

Epidemic outbreaks in two-scale community networks

Stefano Bonaccorsi

*Mathematics Department, University of Trento, via Sommarive 14, 38123 Povo (Trento), Italy*Stefania Ottaviano* and Francesco De Pellegrini
CREATE-NET, via alla Cascata 56/d, 38123 Trento, Italy

Annalisa Socievole

DIMES, University of Calabria, via Ponte P. Bucci, 87036 Rende (Cosenza), Italy

Piet Van Mieghem

EEMCS, Delft University of Technology, Mekelweg 4 2628 CD Delft, The Netherlands

(Received 1 April 2014; published 21 July 2014)

We consider a model for the diffusion of epidemics in a population that is partitioned into local communities. In particular, assuming a mean-field approximation, we analyze a continuous-time susceptible-infected-susceptible (SIS) model that has appeared recently in the literature. The probability by which an individual infects individuals in its own community is different from the probability of infecting individuals in other communities. The aim of the model, compared to the standard, nonclustered one, is to provide a compact description for the presence of communities of local infection where the epidemic process is faster compared to the rate at which it spreads across communities. Ultimately, it provides a tool to express the probability of epidemic outbreaks in the form of a metastable infection probability. In the proposed model, the spatial structure of the network is encoded by the adjacency matrix of clusters, i.e., the connections between local communities, and by the vector of the sizes of local communities. Thus, the existence of a nontrivial metastable occupancy probability is determined by an epidemic threshold which depends on the clusters' size and on the intercommunity network structure.

DOI: [10.1103/PhysRevE.90.012810](https://doi.org/10.1103/PhysRevE.90.012810)

PACS number(s): 89.75.Fb, 89.20.Ff

I. INTRODUCTION

Network models are commonly used to describe interactions which occur between *agents*, e.g., in the digital domain or in a biological environment. The spatial structure of the population and the dynamics of interactions play a central role in determining the evolution of such systems. Apart from recent epidemic studies on interdependent networks [1], the analysis of epidemic diffusion in networks has been focused, primarily, on the behavior of a single *community of agents*, whose spatial interactions are described by the correspondent network structure. Several studies have been characterizing the interplay of interconnected communities working possibly at different space and/or time scales (e.g., [2–9]). In these settings, the whole population is partitioned into communities (also called clouds, households, clusters, or subgraphs).

Several works also account for the effect of migration between households [10–12]. Conversely, the model we are interested in suits better the diffusion of computer viruses or stable social communities, which do not change during the infection period; hence we do not consider migration.

In this paper, we consider the following situation (see Fig. 1): The whole population, consisting of N individuals, is distributed over c communities, or clouds, whose interconnection pattern is described by a $c \times c$ adjacency matrix A . Each community can be thought of as a local population containing two types of individuals: S (susceptible) and I (infected).

There are two infection rates: the global, *intercommunity* infection rate β_G , for infecting individuals among different communities, and the local, *intracommunity* infection rate β_L for infecting individuals in the same community or cloud. The model captures the case when the rate of intracommunity infections among individuals within clouds is much larger than the rate of community infections, e.g., due to the fact that contacts across communities occur at a much smaller rate.

We focus on a continuous-time susceptible-infected-susceptible (SIS) model described via Markov theory. In the exact continuous-time Markovian SIS model [13,14], the state of each agent is described by a Bernoulli random variable $X_i(t)$, such that $X_i(t) = 0$ if the agent is healthy and $X_i(t) = 1$ if it is infected. The probability that an agent is infected is $v_i(t) = \mathbb{P}(X_i(t) = 1)$ and the probability that it is cured is $\mathbb{P}(X_i(t) = 0) = 1 - v_i(t)$. In a simplified, homogeneous model, the curing process of each node is a Poisson process with given rate δ and the infection process is a Poisson process with rate β , each process being independent of the others.

For a network with finite size N , the exact SIS Markov process will always converge towards the absorbing zero state or overall healthy state, in which the virus has disappeared from the network. This absorbing zero state is the unique stationary state of the SIS Markov chain. Nevertheless, prior to absorption [15], the SIS process approaches what appears to be a *quasistationary distribution* that is different from the disease-free equilibrium.

Such a quasistationary distribution is thus obtained by conditioning on the fact that there is no absorption (see [16,17] for a deeper explanation). This distribution can also be regarded as the *limiting conditional distribution*, useful in

*Also at the Mathematics Department, University of Trento, Trento, Italy.

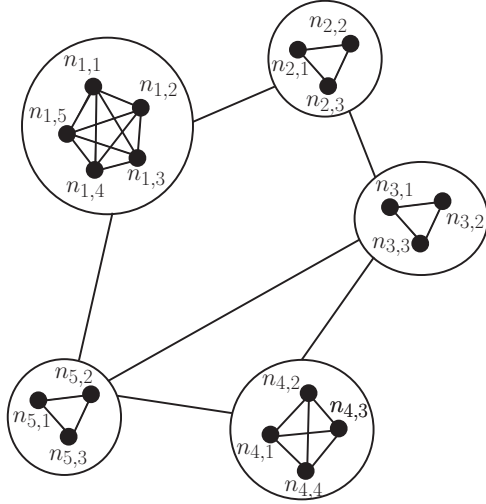


FIG. 1. The two-scale community model: the contagion spreads within each cluster node with intracuster infection rate β_L and among clusters with intercluster infection rate β_G . A link between two clusters means that each node in one cluster is linked with all nodes in the other cluster.

representing the long-term behavior of the process “that in some sense terminates, but appears to be stationary over any reasonable time scale” [18].

After a mean-field approximation, depending on the system operating point, a limiting occupancy probability appears as the second solution of the nonlinear system, which exists, apart from the zero-vector solution that reflects the absorbing state. This nonzero steady-state solution is termed in [13] a *metastable state*. Hence, the metastable state can be understood as the nodes’ ergodic infection probability of a *modified* Markovian SIS model (see [16]), where either the absorbing state is removed or a self-infection probability is imposed as described at the end of this section. Despite the fact that the metastable state is actually a vector of the nodes’ infection probabilities, it can be seen as an apparent steady state, in the sense that, before disappearing, the system remains in that metastable state for a sufficiently long time.

This long-term behavior also justifies the presence of a characteristic epidemic threshold τ_c , mentioned by many authors, (e.g., [19–21]). In a single network with constant infection and curing rates β and δ , respectively, the *effective infection rate* is defined as the ratio $\tau := \beta/\delta$: if the effective infection rate τ is above the epidemic threshold, i.e., $\tau > \tau_c$, a nonzero fraction of the nodes are infected, while, for $\tau \leq \tau_c$, the epidemic dies out.

Strictly speaking, in the exact Markovian SIS model on a finite graph, the epidemic threshold is undefined, since the system is expected to certainly reach the unique absorbing state. In particular, for small networks (i.e., $N < 10$), the time-dependent behavior of the fraction of infected nodes tends towards the absorbing state in an observable time. However, for large N , numerical simulations have revealed that, when $\tau > \tau_c$, the overall healthy state is reached only after an unrealistically long time [22,23], so that the exact steady state is hardly ever reached in real networks, while the metastable state reflects the observed viral behavior fairly

well [24]. Below the critical threshold ($\tau < \tau_c$), the infection vanishes exponentially fast in time.

The continuous-time Markov SIS model on networks can be expressed exactly in terms of a with 2^N states [13,25]. The state space, unfortunately, increases exponentially with N , which poses severe limitations in order to determine the set of solutions for large, real networks. A mean-field approximation for the exact continuous-time SIS model, called the N -intertwined mean-field approximation (NIMFA), was proposed recently in the literature [13,24,26,27]. Essentially, the NIMFA makes one approximation and assumes independence between the infection probability of any pair of nodes in the network; thus, that $\Pr[X_i(t) = 1, X_j(t) = 1] = \Pr[X_i(t) = 1] \Pr[X_j(t) = 1]$. The NIMFA results in a set of N nonlinear equations specifying the time change of the infection probability of a node. Hence, the NIMFA provides a computational advantage compared to the exact model, because it reduces the original system of 2^N linear equations to a system of N nonlinear ones.

The NIMFA model sets the SIS epidemic threshold for the effective infection rate τ at the level $\tau_c^{(1)}$ where

$$\tau_c^{(1)} = \frac{1}{\lambda_1(A)}, \quad (1)$$

and $\lambda_1(A)$ is the principal eigenvalue of the adjacency matrix A of the community network.

Although it was conjectured intuitively already in [13] that an infection somewhere in the network cannot decrease the infection at another place, the recent demonstration in [28] that $\Pr[X_i(t) = 1, X_j(t) = 1] \geq \Pr[X_i(t) = 1] \Pr[X_j(t) = 1]$ in the exact, Markovian SIS process was the missing part in the proof that $\tau_c^{(1)} \leq \tau_c$ or $\tau_c = \alpha \tau_c^{(1)}$ with $\alpha \geq 1$. Moreover, that same demonstration also leads to the property that the NIMFA upper bounds the nodal infection probability [14]. From a practical point of view, computations with the NIMFA thus always result in an upper bound of the infection in a network and basing the engineering of a network on NIMFA values safeguards the design against malware spread. However, for which networks with N nodes the NIMFA is worst, in the sense that $\alpha = \frac{\tau_c}{\tau_c^{(1)}}$ is largest, remains a fundamental open question. A general criterion that assesses the goodness of a mean-field approximation such as the NIMFA for a given graph is believed to be a currently missing, essential building block in the SIS epidemic theory on networks.

Our NIMFA-like approximation is validated here by comparison with the exact SIS model. From the operative standpoint, a method to calculate the actual metastable state of the SIS model is obtained by determining the steady state of the ε -SIS model [25,27]. The ε -SIS model generalizes the SIS model by adding nodal self-infection. Thus, besides receiving the infection from an infected neighbor with rate β , a node can also itself produce a virus with rate ε . For $\varepsilon = 0$, the ε -SIS model corresponds to the classical SIS model.

The ε -SIS model [25] has no absorbing state, corresponding to an irreducible Markov process on a finite state space. Markov theory [14,29] guarantees that irreducibility (and aperiodicity) results in a unique steady state that, by considering appropriate and small values of $\varepsilon > 0$, can be made arbitrarily close to the metastable state of the original SIS model. Hence,

just as in [27], the steady state of the ε -SIS model will be compared here to the metastable state of our model.

II. THE EPIDEMIC MODEL

As introduced before, in this paper we investigate the spread of an epidemic through a network structure with local communities. Each local community contains k_j individuals, for $j = 1, \dots, c$, where c is the number of communities in the corresponding graph. The total number of individuals in the network equals $N = \sum_{j=1}^c k_j$.

When we consider the community level, we refer to the community interconnection network. It is described by an adjacency matrix $A = (a_{ij})$ with $c \times c$ entries. Also, we let node n_j represent the j th community. When we consider the individual level, we denote each individual as a node $n_{i,j}$, where j is the index of the community or cloud and i is the index of the node inside the j th community. Inside each community, we assume all nodes are connected among themselves. Hence, we assume that the internal structure of a cloud is a full mesh or complete graph. A model with a full mesh structure of communities appears natural in the case of small-sized clusters k_j , because the members of a community usually know each other (for example, friends in a certain club, members of the same family, employees in the same department, etc.). Thus, nodes $n_{s,j}$ and $n_{r,j}$ are always connected (for $s \neq r$) and nodes $n_{r,j}$ and $n_{s,l}$ (for $j \neq l$) are connected if $a_{jl} \neq 0$. The full mesh assumption encodes in the model the fact that the contact rate between individuals belonging to the same community is higher than across communities.

The infection probability $v_{i,j}$ of the individual $n_{i,j}$ is described by the mean-field equation

$$\begin{aligned} \frac{d}{dt} v_{i,j}(t) = & \beta_G [1 - v_{i,j}(t)] \sum_{m=1}^c a_{jm} \sum_{l=1}^{k_m} v_{l,m}(t) \\ & + \beta_L [1 - v_{i,j}(t)] \sum_{l=1; l \neq i}^{k_j} v_{l,j}(t) - \delta v_{i,j}(t). \end{aligned} \quad (2)$$

In words, the time derivative of the infection probability of node $n_{i,j}$ consists of two competing processes:

(a) While healthy, with probability $[1 - v_{i,j}(t)]$, all infected neighbors of the node $n_{i,j}$ try to infect it with rate β_L (β_G), according to whether they belong to the same local community or to an external community. The term $\sum_{l=1; l \neq i}^{k_j} v_{l,j}(t)$ computes the expected number of infected neighbors in the local community, while $\sum_{m=1}^c a_{jm} \sum_{l=1}^{k_m} v_{l,m}(t)$ is the expected number of infected individuals in the other clouds connected to cloud j .

(b) While infected, with probability $v_{i,j}(t)$, node $n_{i,j}$ is cured at rate δ .

Our main interest in the analysis of Eq. (2) is to discuss the epidemic threshold for the effective infection rate. However, the notion of effective infection rate will be defined anew, since now it must depend on both β_G and β_L . We show that in regular graphs, the epidemic threshold can be explicitly computed in terms of the infection rates β_G and β_L , the curing rate δ , and the structure of the network, which is specified by the adjacency

matrix A . For more complex networks, we shall find that this value depends also on the distribution of the population among the different clouds. In both cases, we find that above a critical level, there exists a metastable probability distribution, which corresponds to an endemic epidemic state for the system.

We may simplify the problem, using the symmetries of the full mesh assumption, by noticing that all nodes inside a community have the same infection probability. After setting $v_j(t) = v_{i,j}(t) = v_{l,j}(t)$ for all $i, l = 1, \dots, k_j$ in (2), we obtain

$$\begin{aligned} \frac{d}{dt} v_j(t) = & \beta_G [1 - v_j(t)] \sum_{m=1}^c a_{jm} k_m v_m(t) \\ & + (k_j - 1) \beta_L [1 - v_j(t)] v_j(t) - \delta v_j(t). \end{aligned} \quad (3)$$

We now restate the previous equation (3) in matrix form. We introduce the vector $V(t) = (v_1(t), \dots, v_c(t))$; let further $I = I_c$ denote the $c \times c$ identity matrix and $\text{diag}(\alpha) = \text{diag}(\alpha_1, \dots, \alpha_c)$ be the diagonal $c \times c$ matrix whose elements on the principal diagonal are given by the entries of the vector $\alpha = (\alpha_1, \dots, \alpha_c)$. Then we obtain the following representation of (3):

$$\begin{aligned} \frac{d}{dt} V(t) = & -\delta V(t) + \beta_G \{I_c - \text{diag}[V(t)]\} A \text{diag}(k_j) V(t) \\ & + \beta_L \{I_c - \text{diag}[V(t)]\} \text{diag}(k_j - 1) V(t) \\ = & -\delta V(t) + \beta_G \left(A \text{diag}(k_j) + \frac{\beta_L}{\beta_G} \text{diag}(k_j - 1) \right) \\ & \times V(t) - \text{diag}[V(t)] \beta_G \left(A \text{diag}(k_j) \right. \\ & \left. + \frac{\beta_L}{\beta_G} \text{diag}(k_j - 1) \right) V(t). \end{aligned}$$

Introducing the weighted adjacency matrix

$$A_{L,G} = \left(A \text{diag}(k_j) + \frac{\beta_L}{\beta_G} \text{diag}(k_j - 1) \right),$$

whose elements are, in general, not binary (zero or 1), we obtain

$$\frac{d}{dt} V(t) = -\delta V(t) + \beta_G A_{L,G} V(t) - \beta_G \text{diag}[V(t)] A_{L,G} V(t), \quad (4)$$

which has precisely the same form as the NIMFA governing equation introduced first in [26].

III. STEADY STATES

In the following, we focus on the steady state $V_\infty = (v_{j;\infty})_{j=1, \dots, c}$, which is defined by the following conditions:

$$v_{j;\infty} = \lim_{t \rightarrow \infty} v_j(t) \quad \text{and} \quad 0 = \lim_{t \rightarrow \infty} \frac{d}{dt} v_j(t).$$

Passing to the limit in (3) we obtain

$$\begin{aligned} \delta v_{j;\infty} = & \beta_G (1 - v_{j;\infty}) \sum_{m=1}^c a_{jm} k_m v_{m;\infty} \\ & + \beta_L (k_j - 1) v_{j;\infty} (1 - v_{j;\infty}). \end{aligned} \quad (5)$$

It follows that, besides the trivial solution $v_{j;\infty} = 0$, there possibly exist other positive solutions, reflecting the metastable state that we are interested in.

We recall that, with respect to the original SIS Markov model, the NIMFA approximation (2) represents an upper bound to the infection probability of a node. Thus, under the assumption that the graph is connected, a necessary condition for the existence of a nontrivial metastable solution [13] is $v_{j;\infty} > 0, \forall j = 1, \dots, c$.

A. Inhomogeneous systems

We now resort to the framework proposed in [30]: we consider the case of general networks, where the degree of node n_j is denoted by d_j and the size of the population in community n_j is k_j . We write $\kappa = (k_1, \dots, k_n)$.

The metastable state is calculated from (4) as the solution of

$$\delta V_\infty = \beta_G A_{L,G} V_\infty - \beta_G \text{diag}(V_\infty) A_{L,G} V_\infty. \quad (6)$$

With the exclusion of extremal cases (as for instance the case of curing rate $\delta = 0$), the metastable state corresponds to $0 < v_{j;\infty} < 1$. Hence from (6) we get

$$\beta_G A_{L,G} V_\infty = \text{diag}\left(\frac{\delta}{1 - v_{j;\infty}}\right) V_\infty.$$

From the definition of $A_{L,G}$ we can write also

$$A \text{diag}(\kappa) V_\infty = \text{diag}\left(\frac{1}{\frac{\beta_G}{\delta}(1 - v_{j;\infty})} - \frac{(k_j - 1)\beta_L}{\beta_G}\right) V_\infty. \quad (7)$$

The Laplacian of the underlying network is the matrix $Q = \text{diag}(d_j) - A$; Q is a $c \times c$ singular, semipositive defined matrix, and the eigenvector corresponding to the zero eigenvalue is the all-1 vector $u = (1, \dots, 1)^T$. We show how the above construction can be adapted in our case. We denote $A_\kappa := A \text{diag}(\kappa)$, i.e., the matrix obtained from A by multiplying the j th column by $k_j, j = 1, \dots, c$. We introduce the *modified Laplacian* matrix as

$$Q_\kappa(\alpha) = \text{diag}(\alpha) - A_\kappa, \quad \alpha = (\alpha_1, \dots, \alpha_n)^T.$$

It should be noticed that A_κ is not a symmetric matrix, unless all the communities have the same population $k_j = k$.

Next, we write the previous relation (7) by means of the modified Laplacian defined by the vector of entries

$$\alpha_j = \frac{1}{\frac{\beta_G}{\delta}(1 - v_{j;\infty})} - \frac{(k_j - 1)\beta_L}{\beta_G},$$

to get the (nonlinear) equation

$$Q_\kappa\left(\frac{1}{\frac{\beta_G}{\delta}(1 - v_{j;\infty})} - \frac{(k_j - 1)\beta_L}{\beta_G}\right) V_\infty = 0. \quad (8)$$

Hence, if it exists, a nonzero steady-state vector V_∞ is an eigenvector of the eigenvalue zero. In order to find conditions that imply the existence of the zero eigenvalue for the modified Laplacian $Q_\kappa(\alpha)$, we apply Gerschgorin's theorem [29] to both A_κ and A_κ^T .

Corollary 1. Every eigenvalue of the modified Laplacian $Q_\kappa(\alpha)$ lies in (at least) one of the circular disks with center α_j

and radius R_i , where

$$R_i = \max\left(\sum_{i \neq j} k_j a_{ij}, k_i \sum_{j \neq i} a_{ji}\right) = \max\left(\sum_{j \neq i} k_j a_{ji}, k_i d_i\right)$$

and in the previous formula d_i is the degree of community n_i .

B. General graphs with homogeneous partition into households

As mentioned above, in the case when all the households have the same population $k_j = k$, the matrix A_κ , and hence also the matrix $Q_\kappa(\alpha)$ is real symmetric, which implies that all the eigenvalues are real numbers. In this setting, Corollary III A implies that a necessary condition for the existence of a zero eigenvalue is that for some $j = 1, \dots, c$, the inequality

$$\alpha_j - R_j < 0 < \alpha_j + R_j$$

holds. We can further express the above inequality as follows:

$$\frac{1}{\frac{\beta_G}{\delta}(1 - v_{j;\infty})} - \frac{(k - 1)\beta_L}{\beta_G} - R_j < 0,$$

$$\frac{1}{\frac{\beta_G}{\delta}(1 - v_{j;\infty})} - \frac{(k - 1)\beta_L}{\beta_G} + R_j > 0,$$

which, after a little algebra, leads to

$$\frac{1}{(1 - v_{j;\infty})} < (k_j - 1) \frac{\beta_L}{\delta} + R_j \frac{\beta_G}{\delta},$$

$$\frac{1}{(1 - v_{j;\infty})} > (k_j - 1) \frac{\beta_L}{\delta} - R_j \frac{\beta_G}{\delta}.$$

The first inequality implies that

$$v_{j;\infty} < 1 - \frac{\delta}{(k - 1)\beta_L + R_j \beta_G}$$

and since we require $v_{j;\infty} > 0$, we obtain the result that existence of an epidemic state implies

$$\frac{(k - 1)\beta_L + k d_j \beta_G}{\delta} > 1. \quad (9)$$

The second inequality leads to no further conditions. Actually, if the right-hand side is negative, i.e., in the regime $(k - 1)\beta_L < k d_j \beta_G$, the inequality is trivially satisfied, since we require $v_{j;\infty} < 1$. Otherwise, in the regime $(k - 1)\beta_L > k d_j \beta_G$, we get $1 > \frac{\delta}{(k - 1)\beta_L - k d_j \beta_G}$ which leads to the lower bound for the probability $v_{j;\infty} > 1 - \frac{\delta}{(k - 1)\beta_L - k d_j \beta_G}$.

Theorem 1. For a generic network with local clouds of constant cloud size $k_j = k$, a sufficient condition for the uniqueness of the trivial epidemic steady state is the following:

$$\frac{d_{\max} \beta_G + (1 - \frac{1}{k}) \beta_L}{\delta} < \frac{1}{k}.$$

Example: Regular, homogeneous graphs. We consider here the special case of a regular graph, which will enable us to explain the difference of this model with respect to the case when the partition into communities is not considered (compare for instance [13]). A regular graph [29] is a graph where each node has the same degree d .

In this section, we assume that the block size is uniform among the communities, i.e., $k_j = k$. These assumptions

simplify (5), since it follows that $v_{j;\infty}$ is constant in j . Hence, we can compute

$$v_\infty = 1 - \frac{\delta}{(k-1)\beta_L + k\beta_G d},$$

which shows that only the trivial steady state exists provided that

$$(k-1)\beta_L + k\beta_G d < \delta. \quad (10)$$

Let us consider $k = 1$ in our construction: this corresponds to the case when individuals populate a regular network without a community structure. We observe that the condition (10) for the existence of a nontrivial solution corresponds to the steady state given in [24] by

$$v_\infty = 1 - \frac{1}{d\tau}.$$

However, it is important to link the existence of a steady state, which requires $\tau > \frac{1}{d}$, to the structure of the network. This relation is given by the following property (see [29, art. 74]): the largest eigenvalue $\lambda_1(A)$ of the adjacency matrix for a regular graph equals $\lambda_1(A) = d$. Then, from (10), we have the following sufficient condition for the uniqueness of the trivial epidemic steady state for a regular, homogeneous graph:

$$\frac{\lambda_1(A)\beta_G + \left(1 - \frac{1}{k}\right)\beta_L}{\delta} < \frac{1}{k}. \quad (11)$$

We see at once that this formula is equivalent to (1) in the case $k = 1$ and that the influence of the size k of the community is twofold: k increases the value of the left-hand side (increasing the influence of local contacts) and decreases the right-hand side (as it measures the probability of infecting one element of the community).

C. Calculating the metastable state

In order to compute the metastable state, we return to (6) and we write

$$\delta V_\infty = \beta_G \text{diag}(1 - V_\infty) A_{L,G} V_\infty.$$

Recall that the inverse of a positive diagonal matrix is the diagonal matrix with inverse elements on the diagonal; further, since δ is a constant, we have

$$\begin{aligned} \text{diag}(1 - V_\infty)^{-1} V_\infty &= \left(\frac{v_{1;\infty}}{1 - v_{1;\infty}}, \dots, \frac{v_{c;\infty}}{1 - v_{c;\infty}} \right)^T \\ &= u - \left(\frac{1}{1 - v_{1;\infty}}, \dots, \frac{1}{1 - v_{c;\infty}} \right)^T, \end{aligned} \quad (12)$$

where u is the all-1 vector $u = (1, \dots, 1)^T$. After some manipulations, the above equations become

$$\text{diag}(1 - V_\infty)^{-1} u = \left[I - \frac{\beta_G}{\delta} A_{L,G} \text{diag}(V_\infty) \right] u$$

which, in the scalar form, becomes

$$v_{j;\infty} = 1 - \frac{1}{1 + \frac{\beta_G}{\delta} \sum_{h=1}^c (A_{LG})_{jh} v_{h;\infty}}.$$

This formula was already introduced in a similar context in [13]. Also, according to the method proposed there, it is possible to solve the above equation by iteration of the formula

$$z_{j;m+1} = f(z_{1;m}, \dots, z_{c;m}) = 1 - \frac{1}{1 + \frac{\beta_G}{\delta} \sum_{h=1}^c (A_{LG})_{jh} z_{h;m}}.$$

In general, an algorithm based on the above formula converges rapidly to the solution or, equivalently, it is sufficient to perform a limited number of iterations of the above formula in order to obtain a good approximation for the state value.

D. Inhomogeneous cloud distribution

In this section, we extend previous results to the case of different numbers of elements in the communities. The starting point is Eq. (7); however, instead of using Corollary III A directly, we first state the problem in terms of the vector $W_\infty = \text{diag}(\kappa) V_\infty$. We consider the modified Laplacian matrix with respect to the adjacency matrix A :

$$Q(\alpha) = \text{diag}(\alpha) - A,$$

where

$$\alpha_j = \frac{1}{\frac{\beta_G}{k_j \delta} (1 - v_{j;\infty})} - \frac{(k_j - 1) \beta_L}{k_j \beta_G}.$$

The existence of a nontrivial steady state V_∞ requires that W_∞ is a nontrivial eigenvector related to the eigenvalue 0. With the same reasoning as above and taking into account Corollary III A, this implies that, for at least one j ,

$$\alpha_j - R_j < 0 < \alpha_j + R_j.$$

Clearly, in this case the radius R_j is expressed in terms of the adjacency matrix A , and hence

$$R_j = d_j = \text{the degree of community } n_j.$$

Therefore, a sufficient condition for the uniqueness of the trivial epidemic steady state is the following:

$$\frac{1}{\frac{\beta_G}{\delta} (1 - v_{j;\infty})} - (k_j - 1) \frac{\beta_L}{\beta_G} > d_j k_j$$

and, since it must be that $0 < v_{j;\infty} < 1$, we have proved the following result.

Theorem 2. For a generic network with local clouds of nonconstant size $\{k_j\}$, a sufficient condition for the uniqueness of the trivial epidemic steady state is the following:

$$\forall j = 1, \dots, c, \quad \frac{d_j \beta_G + \left(1 - \frac{1}{k_j}\right) \beta_L}{\delta} < \frac{1}{k_j}.$$

The above result confirms the intuition that just the presence of communities implies an increase of the probability of persistence of the epidemics. Actually, it is sufficient to compare the condition in Theorem 2 for $k_j > 1$ with the condition $d_j \beta_G / \delta < 1$ that results in the case $k_j \equiv 1$. Furthermore, the same formula implies that the higher the value of β_L , the smaller is the region of extinction of epidemics.

Finally, the result in Theorem 2 becomes apparent in the limiting cases. Suppose, for instance, that both β_L and β_G converge to zero. Then the inequality in Theorem 2 is trivially

satisfied and the system is in the region of extinction for the epidemics. Conversely, if they both converge to infinity, then the system enters into the region of persistence.

IV. NETWORK WITH INHOMOGENEOUS CURING RATES

In this section, we proceed one step of generalization further and we assume that, apart from the size of the community, we can modify the curing rate for each local community, i.e., we assume that δ_j is the curing rate for the community n_j . In order to handle this case, we introduce the matrix $\Gamma = \text{diag}(\delta_1, \dots, \delta_c)$; then the evolution of the system is governed by the equation [which extends (4)]

$$\frac{d}{dt}V(t) = -\Gamma V(t) + \beta_G A_{L,G}V(t) - \beta_G \text{diag}[V(t)]A_{L,G}V(t). \quad (13)$$

The steady state V_∞ is described by the following governing equation:

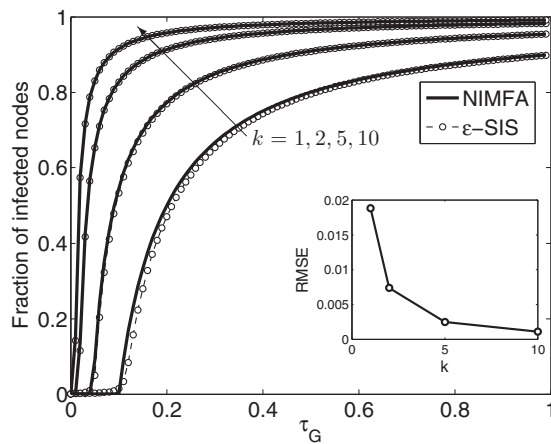
$$\delta_j v_{j;\infty} = \beta_G(1 - v_{j;\infty}) \sum_{m=1}^c a_{jm} k_m v_{m;\infty} + \beta_L(k_j - 1)v_{j;\infty}(1 - v_{j;\infty}). \quad (14)$$

The metastable steady state for the epidemic diffusion is characterized by the solution of the previous formula. Expanding in ε around the critical threshold, the solution is $V_\infty = \varepsilon x$, where x is a vector with non-negative components; then we approximate the generalized Laplacian and (8) becomes, after division by ε ,

$$\mathcal{Q} \left(\frac{1}{\frac{\beta_G}{\delta_j}(1 - \varepsilon x_j) - \frac{(k_j - 1)\beta_L}{\beta_G}} \right) x = 0,$$

which becomes, taking the limit $\varepsilon \rightarrow 0$,

$$\text{diag} \left[\left(\frac{1}{\frac{\beta_G}{\delta_j}} - \frac{(k_j - 1)\beta_L}{\beta_G} \right)^{-1} \right] A_\kappa x = x. \quad (15)$$



(a)

Hence x is the eigenvector of

$$\tilde{A} := \text{diag} \left[\left(\frac{1}{\frac{\beta_G}{\delta_j}} - \frac{(k_j - 1)\beta_L}{\beta_G} \right)^{-1} \right] A_\kappa$$

belonging to the eigenvalue 1.

Let us assume that the following condition holds:

$$\delta_j > (k_j - 1)\beta_L \quad \text{for all } j = 1, \dots, c. \quad (16)$$

Thus \tilde{A} is a non-negative, irreducible matrix (to be precise, we obtain the irreducibility of \tilde{A} from the same property of A , which in turn is equivalent to the fact that the graph is connected, which is an assumption underlying the whole paper).

Under the above assumptions, the Perron-Frobenius Theorem [14,29] states that \tilde{A} has a positive largest eigenvalue $\tilde{\mu}_{\max}$ with a corresponding eigenvector whose elements are all positive and that it is the only eigenvector with only non-negative components. Since x has non-negative components, it must be the eigenvector associated with $\tilde{\mu}_{\max}$.

Theorem 3. The critical threshold is determined by the vectors of the curing rates $\Gamma = (\delta_1, \dots, \delta_n)^T$ which imply $\mu_{\max}(\tilde{A}) = 1$.

Remark 1. In the homogeneous case $\delta_j \equiv \delta$, $k_j \equiv k$, it holds that

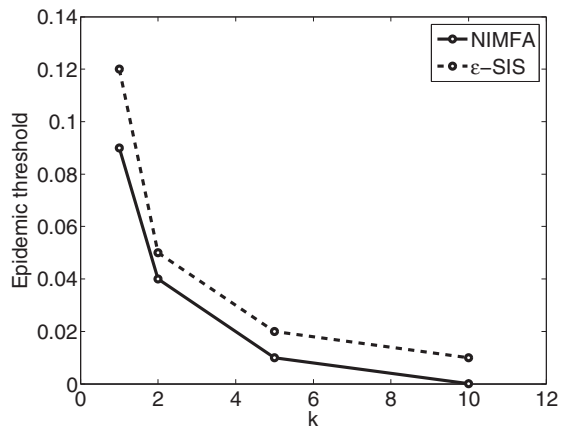
$$\tilde{\mu}_{\max} = \left(\frac{1}{\frac{k\beta_G}{\delta}} - \frac{(k-1)\beta_L}{k\beta_G} \right)^{-1} \mu_{\max}, \quad (17)$$

which leads to the critical threshold for δ

$$\delta = (k-1)\beta_L + \mu_{\max}k\beta_G,$$

where μ_{\max} is the positive largest eigenvalue of A . The computation above is completely analogous to the one in the example of Sec. III B.

In general, Theorem 3 does not lead to an explicit characterization of the epidemic steady state, nor does it state any explicit condition for the uniqueness of the epidemic steady



(b)

FIG. 2. (a) Fraction of infected nodes for different values of k and $\tau_G = \beta_G/\delta$, with fixed ratio $\beta_L/\beta_G = 2$ and the value $\delta = 1$, for a network of regular degree $d = 10$ and $N = 500$. Both the NIMFA and the exact $\varepsilon = 10^{-3}$ SIS model are shown. The inset plot represents the root mean square error between the simulated and the approximated fraction of infected nodes. (b) The corresponding value of the epidemic threshold for the NIMFA and the exact ε -SIS model.

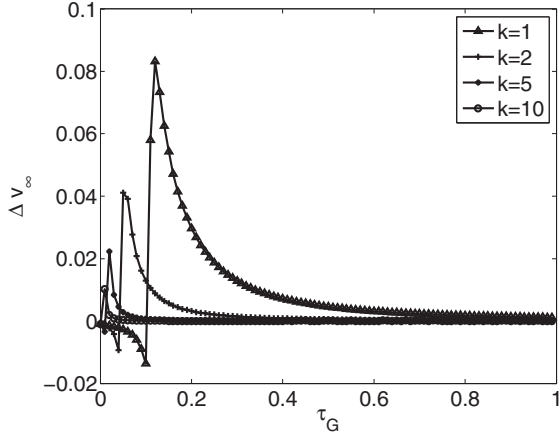


FIG. 3. Difference between the exact ε -SIS model and the NIMFA fractions of infected nodes as a function of τ_G and for different values of k , for a network of regular degree $d = 10$ and $N = 500$.

state. Therefore, we can resort to the computation of Sec. III D to get the following result:

Theorem 5. For a generic network with local clouds of nonconstant size $\{k_j\}$, a sufficient condition for the uniqueness of the trivial epidemic steady state is the following:

$$\forall j = 1, \dots, c, \quad \frac{d_j k_j \beta_G + (1 - \frac{1}{k_j})\beta_L}{\delta_j} < \frac{1}{k_j}.$$

V. NUMERICAL EVALUATIONS

In this section, we describe numerical results for the two-scale community model.

A. Effect of cluster size

We depict first in Fig. 2(a) the impact of the cluster size k on the fraction of infected nodes in the metastable state as derived from (8), and compare the results of our model to the exact ε -SIS model. We consider a range for the normalized

intercommunity infection rate, i.e., $\tau_G := \frac{\beta_G}{\delta}$, for constant ratio $\beta_L/\beta_G = 2$ with respect to the normalized intra-community infection rate $\tau_L := \frac{\beta_L}{\delta}$. The epidemic threshold is measured as the smallest value of τ_G which determines a nonzero epidemic state, while, for the ε -SIS model, it is measured as the value of τ_G where the second derivative of the steady-state fraction of infected nodes equals zero.

The sample network has constant degree $d = 10$ and order $N = 500$. The cluster size k is the same for all communities: curves are drawn for increasing values of k ($k = 1, 2, 5, 10$), where $k = 1$ denotes the absence of local clusters. The threshold effect is well visible in the graphs depicted in Fig. 2 (a). As can be further observed, our model and the exact SIS model are in good agreement and the root mean square error between them decreases with k .

In Fig. 2(b) the corresponding value of the epidemic threshold τ_G for the NIMFA and the exact ε -SIS model is reported. As expected from (11), the critical threshold above which an endogenous infection exists decreases with the size of the clusters. Thus, for large values of the cluster size, a very small value of τ_G is sufficient to cause epidemic outbreaks, irrespective of the actual network structure. Moreover, the NIMFA epidemic threshold well approximates the threshold observed in reality especially for $k > 1$.

Figure 3 illustrates the difference between the NIMFA and the ε -SIS fraction of infected nodes as a function of τ_G and for various k , for the same network sample: we observe that the maximum difference between the two models occurs when τ_G equals the ε -SIS epidemic threshold. This means that for τ_G greater than the ε -SIS epidemic threshold, the difference between the two models decreases and the two models get increasingly closer, according to our definition of the epidemic threshold for the ε -SIS model.

Figures 4(a) and 4(b) have been derived after averaging over 300 instances of Erdős-Rényi random graphs of order $c = 10$ generated according to edge connection probability $p = 0.3$ and setting $\beta_L/\beta_G = 2$. The confidence intervals for the epidemic threshold is set to 98%. In particular, the bottom curve represents the case $k = 1$, i.e., the absence of local

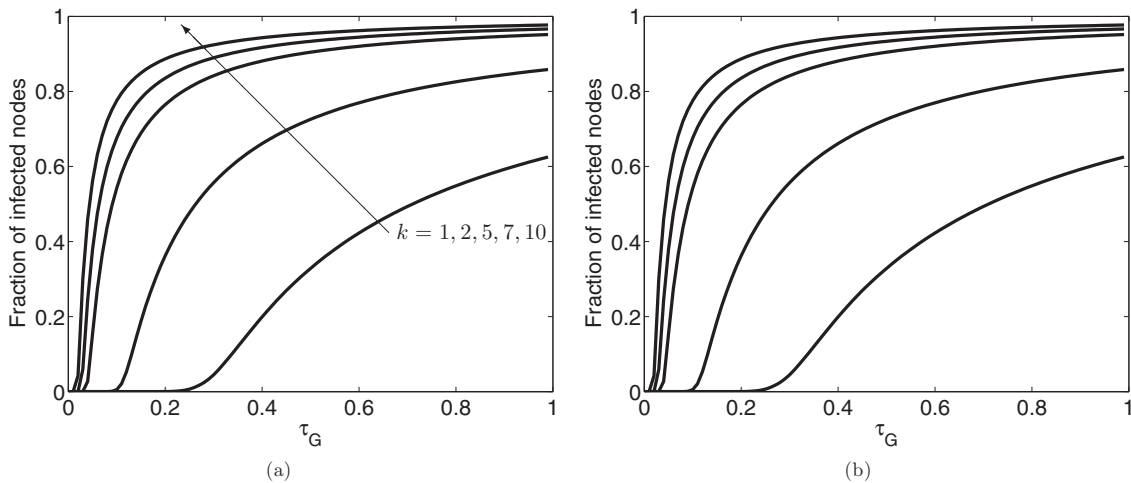


FIG. 4. (a) Fraction of infected nodes for different values of k and τ_G , with fixed ratio $\beta_L/\beta_G = 2$, order $c = 10$, and the value $\delta = 1$. (b) The corresponding value of the epidemic threshold. All graphs have been obtained averaging over 300 instances of Erdős-Rényi random graphs for $p = 0.3$; the level of confidence is set to 98%.

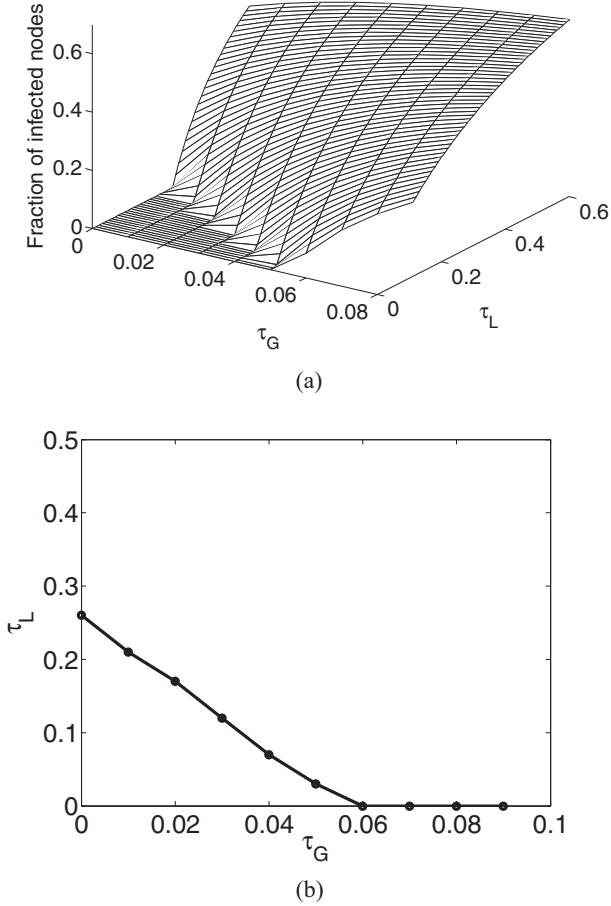


FIG. 5. (a) Fraction of infected nodes as a function of τ_G and τ_L for $k = 5$ and $c = 10$. (b) Detail of the epidemic threshold.

clouds. We interpret the threshold in the case $k = 1$ as the capability of the cluster’s network to sustain a metastable epidemic outbreak. Larger cluster sizes cause the epidemic threshold to drop significantly: in our sample graphs and for

the considered cluster size range, it decreases by one order of magnitude, i.e., it starts above 0.3 for $k = 1$ and it decreases to around 0.1 already for $k = 2$, while it finally drops below 0.03 for $k \geq 7$.

By taking into account one instance of this set of Erdős-Rényi graphs for $k = 5$, we report in Fig. 5(a) the behavior of the fraction of infected nodes as a function of τ_G and τ_L . In that figure, the epidemic threshold is represented by the line of points where the fraction of infected nodes becomes positive: we observe that it depends linearly on the infection rates β_L and β_G , given the cloud dimension k , in agreement with the expression derived in Theorem 5.

Finally, we compare the behavior of our model with respect to the ε -SIS model by considering a sample Erdős-Rényi network with order $c = 20$ and $p = 0.3$. Figures 6(a) and 6(b) show a good agreement between the two models for cluster size $k = 5, 7, 10$. On the contrary, for networks with few individuals, $N = 20$ corresponding to $k = 1$ and $N = 40$ for $k = 2$, the epidemic threshold of the NIMFA model is less close to that of the ε -SIS model. The better agreement between the NIMFA model and the ε -SIS model for larger N is also verified in [25] for nonclustered complete graphs and in [27] for nonclustered bipartite networks, star graphs, and lattice graphs. As further observed from Fig. 7, the maximum difference between the NIMFA and the ε -SIS fractions of infected nodes corresponds to the ε -SIS epidemic threshold only for networks with $k = 5, 7, 10$.

B. Effect of heterogeneity of the cluster size

One interesting question that concerns the two-scale epidemic model is the influence of the cluster size distribution onto the epidemic threshold. In general, it is not obvious whether, with all remaining system parameters fixed, a constant cluster size will lead to a lower or larger epidemic threshold for the same network.

In Fig. 8 we performed a test using a set of 300 sample tree graphs. Each graph is the spanning tree of an Erdős-Rényi

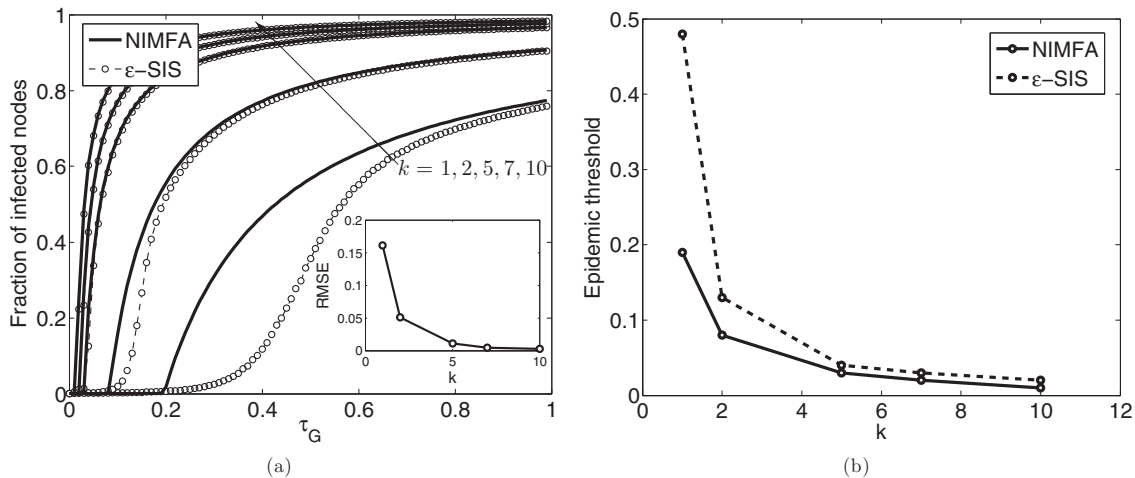


FIG. 6. (a) Fraction of infected nodes for different values of k and $\tau_G = \beta_G/\delta$, with fixed ratio $\beta_L/\beta_G = 2$ and the value $\delta = 1$, for an Erdős-Rényi graph of order $c = 20$ and $p = 0.3$. Both the NIMFA and the exact $\varepsilon = 10^{-3}$ SIS model are shown. The inset plot represents the root mean square error between the simulated and the approximated fraction of infected nodes. (b) The corresponding value of the epidemic threshold for the NIMFA and the exact ε -SIS model.

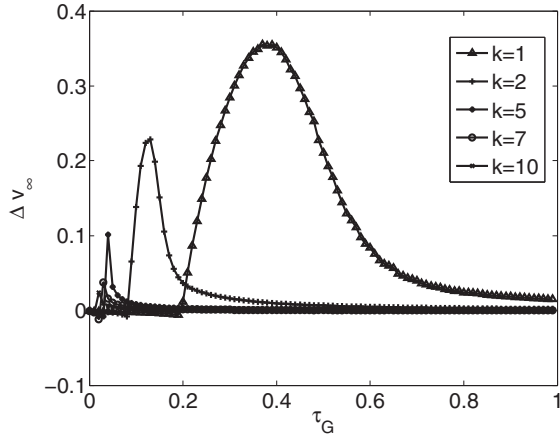


FIG. 7. Difference between the exact ε -SIS model and the NIMFA fractions of infected nodes as a function of τ_G and for different values of k , for an Erdős-Rényi graph of order $c = 20$, $p = 0.3$.

graph of order $c = 10$ and $p = 0.3$. The ratio β_L/β_G is set to 8. Figure 8 draws the difference Δ_{τ_G} between the epidemic thresholds measured for homogeneous and inhomogeneous cloud distributions. In particular, for each sample tree, we considered different values of the average cluster size $k = 5, 10, 15$. In the case of heterogeneous cloud distribution half of the nodes have cluster size 2 and half of them have cluster size $2k - 2$.

Figure 8 exemplifies that heterogeneity of the cluster size lowers the epidemic threshold compared to the case of constant size. This observation agrees with the theorem: among all graphs with N nodes and L links, the regular graph has the lowest spectral radius λ_1 . The theorem follows from the inequality [29, (3.34) on p. 47]:

$$\lambda_1 \geq \frac{2L}{N} \sqrt{1 + \frac{\text{var}[D]}{(E[D])^2}},$$

where D is the degree of a randomly chosen node in the graph. Indeed, we have that the variance $\text{var}[D] = 0$ for a regular graph and equality holds only for a regular graph. Hence,

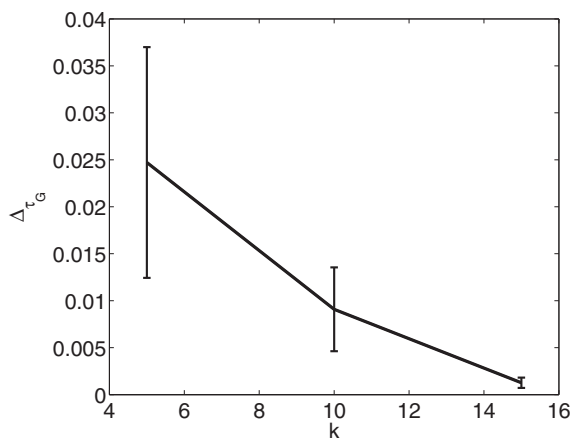


FIG. 8. Difference Δ_{τ_G} between the epidemic threshold in the cases of homogeneous and inhomogeneous cloud distributions for different values of k (5,10,15), with fixed ratio $\beta_L/\beta_G = 8$.

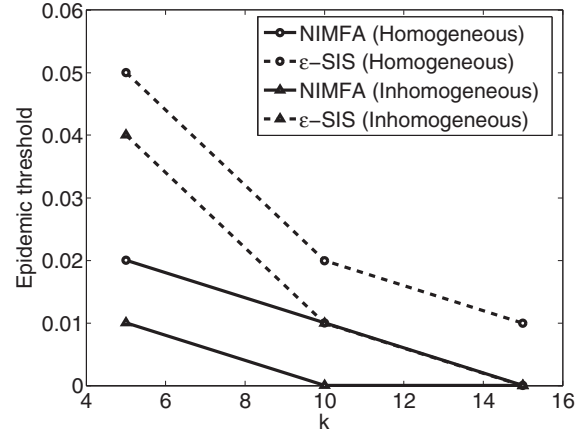


FIG. 9. The epidemic threshold in the cases of homogeneous and inhomogeneous cloud distributions for different values of k , for a spanning tree of an Erdős-Rényi graph of order $c = 10$ and $p = 0.3$. Both the NIMFA and the ε -SIS thresholds are shown.

given that the number N of nodes and the number L of links are fixed, the NIMFA epidemic threshold

$$\tau_c^{(1)} = \frac{1}{\lambda_1} \leq \frac{N}{2L} \frac{1}{\sqrt{1 + \frac{\text{var}[D]}{(E[D])^2}}},$$

implying that the larger the variance in the degree D , the lower the NIMFA epidemic threshold $\tau_c^{(1)}$. Unfortunately, since $\tau_c^{(1)} \leq \tau_c$, we cannot conclude that an increase in $\text{var}[D]$ also always lowers the exact epidemic threshold τ_c .

Figure 9 shows the epidemic thresholds measured for homogeneous and inhomogeneous cluster sizes, by considering one instance of the previous set of spanning trees of an Erdős-Rényi graph. We report both the results obtained for our model and the results obtained for the ε -SIS model: the NIMFA epidemic threshold well estimates the ε -SIS epidemic threshold in both cluster size distributions.

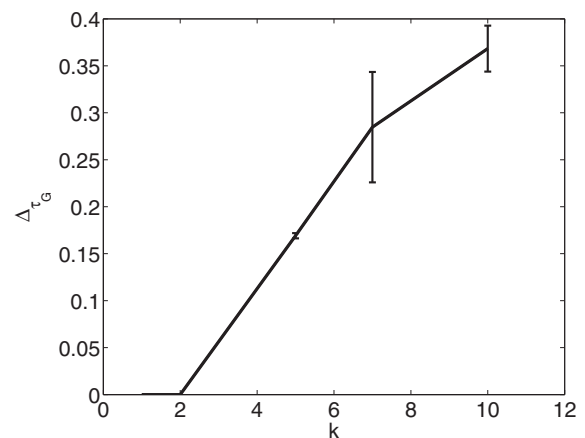


FIG. 10. Ring clusters vs fully connected clusters: the difference Δ_{τ_G} between the epidemic threshold in the case of ring clusters and fully connected clusters for different values of k .

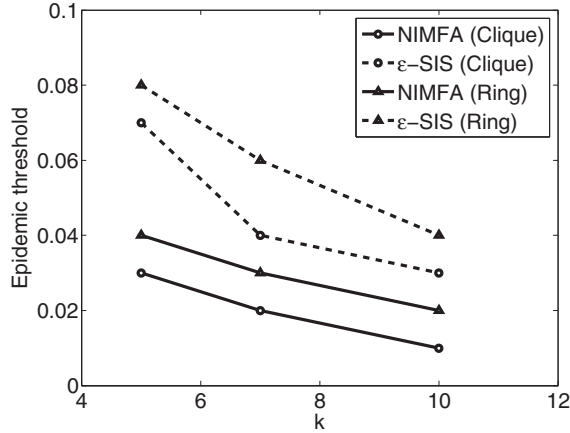


FIG. 11. Epidemic threshold in the case of ring and clique clusters for different values of k , for an instance of Erdős-Rényi random graph of order $c = 10$ and $p = 0.3$. Both the NIMFA and the exact ε -SIS model thresholds are shown.

C. Effect of cluster structure

Finally, we complement the observations on our model with the case when clusters have a structure which is not a clique. The impact of the internal cluster structure is not part of our model: in fact, as we mentioned in Sec. II, we consider small-sized clusters where the full mesh assumption reflects the fact that the contact rate within a community is larger than outside. But if we consider clusters with a large number of elements, assuming a clique as internal structure may not be fully appropriate. Hence we consider hereafter sparser cluster structures, e.g., a ring, and we show that, as we actually expect, this increases the epidemic threshold.

Figure 10 illustrates the results for clusters with a ring topology, compared to fully connected clusters. The figure shows, for different values of k , the difference Δ_{τ_G} between the epidemic threshold in the case of ring clusters and fully connected clusters. The results have been obtained by averaging over 300 instances of Erdős-Rényi random graphs (as before, $c = 10$, $p = 0.3$, and the level of confidence is set to 98%). For $k > 2$, the difference in the value of the epidemic threshold confirms that the cluster structure has a major impact on the epidemic threshold. For those topologies, the values of the threshold that our model generates should be considered as a conservative lower bound. This behavior is further observed in Fig. 11 where we compare the NIMFA and the ε -SIS thresholds for the two cluster topologies. Moreover,

as for the previous test cases, the NIMFA epidemic threshold is close to that of the ε -SIS model.

VI. CONCLUSIONS

SIS epidemic spread in a network of N individuals partitioned into c communities is analysed assuming a NIMFA-like mean-field approximation. The existence of a nontrivial persistent epidemic state is studied in terms of a few parameters that determine the behavior of the network: the infection rate per intracluster β_G and per intercluster β_L and the curing rate for the clusters. Compared with the classical case of a single network, our model satisfies a similar behavior but with an additional feature, that is, the possibility that clusters behave as sources of infections.

We address first the case in which all the communities have the same size k and we discuss the existence of a threshold level for the curing rate. Hence, we describe how the threshold decreases with the size of communities and we show that for a general cluster network structure, the threshold can be estimated in terms of the maximum degree d_{\max} in the community network.

If each community n_j has a different size, then it is possible to provide a relation which ties together the main system parameters, namely, the population size of each community k_j , its connection degree d_j , and curing rate δ_j . In particular, even in the case when intracluster infection rates are small, or the network has small degree, an endemic epidemic state may still occur due to the large order of clusters.

The sufficient condition expressed in Theorem 5 specifies under which choice of parameters the system can be driven to the trivial (i.e., overall healthy) epidemic steady state. The result also indicates that clusters with large values of $d_j k_j$ are most likely to breach the inequality of Theorem 5. As a consequence, an efficient immunization policy for a network partitioned into communities should put the main effort to increasing the curing rate in those communities having a large product $d_j k_j$.

ACKNOWLEDGMENTS

The work has been partially supported by the European Commission within the framework of the CONGAS Project No. FP7-ICT-2011-8-317672 (see www.congas-project.eu) and by the European Commission, European Social Fund, and the Calabria Region-ARUE grant.

[1] H. Wang, Q. Li, G. D'Agostino, S. Havlin, H. E. Stanley, and P. Van Mieghem, *Phys. Rev. E* **88**, 022801 (2013).
 [2] F. Ball, D. Mollison, and G. Scalia-Tomba, *Ann. Appl. Probab.* **7**, 46 (1997).
 [3] F. Ball, D. Sirl, and P. Trapman, *Adv. Appl. Probab.* **41**, 765 (2009).
 [4] L. Pellis, F. Ball, and P. Trapman, *Math. Biosci.* **235**, 85 (2012).
 [5] C. Fraser, *PLoS One* **2**, e758 (2007).

[6] J. V. Ross, T. House, and M. J. Keeling, *PLoS One* **5**, e9666 (2010).
 [7] Z. Liu and B. Hu, *Europhys. Lett.* **72**, 315 (2005).
 [8] L. Huang, K. Park, and Y.-C. Lai, *Phys. Rev. E* **73**, 035103 (2006).
 [9] A. Galstyan and P. Cohen, *Phys. Rev. E* **75**, 036109 (2007).
 [10] V. Colizza and A. Vespignani, *J. Theor. Biol.* **251**, 450 (2008).
 [11] A. Vespignani, *Nat. Phys.* **8**, 32 (2012).

- [12] C. Poletto, S. Meloni, V. Colizza, Y. Moreno, and A. Vespignani, *PLoS Comput. Biol.* **9**, e1003169 (2013).
- [13] P. Van Mieghem, J. Omic, and R. Kooij, *IEEE/ACM Trans. Netw.* **17**, 1 (2009).
- [14] P. Van Mieghem, *Performance Analysis of Complex Networks and Systems* (Cambridge University Press, Cambridge, 2014).
- [15] L. J. S. Allen, *Mathematical Epidemiology* (Springer, Berlin, 2008), pp. 81–130.
- [16] E. Cator and P. Van Mieghem, *Phys. Rev. E* **87**, 012811 (2013).
- [17] J. N. Darroch and E. Seneta, *J. Appl. Probab.* **4**, 192 (1967).
- [18] P. K. Pollett, <http://www.maths.uq.edu.au/pkp/papers/qsds/qsds.pdf>(2008).
- [19] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications*, 2nd ed. (Charles Griffin, London, 1975).
- [20] D. J. Daley, and J. Gani, *Epidemic Modelling: An Introduction* Cambridge University Press, Cambridge, 1999).
- [21] R. Pastor-Satorras, and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).
- [22] M. Draief and L. Massoulié, *Epidemics and Rumours in Complex Networks*, London Mathematical Society Lecture Note Series (Cambridge University Press, Cambridge, 2010), Vol. 369.
- [23] P. Van Mieghem, [arXiv:1310.3980](https://arxiv.org/abs/1310.3980).
- [24] P. Van Mieghem, *Comput. Commun.* **35**, 1494 (2012).
- [25] P. Van Mieghem and E. Cator, *Phys. Rev. E* **86**, 016116 (2012).
- [26] P. Van Mieghem, *Springer Comput.* **93**, 147 (2011).
- [27] C. Li, R. Van De Bovenkamp, and P. Van Mieghem, *Phys. Rev. E* **86**, 026116 (2012).
- [28] E. Cator and P. Van Mieghem, *Phys. Rev. E* **89**, 052802 (2014).
- [29] P. Van Mieghem, *Graph Spectra for Complex Networks* (Cambridge University Press, Cambridge, 2011).
- [30] P. Van Mieghem and J. Omic, [arXiv:1306.2588](https://arxiv.org/abs/1306.2588).