N-strain epidemic model 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;
EaStCHEM 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

In this paper we examine the emergent structures of random networks that have undergone bond percolation an arbitrary, but finite, number of times. We define two types of sequential branching processes: a competitive branching process, in which each iteration performs bond percolation on the residual graph (RG) resulting from previous generations, and a collaborative branching process, where percolation is performed on the giant connected component (GCC) instead. We investigate the behavior of these models, including the expected size of the GCC for a given generation, the critical percolation probability, and other topological properties of the resulting graph structures using the analytically exact method of generating functions. We explore this model for Erdős-Renyi and scale-free random graphs. This model can be interpreted as a seasonal *N*-strain model of disease spreading.

I. INTRODUCTION

A network is a collection of nodes connected by edges. Bond percolation on complex networks is a widely studied binary-state stochastic process. In this model, the edges of a graph are iterated and said to be *occupied* with probability T and remain *unoccupied* with probability $1 - T$. When T is small, the network is composed of many small isolated components of occupied edges. At some critical value, T_c, the small components connect together to form a macroscopic giant connected component (GCC), exhibiting a second-order phase transition. As $T \to 1$ 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; see Fig. 1. The size of the GCC following bond percolation, as well as the value of the critical bond occupation probability, have an equivalence to the absorbing state of the susceptible, infected, removed (SIR) epidemic process [1] when a node's infectious period is drawn from a single-valued distribution [2]. In the SIR model, nodes are either susceptible to infection from their infected neighbors, i.e., infectious, or are in the removed state. Transmission of infection occurs along edges that connect infected nodes to their susceptible neighbors. Once infected, a node remains infectious for a fixed period [2,3], τ, before recovering to the R state. Hence, the absorbing equilibrium of the model is a static binary state that is composed of susceptible nodes that did not contract the disease or removed nodes that did. The size of the GCC following bond percolation is equivalent to the largest outbreak size of the SIR model.

There has been previous work on extending the percolation-SIR equivalence to additional, temporally separate strains of an epidemic for random graphs [4–12]. Such models can be thought of as a model of seasonal diseases, with each pathogen running its course through a population before the next season's strain spreads. Within this context, the nature of how the second strain interacts with nodes that have been previously infected is very important to the resultant disease dynamics. Newman [4,5] studied two strains that compete for hosts by not allowing nodes in the removed state of strain 1 to become infected by strain 2. This model is an example of perfect cross-immunity. Mann *et al.* extended this work to study the role of contact clustering and modularity on the spread of the second pathogen [10]. Conversely, Newman and Ferrario [6] studied the opposite scenario in which only those nodes that have been infected by strain 1 can be later infected by strain 2. This is an example of a perfect coinfection model, and the extension to the clustered case was also studied [9]. In each case, the second strain is a separate SIR process that either spreads exclusively on the RG (cross-immune) or the GCC (coinfection) of the first percolation process; see Fig. 1. Both processes are abstractions of the spread of real diseases, since the strict requirements of perfect cross-immunity and perfect coinfection are the two limits of a spectrum of partial interactions. Partial cross-immunity and partial coinfection passing through a point of complete decoupling is known as a partial immunity model and has recently been studied using bond percolation [13] for the case of two sequential percolation processes (diseases).

Whilst these models are of clear interest to an epidemiology readership, the implications for the understanding of network structure cannot be understated. In particular, the cross-immune model provides us with a model of the RG of a network following bond percolation; similarly, the coinfection model yields insight into the structure of the GCC.

In this paper we extend the percolation models of [4–6] from two sequential strains to an arbitrary, but finite, number $N \in \mathbb{N}$ of sequential strains. We study the topological properties of the graph structures that arise, quantifying the expected
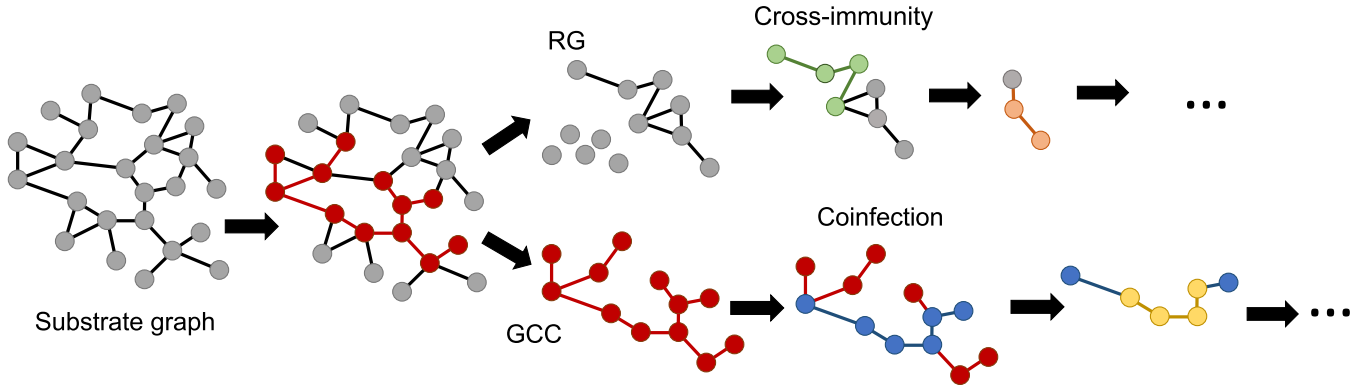
---

[*]pm78@st-andrews.ac.uk

FIG. 1. A substrate network (left) is percolated to yield a GCC (red) and an RG. In the competetive branching process, the RG is percolated again to create another GCC (green), and so on. This is a cross-immunity model. The collaborative branching process instead percolates the GCC of each generation and constitutes a model of complete coinfection.

outbreak sizes for each disease and the critical points of the model as well as their degree distributions and cumulative degree distributions of the graph structures that emerge.

II. GENERATING FUNCTIONS

The study of a bond percolation process over a network using generating functions was pioneered by Newman, Strogatz, and Watts around 20 years ago [1,14,15]. In this section we will review the key components of the generating function formulation that our N-strain model will use. Generating functions are infinite series; in the context of network science the summation is over the degrees of the nodes, k, while the coefficients are the values of the degree distribution, p_k, the probability of choosing a degree k node at random. This is generated by

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k. \quad (1)$$

The coefficients of a generating function can be recovered by repeated differentiation,

$$p_k = \frac{1}{k!} \frac{d^k G_0}{dx^k}\bigg|_{x=0}. \quad (2)$$

It is commonplace to evaluate the derivatives by numerical contour integration using the Cauchy formula

$$p_k = \frac{1}{2\pi i} \oint \frac{G_0(z)}{z^{k+1}} dz. \quad (3)$$

We can obtain the cumulative distribution function (CDF), $p_{k<\nu} = \sum_{k=0}^{\nu} p_k$, defined as the probability that a randomly chosen node has a degree less than or equal to ν, directly from $G_0(x)$. To achieve this, it is convenient to instead calculate one minus the probability of a node having a degree *greater or equal to* ν, $p_{k\geqslant\nu}$, which is called the survival function [16]. The latter distribution is generated by

$$G_{0,\nu}(x) = \frac{G_0(x) - p_0 - \cdots - p_{\nu-1}x^{\nu-1}}{x^{\nu}}. \quad (4)$$

The values of p_k can be obtained from Eq. (3) and evaluating $G_{0,\nu}(x)$ at $x = 1$ yields $p_{k\geqslant\nu}$.

The probability of choosing an edge at random from the network and following it randomly to a particular end to reach a node of degree k is generated by

$$G_1(x) = \frac{G_0'(x)}{G_0'(1)}, \quad (5)$$

where $G_0'(1) = \langle k \rangle$ is the average degree of the network.

For example, one of the simplest degree distributions occurs when the degrees are binomially distributed; this is the Erdős-Renyi degree distribution. In this model, the probability, $\theta = \langle k \rangle / V$, where V is the number of nodes, is the probability that an edge connects any two nodes, for all pairs of nodes in the network. In this instance, we find

$$G_0(x) = \sum_{k=0}^{V} \binom{V}{k} \theta^k (1-\theta)^{V-k} x^k, \quad (6)$$

which, in the limit of large V, is equal to $G_0(x) = e^{\langle k \rangle (x-1)}$ It must be noted, in this special case, that $G_1(x) = G_0(x)$, which simplifies the model significantly.

Cavity method

The cavity method is a statistical technique that is commonly used when dealing with generating functions. The method considers the local environment of a particular node chosen from the network at random (the focal node) once a (percolation) process has reached equilibrium. We suppose that the focal node's degree is k. We then calculate the probability associated with all configurations of the node's neighbors before averaging over the probability that the node we chose did indeed have degree k. As a motivating example of this method, consider a graph whose nodes are in one of two states, A or B, after a binary-state process has reached a static absorbing state. The focal node will have $k_A \leqslant k$ neighbors that are in state A and $k_B = k - k_A$ neighbors that are in the mutually exclusive state B. Assuming that the neighbor-state probabilities are independent and identically distributed (iid) along each edge, the sum over all configurations of the

probability of configuration k_A and k_B given k is

$$g^k = \sum_{k_A=0}^{k} \binom{k}{k_A} (\pi_A x)^{k_A} (1 - \pi_A)^{k - k_A}. \qquad (7)$$

Each combination of k_A and k_B is then wrapped into the kth term of a generating function as we average over the probability that the focal node we chose had degree k, $\sum_k p_k g^k$. This method is key to the analytical description we present in this paper. We will continually examine the permissible neighbours that connect to our focal node and their associated probabilities of occurring along each iid edge. As the number of generations of the disease process increases, the number of permissible neighbor states grows too; and so the local environment that an embedded node experiences evolves into a rich and complex landscape.

III. COMPETITIVE BRANCHING PROCESS

In this section, we define the ith competitive branching process as successive bond percolations occurring on the RG created by the $i - 1$ previous processes for $i = 1, \ldots, N$. This process has been studied previously using generating functions by Newman and Karrer when $N = 2$ [4,5]. The structure of the RG has also been studied for clustered and modular networks [10]. From a network science perspective, this model allows us to study the structure of the RG of sequential bond percolation processes. In particular, we observe how those sequential processes fracture the RG into isolated components and study the phase behavior associated with the sudden inability of the RG to support a GCC. Within the context of the SIR equivalence, this model considers the behavior of N seasonal strains of a disease (or separate diseases) that confer complete cross-immunity to all subsequent pathogens. The model allows us to study the expected outbreak size of each generation and the point of natural burn out due to the shrinking of the susceptible subpopulation.

A. Outbreak size

To the ith process, $i = 1, \ldots, N$, we assign a bond occupation probability T_i and aim to calculate the probability that a randomly chosen node does not belong to the largest percolated component of that generation. From this, we can find the mutually exclusive probability that a node does belong to the ith GCC, $\mathcal{A}_i$. To do this, we define the probability that a neighbor of our randomly selected node is not part of the ith GCC, u_i, given that it does not belong to any of the previous percolated components. Under the SIR equivalence, T_i is the transmissibility of the ith strain and u_i is the probability that a neighbor is not thus far infected.

There are two ways in which an edge emanating from the focal node can fail to connect it to the GCC: first, the neighbor could itself be unconnected, the probability of which by definition is u_i. Second, the neighbor could be connected, $(1 - u_i)$, but the bond is unoccupied $(1 - T_i)$. Therefore, the probability, $\bar{g}_i$, that an edge fails to connect the focal node to the ith GCC given that the neighbor does not belong to any other GCC is

$$\bar{g}_i(T_i, u_i \mid \mathrm{RG}) = u_i + (1 - u_i)(1 - T_i) \qquad (8)$$

The total probability that a neighbor belongs to the RG of the ith percolation can then be found through a set of recursive functions g_i that describe the probability that each iterative percolation failed to occupy this edge as

$$g_i(\boldsymbol{T}, \boldsymbol{u}) = u_{i-1}\bar{g}_i + (1 - u_{i-1})(1 - T_{i-1}) \qquad (9)$$

with $u_0 = 1$. A hierarchy of self-consistent equations can be written to sequentially solve for each u_i value:

$$u_i = \frac{G_1(g_i)}{\prod_j u_j}, \qquad j = 1, \ldots, i - 1. \qquad (10)$$

The size of the ith GCC (epidemic) is then found by

$$\mathcal{A}_i = \prod_{j=1}^{i-1} G_0(g_j) - G_0(g_i). \qquad (11)$$

The total infected fraction of the network is given by $\mathcal{A} = \sum_i \mathcal{A}_i$.

As an example of Eq. (11), we can obtain the expected outbreak size of the first epidemic, $i = 1$, from this system as $\mathcal{A}_1 = 1 - G_0(g_1)$, where $u_1 = G_1(g_1)$ and $g_1 = u_1 + (1 - u_1)(1 - T_1)$, [1]. In the case that $i = 2$ [4], we have $\mathcal{A}_2 = G_0(g_1) - G_0(g_2)$, where $u_2 = G_1(g_2)/u_1$ and

$$g_2 = u_1[u_2 + (1 - u_2)(1 - T_2)] + (1 - u_1)(1 - T_1). \qquad (12)$$

Similarly, for $i = 3$ we have $\mathcal{A}_3 = G_0(g_1)G_0(g_2) - G_0(g_3)$ with $u_3 = G_1(g_3)/(u_1 \cdot u_2)$ and

$$g_3 = u_1\{u_2[u_3 + (1 - u_3)(1 - T_3)]$$
$$+ (1 - u_2)(1 - T_2)\} + (1 - u_1)(1 - T_1). \qquad (13)$$

With these examples, it is hopefully clear how to write further generations of the competitive percolation process.

Given this, we can derive the conditions under which $\mathcal{A}_i$ is maximized given that the previous strains have run their course of the network. To see this, consider that each expression $G_1(g_i)$ has at most two roots for u_i: the trivial case $u_i = 1$, corresponding to no GCC ($\mathcal{A}_i = 0$), and a nonzero root in the unit interval that indicates that a finite fraction of the network is occupied by this strain ($\mathcal{A}_i > \epsilon$ for some nonzero fraction ϵ). Since the coefficients of $G_1(g_i)$ are non-negative, so too are its derivatives (for $u_i \geqslant 0$). Hence, $G_1(g_i)$ is in general positive and convex. Given that each g_i is convex in u_i for $T_i \in [0, 1]$, then, by Jensen's inequality [17], there exists a unique root T_i^* that minimizes $g(u_i)$. Thus, the best case scenario for the ith strain to maximize $\mathcal{A}_i$ is to have $g_j = 1$ and $g_i = 0$ for $j = 1, \ldots, i - 1$. This is obtained by all previous strains adopting the minimizing transmissibility of $T_j^* = 0$ and the ith strain adopting $T_i = 1$.

B. Invasion threshold

The $\mathcal{R}_0$ value is defined as the number of new infections caused by an infected individual. When $\mathcal{R}_0 < 1$, the epidemic fails to infect a significant portion of the network; the GCC comprises $\mathcal{O}(1)$ nodes. When $\mathcal{R}_0 = 1$, the probability that an epidemic infects a macroscopic fraction of the network, $\mathcal{O}(V)$, is nonzero, where V is the number of nodes in the network. This point is also the critical point in bond percolation that marks the smallest value of T that can form a GCC. Within our model, there is an $\mathcal{R}_{0,i}$ value and a critical transmissibility

for each strain. This critical point is a function of both the network topology and the transmissibilities of the previous strains. If the transmissibility of a particular strain is below its critical threshold, it only infects $\mathcal{O}(1)$ nodes before it burns out. Therefore, in the following analysis, we assume that the transmissibility of each strain is greater than its minimum threshold.

The critical point for the ith percolation can be found by applying linear stability analysis on the system given in Eq. (10) around the fixed point $u_i = 1$. This is the point at which the fixed point in u_i bifurcates into two solutions and $\mathcal{A}_i$ becomes nonzero. Performing a Taylor expansion around $\epsilon = 1 - u_i$ we find the following condition:

$$\epsilon_i = 1 - \frac{1}{\prod_j u_j} \frac{\partial G_1(g_i)}{\partial g_i} \frac{\partial g_i}{\partial u_i}\bigg|_{u_i=1} \epsilon_i + (\mathcal{O}^2), \qquad (14)$$

where the derivatives are given by

$$\frac{\partial g_i}{\partial u_i} = T_i \prod_{j=1}^{i-1} u_j. \qquad (15)$$

When evaluated at $u_i = 1$, $G_1'(g_i)$ becomes $G_1'(g_{i-1})$ from Eqs. (8) and (9). The critical transmissibility is found to be

$$T_{i,c} = \frac{1}{G_1'(g_{i-1})}. \qquad (16)$$

Thus, the minimum transmissibility required for each strain to create an epidemic is a function of the network topology and the transmissibilities of the preceding strains. Given that the coefficients of $G_1'(x)$ are non-negative and therefore monotonically increasing on the positive real line (within its radius of convergence), and that $g_j \in [0, 1]$ (since it is a probability) and because $T_j, u_j \leqslant 1$, then it follows that $G_1'(g_i) \leqslant G_1'(g_{i-1})$ $\forall i$. This indicates [from Eq. (14)] that $T_{i,c} \geqslant T_{i-1,c}$. In other words, the epidemic threshold of each strain increases with each generation. This is an intuitive result, since, as each strain passes through the network, nodes with higher degree are preferentially embedded into the GCC of that strain. Therefore, the RG is increasingly composed of lower degree nodes as it fractures with each iteration of the percolation.

Following [5] we can also prove a stronger condition on the minimum bond occupation probability that a subsequent strain must have in order to exhibit an epidemic on the network. It happens that each generation of the disease must have an increasingly higher transmissibility than the last in order to infect $\mathcal{O}(V)$ nodes in the RG. To see this, we note that $g_i(\boldsymbol{T}, \boldsymbol{u})$ is the probability that an edge fails to connect a node to the GCC of the ith epidemic and that this probability can only decrease or stay constant as T_i increases; this implies that $dT_i/dg_i \leqslant 0$. Inverting this quantity such that $T_i = T_i(\boldsymbol{u}, g_i, \boldsymbol{T}\backslash\{T_i\})$, where the notation $\boldsymbol{S}\backslash\{s\}$ excludes element s from set $\boldsymbol{S}$, and performing the derivative we have an expression that involves T_i and $G_1'(g_{i-1})$. This can then be isolated and it can be shown that $T_{i,c} \geqslant T_{i-1}$ $\forall i \in [1, N]$. For example, the critical point of strain 2 is known [4,5] to be greater than the transmissibility of strain 1, T_1. The critical point of strain 3 is given by $T_{3,c} = 1/G_1'(g_2)$ from Eq. (14),

which through the above prescription satisfies

$$T_{3,c} \geqslant \frac{g_1 - g_2}{1 - u_2} = T_2. \qquad (17)$$

This logic can be applied to all adjacent strains to create an ordered set of critical transmissibilities $\{T_{1,c} \leqslant T_{2,c}, \dots, T_{n,c}\}$. This indicates that transmissibility must evolve to increase in order for a given strain to create an epidemic in the presence of others.

The coexistence threshold T_x was defined in [4,5] for two pathogens and marks an additional phase transition in the model. For $N = 2$ it is the largest value of T_1 that still allows the RG to retain sufficient connectivity to support its own GCC for future strains. For instance, when $T_1 > T_x$, the RG fails to be globally connected and strain 2 fails to infect $\mathcal{O}(V)$ nodes even if $T_2 > T_{2,c}$. For our purpose we extend the definition of the coexistence threshold, $T_{i,x}$, in the context of N sequential strains to be the largest transmissibility of strain i that allows the RG to support a GCC for future generations, assuming that they are sufficiently transmissible. Thus, $T_{i,x}$ is a function of the bond occupancy probabilities of all previous percolations, $T_{i,x} = T_{i,x}(\boldsymbol{u}, \boldsymbol{T}\backslash\{T_i\})$. As for $N = 2$ [4,5], Eq. (14) implicitly defines the coexistence threshold of the ith strain and we find that $T_{i,x}$ is the value of T_i for which $G_1'(g_i) = 1$. For instance, for an Erdős-Renyi degree distribution this condition becomes

$$\frac{1}{\langle k \rangle} \ln\left[u_i \prod_{j=1}^{i-1} u_j \right] = g_i - 1, \qquad (18)$$

from which we can solve for $T_{i,c}$ by inverting g_i. For $i = 1$ we have

$$T_{1,x} = \frac{\ln(u_1)}{\langle k \rangle} \frac{1}{(u_1 - 1)}. \qquad (19)$$

In Fig. 2 we plot $\mathcal{A}_i$ for $N = 5$ against $T_1 \in [0, 1]$ and $T_2 = 0.35$, $T_3 = 0.5$, $T_4 = 1.0$, and $T_5 = 1.0$ for a Erdős-Renyi random graph with mean degree $\langle k \rangle = 4$ and $V = 35\,000$ nodes. We observe excellent agreement between experimental bond percolation (scatter points) and the analytical results of Eq. (11) (plotted lines). Below the epidemic threshold of the first strain, $T_1 < T_{1,c}$, strain 1 does not exhibit a GCC. Hence, the RG is large enough to enable the subsequent strains to form their own GCCs, each consuming more of the available space. With $T_4 = 1$, the last edges in the RG are occupied and, despite a supercritical T_5, we have $\mathcal{A}_5 = 0$. Strain 4 is bimodal, exhibiting two turning points as a function of T_1. This is because, at the first turning point, the transmissibility of the previous strains is sufficient to form their own large GCCs in the RG; however, as T_1 increases, strains 2 and 3 fall below their critical thresholds, allowing strain 4 to consume the available sites into its own GCC. The outbreak size of strain 4 then falls to zero through a final turning point as the transmissibility of strain 1 is increased beyond the coexistence threshold. The inset figure shows the total fraction of the network that has become infected, $\mathcal{A}$, versus T_1. Against intuition, the largest fraction of the network that is occupied by (any) disease is not constant. To see this, we understand that the early generations of the disease consume the high degree sites. As these become embedded within the GCC,
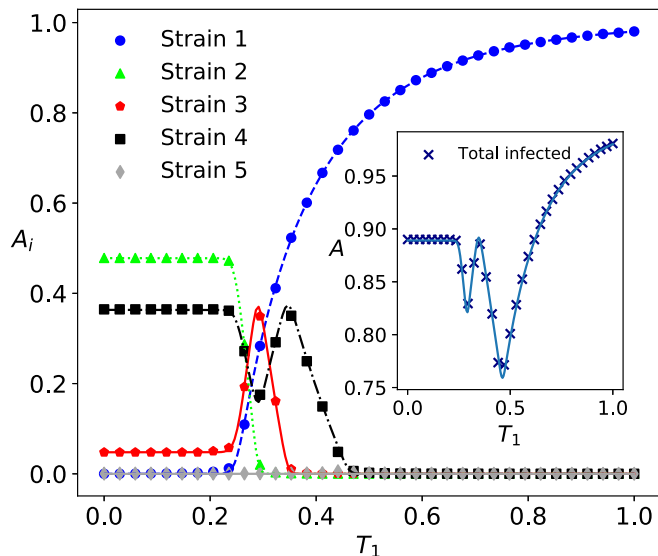
FIG. 2. The outbreak fractions for five generations of the competitive branching process as a function of T_1. Solid lines are the theoretical results from Eq. (11) while scatter points are the average of 50 repetitions of bond percolation over a network with $V = 35\,000$ nodes. The inset shows the total number of infected nodes.

those nodes they connect to can become isolated and thus cannot be incorporated in subsequent GCCs (see Fig. 3). The effect of this is prominent at the onset of the GCC of strain 1 and leads to a local minimum in the total infecteds. Therefore, a disease of low transmissibility early on not only consumes nodes into its own GCC, but also reduces the accessible sites by removing the infection pathways. As the transmissibility of the initial pathogen increases, this effect is reduced and the total infected fraction of the network increases to $\mathcal{A} \approx 0.9$. The global minimum at $T_1 \approx 0.45$ coincides with the coexistence threshold for strain 1; we observe the inability of subsequent strains to create their own epidemic. At this point, strain 1 is sufficiently transmissible to fracture the RG such that it can no longer support a GCC for the other generations of the disease. Beyond this point the total infected fraction follows the number of infected nodes of strain 1.

In a second experiment, we examine the critical points of the first five strains (see Fig. 4) versus disease-1 transmissibility. To do this, we set the transmissibility of each strain to its critical point, $T_{i,c}$, perturbed by a small parameter, δ, such that $T_i = T_{i,c} + \delta$ with $\delta = 1 \times 10^{-4}$. From the logic above, this scenario represents the best-case for the ith strain to form a GCC subject to the constraint that each previous disease did indeed form an epidemic. When $T_1 < T_{1,c}$ the epidemic threshold of the second strain is equal to $1/\langle k \rangle$. As the first process forms a GCC, the substrate RG available to the second percolation process fractures, resulting in an increase in the critical point for the second strain. When there are one or more GCCs in the network, the critical points of subsequent strains are increased. This effect is most significant at the formation of the first GCC, indicating that the RG suddenly "fails" to be well connected. In the presence of multiple GCCs, the epidemic threshold for subsequent strains, $i > 2$, is only slightly higher than the last.
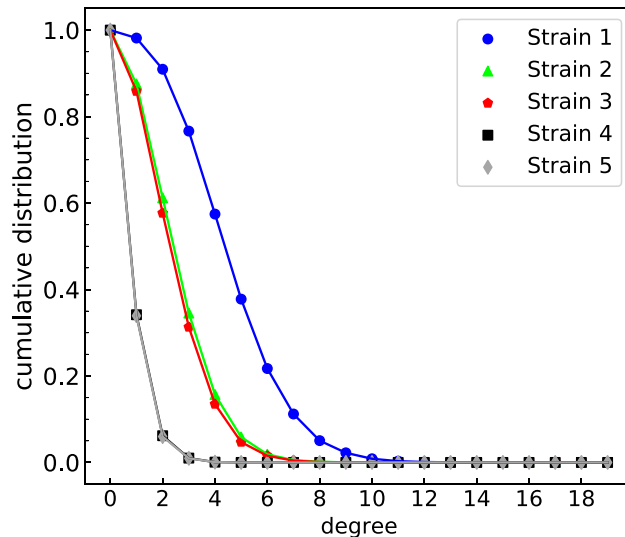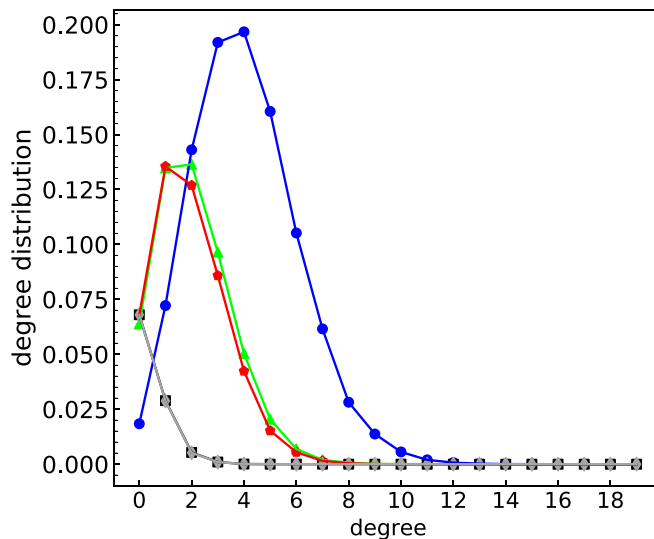
FIG. 3. Top: The degree distribution of the RG created once each strain from Fig. 2 has spread over the network at $T_1 = 0.0$; bottom: the corresponding cumulative degree distribution. Scatter points are the average of 50 repetitions of $N = 35\,000$ Erdős-Renyi networks with $\langle k \rangle = 4$. Curves are the result of Eqs. (3) and (4) acting on the generating functions for the residual graphs of each generation. These plots show how the RG becomes increasingly fractured with each percolation.

In Fig. 5 we examine the outbreak size predicted by the model for the first $N = 75$ strains on Erdős-Renyi graphs as a function of mean degree, finding that the expected outbreak size increases with $\langle k \rangle$. In similar fashion to before, the transmissibility of each strain is set equal to $T_i = T_{i,c} + \delta$. In the left figure, we see $\mathcal{A}_i$ monotonically decreasing with generation index i. Thus, while the transmissibility evolves to increase, the number of infected hosts is expected to decrease with each season of the disease. We numerically solve Eq. (18) to obtain the coexistence threshold and analytically solve Eq. (14) to find the epidemic threshold of the ith generation (Fig. 5 right). The two curves for the first strain are plotted in red, while subsequent strains are lighter grey until the final generation, which has been highlighted green. We

FIG. 4. The critical points of the first five strains of the competitive percolation process as a function of strain 1 transmissibility from Eq. (14). The transmissibility of each strain is set to the critical point plus 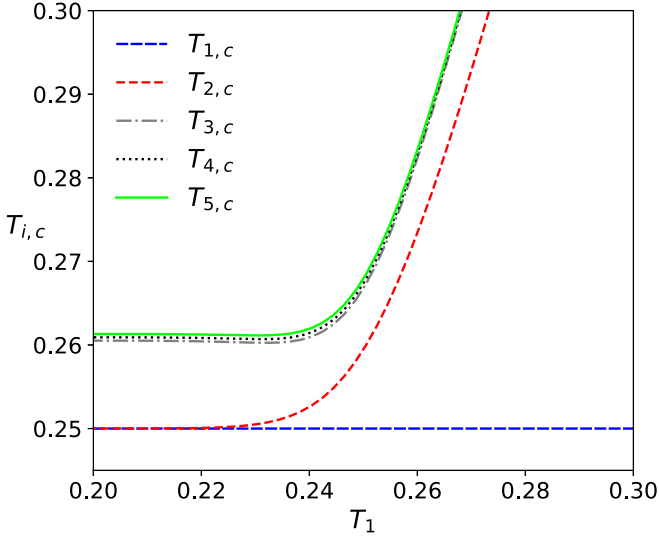a small parameter such that $T_i = T_{i,c} + \delta$ with $\delta = 1 \times 10^{-4}$. The networks are Erdős-Renyi graphs with mean degree $\langle k \rangle = 4$.

observe that both critical points of the model occur at lower transmissibilities with increasing $\langle k \rangle$ as supported by [4] for $N = 2$. The coexistence threshold broadens to a larger extent than the epidemic threshold with increasing i. In other words, $|T_{i-1,x} - T_{i-1,c}| \leqslant |T_{i,x} - T_{i,c}|$, indicating that strain coexistence occupies an increasingly larger area of the model's phase space with greater i.
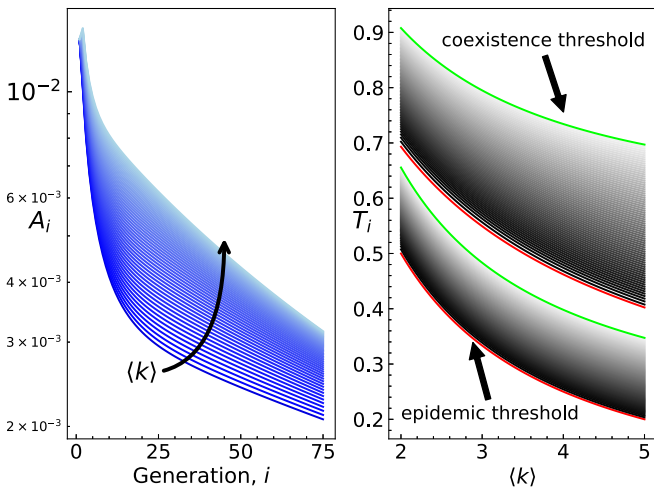

FIG. 5. Left: Theoretical results for the outbreak size of the first $N = 75$ generations of the competitive percolation for Erdős-Renyi graphs with varying mean degree $\langle k \rangle$. Right: The evolution of the epidemic threshold and the coexistence threshold for each generation i. The lines for each subsequent strain are plotted from dark to light grey while the curves belonging to strain 1 have been colored red and those of strain $N = 75$ are highlighted in green.

IV. COLLABORATIVE BRANCHING PROCESS

In this section we define the ith generation of a collaborative branching process as a bond percolation process occurring on the GCC that is created by the $i - 1$ previous processes for $i = 1, \ldots, N$. We impose the strict requirement that only nodes in the GCC created by all of the previous generations are included at the ith generation see Fig. 1. This process has been studied previously by Newman and Ferrario when $N = 2$ [6] as well as for clustered and modular networks [9]. Within the context of the SIR equivalence, this model studies the ability of the ith disease to become an epidemic given that coinfection with all other $i - 1$ strains is a prerequisite for infection with the current strain. If a node fails to become infected with a particular strain, then it cannot become infected with further generations of the disease; therefore, strains that have a low transmissibility significantly reduce the pool of available nodes for future outbreaks.

To describe the model we index the generations $i \in [1, N]$ as before. Consider a node in the GCC after the first disease has reached its equilibrium, but before the second has emerged. There are two additional types of neighbor that this focal node can be surrounded by: infected (but not by the focal node) and, finally, nodes that were directly infected by the focal node in addition to the uninfected substrate nodes. We have to distinguish the particular subset of infected neighbors which the focal node directly infected from those that where infected by one of their other neighbors because the probabilities associated with each state are distinct from one another.

Now let the second disease emerge, spread, and reach its absorbing state, forming its own GCC within the GCC of the first outbreak. Both kinds of infected neighbors from the first process (directly and indirectly infected) give rise to two additional states corresponding to being indirectly or directly infected with strain 2 given direct or indirect infection with strain 1. Therefore, there are now seven possible neighbor states that might surround the focal vertex in the GCC of the second process (including those already present following strain 1 if further infection with strain 2 did not occur). The third process leads to 15 total neighbor states and each one accounts for the specific details of who infected who at each generation. We term a specific combination of direct and indirect infection through the previous generations as an infection history i_h.

The number of new neighbor states at each generation of the branching process is equivalent to the number of leaves of a perfect binary tree, with the total number of neighbor states being the total number of nodes in the tree. This is because each generation branches the current number of maximally coinfected states by a factor of 2 (accounting for directly and indirectly infected neighbors of the coinfected nodes in generation $i - 1$). This means that the total number of neighbor states, η_i, in the ith generation is

$$\eta_i = 1 + \sum_{j=1}^{i} 2^j, \tag{20}$$

comprising all of the states in the previous generation (which did not contract the ith strain) in addition to 2^{i-1} new states that are directly and indirectly infected, plus the uninfected

nodes from the previous generation. Thus, the set of all infection histories for a given generation i, $\{h\}_i$, has cardinality 2^{i-1}; therefore, each generation requires 2^{i-1} new u_{i_h} values, with $i_h \in \{h\}_i$, $h = 1, \ldots, 2^{i-1}$. A visualization of this process is provided in the Appendix.

A. Outbreak size

The aim of this section is to define a prescription to obtain the outbreak size of the ith epidemic of the coinfection model. Each generation requires 2^{i-1} unique u_{i_h} values to be written, each accounting for a particular infection history that a neighboring node could have. Each u_{i_h} value will then be generated by a self-consistent expression in a similar way to those of the competitive model as

$$u_{i_h} = \frac{G_1(\mathcal{P}_{i_h})}{\mathcal{Q}_{i_h}}, \tag{21}$$

where $\mathcal{P}_{i_h}$ is the probability of not obtaining strain i for infection history i_h and $\mathcal{Q}_{i_h}$ is the prior probability that the neighbor has infection history i_h. Note that $\mathcal{P}_{i_h}$ is analogous to g_i from the competitive model. Thus, it remains to calculate both the prior probabilities and $\mathcal{P}_{i_h}$. It happens, for a given generation i, that the probabilities $\mathcal{P}_{i_h}$ can be factored; thus, each $\mathcal{P}_{i_h}$ expression comprises two parts: a multiplying common factor, $\mathcal{C}_i$, and the unique part of the probability of each specific infection history, $\mathcal{H}_{i_h}$, such that

$$\mathcal{P}_{i_h} = \mathcal{C}_i \mathcal{H}_{i_h}. \tag{22}$$

First, we calculate $\mathcal{C}_i$, which is simply all of the common terms belonging to each $\mathcal{P}_{i_h}$, $\forall i_h \in \{h\}_i$. Consider each branch point of the collaborative process from the perspective of an infected node as we progress from generation $j - 1$ to j. Neighbours either do not contract strain j or they do, from either the focal node or one of their other neighbors. Let $f_j(u_{j_h}, v, w)$ be a function that encapsulates the three possible neighbor scenarios and let u_{j_h} be the probability that a neighbor is uninfected by any of its other neighbors by the jth strain at this branch point, given that its infection history is j_h. We have

$$f_j(u_{j_h}, v, w) = \sum_{l=0}^{k} \binom{k}{l} \left[u_{j_h}(1 - T_j) \right]^l \sum_{m_{j_h}=0}^{k-l} \binom{k-l}{m_{j_h}}$$
$$\times \left[(1 - u_{j_h})v \right]^{m_{j_h}} \left[u_{j_h} T_j w \right]^{k-l-m_{j_h}}. \tag{23}$$

We have indexed the variable m with the infection history for later convenience. Despite the complicated form of this expression it is straightforward to construct each term by considering the probabilities associated with each neighboring state. In detail, $u_{j_h}(1 - T_j)$ is the probability that a neighbor was uninfected by its other neighbors given history h and that the focal node did not transmit strain j; $1 - u_{j_h}$ is the probability that a neighbor was already infected and $u_{j_h} T_j$ is the probability that a neighbor was directly infected by the focal node. The arguments v and w are placeholders that allow the further subdivision of the number of neighbors in a given infected state following the next generation.

We construct the common factor $\mathcal{C}_i$ by first constructing a related factor, $\bar{\mathcal{C}}_i$, composing this branch-point logic with itself i times and terminating the composition with $v = w = 1$ at the deepest levels (i.e., the leaves of the branching process) such that

$$\bar{\mathcal{C}}_i = f_1(f_2(\cdots f_i())). \tag{24}$$

The function $\bar{\mathcal{C}}_i$ has 2^{i-1} arguments. The values of j_h in u_{j_h} are given by the particular elements of $\{h\}_j$ $j = 1, \ldots, i - 1$. The common factor for the percolation root (prior to any diseases) is unity, $\bar{\mathcal{C}}_0 = 1$, since all strains belong to the same state. Following strain 1, $\bar{\mathcal{C}}_1$ is given by

$$\bar{\mathcal{C}}_1(1, 1) = f_1(u_{1_1}, 1, 1)$$
$$= u_{1_1}(1 - T_1) + 1 - u_{1_1} + u_{1_1} T_1; \tag{25}$$

for $i = 2$ we have

$$\bar{\mathcal{C}}_2(1, 1, 1, 1) = f_1\big(u_{1_1}, f_2\big(u_{2_1}, 1, 1\big), f_2\big(u_{2_2}, 1, 1\big)\big)$$
$$= u_{1_1}(1 - T_1) + \big(1 - u_{1_1}\big)\big[u_{2_1}(1 - T_2) + 1 - u_{2_1} + u_{2_1} T_2\big] + u_{1_1} T_1\big[u_{2_2}(1 - T_2) + 1 - u_{2_2} + u_{2_2} T_2\big]; \tag{26}$$

similarly, for $i = 3$ we have

$$\bar{\mathcal{C}}_3(\mathbf{1}) = f_1\big(u_1, f_2\big(u_{2_1}, f_3\big(u_{3_1}, 1, 1\big), f_3\big(u_{3_2}, 1, 1\big)\big), f_2\big(u_{2_2}, f_3\big(u_{3_3}, 1, 1\big), f_3\big(u_{3_4}, 1, 1\big)\big)\big), \tag{27}$$

and so on. An interesting observation is that this expression is always unity and that there are always as many terminating 1's as there are unique infection histories required for the next generation. We must define another related probability, $\bar{f}_i$, as

$$\bar{f}_j(u_j, v, w) = \sum_{l=0}^{k} \binom{k}{l} \left[u_j(1 - T_j) \right]^l \sum_{m=0}^{k-l} \binom{k-l}{m}$$
$$\times \left[(1 - u_j)(1 - T_j)v \right]^m \left[u_j T_j w \right]^{k-l-m}, \tag{28}$$

which is equivalent to the $\bar{g}_i$ expression from Sec. III and is the probability that none of the externally infected neighbors transmitted their infection to the focal node. Given these two

functions, we can now build the common terms in the probability that the focal node does not contract the ith strain as

$$\mathcal{C}_i = \bar{\mathcal{C}}_{i-1}(\bar{\boldsymbol{f}}_i). \tag{29}$$

Thus, $\mathcal{C}_i$ contains the common terms in the probability that describes the neighboring states prior to strain i, along with the probability that each of those states then fails to transmit strain i itself. For example, the first few values of $\mathcal{C}_i$ are

$$\mathcal{C}_1 = \bar{f}_1(u_1, 1, 1), \tag{30a}$$
$$\mathcal{C}_2 = \bar{\mathcal{C}}_1\big(\bar{f}(u_{2_1}, 1, 1), \bar{f}(u_{2_2}, 1, 1)\big), \tag{30b}$$

$$\mathcal{C}_3 = \bar{\mathcal{C}}_2\big(\bar{f}_3(u_{3_1}, 1, 1), \bar{f}_3(u_{3_2}, 1, 1),$$
$$\bar{f}_3(u_{3_3}, 1, 1), \bar{f}_3(u_{3_4}, 1, 1)\big). \quad (30c)$$

With a clear prescription to derive $\mathcal{C}_i$ for each generation, we must now calculate the probability associated with each infection history $\mathcal{H}_{i_h}$ in order to finalize the expressions for $\mathcal{P}_{i_h}$, which in turn we require in order to write self-consistent expressions for each u_{i_h} value in Eq. (21). To do this, we consider each pathway from the percolation root to the leaves of the tree created by the collaborative branching process. If, at a particular branching point, we progress via direct infection, we require that the focal node was the node that transmitted infection to the neighbor. For this to occur we require that the neighbors other than the focal node failed to transmit their infection. This occurs with probability $(1 - T_j)^{m_r}$, with reference to Eq. (23). Similarly, the probability that the neighbor *was* infected by a node other than the focal node is $1 - (1 - T_j)^{m_r}$. We now see the utility of subscripting m in Eq. (23) as it allows us to track each particular set of externally infected neighbors over the arguments of $\mathcal{C}_i$. For instance, there are two infection histories at the start of the second process: strain-1 infected nodes have either been externally infected, 2_1, or directly infected by the focal node, 2_2, such that $\{h\}_2 = \{2_1, 2_2\}$. The number of externally 1-infected neighbors is given by m_{1_1} and so we have

$$\mathcal{H}_{2_1} = [1 - (1 - T_1)^{m_{1_1}}], \quad (31)$$

$$\mathcal{H}_{2_2} = (1 - T_1)^{m_{1_1}} \quad (32)$$

Similarly, a node can obtain strain 3 from one of four different neighbor states: externally-1 and externally-2 infected (3_1); externally-1 and directly-2 infected (3_2); directly-1 and externally-2 infected (3_3), or, finally, directly-1 and directly-2 infected (3_4). Thus, there are four infection histories to generate with $\{h\}_4 = \{3_1, 3_2, 3_3, 3_4\}$. We then write

$$\mathcal{H}_{3_1} = \mathcal{H}_{2_1}[1 - (1 - T_2)^{m_{2_1} + m_{2_2}}], \quad (33a)$$

$$\mathcal{H}_{3_2} = \mathcal{H}_{2_1}(1 - T_2)^{m_{2_1} + m_{2_2}}, \quad (33b)$$

$$\mathcal{H}_{3_3} = \mathcal{H}_{2_2}[1 - (1 - T_2)^{m_{2_1} + m_{2_2}}], \quad (33c)$$

$$\mathcal{H}_{3_4} = \mathcal{H}_{2_2}(1 - T_2)^{m_{2_1} + m_{2_2}}. \quad (33d)$$

Each history is constructed from the necessary probabilities to create each scenario. For the next generation, each of these unique infection histories are branched into two to give eight potential sources of strain 4.

With these examples, we now have a prescription to write $\mathcal{P}_{i_h}$ for each potential infection neighboring state. The last component we require in order to calculate the associated u_{i_h} values in Eq. (21) is the prior probabilities that the neighbor was indeed in that particular state following all of the previous strains, but prior to the ith strain itself. It happens that this probability follows a recipe that is simple to compute for each term. We define the following rule: if a neighbor is externally infected at the jth strain, we multiply the prior probability by $(1 - u_{j_h})$; otherwise, for direct infection, we multiply by u_{j_h} instead. The logic behind this rule is simple: to be directly infected by the focal node, a neighbor must not be infected by their other neighbors; the focal node *must* be the successful infection pathway. Therefore, for the first branch point,

the prior probabilities that the neighbor was externally and directly infected, respectively, are given by

$$\mathcal{Q}_{2_1} = (1 - u_{1_1}), \quad (34)$$

$$\mathcal{Q}_{2_2} = u_{1_1}. \quad (35)$$

Similarly following the second strain the prior probabilities for the four infection histories are

$$\mathcal{Q}_{3_1} = \mathcal{Q}_{2_1}(1 - u_{2_1}), \quad (36a)$$

$$\mathcal{Q}_{3_2} = \mathcal{Q}_{2_1}u_{2_1}, \quad (36b)$$

$$\mathcal{Q}_{3_3} = \mathcal{Q}_{2_2}(1 - u_{2_2}), \quad (36c)$$

$$\mathcal{Q}_{3_4} = \mathcal{Q}_{2_2}u_{2_2}. \quad (36d)$$

With this last component we can now construct the self-consistent expressions required to compute the u_{i_h} values in Eq. (21). At this point, a useful check is to ensure that the derived components (priors, histories, and base term) are correct is to set $\bar{f}_i = \mathbf{1}$. When evaluated at unity, the u_{i_h} values should be equal to 1. Both $\mathcal{H}_{i_h}$ and $\mathcal{Q}_{i_h}$ can be constructed by multiplying values associated to the nodes of the binary tree of neighbor states from the root to a given leaf state; this is visualized in the Appendix.

Next, we require the outbreak size of the ith strain, $\mathcal{A}_i$. It happens that this expression is very simple to construct once we have performed the above work; we simply have to take the expression for u_{i_h} for the maximally indirect, maximally coinfected infection history i_X (i.e., the expression for the history where every node state was externally infected), remove the prior denominator, and replace the $G_1(z)$ generating function with a $G_0(z)$ generating function. We then subtract this value from the previous outbreak size such that

$$\mathcal{A}_i = \mathcal{A}_{i-1} - G_0(\mathcal{P}_{i_X}). \quad (37)$$

Since $G_0(\mathcal{P}_{i_X})$ can never be negative, we observe that the outbreak size of each generation can never exceed the size of the previous one. We detail the expressions for the first few generations in the Appendix and show the percolation results for the first four generations spreading over an Erdős-Renyi network in Fig. 6. In Fig. 6(a) we plot the outbreak sizes of each strain; each exhibits a smaller size and a larger percolation threshold with increasing strain index. In Fig. 6(b) the degree distribution of each of the GCC substructures is plotted. The average degree is reduced and the height and variance of each distribution is increasingly reduced and shifted to the left. In Fig. 6(c) the cumulative probability that the degree of a node is larger than k is shown for each strain. These results indicate that eventually the spreading of cooperative processes on Erdős-Renyi graphs will be limited by the fractured topology of the substrate network available to each strain in addition to the transmissibility of the disease.

The complete prescription for solving for the outbreak size of the Nth generation of the cooperative branching process is to hierarchically solve the coupled linear system of equations for each u_{i_j}, $i = 1, \ldots, N$ given by

$$u_{i_j} = u_{i_j}(u_{1_1}, u_{2_1}, \ldots, u_{i_{2^{i-1}}}; T_1, \ldots, T_i) \quad (38)$$

for infection histories $j = 1, \ldots, 2^{i-1}$, and the functional form given by Eq. (21). More details on the structure of these
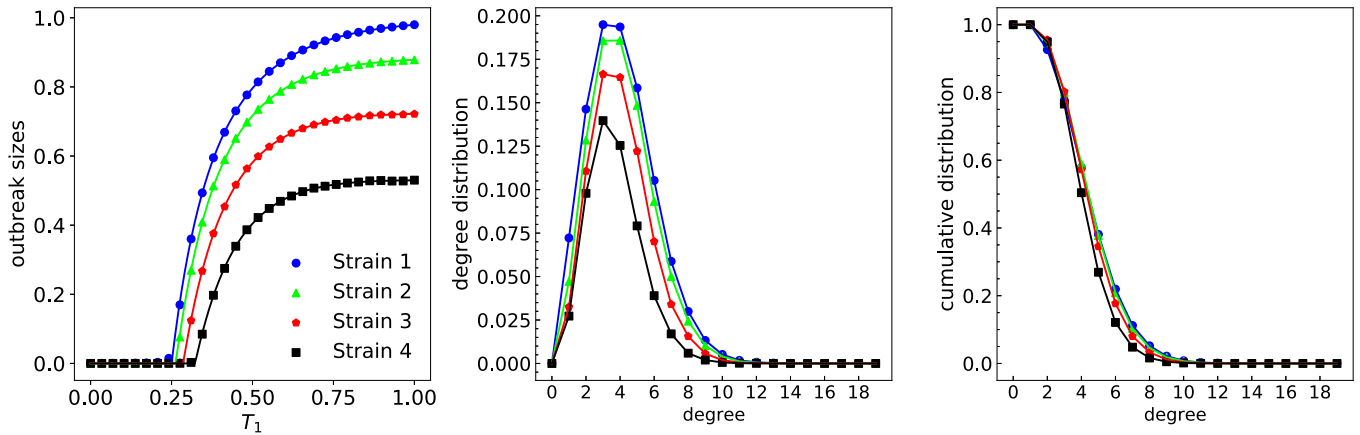
FIG. 6. Four generations of the cooperative branching process with $(T_2, T_3, T_4) = (0.6, 0.5, 0.45)$ and on an Erdős-Renyi network with mean degree $\langle k \rangle = 4$. Scatter points are the average of 35 repeats of Monte Carlo simulations over $V = 30\,000$ node networks; solid lines are the theoretical results. Subplot (a) shows the outbreak sizes; (b) shows the degree distribution at $T_1 = 1$; (c) is the cumulative probability that a node has degree larger than k at $T_1 = 1$.

expressions in terms of a perfect binary tree, as well as an examination of their solutions, are treated in the Appendix.

B. Invasion threshold

In this section we examine the critical points of the co-operative model, generalizing the result of [6] for $N > 2$. As with the competitive percolation, there is an $\mathcal{R}_{0,i}$ value for each generation or strain. If at any point the outbreak size of a generation is subcritical, then there can be no subsequent outbreaks as coinfection is a strong condition on the proliferation of future strains. However, assuming that the previous $i - 1$ strains did indeed cause an $\mathcal{O}(V)$ outbreak, then there is some point $T_{i,c}$ at which the ith strain can also lead to a finite sized propagation if its transmissibility exceeds this value.

The critical point for the ith percolation can be found by applying linear stability analysis on Eq. (21) around the fixed point $\{u_i^h\} = 1$, which is the trivial root of the system of equations for each generation (see the Appendix). The critical point of the first strain is identical to the results from Sec. III; however, it is prudent to review this result. Given that $u_{1_1} = 1$ at the critical point, we perform a Taylor expansion about $\epsilon_{1_1} = 1 - u_{1_1}$ using Eq. (21) and truncate it to first order to obtain

$$\epsilon_{1_1} \approx 1 - G_1(f_1)\big|_{u_{1_1}=1} + G_1'(f_1)f_1'\big|_{u_{1_1}=1}\epsilon_{1_1}. \quad (39)$$

Rearranging this result, with $G_1(1) = 1$ and $f_1' = T_1$, we obtain $T_{1,c} = 1/G_1'(1)$ in accordance with Eq. (16) at $i = 1$. For the second strain, we now have two variables to consider depending on the unique infection history of the neighboring node. The critical point occurs when both u_{2_1} and u_{2_2} are unity (see the Appendix for a graphical motivation of this) and we again perform a first-order Taylor expansion about the small parameter $\epsilon_{2_j} = 1 - u_{2_j}$ to obtain the following coupled system:

$$\epsilon_{2_1} \approx \epsilon_{2_1}\frac{\partial F_{2_1}}{\partial u_{2_1}} + \epsilon_{2_2}\frac{\partial F_{2_1}}{\partial u_{2_2}},$$

$$\epsilon_{2_2} \approx \epsilon_{2_1}\frac{\partial F_{2_2}}{\partial u_{2_1}} + \epsilon_{2_2}\frac{\partial F_{2_2}}{\partial u_{2_2}}, \quad (40)$$

where we have set the functional form of u_{i_j} in Eq. (21) to $u_{i_j} = F_{i_j}$ and evaluate the derivatives at the fixed point $u_{2_1} = u_{2_2} = 1$. The derivatives are

$$\frac{\partial F_{2_1}}{\partial u_{2_1}} = T_2 - G_1'\big(1 - T_1 + u_{1_1}T_1\big)(1 - T_1)T_2, \quad (41)$$

$$\frac{\partial F_{2_1}}{\partial u_{2_2}} = \frac{u_{1_1}T_1}{1 - u_{1_1}}\big[1 - G_1'\big(1 - T_1 + u_{1_1}T_1\big)\big]T_2 \quad (42)$$

and

$$\frac{\partial F_{2_2}}{\partial u_{2_1}} = \frac{G_1'\big(1 - T_1 + u_{1_1}T_1\big)\big(1 - u_{1_1}\big)(1 - T_1)}{u_{1_1}}T_2, \quad (43)$$

$$\frac{\partial F_{2_2}}{\partial u_{2_2}} = G_1'\big(1 - T_1 + u_{1_1}T_1\big)T_1T_2. \quad (44)$$

Thus, we have the following linear system:

$$\boldsymbol{J}\begin{pmatrix} u_{2_1} \\ u_{2_2} \end{pmatrix} = \frac{1}{T_{2,c}}\begin{pmatrix} u_{2_1} \\ u_{2_2} \end{pmatrix}, \quad (45)$$

where $\boldsymbol{J}$ is a Jacobian matrix with eigenvalue $1/T_{2,c}$. Following [6] the system has two eigenvalues, and, by examining the $T_1 \to 1$ limit, the correct eigenvalue is

$$T_{2,c} = \frac{2}{\tau + \sqrt{\tau^2 - 4\Delta}}, \quad (46)$$

where τ is the trace of $\boldsymbol{J}$ and Δ is the determinant.

In the general case we have the following linear system

$$\epsilon_{i_j} \approx \sum_{k=1}^{2^{i-1}}\frac{\partial F_{i_j}}{\partial u_{i_k}}\epsilon_{i_k} \quad (47)$$

with $i \in [1, n]$ and $j \in [1, 2^{i-1}]$. The derivatives are given by

$$\frac{\partial F_{i_j}}{\partial u_{i_k}} = \frac{G_1'(\mathcal{P}_{i_j})}{\mathcal{Q}(i_j)}\frac{\partial\bar{\mathcal{C}}_{i-1}\mathcal{H}_{i_j}}{\partial\bar{f}_i}\frac{\partial\bar{f}_i}{\partial u_{i_k}}\bigg|_{u_{i_k}=1}. \quad (48)$$

The derivative of the final $\bar{f}_i$ term is always $\partial\bar{f}_i = T_i$, meaning that we have a leading factor of T_i multiplying all terms. Thus, we can create the following linear system by simple

rearrangement,

$$\boldsymbol{J}\vec{u} = \frac{1}{T_{i,c}}\vec{u}, \tag{49}$$

where $\vec{u} = \{u_{1_1}, u_{2_1}, \ldots, u_{i_{2^{i-1}}}\}^{\mathsf{T}}$ and $\boldsymbol{J}$ is a Jacobian matrix with elements $\partial_{ik} F_{i_j}/T_i$ evaluated at the fixed point $\boldsymbol{u} = \{1, 1, \ldots, 1\}$. We then find the eigenvalues by solving $\det(\boldsymbol{J} - \frac{1}{T}\mathbb{I}) = 0$, where $\mathbb{I}$ is the identity matrix. The characteristic polynomial of an $n \times n$ matrix can be expressed in terms of powers of the trace; however, roots of polynomials of degree 5 or more are unlikely to yield a closed form solution in general.

V. DISCUSSION

A. $T_i = 1$ limit

Consider the competitive branching process for $N = 2$ strains with $T_1 = T_2 = 1$. The first percolation process occupies the entire GCC in the substrate network such that $\mathcal{A}_1 = 1 - G_0(u_1)$. The outbreak size of the second strain is zero since the fixed point of $\lim_{T_1,T_2 \to 1} u_2 = G_1(u_1 u_2)/u_1$ is unity and $\mathcal{A}_2 = G_0(u_1) - G_0(u_1 u_2)$. This fixed point occurs for all subsequent u_i values for the competitive process and hence all future generations fail to create a finite sized outbreak size.

Consider next the collaborative branching process of N strains with $T_i = 1$ for $i = 1, \ldots, N$. Intuitively, we know that the subsequent generations will occupy all of the edges in the GCC of the first process (the size of which is in turn a function of the topological properties of the substrate network). Thus, we expect

$$\mathcal{A}_1 = \mathcal{A}_i, \quad \forall i \in [1, N]. \tag{50}$$

With reference to Eq. (37) this indicates that $G_0(\mathcal{P}_{i_X}) = 0$, where i_X corresponds to the maximally external-coinfected history. This can be seen from the $T_1, T_2 \to 1$ limit of the outbreak size,

$$\mathcal{A}_2 = \mathcal{A}_1 - G_0\big((1-u_{1_1})u_{2_1} + u_{1_1}u_{2_2}\big) + G_0\big(u_{1_1}u_{2_2}\big) \tag{51}$$

with $\lim_{T_1,T_2 \to 1} u_{2_1} \to 0$ and $\lim_{T_1,T_2 \to 1} u_{2_2} \to 1$. For additional generations under the same limit, we observe that all u_{i_h} values have a fixed point at 0 apart from the history corresponding to the maximally directly coinfected state, which limits to 1; therefore, this result holds for all generations.

B. Scale − free networks

Let us now consider the application of the competitive and collaborative models to the specific example of disease spreading on scale-free networks. Scale-free networks are good representations of the heavy tailed degree distributions found in many empirical networks [1,18]; including human contact networks. Let the degree distribution be given by a power law of exponent $2 < \alpha \leqslant 3$ with an exponential degree cutoff of κ such that

$$p_k = \begin{cases} 0, & k = 0, \\ Ck^{-\alpha}e^{-k/\kappa}, & k \geqslant 1, \end{cases} \tag{52}$$

where C is a constant required to normalize the probability distribution to unity. Using the definition of the polylogarithm

$$\mathrm{Li}_\alpha(z) = \sum_{k=0}^{\infty} \frac{z^k}{k^\alpha} \tag{53}$$

we can write the $G_0(z)$ generating function for this degree distribution as

$$G_0(z) = \frac{\mathrm{Li}_\alpha(ze^{-1/\kappa})}{\mathrm{Li}_\alpha(e^{-1/\kappa})}, \tag{54}$$

while for the excess degree distribution we have

$$G_1(z) = \frac{\mathrm{Li}_{\alpha-1}(ze^{-1/\kappa})}{z\mathrm{Li}_{\alpha-1}(e^{-1/\kappa})}. \tag{55}$$

Considering the competitive percolation process, the critical point for the ith strain is found by applying Eq. (16) to this degree distribution, which gives

$$T_{i,c} = \frac{\mathrm{Li}_{\alpha-1}(e^{-1/\kappa})}{g_{i-1}}\bigg[\mathrm{Li}_{\alpha-2}(g_{i-1}e^{-1/\kappa}) \\ - \frac{1}{g_{i-1}}\mathrm{Li}_{\alpha-1}(g_{i-1}e^{-1/\kappa})\bigg]^{-1}. \tag{56}$$

Thus, with $g_0 = 1$, the critical point of the first strain of the competitive process reduces to the expression found by Newman [1],

$$T_{1,c} = \frac{\mathrm{Li}_{\alpha-1}(e^{-1/\kappa})}{\mathrm{Li}_{\alpha-2}(e^{-1/\kappa}) - \mathrm{Li}_{\alpha-1}(e^{-1/\kappa})}. \tag{57}$$

In the high-degree limit with $\kappa \to \infty$, the epidemic threshold of strain 1 occurs at $T_{1,c} = 0$, since the second moment of the degree distribution diverges [19]. However, the critical point of strain 2 is nonzero and is given by

$$\lim_{\kappa \to \infty} T_{2,c} \approx \frac{1}{\zeta(\alpha)} \frac{\mathrm{Li}_{\alpha-1}(g_1)}{g_1}, \tag{58}$$

where $\zeta(\alpha)$ is the Riemann ζ function. This indicates the rapid fracture of the RG by the first generation and the subsequent loss of the scale-free property [4]. This is because the GCC of the first competitive process targets the high-degree sites of the network, removing the power-law distribution of degrees. In the limit that $T_1 \to 1$, Eq. (58) diverges, exceeding the coexistence threshold, and strain 2 fails to spread on the RG. In Fig. 7 (bottom) we show the cumulative degree distribution (top) and the outbreak sizes (bottom) of the first five strains of the competitive process on a scale-free network with power-law exponent $\alpha = 2$ and degree cutoff $\kappa = 20$ for transmissibilities $(T_2, T_3, T_4, T_5) = (0.3, 0.5, 0.6, 1.0)$. In this experiment, the total infected fraction of the network is fairly constant, ≈ 0.5; however, the outbreak sizes of each strain are strong, multimodal functions of the available space left to spread in the RG of the previous strains. These disease dynamics show that each generation of a disease can generate a finite GCC of its own in the presence of other outbreaks. Figure 7 (top, inset) shows the same data as the main panel on a semilog plot. Each generation leads to an increased fracturing of the network and a transition from scale-free behavior (linearity on a log-log plot) to the emergence of an exponential
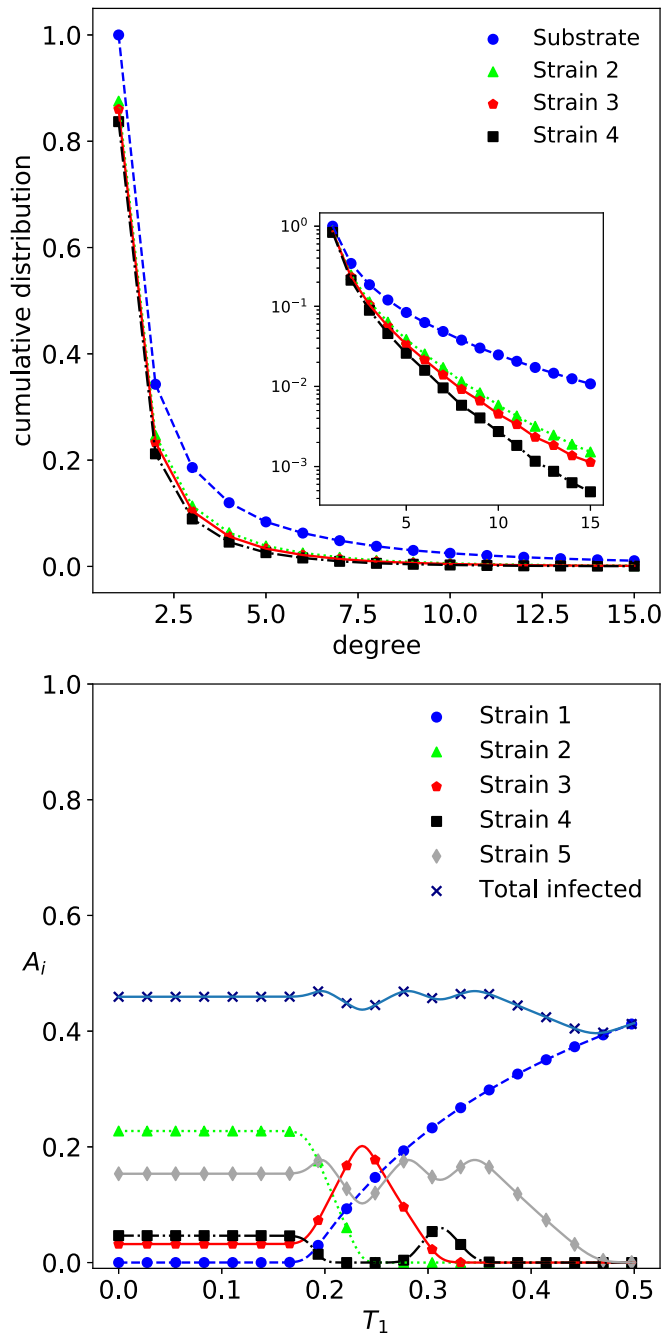
FIG. 7. Top: The cumulative degree distribution of nodes in the substrate network and the GCCs of the RG following competitive bond percolation at $T_1 = 0.0$. Bottom: The outbreak sizes of the competitive branching process. The parameters are $(T_2, T_3, T_4, T_5) = (0.3, 0.5, 0.6, 1.0)$ on a scale-free network with power-law exponent $\alpha = 2$ and $\kappa = 20$. Scatter points are the average of 35 repeats of Monte Carlo simulations over $V = 30\,000$ node networks; solid lines are the theoretical results.

relationship (linearity on a semilog plot) with increasing strain index.

Turning now to the cooperative process, the general expression for the critical transmissibility of strain 2 was given in [6] and is shown in Eq. (46). For the scale-free degree distribution

we have the trace and determinant given by

$$\tau = 1 + (2T_1 - 1)\frac{\text{Li}_{\alpha-1}(e^{-1/\kappa})}{\bar{f}_1}$$

$$\times \left[\text{Li}_{\alpha-2}(\bar{f}_1 e^{-1/\kappa}) - \frac{1}{\bar{f}_1}\text{Li}_{\alpha-1}(\bar{f}_1 e^{-1/\kappa}) \right]^{-1}$$

and

$$\Delta = T_1^2 \frac{\text{Li}_{\alpha-1}(e^{-1/\kappa})}{\bar{f}_1}$$

$$\times \left[\text{Li}_{\alpha-2}(\bar{f}_1 e^{-1/\kappa}) - \frac{1}{\bar{f}_1}\text{Li}_{\alpha-1}(\bar{f}_1 e^{-1/\kappa}) \right]^{-1}$$

respectively. At $\kappa \to \infty$, both $\tau \to \infty$ and $\Delta \to \infty$ and so $T_{2_c} = 0$ from Eq. (46). Since the critical point of subsequent strains will always contain this factor, their critical points also vanish in this limit. Thus, cooperative diseases on scale-free networks will theoretically always spread, regardless of measures that lower the strain transmissibility. In Fig. 8 we show the percolation properties of the first four strains of the cooperative branching process on a scale-free network with power-law exponent $\alpha = 2.0$ and degree cutoff $\kappa = 20$; the transmission probability of each strain is $(T_2, T_3, T_4) = (0.6, 0.5, 0.45)$. Figure 8(a) shows the outbreak fractions become successively smaller with increasing generations. In Fig. 8(b) we see the degree distributions of GCCs formed by each generation of the process at $T_1 = 1.0$. The inset shows (from a repeated experiment with $\kappa = 200$) that the heavy tail of the power law is preserved following each percolation, while the probability of choosing a low degree node becomes increasingly smaller. In Fig. 8(c) the cumulative probability that a node has degree larger than k within the GCC structures shows the removal of the low degree nodes with each generation of the percolation. This is in direct contrast to the fracture process that occurs for Erdős-Renyi networks [see Fig. 6(c)], where the cumulative probability *decreased* with generation index under cooperative percolation. The preservation of the higher-degree core in the scale-free network indicates the self-similarity of the GCC substructures under cooperative percolation.

VI. CONCLUSION

In this paper we have introduced two exact models of generational bond percolation on complex networks. The outbreak sizes as well as the critical points of the models were solved for and topological properties of the resulting graph structures were examined.

Competitive branching processes are defined as a perfect cross-immunity model; each generation spreading on the RG created by the previous generations. It was shown, under the constraint that each disease adopts its critical transmissibility, that there is an evolutionary pressure on each generation to become more transmissible; however, the expected outbreak size becomes increasingly smaller. Collaborative branching processes are defined as a complete coinfection model; each generation spreading on the GCC created by all of the previous strains. It was shown that the outbreak size of each strain is bound by the size of the preceding epidemic.
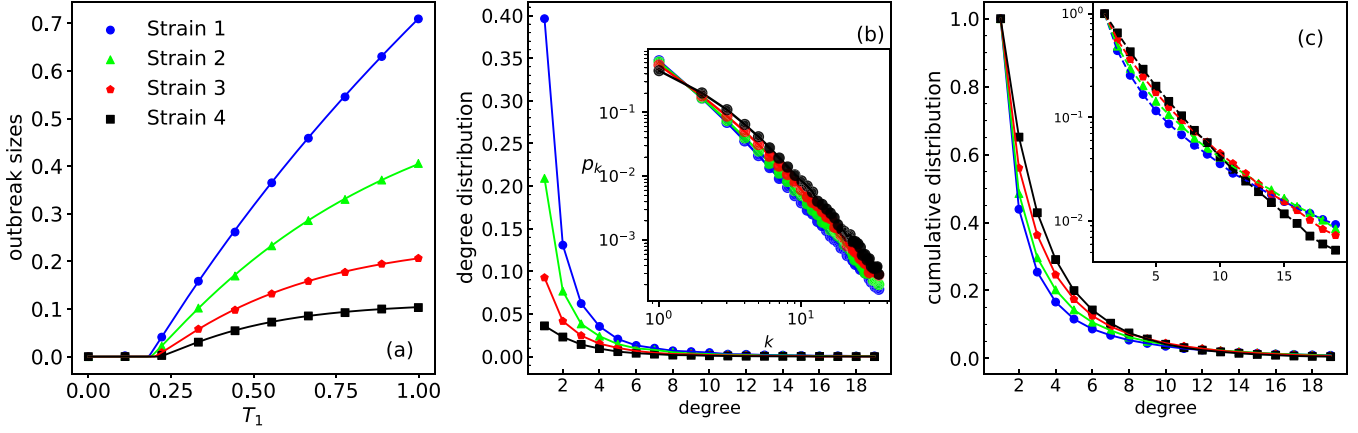
FIG. 8. Four generations of the cooperative branching process with $(T_2, T_3, T_4) = (0.6, 0.5, 0.45)$ on a scale-free network with power-law exponent $\alpha = 2$ and $\kappa = 20$. Scatter points are the average of 35 repeats of Monte Carlo simulations over $V = 500\,000$ node networks; solid lines are the theoretical results. Subplot (a) shows the outbreak sizes, (b) shows the degree distribution at $T_1 = 1$, the inset is a log-log plot with $\kappa = 200$, and (c) is the cumulative probability that a node has degree larger than k in each GCC at $T_1 = 1$; the inset shows the same data on a logarithmic scale for $\kappa = 20$.

Both of the stochastic branching processes studied in this paper increasingly fracture the substrate network as they unfold. This eventually leads to the loss of global connectivity among the nodes of the network and the eventual burnout of each process. We examined the resulting graph structures from both competitive and cooperative percolation types for two common degree distributions: Erdős-Renyi and power-law networks. The RG structures created by the competitive process led to nontrivial outbreak sizes that can exhibit multiple turning points depending on the transmissibility of previous strains. Additionally, for both Erdős-Renyi and scale-free networks the total outbreak size was found to be a nonmonotonic function of the proceeding fracture process.

We found that the GCC structures created through cooperative percolation on Erdős-Renyi networks fracture through a gradual loss of high degree sites. Scale-free networks, in contrast, produce self-similar outbreaks supported by a power-law core.

Whilst these scenarios are illuminating in their own right, this work is an important step towards an N-strain partial immunity model, whereby a strain could infect *all* nodes, regardless of their infection history. Recent work by Mann *et al.* [13] has shown that bond percolation can also be mapped to a partial immunity model for $N = 2$. In this case, the strict criteria concerning the disease couplings required for the *complete* cross-immunity and *complete* coinfection models are relaxed. Importantly, such a model would not burn out. This work can be generalised in many different ways, such as networks with clustering [20], including hypergraphs as well as modular systems or multitype networks [21,22]. These processes could also prove useful to the understanding of other dynamical processes on networks, such as synchronization. Finally, a strong condition on the success of the model is the temporal separation of each generation; we have not addressed any dynamical features of the spreading processes, such as their timescales or growth rates, which we leave for future work.

APPENDIX: OUTBREAK SIZES FOR COLLABORATIVE BRANCHING PROCESSES

In this Appendix we describe how to use the prescription to obtain the outbreak sizes of the first few generations of the collaborative branching process.

1. Strain 1

The prescription for the outbreak size of strain 1 is as follows. First we calculate $\mathcal{C}_1 = \bar{f}_1(u_1, 1, 1)$. A node in the GCC has three neighbor states: uninfected, infected (externally by other nodes), and infected (directly by the focal node); of these, only the two infected states play an active role in the spread of further generations. The probability that an edge between the focal node and an externally infected neighbor failed to transmit strain 1 is $1 - T_1$. Thus, the probability that all m_{1_1} externally infected neighbors failed is $(1 - T_1)^{m_{1_1}}$. The probability that the neighbor was uninfected is then

$$u_{1_1} = \sum_{k=0}^{\infty} q_k \left[u_{1_1}(1 - T_1) + (1 - u_{1_1})(1 - T_1) + u_{1_1} T_1 \right]^k,$$

(A1)

where $q_k = (k + 1)p_{k+1}/\langle k \rangle$ from Eq. (5). With this probability, we now write the outbreak size as

$$\mathcal{A}_1 = 1 - \sum_{k=0}^{\infty} p_k \left[u_{1_1}(1 - T_1) + (1 - u_{1_1})(1 - T_1) + u_{1_1} T_1 \right]^k.$$

(A2)

The bracket reduces to Eq. (8); thus, both competitive and collaborative formulations are in agreement for the first percolation process. The solution u_{1_1} can also be visualized graphically, thus confirming the generating function is convex
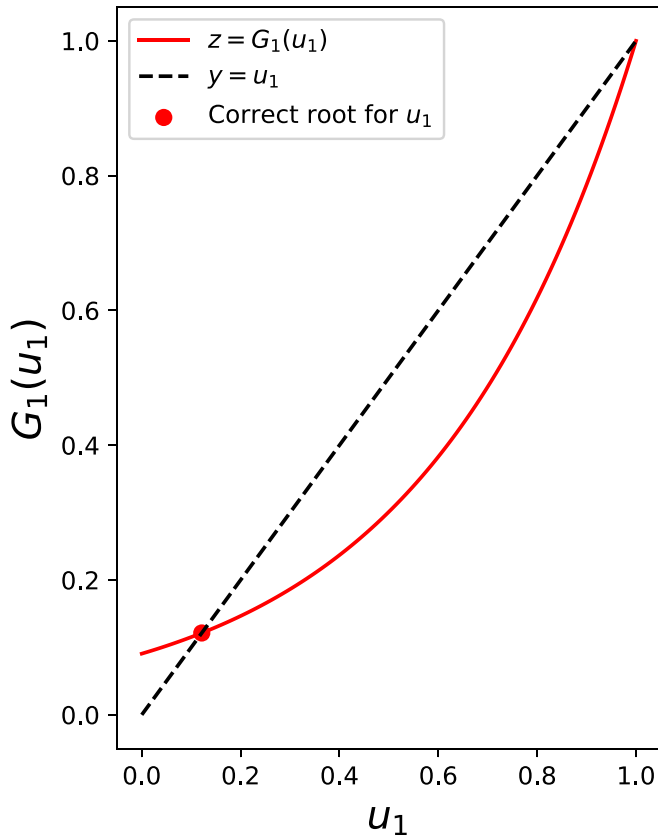
FIG. 9. The graphical solution [15] of the generating function for strain 1 at $T_1 = 0.6$. Plotted are $y = u_{1_1}$ against u_{1_1} as well as the value of the generating function $z_1 = G_1(1 - T_1 + u_{1_1}T_1)$. The intersection of $y = u_{1_1}$ and z_1 corresponds to the root. We notice that the trivial root, u_{1_1}, is also a solution. The scatter point is the result of fixed point iteration.

on the unit interval and that Jensen's inequality can be applied; see Fig. 9.

2. Strain 2

For strain 2 there are two possible infection histories that a neighbor might have: either the focal node infected it directly or it was externally infected. The probabilities that the focal node does not get strain 2 from each of these neighbor states are u_{2_1} and u_{2_2}, respectively. The common factor is given by

$$\mathcal{C}_2 = \bar{\mathcal{C}}_1\big(\bar{f}(u_{2_1}, 1, 1), \bar{f}(u_{2_2}, 1, 1)\big), \tag{A3}$$

which is simply

$$\mathcal{C}_2 = \sum_{l=0}^{k} \binom{k}{l} \big[u_{1_1}(1 - T_1)\big]^l \sum_{m_{1_1}=0}^{k-l} \binom{k-l}{m_{1_1}}$$
$$\times \big[(1 - u_{1_1})\bar{f}(u_{2_1})\big]^{m_{1_1}} \big[u_{1_1}T_1\bar{f}(u_{2_2})\big]^{k-l-m_{1_1}}, \tag{A4}$$

where we have dropped the 1s in the function arguments of the $\bar{f}$ functions. Following the prescription, the history of u_{2_1} is $\mathcal{H}_{2_1} = [1 - (1 - T_1)^{m_{1_1}}]$ while the history of u_{2_2} is $\mathcal{H}_{2_2} = (1 - T_1)^{m_{1_1}}$. Since the u_{2_1} history branches from the $1 - u_{1_1}$ compartment, the prior probability is simply $\mathcal{Q}_{2_1} = 1 - u_{1_1}$, and $\mathcal{Q}_{2_2} = u_{1_1}$. Thus, we have

$$u_{2_1} = \frac{1}{\mathcal{Q}_{2_1}} \sum_{k=0}^{\infty} q_k \mathcal{C}_2 \mathcal{H}_{2_1}, \tag{A5}$$

$$u_{2_2} = \frac{1}{\mathcal{Q}_{2_2}} \sum_{k=0}^{\infty} q_k \mathcal{C}_2 \mathcal{H}_{2_2}. \tag{A6}$$

As we insert the infection histories into this expression, we observe that externally infected histories lead to multiple brackets as the generating function acts on each term. For instance

$$u_{2_1} = \frac{1}{\mathcal{Q}_{2_1}} \sum_{k=0}^{\infty} q_k \mathcal{C}_2[1 - (1 - T_1)^{m_{1_1}}] \tag{A7}$$

$$= \frac{1}{\mathcal{Q}_{2_1}} \sum_{k=0}^{\infty} q_k \mathcal{C}_2 - \frac{1}{\mathcal{Q}_{2_1}} \sum_{k=0}^{\infty} q_k \mathcal{C}_2(1 - T_1)^{m_{1_1}} \tag{A8}$$

$$= \frac{G_1(\mathcal{C}_2)}{\mathcal{Q}_{2_1}} - \frac{G_1(\mathcal{C}_2(1 - T_1)^{m_{1_1}})}{\mathcal{Q}_{2_1}} \tag{A9}$$

$$s = \frac{G_1\big(u_{1_1}(1 - T_1) + (1 - u_{1_1})\bar{f}_2(u_{2_1}) + u_{1_1}T_1\bar{f}_2(u_{2_2})\big)}{1 - u_{1_1}} - \frac{G_1\big(u_{1_1}(1 - T_1) + (1 - u_{1_1})(1 - T_1)\bar{f}_2(u_{2_1}) + u_{1_1}T_1\bar{f}_2(u_{2_2})\big)}{1 - u_{1_1}}. \tag{A10}$$

This does not occur for u_{2_2} since the infection history folds into $\mathcal{C}_2$:

$$u_{2_2} = \frac{G_1\big(u_{1_1}(1 - T_1) + (1 - u_{1_1})(1 - T_1)\bar{f}_2(u_{2_1}) + u_{1_1}T_1\bar{f}_2(u_{2_2})\big)}{u_{1_1}} \tag{A11}$$

We exhibit the graphical solution for these coupled equations in Fig. 10.
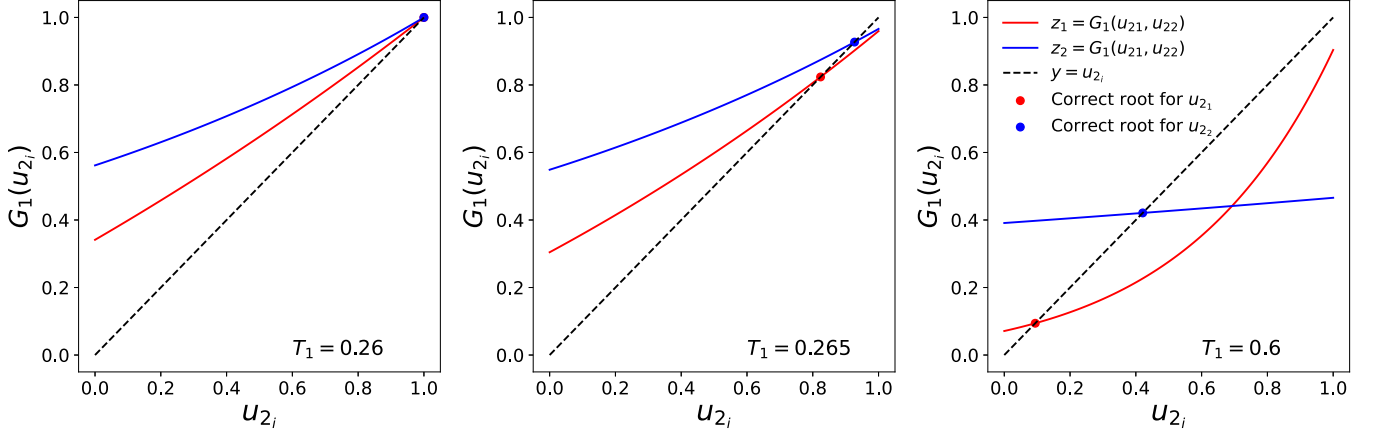
FIG. 10. The graphical solution of the generating functions for strain 2 at three different T_1 values with $T_2 = 0.6$. Plotted are $y = u_{2_i}$ for $i = 1, 2$ against u_{2_i} as well as the value of the generating functions $z_1 = G_1(\mathcal{C}_2\mathcal{H}_{2_1})/(1 - u_{1_1})$ and $z_2 = G_1(\mathcal{C}_2\mathcal{H}_{2_2})/u_{1_1}$. Each z_i varies u_{2_i} while the other value is held fixed at the correct root. The intersection of $y = u_{2_i}$ and z_i corresponds to the root, which is also marked with a scatter point. We notice that the trivial root of $u_{2_1} = u_{2_2} = 1$ is no longer shown as the system moves away from the critical point when the GCC first forms. This is also graphical motivation for finding the critical point from a Taylor series around the trivial root. We also notice the increase (loss) of convexity in u_{2_1} (u_{2_2}) as we increase T_1. This indicates the increasing (decreasing) importance of the u_{2_1} (u_{2_2}) branch to the formation of the GCC at larger transmissibilities.

3. Strain 3

The first step is to write the common factor $\bar{\mathcal{C}}_2(\mathbf{1})$ [see Eq. (30c)]; this is given by

$$
\bar{\mathcal{C}}_2 = \sum_{l=0}^{k} \binom{k}{l} \left[u_{1_1}(1 - T_1)\right]^l \sum_{m_{1_1}=0}^{k-l} \binom{k-l}{m_{1_1}} \left[1 - u_{1_1}\right]^{m_{1_1}} \sum_{a=0}^{m_{1_1}} \binom{m_{1_1}}{a} \left[u_{2_1}(1 - T_2)\right]^a
$$

$$
\times \sum_{m_{2_1}=0}^{m_{1_1}-a} \binom{m_{1_1}-a}{m_{2_1}} \left[1 - u_{2_1}\right]^{m_{2_1}} \left[u_{2_1}T_2\right]^{m_{1_1}-a-m_{2_1}} \left[u_{1_1}T_1\right]^{k-l-m_{1_1}} \sum_{c=0}^{k-l-m_{1_1}} \binom{k-l-m_{1_1}}{c} \left[u_{2_2}(1 - T_2)\right]^c
$$

$$
\times \sum_{m_{2_2}=0}^{k-l-m_{1_1}-c} \binom{k-l-m_{1_1}-c}{m_{2_2}} \left[1 - u_{2_2}\right]^{m_{2_2}} \left[u_{2_2}T_2\right]^{k-l-m_{1_1}-c-m_{2_2}}. \tag{A12}
$$

We then form $\mathcal{C}_3 = \bar{\mathcal{C}}_2(\bar{f}_3(u_{3_1}, 1, 1), \bar{f}_3(u_{3_2}, 1, 1), \bar{f}_3(u_{3_3}, 1, 1), \bar{f}_3(u_{3_4}, 1, 1))$ as

$$
\mathcal{C}_3 = \sum_{l=0}^{k} \binom{k}{l} \left[u_{1_1}(1 - T_1)\right]^l \sum_{m_{1_1}=0}^{k-l} \binom{k-l}{m_{1_1}} \left[1 - u_{1_1}\right]^{m_{1_1}} \sum_{a=0}^{m_{1_1}} \binom{m_{1_1}}{a} \left[u_{2_1}(1 - T_2)\right]^a
$$

$$
\times \sum_{m_{2_1}=0}^{m_{1_1}-a} \binom{m_{1_1}-a}{m_{2_1}} \left[(1 - u_{2_1})\bar{f}_3(u_{3_1})\right]^{m_{2_1}} \left[u_{2_1}T_2\bar{f}_3(u_{3_2})\right]^{m_{1_1}-a-m_{2_1}} \left[u_{1_1}T_1\right]^{k-l-m_{1_1}} \sum_{c=0}^{k-l-m_{1_1}} \binom{k-l-m_{1_1}}{c} \left[u_{2_2}(1 - T_2)\right]^c
$$

$$
\times \sum_{m_{2_2}=0}^{k-l-m_{1_1}-c} \binom{k-l-m_{1_1}-c}{m_{2_2}} \left[(1 - u_{2_2})\bar{f}_3(u_{3_3})\right]^{m_{2_2}} \left[u_{2_2}T_2\bar{f}_3(u_{3_4})\right]^s, \tag{A13}
$$

where we have dropped the 1's from the $\bar{f}$ arguments and simplified $s = k - l - m_{1_1} - c - m_{2_2}$ for brevity. We define the set $\{u_{3_j}\}$, $j = 1, \ldots, 4$, as the set of probabilities that each neighbor fails to transmit strain 3 to the focal node given their particular history. The probabilities $\mathcal{H}_{3_j}$ of each of these unique histories occurring were defined in Eq. (33). Using Eq. (22) we now furnish $\mathcal{C}_3$ with these probabilities to obtain

$$
\mathcal{P}_{3_1} = \sum_{l=0}^{k} \binom{k}{l} \left[u_{1_1}(1 - T_1)\right]^l \sum_{m_{1_1}=0}^{k-l} \binom{k-l}{m_{1_1}} \left[1 - u_{1_1}\right]^{m_{1_1}} \sum_{a=0}^{m_{1_1}} \binom{m_{1_1}}{a} \left[u_{2_1}(1 - T_2)\right]^a \sum_{m_{2_1}=0}^{m_{1_1}-a} \binom{m_{1_1}-a}{m_{2_1}} \left[(1 - u_{2_1})\bar{f}_3(u_{3_1})\right]^{m_{2_1}}
$$

$$\times \left[u_{2_1} T_2 \bar{f}_3(u_{3_2}) \right]^{m_{1_1} - a - m_{2_1}} \left[u_{1_1} T_1 \right]^{k-l-m_{1_1}} \sum_{c=0}^{k-l-m_{1_1}} \binom{k-l-m_{1_1}}{c} \left[u_{2_2}(1-T_2) \right]^c$$

$$\times \sum_{m_{2_2}=0}^{k-l-m_{1_1}-c} \binom{k-l-m_{1_1}-c}{m_{2_2}} \left[(1-u_{2_2})\bar{f}_3(u_{3_3}) \right]^{m_{2_2}} \left[u_{2_2} T_2 \bar{f}_3(u_{3_4}) \right]^s \left[1 - (1-T_1)^{m_{1_1}} \right] \left[1 - (1-T_2)^{m_{2_1}+m_{2_2}} \right], \quad \text{(A14a)}$$

$$\mathcal{P}_{3_2} = \sum_{l=0}^{k} \binom{k}{l} \left[u_{1_1}(1-T_1) \right]^l \sum_{m_{1_1}=0}^{k-l} \binom{k-l}{m_{1_1}} \left[1-u_{1_1} \right]^{m_{1_1}} \sum_{a=0}^{m_{1_1}} \binom{m_{1_1}}{a} \left[u_{2_1}(1-T_2) \right]^a$$

$$\times \sum_{m_{2_1}=0}^{m_{1_1}-a} \binom{m_{1_1}-a}{m_{2_1}} \left[(1-u_{2_1})(1-T_2)\bar{f}_3(u_{3_1}) \right]^{m_{2_1}} \left[u_{2_1} T_2 \bar{f}_3(u_{3_2}) \right]^{m_{1_1}-a-m_{2_1}} \left[u_{1_1} T_1 \right]^{k-l-m_{1_1}}$$

$$\times \sum_{c=0}^{k-l-m_{1_1}} \binom{k-l-m_{1_1}}{c} \left[u_{2_2}(1-T_2) \right]^c \sum_{m_{2_2}=0}^{k-l-m_{1_1}-c} \binom{k-l-m_{1_1}-c}{m_{2_2}} \left[(1-u_{2_2})(1-T_2)\bar{f}_3(u_{3_3}) \right]^{m_{2_2}}$$

$$\times \left[u_{2_2} T_2 \bar{f}_3(u_{3_4}) \right]^s \left[1 - (1-T_1)^{m_{1_1}} \right], \quad \text{(A14b)}$$

$$\mathcal{P}_{3_3} = \sum_{l=0}^{k} \binom{k}{l} \left[u_{1_1}(1-T_1) \right]^l \sum_{m_{1_1}=0}^{k-l} \binom{k-l}{m_{1_1}} \left[(1-u_{1_1})(1-T_1) \right]^{m_{1_1}} \sum_{a=0}^{m_{1_1}} \binom{m_{1_1}}{a} \left[u_{2_1}(1-T_2) \right]^a$$

$$\times \sum_{m_{2_1}=0}^{m_{1_1}-a} \binom{m_{1_1}-a}{m_{2_1}} \left[(1-u_{2_1})\bar{f}_3(u_{3_1}) \right]^{m_{2_1}} \left[u_{2_1} T_2 \bar{f}_3(u_{3_2}) \right]^{m_{1_1}-a-m_{2_1}} \left[u_{1_1} T_1 \right]^{k-l-m_{1_1}} \sum_{c=0}^{k-l-m_{1_1}} \binom{k-l-m_{1_1}}{c} \left[u_{2_2}(1-T_2) \right]^c$$

$$\times \sum_{m_{2_2}=0}^{k-l-m_{1_1}-c} \binom{k-l-m_{1_1}-c}{m_{2_2}} \left[(1-u_{2_2})\bar{f}_3(u_{3_3}) \right]^{m_{2_2}} \left[u_{2_2} T_2 \bar{f}_3(u_{3_4}) \right]^s \left[1 - (1-T_2)^{m_{21}+m_{22}} \right], \quad \text{(A14c)}$$

$$\mathcal{P}_{3_4} = \sum_{l=0}^{k} \binom{k}{l} \left[u_{1_1}(1-T_1) \right]^l \sum_{m_{1_1}=0}^{k-l} \binom{k-l}{m_{1_1}} \left[(1-u_{1_1})(1-T_1) \right]^{m_{1_1}} \sum_{a=0}^{m_{1_1}} \binom{m_{1_1}}{a} \left[u_{2_1}(1-T_2) \right]^a \sum_{m_{2_1}=0}^{m_{1_1}-a} \binom{m_{1_1}-a}{m_{2_1}}$$

$$\times \left[(1-u_{2_1})(1-T_2)\bar{f}_3(u_{3_1}) \right]^{m_{2_1}} \left[u_{2_1} T_2 \bar{f}_3(u_{3_2}) \right]^{m_{1_1}-a-m_{2_1}} \left[u_{1_1} T_1 \right]^{k-l-m_{1_1}} \sum_{c=0}^{k-l-m_{1_1}} \binom{k-l-m_{1_1}}{c} \left[u_{2_2}(1-T_2) \right]^c$$

$$\times \sum_{m_{2_2}=0}^{k-l-m_{1_1}-c} \binom{k-l-m_{1_1}-c}{m_{2_2}} \left[(1-u_{2_2})(1-T_2)\bar{f}_3(u_{3_3}) \right]^{m_{2_2}} \left[u_{2_2} T_2 \bar{f}_3(u_{3_4}) \right]^s. \quad \text{(A14d)}$$

Next, we construct the prior probabilities that the neighbor in question was in the desired state; these are shown in Eq. (36) for each history. Lastly, we use Eq. (21) to finalise the self-consistent expressions for the u_{i_h} values using

$$u_{i_h} = \frac{1}{\mathcal{Q}_{i_h}} \sum_{k=0}^{\infty} q_k \mathcal{P}_{i_h}. \quad \text{(A15)}$$

In some cases, this expression is straightforward to compute; however, those histories obtained from external coinfection lead to multiple $G_1(z)$ terms when the brackets are multiplied out. For instance $\mathcal{H}_{3_1}$ behaves as follows:

$$\left[1 - (1-T_1)^{m_{1_1}} \right] \left[1 - (1-T_2)^{m_{2_1}+m_{2_2}} \right] = 1 - (1-T_1)^{m_{1_1}} - (1-T_2)^{m_{2_1}+m_{2_2}} + (1-T_1)^{m_{1_1}}(1-T_2)^{m_{2_1}+m_{2_2}}. \quad \text{(A16)}$$

When $\mathcal{C}_2$ acts on this expression we generate four brackets. We find

$$u_{3_1} = \frac{1}{(1-u_{1_1})(1-u_{2_1})} G_1\Big(u_{1_1}(1-T_1) + (1-u_{1_1}) \big[u_{2_1}(1-T_2) + (1-u_{2_1})\bar{f}_3(u_{3_1}) + u_{2_1} T_2 \bar{f}_3(u_{3_2}) \big]$$

$$+ u_{1_1} T_1 \big[u_{2_2}(1-T_2) + (1-u_{2_2})\bar{f}_3(u_{3_3}) + u_{2_2} T_2 \bar{f}_3(u_{3_4}) \big] \Big)$$

$$- \frac{1}{(1-u_{1_1})(1-u_{2_1})} G_1\Big(u_{1_1}(1-T_1) + (1-u_{1_1}) \big[u_{2_1}(1-T_2) + (1-u_{2_1})(1-T_2)\bar{f}_3(u_{3_1}) + u_{2_1} T_2 \bar{f}_3(u_{3_2}) \big]$$

$$+ u_{1_1} T_1 \big[u_{2_2}(1-T_2) + (1-u_{2_2})(1-T_2)\bar{f}_3(u_{3_3}) + u_{2_2} T_2 \bar{f}_3(u_{3_4}) \big] \Big)$$

$$- \frac{1}{\left(1 - u_{1_1}\right)\left(1 - u_{2_1}\right)} G_1\left(u_{1_1}(1 - T_1) + \left(1 - u_{1_1}\right)(1 - T_1)\left[u_{2_1}(1 - T_2) + \left(1 - u_{2_1}\right)\bar{f}_3\left(u_{3_1}\right) + u_{2_1}T_2\bar{f}_3\left(u_{3_2}\right)\right]\right.$$

$$+ u_{1_1}T_1\left[u_{2_2}(1 - T_2) + \left(1 - u_{2_2}\right)\bar{f}_3\left(u_{3_3}\right) + u_{2_2}T_2\bar{f}_3\left(u_{3_4}\right)\right]\right)$$

$$+ \frac{1}{\left(1 - u_{1_1}\right)\left(1 - u_{2_1}\right)} G_1\left(u_{1_1}(1 - T_1) + \left(1 - u_{1_1}\right)(1 - T_1)\left[u_{2_1}(1 - T_2) + \left(1 - u_{2_1}\right)(1 - T_2)\bar{f}_3\left(u_{3_1}\right) + u_{2_1}T_2\bar{f}_3\left(u_{3_2}\right)\right]\right.$$

$$+ u_{1_1}T_1\left[u_{2_2}(1 - T_2) + \left(1 - u_{2_2}\right)(1 - T_2)\bar{f}_3\left(u_{3_3}\right) + u_{2_2}T_2\bar{f}_3\left(u_{3_4}\right)\right]\right), \tag{A17a}$$

$$u_{3_2} = \frac{1}{\left(1 - u_{1_1}\right)u_{2_1}} G_1\left(u_{1_1}(1 - T_1) + \left(1 - u_{1_1}\right)\left[u_{2_1}(1 - T_2) + \left(1 - u_{2_1}\right)(1 - T_2)\bar{f}_3\left(u_{3_1}\right) + u_{2_1}T_2\bar{f}_3\left(u_{3_2}\right)\right]\right.$$

$$+ u_{1_1}T_1\left[u_{2_2}(1 - T_2) + \left(1 - u_{2_2}\right)(1 - T_2)\bar{f}_3\left(u_{3_3}\right) + u_{2_2}T_2\bar{f}_3\left(u_{3_4}\right)\right]\right) - \frac{1}{\left(1 - u_{1_1}\right)u_{2_1}} G_1\left(u_{1_1}(1 - T_1)\right.$$

$$+ \left(1 - u_{1_1}\right)(1 - T_1)\left[u_{2_1}(1 - T_2) + \left(1 - u_{2_1}\right)(1 - T_2)\bar{f}_3\left(u_{3_1}\right) + u_{2_1}T_2\bar{f}_3\left(u_{3_2}\right)\right] + u_{1_1}T_1\left[u_{2_2}(1 - T_2)\right.$$

$$+ \left(1 - u_{2_2}\right)(1 - T_2)\bar{f}_3\left(u_{3_3}\right) + u_{2_2}T_2\bar{f}_3\left(u_{3_4}\right)\right]\right), \tag{A17b}$$

$$u_{3_3} = \frac{1}{u_{1_1}\left(1 - u_{2_2}\right)} G_1\left(u_{1_1}(1 - T_1) + \left(1 - u_{1_1}\right)(1 - T_1)\left[u_{2_1}(1 - T_2) + \left(1 - u_{2_1}\right)\bar{f}_3\left(u_{3_1}\right) + u_{2_1}T_2\bar{f}_3\left(u_{3_2}\right)\right]\right.$$

$$+ u_{1_1}T_1\left[u_{2_2}(1 - T_2) + \left(1 - u_{2_2}\right)\bar{f}_3\left(u_{3_3}\right) + u_{2_2}T_2\bar{f}_3\left(u_{3_4}\right)\right]\right) - \frac{1}{u_{1_1}\left(1 - u_{2_2}\right)} G_1\left(u_{1_1}(1 - T_1)\right.$$

$$+ \left(1 - u_{1_1}\right)(1 - T_1)\left[u_{2_1}(1 - T_2) + \left(1 - u_{2_1}\right)(1 - T_2)\bar{f}_3\left(u_{3_1}\right) + u_{2_1}T_2\bar{f}_3\left(u_{3_2}\right)\right]$$

$$+ u_{1_1}T_1\left[u_{2_2}(1 - T_2) + \left(1 - u_{2_2}\right)(1 - T_2)\bar{f}_3\left(u_{3_3}\right) + u_{2_2}T_2\bar{f}_3\left(u_{3_4}\right)\right]\right), \tag{A17c}$$

$$u_{3_4} = \frac{1}{u_{1_1}u_{2_2}} G_1\left(u_{1_1}(1 - T_1) + \left(1 - u_{1_1}\right)(1 - T_1)\left[u_{2_1}(1 - T_2) + \left(1 - u_{2_1}\right)(1 - T_2)\bar{f}_3\left(u_{3_1}\right) + u_{2_1}T_2\bar{f}_3\left(u_{3_2}\right)\right]\right.$$

$$+ u_{1_1}T_1\left[u_{2_2}(1 - T_2) + \left(1 - u_{2_2}\right)(1 - T_2)\bar{f}_3\left(u_{3_3}\right) + u_{2_2}T_2\bar{f}_3\left(u_{3_4}\right)\right]\right). \tag{A17d}$$

The outbreak size of the third strain is then given by

$$\mathcal{A}_3 = \mathcal{A}_2 - \left[G_0\left(u_{1_1}(1 - T_1) + \left(1 - u_{1_1}\right)\left[u_{2_1}(1 - T_2) + \left(1 - u_{2_1}\right)\bar{f}_3\left(u_{3_1}\right) + u_{2_1}T_2\bar{f}_3\left(u_{3_2}\right)\right]\right.\right.$$

$$+ u_{1_1}T_1\left[u_{2_2}(1 - T_2) + \left(1 - u_{2_2}\right)\bar{f}_3\left(u_{3_3}\right) + u_{2_2}T_2\bar{f}_3\left(u_{3_4}\right)\right]\right)$$

$$- G_0\left(u_{1_1}(1 - T_1) + \left(1 - u_{1_1}\right)\left[u_{2_1}(1 - T_2) + \left(1 - u_{2_1}\right)(1 - T_2)\bar{f}_3\left(u_{3_1}\right) + u_{2_1}T_2\bar{f}_3\left(u_{3_2}\right)\right]\right.$$

$$+ u_{1_1}T_1\left[u_{2_2}(1 - T_2) + \left(1 - u_{2_2}\right)(1 - T_2)\bar{f}_3\left(u_{3_3}\right) + u_{2_2}T_2\bar{f}_3\left(u_{3_4}\right)\right]\right)$$

$$- G_0\left(u_{1_1}(1 - T_1) + \left(1 - u_{1_1}\right)(1 - T_1)\left[u_{2_1}(1 - T_2) + \left(1 - u_{2_1}\right)\bar{f}_3\left(u_{3_1}\right) + u_{2_1}T_2\bar{f}_3\left(u_{3_2}\right)\right]\right.$$

$$+ u_{1_1}T_1\left[u_{2_2}(1 - T_2) + \left(1 - u_{2_2}\right)\bar{f}_3\left(u_{3_3}\right) + u_{2_2}T_2\bar{f}_3\left(u_{3_4}\right)\right]\right)$$

$$+ G_0\left(u_{1_1}(1 - T_1) + \left(1 - u_{1_1}\right)(1 - T_1)\left[u_{2_1}(1 - T_2) + \left(1 - u_{2_1}\right)(1 - T_2)\bar{f}_3\left(u_{3_1}\right) + u_{2_1}T_2\bar{f}_3\left(u_{3_2}\right)\right]\right.$$

$$+ u_{1_1}T_1\left[u_{2_2}(1 - T_2) + \left(1 - u_{2_2}\right)(1 - T_2)\bar{f}_3\left(u_{3_3}\right) + u_{2_2}T_2\bar{f}_3\left(u_{3_4}\right)\right]\right)\right], \tag{A18}$$

which is simply u_{3_1}, the maximally externally coinfected state, without the prior probabilities and with $G_1(z)$ replaced by $G_0(z)$.

4. Strain 4

In this section we evaluate the expressions for the outbreak size of strain 4 for collaborative percolation. The first step is to find the base probability. To do this, we require $\mathcal{C}_4 = \bar{\mathcal{C}}_3(\bar{f}_4)$, which is quite a complicated nested function. However, Figs. 11 and 12 show how this expression is formed graphically as a perfect binary tree. First,

$$\bar{\mathcal{C}}_3(\mathbf{1}) = \sum_{l=0}^{k} \binom{k}{l}\left[u_{1_1}(1 - T_1)\right]^l \sum_{m_{1_1}=0}^{k-l} \binom{k - l}{m_{1_1}}\left[1 - u_{1_1}\right]^{m_{1_1}} \sum_{a=0}^{m_{1_1}} \binom{m_{1_1}}{a}\left[u_{2_1}(1 - T_2)\right]^a \sum_{m_{2_1}=0}^{m_{1_1}-a} \binom{m_{1_1} - a}{m_{2_1}}\left[\left(1 - u_{2_1}\right)\right]^{m_{2_1}}$$

$$\times \sum_{b=0}^{m_{2_1}} \binom{m_{2_1}}{b}\left[u_{3_1}(1 - T_3)\right]^b \sum_{m_{3_1}=0}^{m_{2_1}-b} \binom{m_{2_1} - b}{m_{3_1}}\left[1 - u_{3_1}\right]^{m_{3_1}}\left[u_{3_1}T_3\right]^{m_{2_1}-b-m_{3_1}}\left[u_{2_1}T_2\right]^{m_{1_1}-a-m_{2_1}}$$
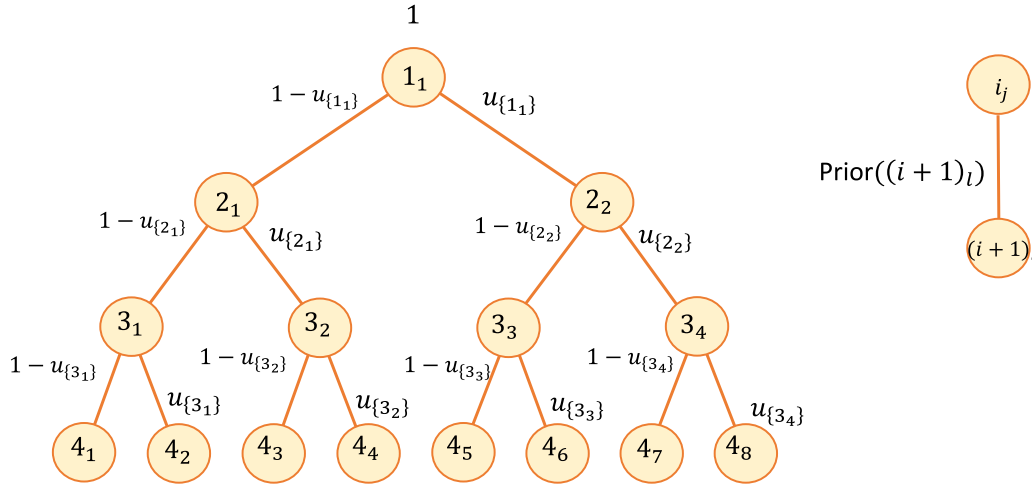
FIG. 11. A visualization of the perfect binary tree of coinfected neighbors which a focal node could be surrounded by after strain 3. The states are marked inside the nodes, while the priors can be constructed by the product of the terms associated to each edge from the root to each leaf. External infection occurs on the left child, while direct infection occurs on the right.

$$
\times \sum_{c=0}^{m_{1_1}-a-m_{2_1}} \binom{m_{1_1}-a-m_{2_1}}{c} \left[u_{3_2}(1-T_3)\right]^c \sum_{m_{3_2}=0}^{m_{1_1}-a-m_{2_1}-c} \binom{m_{1_1}-a-m_{2_1}-c}{m_{3_2}} \left[1-u_{3_2}\right]^{m_{3_2}} \left[u_{3_2}T_3\right]^{m_{1_1}-a-m_{2_1}-c-m_{3_2}}
$$

$$
\times \left[u_{1_1}T_1\right]^{k-l-m_{1_1}} \sum_{e=0}^{k-l-m_{1_1}} \binom{k-l-m_{1_1}}{e} \left[u_{2_2}(1-T_2)\right]^e \sum_{m_{2_2}=0}^{k-l-m_{1_1}-e} \binom{k-l-m_{1_1}-e}{m_{2_2}} \left[1-u_{2_2}\right]^{m_{2_2}}
$$

$$
\times \sum_{f=0}^{m_{2_2}} \binom{m_{2_2}}{f} \left[u_{3_3}(1-T_3)\right]^f \sum_{m_{33}=0}^{m_{2_2}-f} \binom{m_{2_2}-f}{m_{3_3}} \left[1-u_{3_3}\right]^{m_{3_3}} \left[u_{3_3}T_3\right]^{m_{2_2}-f-m_{3_3}} \left[u_{2_2}T_2\right]^{k-l-m_{1_1}-e-m_{2_2}}
$$

$$
\times \sum_{g=0}^{k-l-m_{1_1}-e-m_{2_2}} \binom{k-l-m_{1_1}-e-m_{2_2}}{g} \left[u_{3_4}(1-T_3)\right]^g \sum_{m_{3_4}=0}^{k-l-m_{1_1}-e-m_{2_2}-g} \binom{k-l-m_{1_1}-e-m_{2_2}-g}{m_{3_4}}
$$

$$
\times \left[1-u_{3_4}\right]^{m_{3_4}} \left[u_{3_4}T_3\right]^{k-l-m_{1_1}-e-m_{2_2}-g-m_{3_4}}. \tag{A19}
$$

We then have

$$
\mathcal{C}_4 = f_1\big(u_1, f_2\big(u_{2_1}, f_3\big(u_{3_1}, \bar{f}_4(u_{4_1}), \bar{f}_4(u_{4_2})\big), f_3\big(u_{3_2}, \bar{f}_4(u_{4_3}), \bar{f}_4(u_{4_4})\big)\big),
$$
$$
f_2\big(u_{2_2}, f_3\big(u_{3_3}, \bar{f}_4(u_{4_5}), \bar{f}_4(u_{4_6})\big), f_3\big(u_{3_4}, \bar{f}_4(u_{4_7}), \bar{f}_4(u_{4_8})\big)\big)\big) \tag{A20}
$$
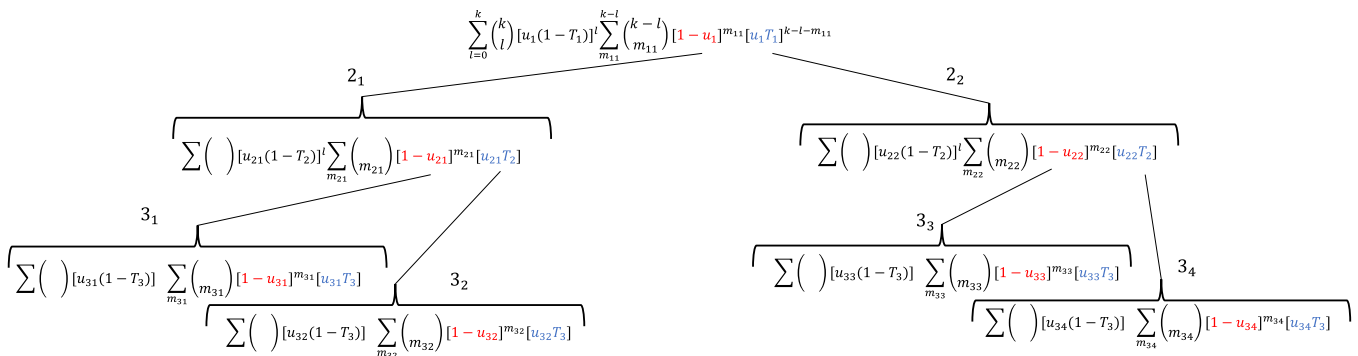
FIG. 12. A visualization of the perfect binary tree of coinfected neighbors which a focal node could be surrounded by and the equations that generate the base probability $\bar{\mathcal{C}}_3(\mathbf{1})$. In this notation, the left child always represents external infection while the right child represents direct infection, as in Fig. 11. Thus, the leaves of the tree represent all possible infection histories for the fourth strain.
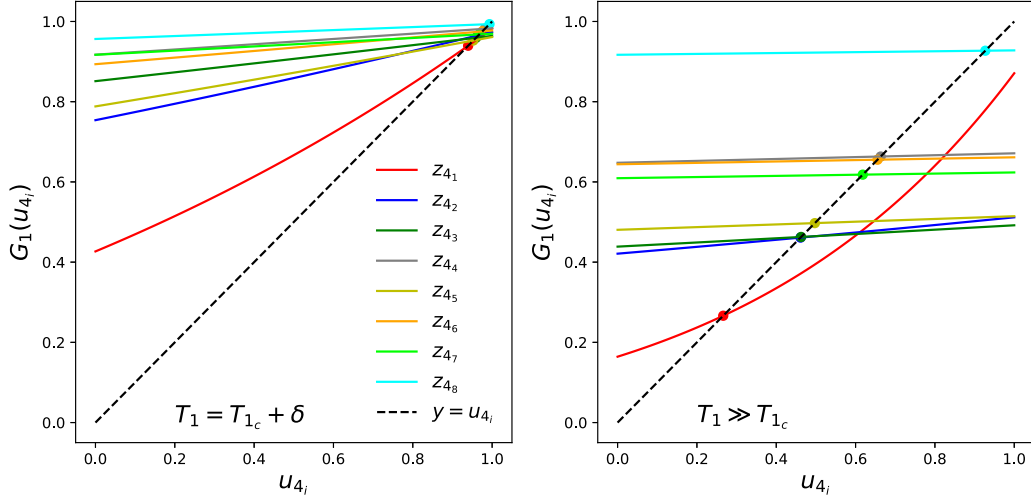
FIG. 13. The graphical solution of the generating functions for strain 4 around the critical point of strain 1 (left) with $T_1 = 0.325$, $T_2 = 0.6$, $T_3 = 0.5$, $T_4 = 0.45$ and away from the critical point at $T_1 = T_2 = 0.6$, $T_3 = 0.5$, $T_4 = 0.45$ (right). Plotted are $y = u_{4_i}$ for $i = 1, \ldots, 8$ against u_{4_i} as well as the value of the generating functions $z_i = G_1(C_2 \mathcal{H}_{4_i})/Q_{4_i}$. Each z_i varies u_{4_i} while the other value held fixed at the correct root. The intersection of $y = u_{4_i}$ and z_i corresponds to the root, which is also marked with a scatter point from a nonlinear solve. We notice that u_{4_1} varies convexly over almost the entire unit interval while the gradient of the other generating functions is increasingly flat in this region of the parameter space. This indicates that the contribution of these values is less important than that u_{4_1}. Thus, we can expect that this infection history is the dominant term in the nonlinear system that describes strain 4.

The next step is to formulate the eight u_{4_i} values given each unique infection history. These are given by

$$u_{4_1} = \frac{1}{(1 - u_{1_1})(1 - u_{2_1})(1 - u_{3_1})} \sum_{k=0}^{\infty} q_k C_4 \mathcal{H}_{3_1} [1 - (1 - T_3)^{m_{3_1} + m_{3_2} + m_{3_3} + m_{3_4}}], \tag{A21a}$$

$$u_{4_2} = \frac{1}{(1 - u_{1_1})(1 - u_{2_1}) u_{3_1}} \sum_{k=0}^{\infty} q_k C_4 \mathcal{H}_{3_1} (1 - T_3)^{m_{3_1} + m_{3_2} + m_{3_3} + m_{3_4}}, \tag{A21b}$$

$$u_{4_3} = \frac{1}{(1 - u_{1_1}) u_{2_1} (1 - u_{3_2})} \sum_{k=0}^{\infty} q_k C_4 \mathcal{H}_{3_2} [1 - (1 - T_3)^{m_{3_1} + m_{3_2} + m_{3_3} + m_{3_4}}], \tag{A21c}$$

$$u_{4_4} = \frac{1}{(1 - u_{1_1}) u_{2_1} u_{3_2}} \sum_{k=0}^{\infty} q_k C_4 \mathcal{H}_{3_2} (1 - T_3)^{m_{3_1} + m_{3_2} + m_{3_3} + m_{3_4}}, \tag{A21d}$$

$$u_{4_5} = \frac{1}{u_{1_1} (1 - u_{2_2})(1 - u_{3_3})} \sum_{k=0}^{\infty} q_k C_4 \mathcal{H}_{3_3} [1 - (1 - T_3)^{m_{3_1} + m_{3_2} + m_{3_3} + m_{3_4}}], \tag{A21e}$$

$$u_{4_6} = \frac{1}{u_{1_1} (1 - u_{2_2}) u_{3_3}} \sum_{k=0}^{\infty} q_k C_4 \mathcal{H}_{3_3} (1 - T_3)^{m_{3_1} + m_{3_2} + m_{3_3} + m_{3_4}}, \tag{A21f}$$

$$u_{4_7} = \frac{1}{u_{1_1} u_{2_2} (1 - u_{3_4})} \sum_{k=0}^{\infty} q_k C_4 \mathcal{H}_{3_4} [1 - (1 - T_3)^{m_{3_1} + m_{3_2} + m_{3_3} + m_{3_4}}], \tag{A21g}$$

$$u_{4_8} = \frac{1}{u_{1_1} u_{2_2} u_{3_4}} \sum_{k=0}^{\infty} q_k C_4 \mathcal{H}_{3_4} (1 - T_3)^{m_{3_1} + m_{3_2} + m_{3_3} + m_{3_4}}, \tag{A21h}$$

We show the solution to these expressions graphically in Fig. 13 around the critical point and for large T values. The outbreak size is then given by

$$\mathcal{A}_4 = \mathcal{A}_3 - \sum_{k=0}^{\infty} p_k C_4 \mathcal{H}_{3_1} [1 - (1 - T_3)^{m_{3_1} + m_{3_2} + m_{3_3} + m_{3_4}}]. \tag{A22}$$

5. Strain 5

Following the recipe, the outbreak size of strain 5 is calculated as follows:

$$\mathcal{C}_5 = f_1\big(u_1, f_2\big(u_{2_1}, f_3\big(u_{3_1}, f_4\big(u_{4_1}, \bar{f}_5(u_{5_1}), \bar{f}_5(u_{5_2})\big), f_4\big(u_{4_2}, \bar{f}_5(u_{5_3}), \bar{f}_5(u_{5_4})\big)\big), f_3\big(u_{3_2}, f_4\big(u_{4_3}, \bar{f}_5(u_{5_5}), \bar{f}_5(u_{5_6})\big),$$
$$f_4\big(u_{4_4}, \bar{f}_5(u_{5_7}), \bar{f}_5(u_{5_8})\big)\big)\big), f_2\big(u_{2_2}, f_3\big(u_{3_3}, f_4\big(u_{4_5}, \bar{f}_5(u_{5_9}), \bar{f}_5(u_{5_{10}})\big), f_4\big(u_{4_6}, \bar{f}_5(u_{5_{11}}), \bar{f}_5(u_{5_{12}})\big)\big),$$
$$f_3\big(u_{3_4}, f_4\big(u_{4_7}, \bar{f}_5(u_{5_{13}}), \bar{f}_5(u_{5_{14}})\big), f_4\big(u_{4_8}, \bar{f}_5(u_{5_{15}}), \bar{f}_5(u_{5_{16}})\big)\big)\big)\big), \tag{A23}$$

$$u_{5_1} = \frac{1}{(1-u_{1_1})(1-u_{2_1})(1-u_{3_1})(1-u_{4_1})} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_1}[1-(1-T_4)^M], \tag{A24a}$$

$$u_{5_2} = \frac{1}{(1-u_{1_1})(1-u_{2_1})(1-u_{3_1})u_{4_1}} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_1}(1-T_4)^M, \tag{A24b}$$

$$u_{5_3} = \frac{1}{(1-u_{1_1})(1-u_{2_1})u_{3_1}(1-u_{4_2})} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_2}[1-(1-T_4)^M], \tag{A24c}$$

$$u_{5_4} = \frac{1}{(1-u_{1_1})(1-u_{2_1})u_{3_1}u_{4_2}} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_2}(1-T_4)^M, \tag{A24d}$$

$$u_{5_5} = \frac{1}{(1-u_{1_1})u_{2_1}(1-u_{3_2})(1-u_{4_3})} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_3}[1-(1-T_4)^M], \tag{A24e}$$

$$u_{5_6} = \frac{1}{(1-u_{1_1})u_{2_1}(1-u_{3_2})u_{4_3}} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_3}(1-T_4)^M, \tag{A24f}$$

$$u_{5_7} = \frac{1}{(1-u_{1_1})u_{2_1}u_{3_2}(1-u_{4_4})} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_4}[1-(1-T_4)^M], \tag{A24g}$$

$$u_{5_8} = \frac{1}{(1-u_{1_1})u_{2_1}u_{3_2}u_{4_4}} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_4}(1-T_4)^M, \tag{A24h}$$

$$u_{5_9} = \frac{1}{u_{1_1}(1-u_{2_2})(1-u_{3_3})(1-u_{4_5})} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_5}[1-(1-T_4)^M], \tag{A24i}$$

$$u_{5_{10}} = \frac{1}{u_{1_1}(1-u_{2_2})(1-u_{3_3})u_{4_5}} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_5}(1-T_4)^M, \tag{A24j}$$

$$u_{5_{11}} = \frac{1}{u_{1_1}(1-u_{2_2})u_{3_3}(1-u_{4_6})} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_6}[1-(1-T_4)^M], \tag{A24k}$$

$$u_{5_{12}} = \frac{1}{u_{1_1}(1-u_{2_2})u_{3_3}u_{4_6}} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_6}(1-T_4)^M, \tag{A24l}$$

$$u_{5_{13}} = \frac{1}{u_{1_1}u_{2_2}(1-u_{3_4})(1-u_{4_7})} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_7}[1-(1-T_4)^M], \tag{A24m}$$

$$u_{5_{14}} = \frac{1}{u_{1_1}u_{2_2}(1-u_{3_4})u_{4_7}} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_7}(1-T_4)^M, \tag{A24n}$$

$$u_{5_{15}} = \frac{1}{u_{1_1}u_{2_2}u_{3_4}(1-u_{4_8})} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_8}[1-(1-T_4)^M], \tag{A24o}$$

$$u_{5_{16}} = \frac{1}{u_{1_1}u_{2_2}u_{3_4}u_{4_8}} \sum_{k=0}^{\infty} q_k \mathcal{C}_5 H_{4_8}(1-T_4)^M, \tag{A24p}$$

with $M = \sum_{j=1}^{8} m_{4_j}$. The outbreak size is then given by

$$\mathcal{A}_5 = \mathcal{A}_4 - \sum_{k=0}^{\infty} p_k \mathcal{C}_5 H_{4_1}[1 - (1 - T_4)^M]. \tag{A25}$$

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

[2] E. Kenah and J. M. Robins, Second look at the spread of epidemics on networks, Phys. Rev. E **76**, 036113 (2007).

[3] J. C Miller, Epidemic size and probability in populations with heterogeneous infectivity and susceptibility, Phys. Rev. E **76**, 010101(R) (2007).

[4] M. E. J. Newman, Threshold Effects for Two Pathogens Spreading on a Network, Phys. Rev. Lett. **95**, 108701 (2005).

[5] B. Karrer and M. E. J. Newman Competing epidemics on complex networks, Phys. Rev. E **84**, 036106 (2011).

[6] M. E. J. Newman and C. R Ferrario, Interacting epidemics and coinfection on contact networks, PLoS ONE **8**, 8 (2013).

[7] S. Funk and V. A. A. Jansen, Interacting epidemics on overlay networks, Phys. Rev. E **81**, 036118 (2010).

[8] S. Bansal and L. A. Meyers, The impact of past epidemics on future disease dynamics, J. Theor. Biol. **309**, 176 (2012).

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

[10] 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).

[11] W. Cai, L. Chen, F. Ghanbarnejad, and P. Grassberger, Avalanche outbreaks emerging in cooperative contagions, Nat. Phys. **11**, 936 (2015).

[12] S. Bansal, B. Pourbohloul, N. Hupert, B. Grenfell, and L. A. Meyers, The shifting demographic landscape of pandemic influenza, PLoS ONE **5**, e9360 (2010).

[13] P. Mann, V. A. Smith, J. B. O. Mitchell, and S. Dobson, Symbiotic and antagonistic disease dynamics on networks using bond percolation, Phys. Rev. E **104**, 024303 (2021).

[14] 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).

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

[16] H. S Wilf, *Generatingfunctionology* (Academic, New York, 1990).

[17] J. L. W. V. Jensen, Sur les fonctions convexes et les inégalités entre les valeurs moyennes, Acta Math. **30**, 175 (1906).

[18] M. E. J. Newman, Power laws, Pareto distributions and Zipfs law, Contemp. Phys. **46**, 323 (2005).

[19] R. Pastor-Satorras, and A. Vespignani, Epidemic Spreading in Scale-Free Networks, Phys. Rev. Lett. **86**, 3200 (2001).

[20] P. Mann, V. A. Smith, J. B. O. Mitchell, C. Jefferson, and S. Dobson, Exact formula for bond percolation on cliques, Phys. Rev. E **104**, 024304 (2021).

[21] Y. Oz, I. Rubinstein, and M. Safra, Multivariate generating functions for information spread on multi-type random graphs, J. Stat. Mech.: Theory Exp. (2022) 033501.

[22] 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).