M. E. Newman, *Networks* (Oxford University Press, Oxford, 2019).

[29] M. Newman, Power laws, Pareto distributions and Zipf's law, Contemp. Phys. **46**, 323 (2005).

[30] P. Mann, V. A. Smith, J. B. O. Mitchell, and S. Dobson, Percolation in random graphs with higher-order clustering, Phys. Rev. E **103**, 012313 (2021).

[31] P. Mann, V. A. Smith, J. B. O. Mitchell, and S. Dobson, Random graphs with arbitrary clustering and their applications, Phys. Rev. E **103**, 012309 (2021).

[32] J. Sanz, C.-Y. Xia, S. Meloni, and Y. Moreno, Dynamics of Interacting Diseases, Phys. Rev. X **4**, 041005 (2014).

[33] E. M. Volz, J. C. Miller, A. Galvani, and L. Ancel Meyers, Effects of heterogeneous and clustered contact patterns on infectious disease dynamics, PLoS Comput. Biol. **7**, 1 (2011).