Spread of infectious disease and social awareness as parasitic contagions on clustered networks

Laurent Hébert-Dufresne,[1,2] Dina Mistry [3] and Benjamin M. Althouse[3,4,5]

[1]*Department of Computer Science, University of Vermont, Burlington, Vermont 05405, USA*
[2]*Vermont Complex Systems Center, University of Vermont, Burlington, Vermont 05405, USA*
[3]*Institute for Disease Modeling, Global Health, Bill & Melinda Gates Foundation, Seattle, Washington 98109, USA*
[4]*University of Washington, Seattle, Washington 98105, USA*
[5]*New Mexico State University, Las Cruces, New Mexico 88003, USA*

There is a rich history of models for the interaction of a biological contagion like influenza with the spread of related information such as an influenza vaccination campaign. Recent work on the spread of interacting contagions on networks has highlighted that these interacting contagions can have counterintuitive interplay with network structure. Here, we generalize one of these frameworks to tackle three important features of the spread of awareness and disease: one, we model the dynamics on highly clustered, cliquish, networks to mimic the role of workplaces and households; two, the awareness contagion affects the spread of the biological contagion by reducing its transmission rate where an aware or vaccinated individual is less likely to be infected; and three, the biological contagion also affects the spread of the awareness contagion but by increasing its transmission rate where an infected individual is more receptive and more likely to share information related to the disease. Under these conditions, we find that increasing network clustering, which is known to hinder disease spread, can actually allow them to sustain larger epidemics of the disease in models with awareness. This counterintuitive result goes against the conventional wisdom suggesting that random networks are justifiable as they provide worst-case scenario forecasts. To further investigate this result, we provide a closed-form criterion based on a two-step branching process (i.e., the numbers of expected tertiary infections) to identify different regions in parameter space where the net effect of clustering and coinfection varies. Altogether, our results highlight once again the need to go beyond random networks in disease modeling and illustrate the type of analysis that is possible even in complex models of interacting contagions.

I. INTRODUCTION

Models of contagion are used to study the transmission dynamics of a pathogen or information being transmitted through a structured population. Most of these are defined as compartmental models [1], which mathematically distinguishes individuals based on their state; i.e., whether they are susceptible to a contagion or infectious with that contagion. Using this approach, coupling different contagions to model their interactions is straightforward as we can then simply distinguish individuals based on all possible combinations of states for the different contagions. Of particular interest is the coupling of an infectious disease with the spread of preventative information related to the disease [2]. We typically expect this "disease awareness" to at least hinder, if not completely stop, the spread of the pathogen. Since both the pathogen and the information spread through standard contagion mechanisms, their coupled dynamics is often referred to as either *antagonistic* or *dueling* contagions. Here, we assume that the coupling between awareness and disease is twofold: Awareness reduces the transmission rate of a disease, but sees its own transmission rate increased by the presence of the disease. Studies in health psychology suggest that individuals are more likely to adopt preventative behaviors related to a disease—hand washing, self-isolation, vaccination, treatment-seeking—if people around them adopt that behavior and/or if they perceive themselves to be at high risk of infection [3,4]. Indeed, data collected during and after the 2003 SARS coronavirus outbreak and the 2009 influenza A (H1N1) pandemic indicate the adoption of preventative behaviors such as increased hand washing, the use of face masks, and avoiding crowded public spaces in response to these outbreaks and the increased perception of infection risk [5–7]. We therefore refer to this coupled awareness-disease system as one of *parasitic contagions*, since one contagion (awareness) both hurts and benefits from the other (the disease).

Research on dueling or parasitic contagions tend to assume well-mixed populations [2,8], in line with classic epidemic models. However, there are a number of generalized frameworks that account for the fact that information and disease both spread over a network structure [9–11], and these frameworks have become more common in recent years with the rise in popularity of multiplex or multilayer network models

[12–16]. In the current work, we relax the well-mixed assumption to consider the community structure of most networks where connections are often grouped in dense environments such as households and workplaces. To do so, we rely on two recent mathematical approaches: (i) a compartmental model coupled to a clique-based master equation formalism [17,18] that explicitly tracks the state of different groups; and (ii) a more straightforward two-step branching process analysis [18] that approximates network clustering.

In Sec. II we present our general model of interacting contagion and provide the mean-field formalism to follow its evolution. We find that the impact of network clustering on the interaction of disease and awareness can be nontrivial to predict, and can even accelerate the spread of the disease. In Sec. II E, we then develop a purely analytical tool based on a two-step branching process to better inform us of the net effect of clustering on the dynamics. The two approaches are leveraged together in a case study presented in Sec. III where we show that there can be finite regions in parameter space where the disease can actually benefit from network clustering. Section IV provides conclusions and outlines potential areas of future work. Altogether, our results raise potential questions about optimal coupling between the epidemiological parameters of a disease, the behavioral parameters of awareness, and the network structure of the population.

II. AWARENESS AND DISEASE AS PARASITIC INFECTIONS

A. Network structures

To study parasitic contagions on clustered contact networks, we use a general definition of community structure where every network is decomposed in terms of groups [19]. The contact network between individuals can thus be interpreted as the projection of a bipartite networks where nodes are connected to social groups of different sizes. In this context, even random links are interpreted as groups of size two. The network topology of our model is illustrated in Fig. 1. In order to highlight the effects of community structure (CS) versus random network, the CS network will be compared with its equivalent random network (ERN): a network with exactly the same degree distribution, but with randomly connected nodes. Both topologies will be studied analytically and numerically.

Typical network datasets are often only available as a collection of pairwise edges rather than higher-order structure-like groups. One then has to rely on some numerical methods such as community detection to infer group structure [20]. Likewise, for theoretical models, one can simply rely on known distributions of groups per node (membership) and of nodes per group (group size) from previous studies on overlapping communities [21,22]. We here use the simplest possible distributions in order to avoid confounding the impact of group structure from that of degree heterogeneity or degree correlations [23].

The dynamics of a single contagion on this community structure model was studied in [17]. Using a mean-field description, it was shown that the clustering of links in groups slowed down propagation as links are wasted on redundant
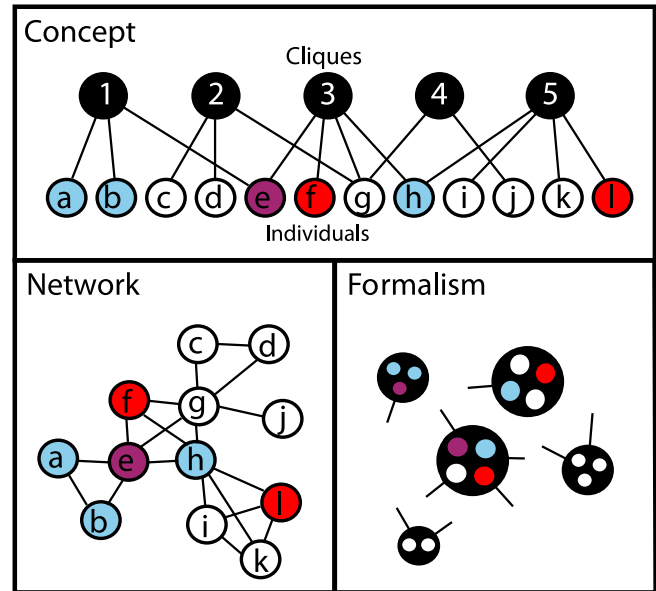

FIG. 1. Schematization of the particular topology and dynamics studied in this paper. An open circle represents a susceptible individual; a shaded one, a contagious individual (infected with the disease, awareness, or both); and a black circle represents a group (or clique). The topology is constructed by allowing individuals to belong to a given number of cliques where they can be linked to other participants (solid lines). Note that in the formalism, the cliques are distinguished by their exact population and state, while the precise connections between them remain unspecified. Modified from Ref. [17].

connections instead of reaching new individuals. Expanding on this study, we more recently introduced a similar mean-field description for two synergistic disease [18], which is the model that we here generalize to interactions of other nature.

B. Dynamical process

To model the concurrent spread of an infectious disease and awareness of it on clustered networks, we will introduce a generalization of the model of interacting contagions used in Ref. [18]. We study the coevolution of two susceptible-infectious-susceptible (SIS) processes such that, at any given time, the state of each individual is determined by their status regarding the two contagion processes. Without interaction with the other contagion, an individual with contagion i would infect its susceptible neighbors at a rate β_i and recover at a rate α_i, but we will here introduce a parametrization scheme to modify these rates and model possible interactions as generally as possible. Our model is flexible and could be applied to any type of pairwise interaction between two SIS processes. However, as the notation will become quite involved we will ground our presentation by referring to the first contagion as the disease (with natural parameters β_D and α_D) and to the second contagion as awareness (with natural parameters β_A and α_A). All parameters are summarized in Table I.

To keep track of both contagions simultaneously, we distinguish nodes by their state $[XY]_m$, where m is their membership number, $X \in \{S_D, I_D\}$ corresponds to their state regarding the

TABLE I. Description of all parameters in the model.

Symbol	Definition
$\{g_m\}$	Distribution of groups per node (memberships)
$\{p_n\}$	Distribution of nodes per group (sizes)
β_D	Transmission rate of the disease ($S_D \rightarrow I_D$)
β_A	Transmission rate of awareness ($S_A \rightarrow I_A$)
α_D	Recovery rate of the disease ($I_D \rightarrow S_D$)
α_A	Recovery rate of awareness ($I_A \rightarrow S_A$)
d_{SS}^{II}	Factor of β_D when the infector is aware
d_{SI}^{IS}	Factor of β_D when the infectee is aware
d_{SI}^{II}	Factor of β_D when infector *and* infectee are aware
a_{SS}^{II}	Factor of β_A when the infector is sick
a_{IS}^{SI}	Factor of β_A when the infectee is sick
a_{IS}^{II}	Factor of β_A when infector *and* infectee are sick
d^{II}	Factor of α_D when the infected is also aware
a^{II}	Factor of α_A when the infected is also sick

disease (susceptible or infectious), and $Y \in \{S_A, I_A\}$ their state regarding the awareness ("susceptible to awareness" meaning naive, and "infectious to awareness" meaning aware). Similarly, we will distinguish groups by their size n and the states of the nodes they contain. For instance, $[ijk]_n$, where i is the number of $[I_D S_A]$ nodes, j is the number of $[S_D I_A]$, and k is the number of $[I_D I_A]$; such that $n - i - j - k$ yields the number of $[S_D S_A]$. Keeping track of the number of nodes with both contagions is critical considering that we are interested in the effect of coinfection.

In the original model of Ref. [18], coinfection had symmetric effect on both contagions, embodied in a single interaction parameter. For parasitic contagions, we want one contagion—the awareness—to benefit from being in the neighborhood of the other, the disease. Individuals might be more likely to listen to an awareness campaign (an increase in β_A) if they are themselves sick or if the message comes from a sick individual [3,4]. Likewise, they might be less likely to forget important information related to a disease if they are currently infected (a decrease in α_A). Second, we also want the disease to be hindered whenever nodes in a given neighborhood are aware of transmission risks or treatment options. Awareness can for example lead both healthy and sick individual to wear a mask (a decrease in β_D) or infectious individuals to treat themselves with antibiotics if appropriate (an increase in α_D) [24]. Overall, we thus expect an increase in awareness transmission rate around infected individuals, a decrease in loss of awareness for infected individuals, a decrease in disease transmission rate around aware individuals, and/or an increase in disease recovery rate for aware individuals who might avoid contacts and seek treatment.

To track all these possible interactions, we therefore need to distinguish each possible infection by the state $[XY]$ of the infector and the state $[UV]$ of the infectee, where we omit the subscript A and D for clarity but use the first state to refer to the disease and the second for awareness. The interaction of these states are embodied in a set of parameters, a_{UV}^{XY}, d_{UV}^{XY}, d^{II}, and a^{II}. While the notation gets complicated because of the many possible interaction mechanisms, the logic goes as follows: Greek letters are used for the natural rates of transmission and recovery of both awareness and

disease (β_A, α_A, β_D, α_D) while a and d factors affect either transmission rates of awareness and disease from state $[XY]$ to $[UV]$ when both superscript and subscript are specified or the recovery rate of coinfected individuals when only the superscript is given. For example, a $[I_1 I_2]$ individual will transmit the disease to a $[S_1 I_2]$ individual at a rate $d_{SI}^{II} \beta_D$. Of course, $d_{UV}^{XY} = 0$ whenever $U \equiv I$ and $a_{UV}^{XY} = 0$ whenever $V \equiv I$ as these individuals are already infected with the corresponding contagion; similarly $d_{UV}^{XY} = 0$ whenever $X \equiv S$ and $a_{UV}^{XY} = 0$ whenever $Y \equiv S$ as only infected individuals can transmit the contagion. We also assume that $d_{SS}^{IS} = a_{SS}^{SI} = 1$ to preserve the natural transmission rate of each contagion, although we still use this term in the general equations to make the type of transmission explicit. Finally, d^{II} and a^{II} give the factor by which the recovery rate of the disease or awareness are modified if the individual is also aware or sick, respectively, but we do not use d^{IS} or a^{SI} for clarity as it is always clear that the terms proportional to α refer to recovery.

C. Mean-field description

A mean-field description of the time evolution of our general model can be written in the spirit of previous formalisms. Leaving out all explicit mention of time dependencies as all variables and mean-fields vary in time, the population density within each node state evolves as

$$\frac{d}{dt}[S_D S_D]_m = -m\big(\beta_D B_{SS}^{(D)} + \beta_A B_{SS}^{(A)}\big)[S_D S_A]_m$$
$$+ \alpha_D [I_D S_A]_m + \alpha_A [S_D I_A]_m, \qquad (1)$$

$$\frac{d}{dt}[I_D S_A]_m = m\big(\beta_D B_{SS}^{(D)}[S_D S_A]_m - \beta_A B_{IS}^{(A)}[I_D S_A]_m\big)$$
$$+ a^{II}\alpha_A [I_D I_A]_m - \alpha_D [I_D S_A]_m, \qquad (2)$$

$$\frac{d}{dt}[S_D I_A]_m = m\big(\beta_A B_{SS}^{(A)}[S_D S_A]_m - \beta_D B_{SI}^{(D)}[S_D I_A]_m\big)$$
$$+ d^{II}\alpha_D [I_D I_A]_m - \alpha_A [S_D I_A]_m, \qquad (3)$$

$$\frac{d}{dt}[I_D I_D]_m = m\big(\beta_D B_{SI}^{(D)}[S_D I_A]_m + \beta_A B_{IS}^{(A)}[I_D S_A]_m\big)$$
$$- (d^{II}\alpha_D + a^{II}\alpha_A)[I_D I_A]_m, \qquad (4)$$

where $B_{UV}^{(i)}$ is a mean-field value of interactions representing the expected number of interactions with contagion i, per membership, for a node in state $[UV]$ and must therefore always be paired with a transmission rate β_i and a node in state S_i with $i \in \{A, D\}$. Notice that in each equation, the first row of terms correspond to infection events and the second row the recoveries. The challenge in correctly writing these equations resides mostly in identifying to which state each event transfers some population density. Conservation of total population density (i.e., the sum over all state densities remains equal to one) is easily verified since the sum of Eqs. (1) to (4) is zero.

Let us assume that we know the density $[ijk]_n$ of cliques that contain n individuals with i nodes contagious with the disease only, j contagious with awareness only, and k with both contagions. We could use this information to write the interaction mean fields for the average level of interaction with

contagious individuals within a given group by summing over the set of possible group states $\{[ijk]_n\}$. For the disease, we write

$$B_{SS}^{(D)} = \frac{\sum_{\{[ijk]_n\}} (n-i-j-k)(id_{SS}^{IS} + kd_{SS}^{II})[ijk]_n}{\sum_{\{[ijk]_n\}} (n-i-j-k)[ijk]_n}, \tag{5}$$

$$B_{SI}^{(D)} = \frac{\sum_{\{[ijk]_n\}} j(id_{SI}^{IS} + kd_{SI}^{II})[ijk]_n}{\sum_{\{[ijk]_n\}} j[ijk]_n}, \tag{6}$$

and for awareness,

$$B_{SS}^{(A)} = \frac{\sum_{\{[ijk]_n\}} (n-i-j-k)(ja_{SS}^{SI} + ka_{SS}^{II})[ijk]_n}{\sum_{\{[ijk]_n\}} (n-i-j-k)[ijk]_n}, \tag{7}$$

$$B_{IS}^{(A)} = \frac{\sum_{\{[ijk]_n\}} i(ja_{IS}^{SI} + ka_{IS}^{II})[ijk]_n}{\sum_{\{[ijk]_n\}} i[ijk]_n} . \tag{8}$$

These expressions can be understood with the following logic. For instance, in the case of $B_{SS}^{(D)}$, a fully susceptible individual in $[S_D S_A]$ is twice as likely to be part of a clique with twice as many susceptible nodes, which is what the $(n-i-j-k)$ factor takes into account. We then simply calculate the average infection terms within each possible clique, i.e., $id_{SS}^{IS} + kd_{SS}^{II}$, and use a similar sum over the clique states in the denominator to normalize the biased distribution.

We now need to follow the evolution of clique states using a general, but complicated, equation:

$$\begin{aligned}
\frac{d}{dt}[ijk]_n &= (i+1)\alpha_D[(i+1)jk]_n + (j+1)\alpha_A[i(j+1)k]_n + (k+1)\{d^{II}\alpha_D[i(j-1)(k+1)]_n + a^{II}\alpha_A[(i-1)j(k+1)]_n\} \\
&\quad - (i\alpha_D + j\alpha_A + kd^{II}\alpha_D + ka^{II}\alpha_A)[ijk]_n + \beta_D(n-i+1-j-k)\{(i-1)d_{SS}^{IS} + kd_{SS}^{II} + \tilde{B}_{SS}^{(D)}\}[(i-1)jk]_n \\
&\quad - \beta_D(n-i-j-k)\{id_{SS}^{IS} + kd_{SS}^{II} + \tilde{B}_{SS}^{(D)}\}[ijk]_n + \beta_A(n-i-j+1-k)\{(j-1)a_{SS}^{SI} + ka_{SS}^{II} + \tilde{B}_{SS}^{(A)}\}[i(j-1)k]_n \\
&\quad - \beta_A(n-i-j-k)\{ja_{SS}^{SI} + ka_{SS}^{II} + \tilde{B}_{SS}^{(A)}\}[ijk]_n + \beta_A(i+1)\{ja_{IS}^{SI} + (k-1)a_{IS}^{II} + \tilde{B}_{IS}^{(A)}\}[(i+1)j(k-1)]_n \\
&\quad - \beta_A i\{ja_{IS}^{SI} + ka_{IS}^{II} + \tilde{B}_{IS}^{(A)}\}[ijk]_n + \beta_D(j+1)\{id_{SI}^{IS} + (k-1)d_{SI}^{II} + \tilde{B}_{SI}^{(D)}\}[i(j+1)(k-1)]_n \\
&\quad - \beta_D j\{id_{SI}^{IS} + kd_{SI}^{II} + \tilde{B}_{SI}^{(D)}\}[ijk]_n,
\end{aligned} \tag{9}$$

which is defined over all non-negative integers $n \geqslant 2$ and $i + j + k \leqslant n$. Equation (9) is coupled to the previous system of ODEs through the mean-field values of *excess* interactions $\tilde{B}_{UV}^{(x)}$, representing interactions with *outside* groups, given by

$$\tilde{B}_{UV}^{(x)} = \left(\frac{\sum_m m(m-1)[UV]_m}{\sum_m m[UV]_m} \right) B_{UV}^{(x)} . \tag{10}$$

The first four terms of Eq. (9) are those corresponding to recoveries; positive for those corresponding to cliques relaxing into $[ijk]_n$ and negative for those where $[ijk]_n$ relaxes into a less infected state. The other terms represent each possible infection event. Notice that creating a k individual implies either removing a i or j, through their infection with contagion 2 or 1, respectively; just as recoveries can create i or j individuals when a k individual recovers from a contagion.

D. Validation

To validate the accuracy of our mean-field description, we run simulations on highly clustered networks where every node belongs to two cliques of size 10. We use this network for two reasons: First, to avoid degree-degree correlations, such that we know that the effect of clustering will be the main structural effect. Second, to feature a realistic local clustering coefficient, C, i.e., the ratio of triangles to pairs of links around a given node, which is here $C = 0.47$. In Fig. 2, we show prevalence for the disease and awareness

over time on a clustered (CS) and an equivalent random graph (ERN), using both Monte Carlo simulations and our ODE system. The accuracy of the mean-field approximations were expected given Refs. [17,18], and for the rest of the paper we therefore rely on the ODE system rather than slower Monte Carlo simulations.

Most importantly, while we know that an awareness campaign or network clustering can both hinder the spread of a disease, it appears that network clustering can actually help a disease spread further when it is competing against a second contagion such as an awareness campaign. This result shows once more that the impacts of different dynamical or structural features can combine in nontrivial ways in models of contagion on networks.

E. Two-step branching process

In Ref. [18], we also introduced a simple criterion to determine whether a clustered structure would spread two synergistic contagions faster. This criterion can conceptually be thought of as a generalization of the basic reproductive number (R_0, the number of secondary infections from an average infectious individual in a completely susceptible population), which is often used to characterize the initial speed of epidemics, which in our case considers two infection steps in order to include clustering. Physically, it can be interpreted as a two-step branching process, as we count the number of
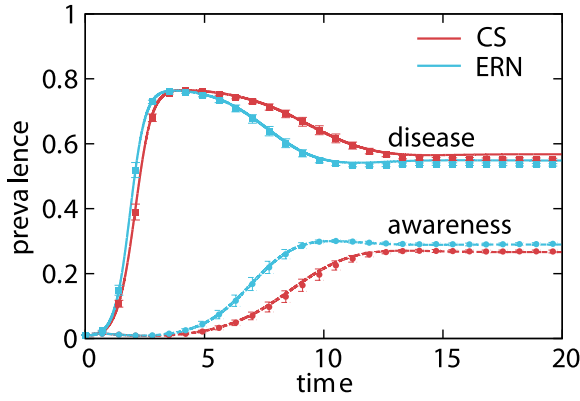
FIG. 2. Parasitic interaction in a population where nodes all belong to two groups of ten nodes. Markers represent average results of Monte Carlo simulations with error bars representing the standard deviation of over 100 runs on networks of 50 000 nodes. Solid curves are obtained by integrating our formalism. Results on the clustered networks are shown in red and those on their equivalent random network are shown in light blue. The top two curves correspond to disease prevalence, and bottom two to awareness prevalence. The dynamics follow $\beta_D = 0.02$, $\beta_A = 0.25$, $\alpha_D = \alpha_A = d^{II} = a^{II} = 1.0$, $d_{UV}^{IS} = 1$, $d_{UV}^{II} = 10$, and $a_{UV}^{XY} = 0.05$ except $d_{SS}^{IS} = a_{SS}^{SI} = 1.0$.

tertiary infections caused by a coinfected individual (i.e., how many second neighbors will be infected).

However, the analogy is imperfect: Different generations of this "branching process" are not independent since we do not distinguish which contagion(s) caused those tertiary infections, and therefore the average branching factor can change from generation to generation. Yet, it proved to be a useful tool in Ref. [18] to identify the net effect of clustering across parameter space; i.e., to determine whether clustering speeds up or slows down disease spread.

We start with a single node infected with both contagions, and denote the average excess degree of recently infected nodes as z_1 (i.e., we assume this node received the contagions from a single neighbor and z_1 describes the average number of *other* neighbors this node is expected to have). For the first step of our criterion, we need to distinguish the probability of transmitting only the disease, only the awareness, or both. Since we ignore the possibility of recovery followed by re-infection, the latter scenario can occur in two ways: either by transmitting both while coinfected; or transmitting the first (or second) while coinfected before recovering from it and then transmitting the second (or first). Summing the two events yields the probability T_{II} of a coinfected transmitting both contagions to a given first neighbor, i.e.,

$$T_{II} = \frac{d_{SS}^{II}\beta_D}{d_{SS}^{II}\beta_D + a_{SS}^{II}\beta_A + d^{II}\alpha_D + a^{II}\alpha_A}\left[\frac{a_{IS}^{II}\beta_A}{a_{IS}^{II}\beta_A + d^{II}\alpha_D + a^{II}\alpha_A} + \frac{d^{II}\alpha_D}{a_{IS}^{II}\beta_A + d^{II}\alpha_D + a^{II}\alpha_A}\left(\frac{a_{IS}^{SI}\beta_A}{a_{IS}^{SI}\beta_A + \alpha_A}\right)\right]$$

$$+ \frac{a_{SS}^{II}\beta_A}{d_{SS}^{II}\beta_D + a_{SS}^{II}\beta_A + d^{II}\alpha_D + a^{II}\alpha_A}\left[\frac{d_{SI}^{II}\beta_D}{d_{SI}^{II}\beta_D + d^{II}\alpha_D + a^{II}\alpha_A} + \frac{a^{II}\alpha_A}{d_{SI}^{II}\beta_D + d^{II}\alpha_D + a^{II}\alpha_A}\left(\frac{d_{SI}^{IS}\beta_D}{d_{SI}^{IS}\beta_D + \alpha_D}\right)\right]. \quad (11)$$

Similarly, a coinfected node can transmit only the disease in two ways, either by infecting while coinfected then recovering before the transmitting awareness or by recovering from awareness before transmitting the disease. Again, summing these events gives the probability T_{IS} of a coinfected transmitting only the disease. The same logic applies to the probability T_{SI} of transmitting only the second contagion. We can thus write

$$T_{IS} = \frac{d_{SS}^{II}\beta_D}{d_{SS}^{II}\beta_D + a_{SS}^{II}\beta_A + d^{II}\alpha_D + a^{II}\alpha_A}\left[1 - \frac{a_{IS}^{II}\beta_A}{a_{IS}^{II}\beta_A + d^{II}\alpha_D + a^{II}\alpha_A} - \frac{d^{II}\alpha_D}{a_{IS}^{II}\beta_A + d^{II}\alpha_D + a^{II}\alpha_A}\left(\frac{a_{IS}^{SI}\beta_A}{a_{IS}^{SI}\beta_A + \alpha_A}\right)\right]$$

$$+ \frac{a^{II}\alpha_A}{d_{SS}^{II}\beta_D + a_{SS}^{II}\beta_A + d^{II}\alpha_D + a^{II}\alpha_A}\left(\frac{d_{SS}^{IS}\beta_D}{d_{SS}^{IS}\beta_D + \alpha_D}\right), \quad (12)$$

$$T_{SI} = + \frac{a_{SS}^{II}\beta_A}{d_{SS}^{II}\beta_D + a_{SS}^{II}\beta_A + d^{II}\alpha_D + a^{II}\alpha_A}\left[1 - \frac{d_{SI}^{II}\beta_D}{d_{SI}^{II}\beta_D + d^{II}\alpha_D + a^{II}\alpha_A} - \frac{a^{II}\alpha_A}{d_{SI}^{II}\beta_D + d^{II}\alpha_D + a^{II}\alpha_A}\left(\frac{d_{SI}^{IS}\beta_D}{d_{SI}^{IS}\beta_D + \alpha_D}\right)\right]$$

$$+ \frac{d^{II}\alpha_D}{d_{SS}^{II}\beta_D + a_{SS}^{II}\beta_A + d^{II}\alpha_D + a^{II}\alpha_A}\left(\frac{a_{SS}^{SI}\beta_A}{a_{SS}^{SI}\beta_A + \alpha_A}\right). \quad (13)$$

In its first neighborhood, we now know that a single coinfected individual will on average cause $z_1 T_{II}$ coinfections, $z_1 T_{IS}$ transmissions of the disease only, and $z_1 T_{SI}$ transmissions of the awareness only. We call those secondary infections. Our two-step branching process then looks at the number of tertiary infections, i.e., the number of transmission events of either contagions in the second neighborhood.

In a clustered network, there is an overlap between the second neighborhood and the first, such that neighbors of the original coinfection can be infected during the second step

of the process if they were not already. Let us consider one of the $z_1 T_{IS}$ first neighbors infected only with the disease and now trying to infect a susceptible node. We know that in its own first neighborhood, a number $(z_1 - 1)C(T_{II} + T_{IS})$ of them are already infected with the same contagion ($z_1 - 1$ is an approximation, equal to its excess degree minus the targeted susceptible node). The fact that a fraction of its neighborhood is already infected by the root node is the negative impact of clustering on the dynamics. However, a number $(z_1 - 1)C(T_{II} + T_{SI})$ are now also aware, such that

they could transmit it to the node of interest and change its transmissibility. This is a potentially positive impact of clustering depending on the nature of the coupling between contagions (e.g., positive for the spread of awareness, negative for the disease itself).

Still considering the same first neighbor infected with the disease only, we need to know the rate at which it is coinfected by one of its $(z_1 - 1)C(T_{II} + T_{SI})$ aware neighbors. Assume that we know the value of that rate, denoted x_D for a coinfection to a diseased node, then the probability of recovery before coinfection would simply be $\alpha_D/(x_D + \alpha_D)$. Since we can also write that probability as every node involved recovering before coinfection, we can require the following equality:

$$\frac{\alpha_D}{x_D + \alpha_D} = \left(\frac{\alpha_D + \alpha_A}{a_{IS}^{SI}\beta_A + \alpha_D + \alpha_A}\right)^{(z_1-1)CT_{SI}} \left(\frac{\alpha_D + a^{II}\alpha_A}{a_{IS}^{II}\beta_A + \alpha_D + a^{II}\alpha_A}\right)^{(z_1-1)CT_{II}}. \tag{14}$$

Note that this self-consistency condition for the effective rate of coinfection is a corrected version of the equivalent (but simpler) condition found in Ref. [18] which contained a transcription error. Furthermore, the same logic here also applies around awareness to describe coinfection events involving nodes that are aware but not sick. We can solve for the effective rates of coinfection through clustering, i.e., x_D and x_A (which mean coinfections rates for nodes that currently have the disease or awareness only, respectively, and obtain

$$x_D = \alpha_D \left[\left(\frac{\alpha_D + \alpha_A}{a_{IS}^{SI}\beta_A + \alpha_D + \alpha_A}\right)^{-(z_1-1)CT_{SI}} \left(\frac{\alpha_D + a^{II}\alpha_A}{a_{IS}^{II}\beta_A + \alpha_D + a^{II}\alpha_A}\right)^{-(z_1-1)CT_{II}} - 1\right], \tag{15}$$

$$x_A = \alpha_A \left[\left(\frac{\alpha_D + \alpha_A}{d_{SI}^{IS}\beta_D + \alpha_D + \alpha_A}\right)^{-(z_1-1)CT_{IS}} \left(\frac{\alpha_A + d^{II}\alpha_D}{d_{SI}^{II}\beta_D + d^{II}\alpha_D + \alpha_A}\right)^{-(z_1-1)CT_{II}} - 1\right]. \tag{16}$$

With these effective rates, we can write the probabilities of a tertiary transmission of either disease or awareness, respectively, $T_{IS}^{(D)}$, $T_{IS}^{(A)}$ if coming from a node initially infected only with the disease and $T_{SI}^{(D)}$ and $T_{SI}^{(A)}$ if coming from a node initially infected only with awareness. We write

$$T_{IS}^{(D)} = [1 - C(T_{IS} + T_{II})]\left[\frac{d_{SS}^{IS}\beta_D}{d_{SS}^{IS}\beta_D + \alpha_D + x_D} + \frac{x_D}{d_{SS}^{IS}\beta_D + \alpha_D + x_D}(T_{IS} + T_{II})\right], \tag{17}$$

$$T_{IS}^{(A)} = [1 - C(T_{SI} + T_{II})]\left[0 + \frac{x_D}{d_{SS}^{IS}\beta_D + \alpha_D + x_D}(T_{SI} + T_{II})\right], \tag{18}$$

$$T_{SI}^{(D)} = [1 - C(T_{IS} + T_{II})]\left[0 + \frac{x_A}{a_{SS}^{SI}\beta_A + \alpha_A + x_A}(T_{IS} + T_{II})\right], \tag{19}$$

$$T_{SI}^{(A)} = [1 - C(T_{SI} + T_{II})]\left[\frac{a_{SS}^{SI}\beta_A}{a_{SS}^{SI}\beta_A + \alpha_A + x_A} + \frac{x_A}{a_{SS}^{SI}\beta_A + \alpha_A + x_A}(T_{SI} + T_{II})\right], \tag{20}$$

where the initial factor is the probability that a given neighbor is currently susceptible and where the two terms in brackets are, respectively, the probability of directly passing the correct contagion before coinfection (sometimes zero if not initially contagious) or of passing it after coinfection. More directly, we can write the probability of a tertiary infection of either disease or awareness from an individual who received both contagion from the original coinfected:

$$T_{II}^{(D)} = [1 - C(T_{IS} + T_{II})](T_{IS} + T_{II}), \tag{21}$$

$$T_{II}^{(A)} = [1 - C(T_{SI} + T_{II})](T_{SI} + T_{II}). \tag{22}$$

From all of these, we write the number of tertiary infections of disease or awareness caused by an original coinfected individual as

$$R_1^{(D)} = z_1^2(T_{IS}T_{IS}' + T_{II}T_{II}'), \tag{23}$$

$$R_1^{(A)} = z_1^2(T_{SI}T_{SI}' + T_{II}T_{II}'). \tag{24}$$

These two R_1 quantities are not generative numbers per se, as the process is not multiplicative for two reasons: (i) we count all tertiary infections, not only nodes in the same coinfected state as our original node and (ii) the next step in the process would imply facing clustering in both the second *and* first neighborhood. Nevertheless, in Ref. [18], it was shown that comparing R_1 values obtained using $C > 0$ to that of an ERN with $C = 0$ allows us to determine whether clustering slows down the dynamics [$R_1(C > 0) < R_1(C = 0)$] or speeds it up [$R_1(C > 0) > R_1(C = 0)$].

III. RESULTS

In this section, we investigate the final epidemic sizes and peak prevalence values within the mean-field model, but also test the usefulness of the branching factor approach in identifying the epidemic threshold of the model as well as the net impact of clustering. Indeed, we can use the previous analysis to (i) identify whether the disease can maintain an outbreak despite clustering and awareness (i.e., if $R_1^{(D)} > 1$) and (ii) evaluate whether a clustered network structure will lead to larger epidemic peaks than an equivalent random network (i.e., if $R_1^{(D)}/R_1^{(A)}$ is larger with $C > 0$). We focus our results on regimes where awareness is subcritical on its own, since we assume that awareness of an extinct disease would
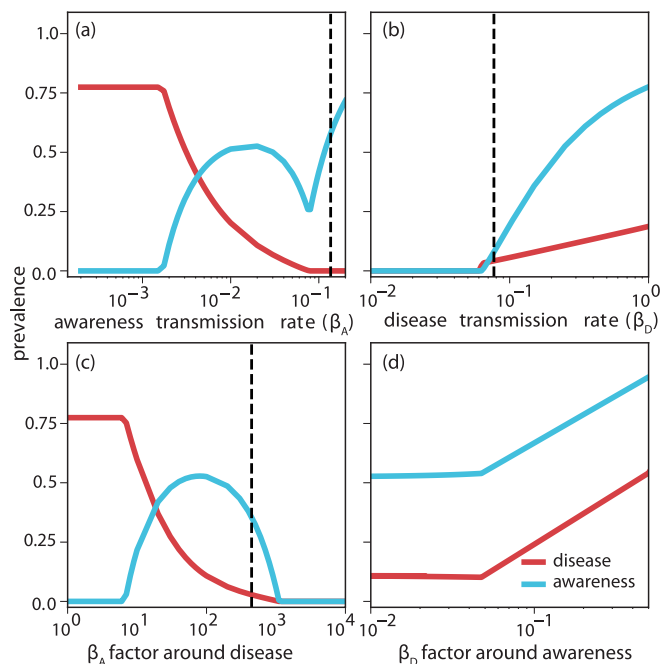
FIG. 3. Using the same clustered network structure and parametrization as in Fig. 2, we now plot the final prevalence (i.e., final steady-state size) of disease and awareness across a range of parameters. (a) We vary the basic transmission rate of awareness. (b) We vary the basic transmission rate of the disease. (c) We vary the factor a affecting awareness transmission around sick individuals. d) We vary the factor d affecting disease transmission around aware individuals. The dotted vertical line marks the analytical epidemic threshold (if any) as approximated by $R_1^{(D)} = 1$. All parameters are fixed to the following values unless we explicitly vary them: $\beta_A = 0.02$, $\beta_D = 0.25$, $\alpha_i = a^{II} = d^{II} = 1$, $a_{UV}^{SI} = 1$, $a_{UV}^{II} = a = 100$, and $d_{UV}^{XY} = d = 0.005$ except $d_{SS}^{IS} = a_{SS}^{SI} = 1.0$.

not propagate through a social network, but is able to spread to a large fraction of the population because of its interaction with the disease.

A. Final sizes

We investigate the robustness of the endemic disease state on clustered networks in Fig. 3. The most interesting case occurs when increasing the transmission rate β_A of awareness can cause disease eradication by depleting the pool of susceptible individuals. At very low β_A, the awareness contagion fails to spread and the disease is left unhindered. At intermediate values of β_A, awareness is able to spread mostly due to its interaction with the disease, meaning it will reach a fraction of those already reached by the disease and fail to invade the susceptible population. In this regime, we find a nonmonotonous relationship between the prevalence and transmission rate of awareness because increasing β_A increases the probability of awareness reaching sick neighbors, while also decreasing the global fraction of sick individuals. After a certain threshold in β_A, the prevalence of the disease falls to zero and awareness then spreads as a regular contagion.

As shown in Fig. 3, the epidemic thresholds predicted by $R_1^{(D)} = 1$ are typically within a factor 2 of the true epidemic threshold. This error is most likely due to the fact

that the analysis is seeded with a coinfected individual and we ignore the fact that awareness and disease are likely to drift apart, benefiting the disease. Our main approximation is in focusing our branching factor analysis on the number of tertiary infections caused by coinfected individuals (both infectious with the disease and aware). However, because our results focus on regimes where awareness is subcritical in the absence of disease, both disease tend to have very similar epidemic thresholds where disease can easily only spread through individuals in $[I_D S_A]$ and not the $[I_D I_A]$ considered by the branching process. A next-order, more involved, analysis would have to consider heterogeneous branching factors for the three types of infectious individuals: $[I_D S_A]$, $[S_D I_A]$, and $[I_D I_A]$.

B. Peak values

In Ref. [18], R_1 was used to determine whether two synergistic diseases would spread faster on a clustered or random network. The idea being that while clustering typically slows down dynamics, there can be an acceleration associated with the synergistic interactions and the benefit of being together by clustering. Here, both clustering in network structure and the interaction with awareness slow down the spread of the disease. We therefore do not expect to find a regime of accelerated disease spread. However, it is possible that clustering slows down awareness more than it slows down the disease, in which case a slower dynamics might still lead to a higher epidemic peak.

In Fig. 4, we vary the transmission rate of the disease and its interaction with awareness while tracking whether $R_1^{(D)}/R_1^{(A)}$ is larger with $C > 0$ (larger epidemic peak on a clustered structure) or with $C = 0$ (larger peak on the equivalent random network). We find that there can indeed be two separate regimes and the branching factor analysis provides a good approximation of where this crossover can occur. For example, Fig. 4(a) presents a parametrization close to this crossover where the clustered network leads to a higher epidemic peak than the equivalent random network, while Fig. 4(b) and 4(c) shows two cases on the other side of the crossover where the peaks on random networks are progressively larger than on the clustered version as we move away from the crossover interface. Finally, Fig. 4(d) shows the ratio of peak values for a whole range of parametrization as well as the parameters used in Fig. 4(a)–4(c) showing a good but approximate agreement between the branching factor analysis and integration of the mean-field formalism.

IV. CONCLUSION

With disease transmission comes the possibility for awareness of the disease and of the risk factors associated with its transmission. Awareness of the disease may cause individuals to respond by reducing their own transmissibility or adopting preventative behaviors. Here, we explored the effects of awareness in a model that looks at both disease and awareness as cocontagions in a parasitic relationship: spread of the disease leads to transmission of awareness which in turn leads to decreased disease prevalence as a result of reduced disease transmission around aware individuals. Our results
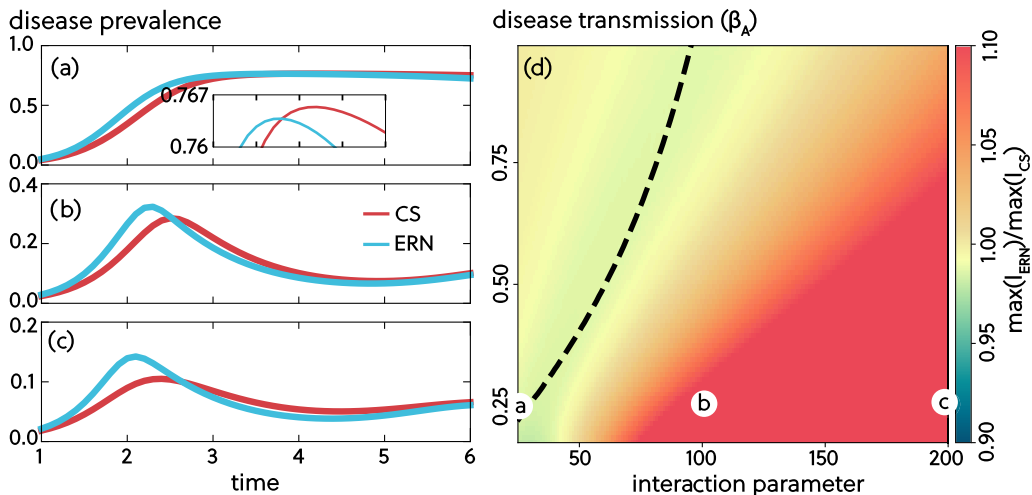
FIG. 4. Differences in peak prevalence values between network structure. We again use the same network structure and parametrization as in Fig. 2, with a fixed $\beta_A = 0.02$, $\alpha_i = a^{II} = d^{II} = 1$, and with other interaction factors parametrized using a single interaction parameter $a_{UV}^{II} = a$ and setting $d_{UV}^{XY} = 1/(2a)$ except $d_{SS}^{IS} = a_{UV}^{SI} = 1.0$. (a)–(c) show the time series of disease prevalence for three sets of parameter values while (d) study a large range of variations in β_2 and ρ. The inset of (a) provides a closer look at the epidemic peak of both time series, showing that the clustered network indeed leads to a slightly higher epidemic peak with an interaction parameter a just below its critical value shown in (d). The color axis in (d) corresponds to the ratio of peak values obtained by integrating the mean-field system on clustered and equivalent random networks. The dotted line shows the crossover regime predicted by our analytical branching factor analysis; it defines the set of parameters for which $R_1^{(D)}/R_1^{(A)}$ is the same on both clustered and random networks. Finally, parameter values used in (a)–(c) are shown with markers on (d).

show that interacting cocontagion models lead to different dynamics depending on the network structure on which they unfold. Characteristic measures such as the final outbreak size and the peak incidence exhibit regimes where they can be higher in networks exhibiting clustering than on equivalent but random network structures.

Altogether, our study highlights once again the need for disease models to go beyond random networks as social clustering can lead to either smaller or larger forecasts depending on the dynamics at play. We showed how interactions between contagions can combine with network structure in nontrivial ways and are therefore especially important to include in disease models. To this end, we have generalized the tools of Ref. [18] to account for more complicated interaction mechanisms. Other generalizations could be considered in future work. We previously mentioned the need for heterogeneous branching factors to account for the fact that individuals infectious with only the disease, awareness, or both, can spread the contagions very differently. Tracking this heterogeneity through multiple epidemic generations to account for clustering would be a complicated calculation, but one should that should lead to much better approximations of the epidemic thresholds. Finally, we focused on homogeneous but clustered contact networks to isolate the impact of clustering on parasitic contagions, but the role of correlations and of the internal structure of communities could also have nontrivial consequences [25,26]. Are nodes with higher membership part of larger or smaller communities than expected? Do they have larger degree within these communities? New heterogeneous mean-field approaches need to be developed to account for these network structures.

Our current approach lead to useful analytical tools, but their development becomes so involved and complicated as

to be almost intractable. This raises the important problem of developing effective models for interacting contagions whose complexity does not grow exponentially with the number of contagions or with the number of interaction mechanisms. Indeed, not only do infectious diseases interact with social contagions such as vaccination and other preventative behaviors, but they also interact, often synergistically, with other biological infections [27–29]. New tools are therefore needed to account for all of these interactions in a tractable and insightful analytical framework [30].

Finally, with these theoretical advances also comes the need for improved data collection on the dynamics of awareness spreading and methods to measure how this materializes into effective preventative behaviors. With most of it now shared on online social media, information and messages regarding public health crises are increasingly important in shaping human behavior during epidemics. Unfortunately, data surrounding that messaging are not readily available to researchers and public health officials, even if we know it interacts in critical ways with our models and forecasts. The parallel development of theoretical frameworks and of data sharing protocols for social messaging related to public health crises will be invaluable going forward. Public awareness is an integral part of public health and advances to address its social media dimension should be integrated into existing public health surveillance systems.

[1] R. M. Anderson, R. M. May, and B. Anderson, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1992).

[2] S. Funk, M. Salathé, and V. A. Jansen, Modelling the influence of human behaviour on the spread of infectious diseases: A review, J. R. Soc. Interface **7**, 1247 (2010).

[3] A. M. Parker, R. Vardavas, C. S. Marcum, and C. A. Gidengil, Conscious consideration of herd immunity in influenza vaccination decisions, Am. J. Prev. Med. **45**, 118 (2013).

[4] W. Bruine de Bruin, A. M. Parker, M. Galesic, and R. Vardavas, Reports of social circles' and own vaccination behavior: A national longitudinal survey, Health Psychol. **38**, 975 (2019).

[5] J. T. F. Lau, X. Yang, H. Tsui, and J. H. Kim, Monitoring community responses to the SARS epidemic in Hong Kong: From day 10 to day 62, J. Epidemiol. Community Health **57**, 864 (2003).

[6] J. T. F. Lau, X. Yang, H. Y. Tsui, and E. Pang, SARS related preventive and risk behaviours practised by Hong Kong-mainland China cross border travellers during the outbreak of the SARS epidemic in Hong Kong, J. Epidemiol. Community Health **58**, 988 (2004).

[7] G. J. Rubin, R. Amlôt, L. Page, and S. Wessely, Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: Cross sectional telephone survey, BMJ **339**, b2651 (2009).

[8] S. Funk, E. Gilad, C. Watkins, and V. A. Jansen, The spread of awareness and its impact on epidemic outbreaks, Proc. Natl. Acad. Sci. USA **106**, 6872 (2009).

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

[10] V. Marceau, P.-A. Noël, L. Hébert-Dufresne, A. Allard, and L. J. Dubé, Modeling the dynamical interaction between epidemics on overlay networks, Phys. Rev. E **84**, 026105 (2011).

[11] F. Fu, N. A. Christakis, and J. H. Fowler, Dueling biological and social contagions, Sci. Rep. **7**, 43634 (2017).

[12] C. Granell, S. Gómez, and A. Arenas, Competing spreading processes on multiplex networks: Awareness and epidemics, Phys. Rev. E **90**, 012808 (2014).

[13] C.-j. Fan, Y. Jin, L.-a. Huo, C. Liu, Y.-p. Yang, and Y.-q. Wang, Effect of individual behavior on the interplay between awareness and disease spreading in multiplex networks, Physica A **461**, 523 (2016).

[14] M. Scatà, A. Di Stefano, P. Liò, and A. La Corte, The impact of heterogeneity and awareness in modeling epidemic spreading on multiplex networks, Sci. Rep. **6**, 37105 (2016).

[15] W. Wang, Q.-H. Liu, S.-M. Cai, M. Tang, L. A. Braunstein, and H. E. Stanley, Suppressing disease spreading by using information diffusion on multiplex networks, Sci. Rep. **6**, 29259 (2016).

[16] C. Zheng, C. Xia, Q. Guo, and M. Dehmer, Interplay between SIR-based disease spreading and awareness diffusion on multiplex networks, J. Parallel Distrib. Comput. **115**, 20 (2018).

[17] L. Hébert-Dufresne, P.-A. Noël, V. Marceau, A. Allard, and L. J. Dubé, Propagation dynamics on networks featuring complex topologies, Phys. Rev. E **82**, 036115 (2010).

[18] L. Hébert-Dufresne and B. M. Althouse, Complex dynamics of synergistic coinfections on realistically clustered networks, Proc. Natl. Acad. Sci. USA **112**, 10551 (2015).

[19] M. E. J. Newman, Properties of highly clustered networks, Phys. Rev. E **68**, 026121 (2003).

[20] S. Fortunato and D. Hric, Community detection in networks: A user guide, Phys. Rep. **659**, 1 (2016).

[21] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, Link communities reveal multiscale complexity in networks, Nature **466**, 761 (2010).

[22] J. Xie, S. Kelley, and B. K. Szymanski, Overlapping community detection in networks: The state-of-the-art and comparative study, ACM Comput. Surv. **45**, 1 (2013).

[23] I. Z. Kiss and D. M. Green, Comment on "Properties of highly clustered networks", Phys. Rev. E **78**, 048101 (2008).

[24] I. Z. Kiss, J. Cassell, M. Recker, and P. L. Simon, The impact of information transmission on epidemic outbreaks, Math. Biosci. **225**, 1 (2010).

[25] J.-G. Young, L. Hébert-Dufresne, A. Allard, and L. J. Dubé, Growing networks of overlapping communities with internal structure, Phys. Rev. E **94**, 022317 (2016).

[26] D. Vasques Filho and D. R. J. O'Neale, Transitivity and degree assortativity explained: The bipartite structure of social networks, Phys. Rev. E **101**, 052305 (2020).

[27] S. B. Halstead, Neutralization and antibody-dependent enhancement of dengue viruses, Adv. Virus Res. **60**, 421 (2003).

[28] S. Shrestha, B. Foxman, D. M. Weinberger, C. Steiner, C. Viboud, and P. Rohani, Identifying the interaction between influenza and pneumococcal pneumonia using incidence data, Sci. Transl. Med. **5**, 191ra84 (2013).

[29] S. Nickbakhsh, C. Mair, L. Matthews, R. Reeve, P. C. Johnson, F. Thorburn, B. von Wissmann, A. Reynolds, J. McMenamin, R. N. Gunson *et al.*, Virus–virus interactions impact the population dynamics of influenza and the common cold, Proc. Natl. Acad. Sci. USA **116**(52), 27142 (2019).

[30] L. Hébert-Dufresne, S. V. Scarpino, and J.-G. Young, Macroscopic patterns of interacting contagions are indistinguishable from social reinforcement, Nat. Phys. **16**, 426 (2020).