

## Epidemics in networks with nodal self-infection and the epidemic threshold

Piet Van Mieghem\* and Eric Cator

*Delft University of Technology, Faculty of Electrical Engineering, Mathematics and Computer Science, P.O. Box 5031, 2600 GA Delft, The Netherlands*

(Received 10 May 2012; published 30 July 2012)

Since the Susceptible-Infected-Susceptible (SIS) epidemic threshold is not precisely defined in spite of its practical importance, the classical SIS epidemic process has been generalized to the  $\varepsilon$ -SIS model, where a node possesses a self-infection rate  $\varepsilon$ , in addition to a link infection rate  $\beta$  and a curing rate  $\delta$ . The exact Markov equations are derived, from which the steady state can be computed. The major advantage of the  $\varepsilon$ -SIS model is that its steady state is different from the absorbing (or overall-healthy state) and approximates, for a certain range of small  $\varepsilon > 0$ , the in reality observed phase transition, also called the “metastable” state, that is characterized by the epidemic threshold. The exact steady-state analysis for the complete graph illustrates the effect of small  $\varepsilon$  and the quality of the first-order mean-field approximation, the  $N$ -intertwined model, proposed earlier. Apart from duality principles, often used in the mathematical literature, we present an exact recursion relation for the Markov infinitesimal generator.

DOI: [10.1103/PhysRevE.86.016116](https://doi.org/10.1103/PhysRevE.86.016116)

PACS number(s): 89.75.Hc, 89.20.-a

### I. INTRODUCTION

We consider virus spread in an undirected graph  $G(N, L)$ , with  $N$  nodes and  $L$  links, characterized by a symmetric adjacency matrix  $A$ . The graph  $G$  is fixed and does not change over time. The viral state of a node  $i$  at time  $t$  is specified by a Bernoulli random variable  $X_i(t) \in \{0, 1\}$ :  $X_i(t) = 0$  for a healthy node and  $X_i(t) = 1$  for an infected node. A node  $i$  at time  $t$  can be in one of the two states: *infected*, with probability  $v_i(t) = \text{Prob}[X_i(t) = 1]$  or *healthy*, with probability  $1 - v_i(t)$ , but susceptible to the virus. We assume that the curing process per node  $i$  is a Poisson process with rate  $\delta$  and that the infection rate per link is a Poisson process with rate  $\beta$ . Obviously, only when a node is infected can it infect its direct neighbors that are still healthy. Both the curing and infection Poisson process are independent. This is the general continuous-time description of the simplest type of a Susceptible-Infected-Susceptible (SIS) virus process on a network.

In this paper we generalize the SIS process by adding a nodal component to the infection. We assume that each node  $i$  can be infected with a rate  $\varepsilon$ . Hence, besides receiving the infection over links from infected neighbors with rate  $\beta$ , the node  $i$  can also itself produce a virus with rate  $\varepsilon$ . Again, all involved Poisson processes are independent. The motivation to consider a nodal infection component stems from the analogy of epidemics with information spread in social networks, where nodes can generate themselves information, which is spread over links to neighbors. This generalization is here called the  $\varepsilon$ -SIS model, which clearly reduces to the “classical” SIS model when  $\varepsilon = 0$ . This SIS generalization is not new: The  $\varepsilon$ -SIS model was already proposed recently by Hill *et al.* [1], who have considered the happiness of persons as a form of infection over a social contact network. However, in contrast to Hill *et al.* [1], we present an exact analysis of the  $\varepsilon$ -SIS model and deduce insights in the SIS epidemic process.

Apart from the greater flexibility of the  $\varepsilon$ -SIS epidemic to model practical cases of information diffusion, there is

a second, more fundamental motivation to consider in this generalization of the “classical” SIS model. Many authors (see, e.g., Refs. [2–7]) mention the existence of an epidemic threshold  $\tau_c$ . If the effective spreading rate  $\tau = \frac{\beta}{\delta} > \tau_c$ , the virus persists and a nonzero fraction of the nodes are infected, whereas for  $\tau \leq \tau_c$ , the epidemic dies out and the network is virus free in the steady state. From the point of view of network protection against viral infections, the epidemic threshold  $\tau_c$  is the key parameter in the design of immunization strategies in networks [8–10]. Many approximate methods applied to the SIS model have proposed various types of estimates for  $\tau_c$ . The most successful and general approximation is the lower bound  $\tau_c \geq \tau_c^{(1)} = \frac{1}{\lambda_1(A)}$ , first proposed by Wang *et al.* [11], but rigorously proved in Refs. [12,13], where  $\lambda_1(A)$  is the largest eigenvalue of the adjacency matrix  $A$  of the network. Higher order mean-field approximations have been derived in Ref. [14] illustrating the existence of a sequence of more accurate lower bounds  $\tau_c \geq \dots \geq \tau_c^{(m)} \geq \tau_c^{(m-1)} \geq \dots \geq \tau_c^{(2)} \geq \tau_c^{(1)}$ . In particular, a similar largest eigenvalue expression as for  $\tau_c^{(1)}$  was deduced, namely,  $\tau_c^{(2)} = \frac{1}{\lambda_1(H)}$ , where the  $N^2 \times N^2$  matrix  $H$  contains elements of the adjacency matrix  $A$  and reflects network properties.

However, when the SIS model is exactly described via Markov theory as shown in Ref. [12] and its generalization in Sec. II, the observation that this Markov chain (with a finite number of states) possesses an absorbing state equal to the overall healthy state contradicts the existence of any threshold. For, in an irreducible Markov chain, where all states are reachable from each other, the existence of an absorbing state implies that all other states are transient states and that the steady state is the absorbing state. Moreover, the probability that the process is in a transient state exponentially tends to zero with time. However, the convergence time  $T$  to the steady state can be very large [12,15,16]. When the number of states grows unboundedly, major complications arise because an infinite state Markov process is considerably more complex than a finite state Markov chain. Most of the epidemic work in mathematics so far (see, e.g., Ref. [17]) witnesses the difficulty

\*P.F.A.VanMieghem@tudelft.nl

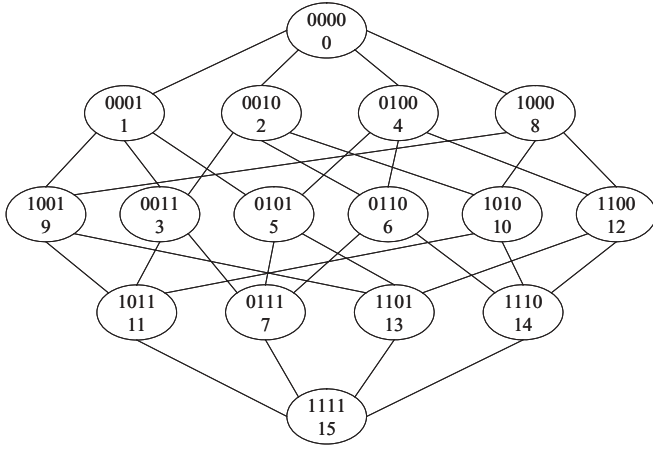


FIG. 1. The state diagram in a graph with  $N = 4$  nodes and the binary numbering of the states.

to analyze infinitely large networks. Although there is an absorbing state, in an infinite-state Markov process, there is a nonzero chance that the process never dies out. Since the exact Markov chain [12] consists of  $2^N$  states in a network of  $N$  nodes, features of the infinite-state Markov process rapidly pop up. The apparent steady state connected with the observation of an epidemic threshold is often termed the “metastable state” since, on a sufficiently long timescale for finite-state systems, it disappears.

A major consequence of the theoretical nonexistence of the SIS epidemic threshold in a *finite* network confuses the epidemiological field greatly. While simulations in “sufficiently” large networks (already of the order of magnitude for  $N$  of 100) show the epidemic phase transition, the mathematical precise definition obscures any observation, because simulating or measuring longer alters the precise position of the epidemic threshold as eventually the SIS epidemics will disappear. In particular, in relatively small networks ( $N$  about 10 or 20), the time-dependent evolution towards the absorbing state of the SIS epidemics is clearly observed, for any value of the effective spreading rate  $\tau$ . The probability to reach the absorbing state in which, eventually, all the nodes in the network will be healthy so that the virus disappears decreases with  $N$ . Figure 1 shows this decrease with  $N$  immediately: The number of paths in the Markov graph from a certain state towards the absorbing state dramatically increases with  $N$ . Thus, the major complication in the analysis of the SIS model is the absorbing state, not only theoretical, but also simulative. In order to determine the epidemic threshold  $\tau_c$  via computer simulations in sufficiently large networks, the occurrence of hitting the absorbing state is large because the effective infection rate  $\tau$  is so low that infection often vanishes, yet  $\tau$  should be high enough to avoid exponential dying out of the virus. Such simulations do not only take a long time, but also possess high variability, which questions how accurately computer simulations can determine the epidemic SIS threshold.

The second major motivation to study the  $\varepsilon$ -SIS model is that, for  $\varepsilon > 0$ , there is no absorbing state and the steady state can be well determined by ordinary Markov theory. The absence of the absorbing state follows from the fact that, if all nodes are healthy, the network does not permanently

remain healthy but gets infected with rate  $N\varepsilon$  (due to nodal self-infection with rate  $\varepsilon$ ), implying that the probability that the network is healthy (all nodes are in state 0) is less than 1. This also means that, for a given small  $\varepsilon$ , the epidemics as a function of the infection rate  $\tau$  can be computed as well as the boundary between low infection (due to the immanent self-infection with rate  $\varepsilon$ ) and the high infection region due to neighbor infection with rate  $\beta$ . The precise SIS epidemic threshold is now defined as the sharp transition between those two regimes for a given small  $\varepsilon$ . In contrast to the SIS model, for small values of  $\varepsilon = 10^{-a}$  with  $10 \leq a \leq 3$  in the  $\varepsilon$ -SIS model, computer simulations on finite, sufficiently large networks are accurate, as shown by extensive simulations in Ref. [18] and in Sec. IV. In particular, Fig. 6 illustrates that even for unrealistically low values of  $\varepsilon = 10^{-a}$  with  $a > 20$ , a nonzero self-infection rate, though nearly arbitrarily small, results in a clear phase transition. Also, the shift of the phase transition when  $\varepsilon = 10^{-a}$  varies from  $a = 6$  to  $a = 10$ , is rather small.

From a biological point of view, the phase transition is important because of self-organization (see, e.g., Ref. [19]). When the virus dies out, the population will start losing immunity (i.e., the curing rate will decrease with time) against that virus, and when it suddenly appears again, the whole population may be wiped out (since  $\delta$  has become very low). Hence, the existence of a very few infected nodes on average keeps the population fit against the virus (i.e.,  $\delta$  remains more or less constant) because their immunity system is constantly challenged. The “background” presence of viruses can be modeled by a low nodal self-infection rate  $\varepsilon$ . The most obvious equilibrium situation of such a population is to operate with  $\tau$  very close to the epidemic threshold  $\tau_c$ . This argument of self-organized criticality underlines the desire to determine the SIS epidemic threshold  $\tau_c$  accurately.

Here we extend the analysis in Ref. [12] to  $\varepsilon$ -SIS epidemics (Sec. II). In Sec. III we provide a new matrix recursion for the infinitesimal generator and show that the Markov graph exhibits a regular bipartite structure. Simulations and exact results for the steady-state fraction of infected nodes in the complete graph are presented in Sec. IV and in the Appendix to show the effect of the self-infection rate  $\varepsilon$ .

## II. THE $\varepsilon$ -SIS $2^N$ -STATE MARKOV CHAIN

A “physical” description of the  $\varepsilon$ -SIS epidemic process is as follows. Let  $I$  denote the set of infected nodes in the graph  $G$ . Then the Markov transitions

$$I \mapsto I \cup \{j\} \quad (j \notin I) \quad \text{at rate} \quad \beta \sum_{k \in I} a_{kj} + \varepsilon,$$

$$I \mapsto I \setminus \{i\} \quad (i \in I) \quad \text{at rate} \quad \delta$$

detail the dynamics<sup>1</sup> between the infected subgraph  $I$  and its complement  $I^c = G \setminus I$ . Computationally we need to enumerate the subgraphs  $I$  in  $G$ , which leads to another description. The state  $Y(t)$  of the network at time  $t$  is defined

<sup>1</sup>A mean-field analysis based on infected subgraphs is given in Ref. [20], pp. 48–51.

by all possible combinations of states in which the  $N$  nodes can be at time  $t$ ,

$$Y(t) = [Y_0(t) Y_1(t) \dots Y_{2^N-1}(t)]^T$$

and

$$Y_i(t) = \begin{cases} 1, & i = \sum_{k=1}^N X_k(t)2^{k-1} \\ 0, & i \neq \sum_{k=1}^N X_k(t)2^{k-1}. \end{cases}$$

Hence, the state space of the Markov chain is organized with  $x_k \in \{0, 1\}$ , where  $1 \leq k \leq N$  refers to the node with label  $k$  in  $G$ , as

State number $i$	$x_N x_{N-1} \dots x_2 x_1$
0	00...000
1	00...001
2	00...010
3	00...011
...	...
$2^N - 1$	11...11

The number of the states with  $j$  infected nodes is  $\binom{N}{j}$ . Figure 1 shows an example of the Markov state graph with  $N = 4$  nodes. A state with  $j$  infected nodes needs to make at least  $j$  transitions to reach the overall-healthy state 0. The total number of transitions from any state to state 0 is at least  $\sum_{j=1}^N j \binom{N}{j} = N2^{N-1}$ , so that the average number of transitions from an arbitrary state to state 0 is at least  $\frac{N}{2}$ . Likewise the corresponding variance equals  $\sum_{j=1}^N j^2 \binom{N}{j} 2^{-N} - (\frac{N}{2})^2 = \frac{N}{4}$ . Hence, for large  $N$ , it requires for a randomly infected network,

in the best possible case,<sup>2</sup> on average  $\frac{N}{2} + O(\sqrt{N})$  hops to eradicate the virus.

The representation of a integer number  $n$  in a binary base is

$$n = \sum_{k=0}^{\log_2(n)} w_k(n)2^k,$$

where the binary  $k$ th digit  $w_k(n) = \frac{1}{2}(1 - (-1)^{\lfloor \frac{n}{2^k} \rfloor})$  is<sup>3</sup> either 0 or 1. Here we slightly modify the representation for an integer  $i$  as

$$i = \sum_{k=1}^N x_k(i)2^{k-1}$$

because the binary  $k$ th digit  $x_k(i)$  represents the infectious state of a node  $k$  in the network, while in state  $i$ , and node labels in a graph  $G$  range from 1 to  $N$ , rather than from 0 to  $N - 1$ .

In any continuous-time Markov process [21], there is only one event possible during an arbitrary small time interval, which implies for the  $\varepsilon$ -SIS process that a transition out of state  $i$  can only end in a state  $j$ , whose binary representation  $x_N x_{N-1} \dots x_2 x_1$  has either one more 1 bit<sup>4</sup> or one less than that of state  $i$ . If the state  $j$  has less 1 bits, then precisely one of the 1 bits (corresponding with one node in the network) of state  $i$  has been cured with rate  $\delta$ . If the state  $j$  has precisely one additional 1 bit, then one of the nodes, say,  $m$ , with  $x_m(i) = 0$  has become infected [so that  $x_m(j) = 1$ ] with rate  $\varepsilon + \beta \sum_{k=1}^N a_{mk} x_k(i)$ : The strength of infection is due to all neighbors of node  $m$  that are infected, i.e.,  $x_k(i) = 1$ , augmented with the nodal self-infection rate  $\varepsilon$  of node  $m$ . The defined virus infection process is a continuous-time Markov chain with  $2^N$  states specified by the infinitesimal generator  $Q$  with elements

$$q_{ij} = \begin{cases} \delta & \text{if } j = i - 2^{m-1}; \quad m = 1, 2 \dots N \quad \text{and } x_m(i) = 1 \\ \varepsilon + \beta \sum_{k=1}^N a_{mk} x_k(i) & \text{if } j = i + 2^{m-1}; \quad m = 1, 2 \dots N \quad \text{and } x_m(i) = 0 \\ -\sum_{k=1; k \neq j}^N q_{kj} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}, \quad (1)$$

where  $i = \sum_{k=1}^N x_k(i)2^{k-1}$ . For example, if  $i = 0$ , then all  $x_k = 0$ , and the transition probability rates out of the network overall-healthy state 0 are  $q_{0j} = \varepsilon$  for  $j = 2^{m-1}$  with  $1 \leq m \leq N$  (ranging over all network states with 1 infected node), while  $q_{00} = -N\varepsilon$  and  $q_{0j} = 0$  for all other  $j$ . Hence, for  $\varepsilon > 0$ , the first row in the infinitesimal generator  $Q$  is nonzero, whereas it is zero in the ‘‘classical’’ SIS model ( $\varepsilon = 0$ ), corresponding to the absorbing state. When  $\beta = 0$ , there are no link-based infections, only nodal infections. Thus, locally, the infection process per node is a two-state continuous-time Markov

process with self-infection rate  $\varepsilon$  and curing rate  $\delta$ , from which we know (see Ref. [21], p. 196) that the steady-state infection probability for each node  $k \in G$  equals  $\text{Prob}[X_{k\infty} = 1] = \frac{\varepsilon}{\varepsilon + \delta}$ .

The time dependence of the probability state vector  $s(t)$ , with components

$$s_i(t) = \text{Prob}[Y(t) = i] \\ = \text{Prob}[X_1(t) = x_1(i), X_2(t) = x_2(i), \dots, X_N(t) = x_N(i)],$$

<sup>2</sup>Only curing transitions occur, and all transition links are equally weighted.

<sup>3</sup> $[x]$  denotes the integral part of the real number  $x$ .

<sup>4</sup>Each  $x_k$  is called a bit or binary digit, and a 1 bit means that  $x_k = 1$ , while a zero bit means that  $x_k = 0$ .

and normalization  $\sum_{i=0}^{2^N-1} s_i(t) = 1$ , obeys (see Ref. [21], p. 182) the differential equation

$$\frac{ds^T(t)}{dt} = s^T(t)Q,$$

whose solution is

$$s^T(t) = s^T(0)e^{Qt}.$$

The definition of  $s_i(t)$  as a joint probability distribution shows that, if we sum over all the states of all nodes except for the node  $j$ , we obtain the probability that a node  $j$  is either healthy  $y_j = 0$  or infected  $y_j = 1$ ,

$$\text{Prob}[X_j(t) = y_j] = \sum_{i=0; i \neq j}^{2^N-1} s_i(t),$$

where, in the index  $i = \sum_{k=1}^N x_k(i)2^{k-1}$  in the sum above, every  $x_k(i)$  with  $k \neq j$  takes both values from the set  $\{0,1\}$ , while for  $k = j$ ,  $x_k(i) = y_j$  is either 0 (healthy) or 1 (infected). Defining the nodal viral infection probability as  $v_j(t) = \text{Prob}[X_j(t) = 1]$ , then the relation between the vectors  $s(t)$  and  $v(t)$  is

$$v^T(t) = s^T(t)M,$$

where the  $2^N \times N$  matrix  $M$  contains the states in binary notation, but bit reversed:

$$M = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Thus, the element  $m_{ij} = x_{j-1}(i-1)$ . The average fraction of infected nodes in  $G$  at time  $t$  equals  $y(t) = E[\frac{1}{N} \sum_{j=1}^N X_j(t)] = \frac{1}{N} u^T v(t)$  and

$$y(t) = \frac{1}{N} s^T(0) e^{Qt} M u,$$

where  $u$  is the all-one vector. We denote the steady-state average fraction of infected nodes for an effective infection rate  $\tau$  by  $y_\infty(\tau, \varepsilon)$ . For each graph, the steady-state average fraction of infected nodes in  $G$  at  $\beta = \tau = 0$  equals  $y_\infty(0, \varepsilon) = \frac{\varepsilon}{\varepsilon + \delta}$  and, obviously,  $1 \geq y_\infty(\tau, \varepsilon) \geq \frac{\varepsilon}{\varepsilon + \delta}$  for  $\tau \geq 0$ .

The joint probabilities  $z_{ij} = \text{Prob}[X_i = 1, X_j = 1]$  can be organized in the matrix  $Z$ . This matrix can be rewritten as a  $N^2 \times 1$  vector  $\text{vec}(Z)$ , consisting of all elements of  $Z$  columnwise (see Ref. [22], p. 254),

$$\text{vec}(Z_\infty) = (z_{11}, \dots, z_{N1}, z_{12}, \dots, z_{N2}, \dots, z_{1N}, \dots, z_{NN}),$$

which is obtained from the state vector  $s$  as

$$\text{vec}(Z) = Ws,$$

where the elements of the  $N^2 \times 2^N$  matrix  $W$  are  $w_{ij} = x_{l-1}(j-1)x_{m-1}(j-1)$ , where  $l + (m-1)N = i$ . The vari-

ance  $\text{Var}[\sum_{j=1}^N X_j]$  follows from Ref. [21], p. 30, as

$$\begin{aligned} \text{Var} \left[ \sum_{j=1}^N X_j \right] &= \sum_{j=1}^N E[X_j] - \sum_{j=1}^N (E[X_j])^2 \\ &\quad + \sum_{k=1}^N \sum_{j=1}^N \{E[X_k X_j] - E[X_k]E[X_j]\} \\ &= \sum_{j=1}^N z_{jj} - \sum_{j=1}^N z_{jj}^2 + \sum_{k=1}^N \sum_{j=1}^N z_{kj} - \left( \sum_{j=1}^N z_{jj} \right)^2 \\ &= \text{tr}(Z) - v^T v + u^T Z u - [\text{tr}(Z)]^2. \end{aligned}$$

Hence, the variance of the fraction of infected nodes equals  $\sigma_y^2 = \frac{1}{N^2} \text{Var}[\sum_{j=1}^N X_j]$  and, written in terms of the vectors  $v$  and  $\text{vec}(Z)$ :

$$\sigma_y^2 = \frac{1}{N^2} [u^T v - v^T v + u^T_{N^2 \times 1} \text{vec}(Z) - (u^T v)^2].$$

The sum of the rows in any infinitesimal general  $Q$  (similar to the Laplacian of any graph  $G$ ) is zero,  $Qu = 0$ , which shows that the  $2^N \times 1$  all one vector  $u$  is the right-eigenvector belonging to the largest eigenvalue  $\mu = 0$ . The steady-state vector  $\pi$  in the  $\varepsilon$ -SIS model [21] obeys<sup>5</sup>  $\pi^T Q = 0$ , so that  $\pi$  equals the left-eigenvector, normalized as  $\sum_{n=0}^{2^N-1} \pi_n = 1$ , belonging to the zero eigenvalue.

#### A. Example: The case when $N = 2$

Let us consider the simplest possible case of the matrix  $Q$ , where  $N = 2$ ,  $\varepsilon^* = \frac{\varepsilon}{\delta}$ , and  $Q_2^* = \frac{Q_2}{\delta}$ :

$$Q_2^* = \begin{bmatrix} -2\varepsilon^* & \varepsilon^* & \varepsilon^* & 0 \\ 1 & -1 - \varepsilon^* - \tau a_{21} & 0 & \varepsilon^* + \tau a_{21} \\ 1 & 0 & -1 - \varepsilon^* - \tau a_{12} & \varepsilon^* + \tau a_{12} \\ 0 & 1 & 1 & -2 \end{bmatrix},$$

which can be evaluated analytically. The graph  $G$  consists of two nodes that are possibly connected by a link (if  $a_{12} = 1$ ). The steady-state vector  $\pi$  has the vector components

$$\begin{aligned} \pi_0 &= \frac{1}{(1 + \varepsilon^*)^2 + \tau \varepsilon^* a_{12}}, \\ \pi_1 &= \pi_2 = \frac{\varepsilon^*}{(1 + \varepsilon^*)^2 + \tau \varepsilon^* a_{12}}, \\ \pi_3 &= \frac{\varepsilon^*(\varepsilon^* + \tau a_{12})}{(1 + \varepsilon^*)^2 + \tau \varepsilon^* a_{12}}, \end{aligned}$$

illustrating that for  $\varepsilon^* = 0$ , we find the absorbing state, while steady-state symmetry of state 1 and 2, corresponding to the node configuration 01 and 10, is expected. The vector  $v_\infty = M\pi$  has components  $v_{1\infty} = v_{2\infty} = \frac{\varepsilon^*(1 + \varepsilon^* + \tau a_{12})}{(1 + \varepsilon^*)^2 + \tau \varepsilon^* a_{12}}$ , showing that for  $N = 2$  there is no indication of an epidemic threshold.

<sup>5</sup>Here we write a vector as a column vector, as usual in linear algebra, but as opposed to Markov theory [21].

In order to compute the eigenvalues of  $Q_2^*$ , we compute the characteristic polynomial  $c_{Q_2^*}(\lambda) = \det(\frac{Q_2}{\delta} - \lambda I) = \sum_{k=1}^4 c_k \lambda^k$  and find, with  $c_4 = 1$ :

$$\begin{aligned} c_1 &= 2\{\varepsilon^* \tau^2 a_{12} + \tau a_{12}(1 + 3\varepsilon^* + \varepsilon^{*2}) + (1 + \varepsilon^*)^3\}, \\ c_2 &= \tau^2 a_{12} + 2\tau a_{12}(2 + 3\varepsilon^*) + 5(1 + \varepsilon^*)^2, \\ c_3 &= 2[\tau a_{12} + 2(1 + \varepsilon^*)]. \end{aligned}$$

The zeros of  $c_{Q_2^*}(\mu) = 0$  or eigenvalues  $0 = \mu_1 \geq \mu_2 \geq \mu_3 \geq \mu_4$  of  $Q_2^*$  can be found analytically:

$$\mu_{2,4} = -\frac{3(1 + \varepsilon^*) + \tau a_{12}}{2} \pm \frac{1}{2} \sqrt{(1 + \varepsilon^* - \tau a_{12})^2 + 8\tau a_{12}}$$

and

$$\mu_3 = -1 - \tau a_{21} - \varepsilon^*.$$

We note that the two eigenvalues  $\mu_{2,4}$  can never be complex. Hence, all eigenvalues of  $Q_2^*$  are real and negative. For larger  $N$ , we found numerically that eigenvalues of  $Q_N^*$  can be complex (with negative real part).

The time dependence of the number of infected nodes in the connected graph ( $a_{12} = 1$ ) is

$$y(t) = y_1 + y_2 e^{-\mu_2 t^*} + y_3 e^{-\mu_3 t^*} + y_4 e^{-\mu_4 t^*},$$

where  $t^* = \delta t$  is the normalized time and where  $y_1 = v_{1\infty} = \frac{\varepsilon^*(1 + \varepsilon^* + \tau a_{12})}{(1 + \varepsilon^*)^2 + \tau \varepsilon^* a_{12}}$ . Using

$$(1 + \varepsilon^* - \tau a_{12})^2 + 8\tau a_{12} = (\tau a_{12} + 3 - \varepsilon^*)^2 - 8(1 + \varepsilon^*),$$

we observe that, for large  $\tau$ ,

$$\begin{aligned} \mu_2 &= -\frac{3(1 + \varepsilon^*) + \tau a_{12}}{2} + \frac{1}{2} \sqrt{(\tau a_{12} + 3 - \varepsilon^*)^2 - 8(1 + \varepsilon^*)} \\ &= -2\varepsilon^* - \frac{2(1 + \varepsilon^*)}{(\tau a_{12} + 3 - \varepsilon^*)} + O(\tau^{-3}) \end{aligned}$$

so that the number of infected nodes  $y(t) \approx y_1 + y_2 e^{-[2\varepsilon^* + O(\tau^{-1})]t^*}$  tends to  $y_1$  in normalized time  $t^*$  with timing constant  $\frac{1}{2\varepsilon^* + O(\tau^{-1})}$ . Hence, when  $\varepsilon^* \rightarrow 0$ , the timing constant increases or the tendency towards the steady state slows down. Although computed for  $N = 2$ , the property of slow convergence towards the steady state (above  $\tau_c$ ) is a general characteristic for any  $N$  (see, e.g., Ref. [16]).

### III. PROPERTIES OF SIS EPIDEMICS

#### A. Recursive structure of $Q_N^*$

By inspection, we found the recursive structure

$$Q_{N+1}^* = \begin{bmatrix} Q_N^* - B_N & B_N \\ I_{2^N} & Q_N^* + R_N - D_N \end{bmatrix}, \quad (2)$$

where  $I_{2^N}$  is the  $2^N \times 2^N$  identity matrix and  $Q_N^* = \frac{Q_N}{\delta}$ , when in a graph  $G$  a node is added in an arbitrarily fashion specified by the adjacency row vector  $a_{N+1}$ . Moreover,  $B_n = \text{diag}(b_k)$  and, for  $1 \leq k \leq 2^n$ ,

$$b_k = \varepsilon^* + \sum_{j=1}^{2^N} a_{N+1;j}^* x_j(k-1)$$

reflects the binary expansion of the number  $k-1$  in which each 1 bit is replaced by  $a_{N+1;j}^*$  and added. The first row of  $B_n$

thus equals  $\varepsilon^* e_1$ , where  $e_1 = (1, 0, \dots, 0)$  is the first  $(2^N \times 1)$  basic vector. The matrix  $R_N$  is a  $2^N \times 2^N$  upper triangle matrix with zero diagonal elements and  $D_N$  is a diagonal matrix with elements equal to the corresponding row sum of  $R_N + 1$ . The upper triangle matrix  $R_N$  has elements along diagonals

$$(R_N)_{i,i+2^{m-1}} = a_{N+1;m} \left\{ \frac{1 - (-1)^{\lfloor \frac{i+2^{m-1}-1}{2^{m-1}} \rfloor}}{2} \right\}$$

for  $m = 1, 2, \dots, N$  and  $i = 1$  up to  $2^N - 2^{m-1}$ . When plotting  $R_N$ , we observe that the upper triangle matrix  $R_N$  has a nested structure of scaled identity matrices  $a_{N+1;m} I_{2^{m-1}}$ , where the largest of size  $2^{N-1} \times 2^{N-1}$  appears once in the upper corner, those of size  $2^{N-2} \times 2^{N-2}$  appear twice, and the smallest of size  $1 \times 1$  appears  $2^{N-1}$  times on diagonal ( $i, i+1$ ) for  $i = 1$  to  $2^N - 1$ . Specifically, we divide the  $2^N \times 2^N$  matrix in 4 block matrices, each of size  $2^{N-1} \times 2^{N-1}$ , and we fill the right upper block with  $a_{N+1,N} I_{2^{N-1}}$ . The left-upper and right-lower block matrices are again divided in 4 equal size block matrices. Only the right upper block matrices of size  $2^{N-2} \times 2^{N-2}$  are filled with  $a_{N+1,N-1} I_{2^{N-2}}$ . The remaining left-upper and right-lower block matrices are again divided in 4 equal size block matrices, and we repeat the scheme of filling in only the right-upper block matrices and so on until the size of the block matrices is equal to  $2 \times 2$  and, in each of the  $2^{N-1}$  such remaining block matrices with fill the right-upper element by  $a_{N+1;1}$  and stop.

Let us consider the steady-state vector  $\pi_{N+1} = \begin{bmatrix} y_N \\ w_N \end{bmatrix}$  of  $Q_{N+1}^*$ , obeying  $\pi_{N+1}^T Q_{N+1}^* = 0$  or with the recursive structure

$$\begin{aligned} 0 &= \begin{bmatrix} y_N^T & w_N^T \end{bmatrix} \begin{bmatrix} Q_N^* - B_N & B_N \\ I_{2^N} & Q_N^* + R_N - D_N \end{bmatrix} \\ &= \begin{bmatrix} y_N^T(Q_N^* - B_N) + w_N^T y_N^T B_N & w_N^T(Q_N^* + R_N - D_N) \end{bmatrix}. \end{aligned}$$

Thus, we find that

$$\begin{cases} w_N^T = y_N^T (B_N - Q_N^*) \\ y_N^T B_N = -w_N^T (Q_N^* + R_N - D_N) \end{cases}$$

Substituting the first vector into the second vector equation yields

$$\begin{aligned} 0 &= y_N^T B_N + w_N^T (Q_N^* + R_N - D_N) \\ &= y_N^T B_N + y_N^T (B_N - Q_N^*) (Q_N^* + R_N - D_N), \end{aligned}$$

illustrating that  $y_N$  is the left-eigenvector belonging to the zero eigenvalue of the matrix

$$H = B_N (I_{2^N} + Q_N^* + R_N - D_N) - Q_N^* (Q_N^* + R_N - D_N)$$

and that  $y_N$  and  $w_N$  are not easily related to  $\pi_N$ , that obeys  $\pi_N^T Q_N^* = 0$ . Hence, the matrix recursion (2) for  $Q_N^*$  does not translate (in an obvious way) to a recursion for the steady-state vector  $\pi_{N+1}$  in terms of  $\pi_N$ .

#### B. Regular bipartite structure of the Markov graph

Figure 1 reveals that the Markov state graph is a bipartite graph. Indeed, we apply the same type of folding as explained in Ref. [22], p. 132, to map the level set of any tree into a bipartite graph. In particular, we reorganize the nodes in the Markov state graph into two sets  $S_0$  and  $S_1$ . We start by placing

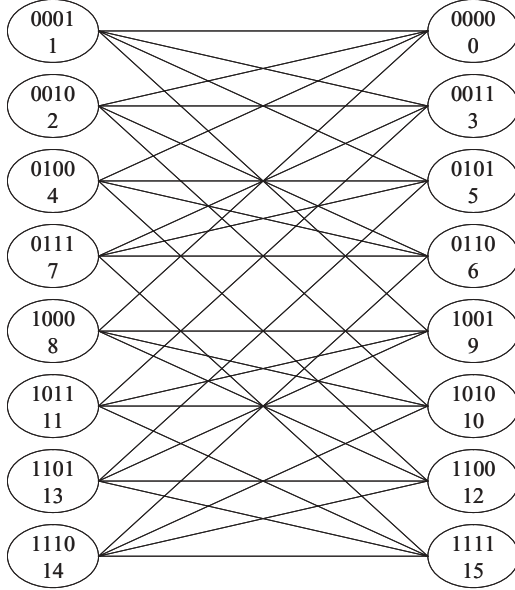


FIG. 2. The Markov graph of a network with  $N = 4$  nodes arranged as bipartite graph.

the states with zero 1 bits into  $S_0$ . Next, we place all states with one 1 bit into  $S_1$ , with two 1 bits into  $S_2$  and so on. Since there are no links between states with a same number of bits, none of the nodes in  $S_0$  and  $S_1$  is connected, so that a bipartite graph is thus constructed. The construction shows that  $S_0$  contains all states with an even number of 1 bits, whereas  $S_1$  contains all states with an odd number of 1 bits. This relabeling of states creates an infinitesimal generator with bipartite structure as illustrated in Fig. 2, where the degree of each node or state is precisely  $N$ . The constant degree of  $N$  arises from the Markov property that only 1 bit can change per event, and the binary representation of each state  $i = \sum_{k=1}^N x_k(i)2^{k-1}$  reveals that each state has  $N$  bits. Hence, the Markov graph is a regular bipartite<sup>6</sup> graph  $B_{2^{N/2}, 2^{N/2}}$  with degree  $N$ .

For a regular Markov graph, the largest eigenvalue [22] of its corresponding adjacency matrix equals the degree  $N$ . Moreover, all eigenvalues of its corresponding Laplacian follow directly from those of the adjacency matrix as  $\mu_j(Q) = N - \lambda_{N+1-j}(A)$ . Consequently, the second smallest eigenvalue of the Laplacian, the algebraic connectivity, equals the spectral gap, the difference between the largest and second largest adjacency eigenvalue and  $\mu_{N-1}(Q) \leq N$  [see Ref. [22], Eq. (4.23)]. However, the infinitesimal generator  $Q_N$  is a weighted Laplacian, for which the above nice properties do not immediately apply.

The relabeled infinitesimal generator  $\tilde{Q}_{N+1}$  has a bipartite structure

$$\tilde{Q}_{N+1} = \begin{bmatrix} \text{diag}((M_N u)_i) & M_N \\ M_N^T & \text{diag}((M_N u)_i) \end{bmatrix}, \quad (3)$$

<sup>6</sup>The complete regular bipartite graph  $K_{m,m}$  has degree equal to  $m$ .

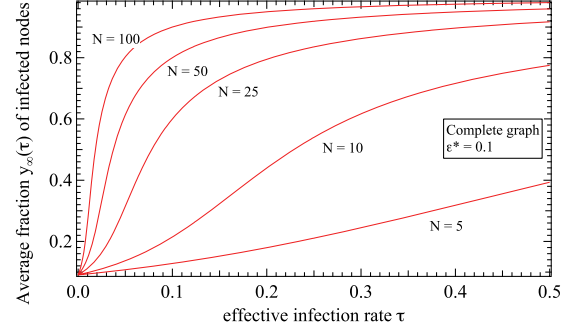


FIG. 3. (Color online) The average steady-state fraction  $y_\infty(\tau)$  of infected nodes in the complete graph versus the effective infection rate  $\tau$  for  $\varepsilon^* = 10^{-1}$  and for various sizes  $N$ .

where  $M_N$  is a  $2^N \times 2^N$  matrix with  $N$  elements per row and the  $i$ th row sum is  $(M_N u)_i$ . The relabeled steady-state vector  $\tilde{\pi}_{N+1} = \begin{bmatrix} y_N \\ w_N \end{bmatrix}$  of  $\tilde{Q}_{N+1}$  obeys  $\tilde{\pi}_{N+1}^T \tilde{Q}_{N+1} = 0$ ,

$$0 = \begin{bmatrix} y_N^T & w_N^T \end{bmatrix} \begin{bmatrix} \Upsilon & M_N \\ M_N^T & \Upsilon \end{bmatrix} \\ = \begin{bmatrix} y_N^T \Upsilon + w_N^T M_N^T & y_N^T M_N + w_N^T \Upsilon \end{bmatrix},$$

where  $\Upsilon = \text{diag}((M_N u)_i)$ . Hence,

$$\begin{cases} y_N^T = -w_N^T M_N^T \text{diag}(\frac{1}{(M_N u)_i}) \\ w_N^T = -y_N^T M_N \text{diag}(\frac{1}{(M_N u)_i}) \end{cases}$$

or

$$\begin{cases} y_N^T = -w_N^T M_N^T \text{diag}(\frac{1}{(M_N u)_i}) \\ w_N^T = -w_N^T M_N^T \text{diag}(\frac{1}{(M_N u)_i}) M_N \text{diag}(\frac{1}{(M_N u)_i}) \end{cases}.$$

Alternatively,

$$\begin{cases} w_N^T = -y_N^T M_N \text{diag}(\frac{1}{(M_N u)_i}) \\ y_N^T = -y_N^T M_N \text{diag}(\frac{1}{(M_N u)_i}) M_N^T \text{diag}(\frac{1}{(M_N u)_i}) \end{cases},$$

illustrating that the vector  $w_N$  and  $y_N$  are related, but, again in a less obvious way, because both involved matrices are different and related as  $H_1 = C^T D$  and  $H_2 = D C^T$ , where  $C = \text{diag}(\frac{1}{(M_N u)_i}) M_N$  and  $D = M_N \text{diag}(\frac{1}{(M_N u)_i})$ .

While we believe that the recursion (2) and the related bipartite structure (Fig. 2) point to fundamental properties of

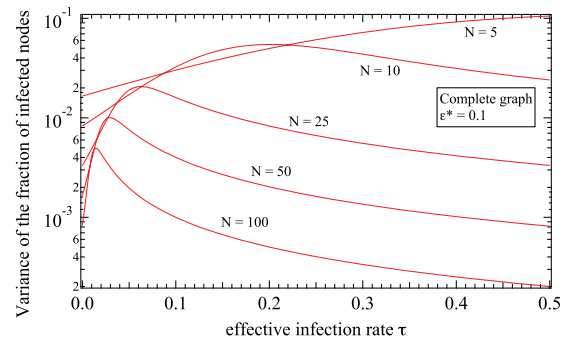


FIG. 4. (Color online) The variance of the steady-state fraction of infected nodes in the complete graph versus the effective infection rate  $\tau$  for  $\varepsilon^* = 10^{-1}$  and for various sizes  $N$ .

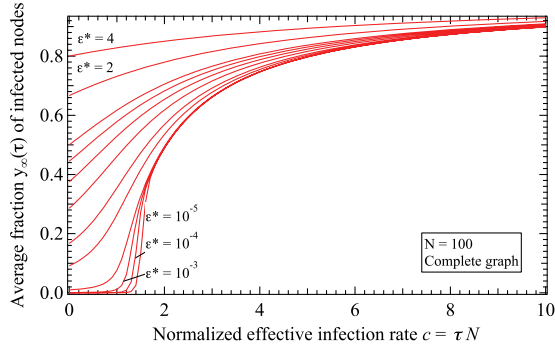


FIG. 5. (Color online) The average steady-state fraction  $y_\infty(\tau)$  of infected nodes in the complete graph with  $N = 100$  nodes versus  $\tau = \frac{c}{N}$  and various values of  $\epsilon^*$ . The unlabeled curves correspond to  $\epsilon^* = \{1, 0.8, 0.6, 0.6, 0.4, 0.2, 0.1, 10^{-2}\}$ .

the  $\epsilon$ -SIS process on any graph, we have failed so far to deduce useful insights.

#### IV. EPIDEMICS ON THE COMPLETE GRAPH $K_N$

We confine to the complete graph because an exact analysis of the steady-state fraction  $y_{\infty;N}(\tau)$  of infected nodes is possible, as presented in Appendix A 1.

Figures 3 and 4 show the average (A5) and variance (A9) of the steady-state fraction  $y_\infty(\tau) = \lim_{t \rightarrow \infty} y(t; \tau)$  of infected nodes for various complete graphs  $K_N$  as a function of the effective infection rate  $\tau$ . Figure 4 illustrates that the largest variance occurs in a region around the “epidemic threshold” and that the variance decreases with  $N$ .

Figure 5 for  $N = 100$  and Fig. 6 for  $N = 500$  illustrate the rate at which  $y_\infty(\tau)$  tends to the absorbing state with  $\epsilon^*$ . When  $\epsilon^* \rightarrow 0$ , the fraction of infected nodes  $y_\infty \rightarrow 0$ , and no obvious interpretation of an epidemic threshold can be deduced. Figure 6 also illustrates that  $\epsilon^*$  is not a perturbation parameter, because arbitrary small, positive values of  $\epsilon^*$  have a significant effect on  $y_\infty$ . However, for large  $N$  and  $\epsilon^* = 10^{-a}$  with  $3 \leq a \leq 10$ , but  $\epsilon^* < \frac{1}{N}$  as shown in Appendix A 2, Fig. 6 on the linear scale leads us to “estimate” the in reality

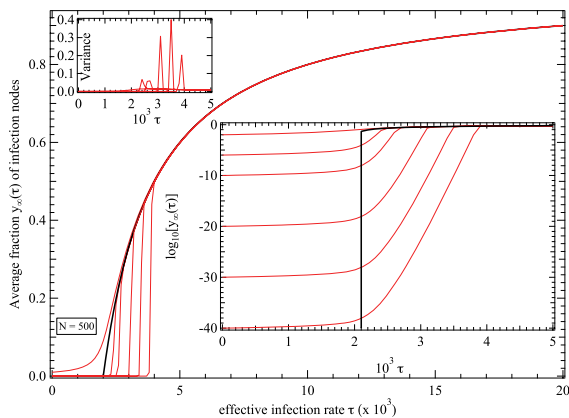


FIG. 6. (Color online) The average fraction  $y_\infty(\tau)$  of infected nodes in  $K_{500}$  as a function of  $\tau$  for various small values of  $\epsilon^* = \{10^{-2}, 10^{-6}, 10^{-10}, 10^{-20}, 10^{-30}, 10^{-40}\}$ , that can be read of in the inserted log-scale plot because  $y_\infty(0) = \frac{\epsilon^*}{1+\epsilon^*} \simeq \epsilon^*$ . The thick line represent the  $N$ -intertwined mean-field approximation.

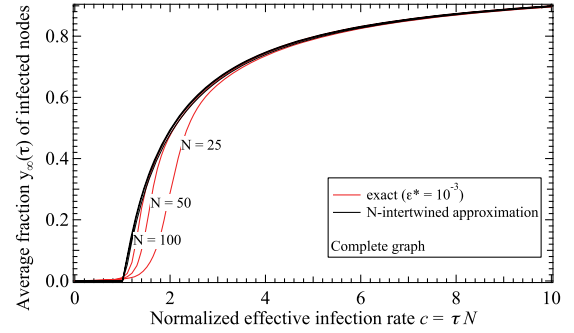


FIG. 7. (Color online) The steady-state average fraction  $y_\infty(\tau)$  of infected nodes versus the normalized effective infection rate  $c = \tau N$  for  $N = 25, 50$ , and  $100$ . Both the exact  $\epsilon^* = 10^{-3}$  SIS model and the  $N$ -intertwined mean-field approximation are shown.

observed threshold fairly well. The left-top insert shows the corresponding highly peaked variance. Figure 6 indicates that the epidemic threshold deduced for  $\epsilon^* = 10^{-6}$  and  $\epsilon^* = 10^{-10}$  differs only moderately. Moreover, much lower values of  $\epsilon^* = 10^{-10}$  are difficult to simulate and seem to be unrealistically small to occur in nature. All these considerations suggested to us to propose to *define* the epidemic threshold for a small  $\epsilon^* = 10^{-a} < \frac{1}{N}$  with  $3 \leq a \leq 10$  as the  $\tau$  value where  $y_\infty(\tau)$  jumps from practically zero to some positive value.

For large  $\tau$  and larger  $N$ , the  $N$ -intertwined mean-field model, yielding  $y_{\infty;N}(\tau) \lesssim 1 - \frac{1}{(N-1)\tau}$ , is accurate as verified from Figs. 6 and 7. The  $N$ -intertwined mean-field approximation and the exact SIS model are further compared for different graph types in Ref. [18].

#### V. CONCLUSIONS

The classical SIS epidemic process has been generalized to the  $\epsilon$ -SIS model: Besides link infections from infected neighbors with rate  $\beta$ , a node itself produces a virus with rate  $\epsilon$  and can be cured with rate  $\delta$ . The exact Markov chain is derived, from which the steady state can be computed, and an exact recursion relation for the Markov infinitesimal generator is presented. The major advantage of the  $\epsilon$ -SIS model is that its steady state is different from the absorbing (or overall-healthy state) and approximates, for a certain range of small  $\epsilon > 0$ , the in reality observed phase transition, that is characterized by the epidemic threshold. The complete graph allows an exact analysis, that illustrates the effect of small  $\epsilon$  and the quality of the first-order mean-field approximation, the  $N$ -intertwined model, proposed earlier.

#### APPENDIX: THE NUMBER OF INFECTED NODES IN $K_N$

The number of infected nodes  $M(t)$  at time  $t$  in the complete graph  $K_N$  is a continuous-time Markov process on  $\{0, 1, \dots, N\}$  with the following rates:

$$\begin{aligned} M &\mapsto M + 1 & \text{at rate} & \quad (\beta M + \epsilon)(N - M), \\ M &\mapsto M - 1 & \text{at rate} & \quad \delta M. \end{aligned}$$

Every infected node heals with rate  $\delta$ , whereas every healthy node (of which there are  $N - M$ ) has exactly  $M$  infected neighbors, each actively transferring the virus with rate  $\beta$  in

addition to the self-infection rate  $\varepsilon$ . This Markov process  $M(t)$  is, in fact, a birth and death process with birth rate  $\lambda_j = (\beta j + \varepsilon)(N - j)$  and death rate  $\mu_j = j\delta$ , whose steady-state probabilities  $\pi_0, \dots, \pi_N$ , where  $\pi_j = \lim_{t \rightarrow \infty} \text{Prob}[M(t) = j]$ , can be computed exactly [21], p. 209] as

$$\pi_0 = \frac{1}{1 + \sum_{k=1}^N \prod_{m=0}^{k-1} \frac{(\beta m + \varepsilon)(N - m)}{(m+1)\delta}}, \quad (\text{A1})$$

$$\pi_j = \pi_0 \prod_{m=0}^{j-1} \frac{(\beta m + \varepsilon)(N - m)}{(m+1)\delta} \quad 1 \leq j \leq N. \quad (\text{A2})$$

Using the Gamma function's basic property  $\Gamma(z+1) = z\Gamma(z)$ , we have

$$\begin{aligned} \pi_j &= \pi_0 \prod_{m=0}^{j-1} \frac{\tau m + \varepsilon^*}{m+1} \prod_{m=0}^{j-1} (N - m) \\ &= \frac{\pi_0}{j!} \tau^j \prod_{m=1}^j \left(m - 1 + \frac{\varepsilon^*}{\tau}\right) \prod_{m=1}^j (N + 1 - m) \\ &= \frac{\pi_0}{j!} \tau^j \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)} \frac{N!}{(N-j)!} = \pi_0 \binom{N}{j} \tau^j \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)} \\ &= \pi_0 \binom{N}{j} \varepsilon^* \tau^{j-1} \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau} + 1\right)}, \end{aligned} \quad (\text{A3})$$

which shows that, if  $\varepsilon^* = 0$ , then  $\pi_j = 0$  and consequently,  $\pi_0 = 1$ , because  $\sum_{i=0}^N \pi_i = 1$  (by the conservation of probability). Further, (A1) becomes

$$\pi_0 = \frac{1}{\sum_{k=0}^N \binom{N}{k} \tau^k \frac{\Gamma\left(\frac{\varepsilon^* + k}{\tau}\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)}}. \quad (\text{A4})$$

For  $\tau = 0$  [and approximately for large  $\frac{\varepsilon^*}{\tau}$  because then  $\frac{\Gamma\left(\frac{\varepsilon^* + j}{\tau}\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)} \sim \left(\frac{\varepsilon^*}{\tau}\right)^{j-1}$ ], we find that  $\pi_j = \pi_0 \binom{N}{j} (\varepsilon^*)^j$  and  $\pi_0 = \frac{1}{\sum_{k=0}^N \binom{N}{k} (\varepsilon^*)^k} = \frac{1}{(1 + \varepsilon^*)^N}$ , from which  $y_{\infty;N}(0) = \frac{\varepsilon^*}{1 + \varepsilon^*}$ , in agreement with the general theory above. When  $\frac{\varepsilon^*}{\tau} = \frac{\varepsilon}{\beta} = 1$ , then  $\pi_j = \pi_0 \frac{N!}{(N-j)!} \tau^j$ , which almost reduces to the steady state of a Markov model with a forbidden absorbing state.

For large  $\tau$ , we find from (A3) that

$$\pi_j \sim \pi_0 \frac{N!}{(N-j)! j} \varepsilon^* \tau^{j-1},$$

and from (A4) that

$$\begin{aligned} \pi_0^{-1} &= 1 + \varepsilon^* \sum_{k=1}^N \binom{N}{k} \tau^{k-1} \frac{\Gamma\left(\frac{\varepsilon^* + k}{\tau}\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)} \\ &\sim 1 + \varepsilon^* \sum_{k=1}^N \frac{N!}{(N-k)! k} \tau^{k-1} \\ &= 1 + N! \tau^{N-1} \varepsilon^* \sum_{j=0}^{N-1} \frac{\tau^{-j}}{j!(N-j)} \\ &\sim (N-1)! \tau^{N-1} \varepsilon^*. \end{aligned}$$

Hence, for large  $\tau$  and fixed  $\varepsilon^*$ , it holds that

$$\pi_j \sim \frac{N}{(N-j)! j} \tau^{j-N},$$

illustrating that the steady-state probability that  $j$  nodes are infected increases with  $j$  and ultimately that  $\lim_{\tau \rightarrow \infty} \pi_j = \delta_{0N} = 1_{\{j=N\}}$  and  $\lim_{\tau \rightarrow \infty} y_{\infty;N}(\tau) = 1$ .

### 1. The average steady-state fraction $y_{\infty;N}(\tau)$ of infected nodes

The average steady-state fraction of infected nodes is

$$y_{\infty;N}(\tau) = \frac{1}{N} \sum_{j=0}^N j \pi_{j;N} = \frac{\pi_{0;N}}{N} \sum_{j=1}^N j \binom{N}{j} \tau^j \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)}, \quad (\text{A5})$$

and, using  $j \binom{N}{j} = N \binom{N-1}{j-1}$ ,

$$\begin{aligned} \frac{y_{\infty;N}(\tau)}{\pi_{0;N}} &= \sum_{j=1}^N \binom{N-1}{j-1} \tau^j \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)} \\ &= \tau \sum_{j=0}^{N-1} \binom{N-1}{j} \tau^j \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j + 1\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)} \\ &= \varepsilon^* + \tau \sum_{j=1}^{N-1} \binom{N-1}{j} \tau^j \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j + 1\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)}. \end{aligned}$$

Further, using the Gamma function's basic property  $\Gamma(z+1) = z\Gamma(z)$  in the last sum gives

$$\begin{aligned} R &= \sum_{j=1}^{N-1} \binom{N-1}{j} \tau^j \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)} \\ &= \frac{\varepsilon^*}{\tau} \sum_{j=1}^{N-1} \binom{N-1}{j} \tau^j \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)} \\ &\quad + \sum_{j=1}^{N-1} j \binom{N-1}{j} \tau^j \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)}. \end{aligned}$$

We recognize from (A4) that the first sum equals

$$\sum_{j=1}^{N-1} \binom{N-1}{j} \tau^j \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)} = \frac{1}{\pi_{0;N-1}} - 1,$$

while (A5) indicates that the last sum equals

$$\sum_{j=1}^{N-1} j \binom{N-1}{j} \tau^j \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)} = (N-1) \frac{y_{\infty;N-1}(\tau)}{\pi_{0;N-1}},$$

so that a recursion relation for  $\frac{y_{\infty;N}(\tau)}{\pi_{0;N}}$

$$\frac{y_{\infty;N}(\tau)}{\pi_{0;N}} = \frac{\varepsilon^*}{\pi_{0;N-1}} + (N-1)\tau \frac{y_{\infty;N-1}(\tau)}{\pi_{0;N-1}} \quad (\text{A6})$$

is found.

Next, we use the binomial recursion  $\binom{N}{k} = \binom{N-1}{k} + \binom{N-1}{k-1}$  in the denominator  $F_d$  of

$$y_{\infty;N}(\tau) = \frac{\sum_{k=1}^N \binom{N-1}{k-1} \tau^k \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + k\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)}}{\sum_{k=0}^N \binom{N}{k} \tau^k \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + k\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)}}$$



and obtain, invoking the recursion (A6) and (A4),

$$\begin{aligned} F_d &= \sum_{k=0}^N \left\{ \binom{N-1}{k-1} + \binom{N-1}{k} \right\} \tau^k \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + k\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau}\right)} \\ &= \frac{\varepsilon^*}{\pi_{0;N-1}} + (N-1)\tau \frac{y_{\infty;N-1}(\tau)}{\pi_{0;N-1}} + \frac{1}{\pi_{0;N-1}}, \end{aligned}$$

which leads to a recursion for  $y_{\infty;N}(\tau)$ :

$$y_{\infty;N}(\tau) = \frac{1}{1 + \frac{1}{\varepsilon^* + (N-1)\tau y_{\infty;N-1}(\tau)}}, \quad (\text{A7})$$

from which we finally deduce

$$y_{\infty;N}(\tau) = 1 - \frac{1}{\frac{\varepsilon^*}{y_{\infty;N}(\tau)} + (N-1)\tau \frac{y_{\infty;N-1}(\tau)}{y_{\infty;N}(\tau)}}. \quad (\text{A8})$$

For sufficiently large  $N$ , there holds that  $\frac{y_{\infty;N-1}(\tau)}{y_{\infty;N}(\tau)} \lesssim 1$  and (A8) indicates that, when  $\frac{\varepsilon^*}{y_{\infty;N}(\tau)} \leq 1 + \varepsilon^*$  is small enough to neglect, we find that  $y_{\infty;N}(\tau) \approx 1 - \frac{1}{(N-1)\tau}$ , which supports the simulations in Fig. 7. Recall [12,23] that the steady-state fraction of infected nodes in the  $N$ -intertwined model (for  $\varepsilon^* = 0$ ) for the complete graph  $K_N$  equals, for  $\tau \geq \tau_c^{(1)} = \frac{1}{N-1}$ ,

$$y_{\infty;N}^{(1)}(\tau) = 1 - \frac{1}{(N-1)\tau},$$

illustrating how good the first-order mean-field approximation for the complete graph is. When iterating (A7), a continued fraction for  $y_{\infty;N}(\tau)$  is found, which bears resemblance with the continued fraction of  $v_{j\infty}$  in the  $N$ -intertwined model. Since  $y_{\infty;N-1}(\tau) \geq y_{\infty;N-1}(0) = \frac{\varepsilon^*}{1+\varepsilon^*}$ , each convergent can be used as a lower bound. For example, the first convergent is

$$1y_{\infty;N}(\tau) = \frac{1}{1 + \frac{1}{\varepsilon^* + (N-1)\tau \frac{\varepsilon^*}{1+\varepsilon^*}}} < y_{\infty;N}(\tau),$$

and the second convergent is

$$2y_{\infty;N}(\tau) = \frac{1}{1 + \frac{1}{\varepsilon^* + \frac{1}{1 + \frac{1}{\varepsilon^* + (N-1)\tau \frac{\varepsilon^*}{1+\varepsilon^*}}}}} < y_{\infty;N}(\tau).$$

The variance  $\sigma_y^2 = \frac{1}{N^2} \text{Var}[\sum_{j=1}^N X_j]$  of the steady-state fraction of infected nodes equals

$$\sigma_y^2 = \frac{1}{N^2} \sum_{j=0}^N j^2 \pi_{j;N} - y_{\infty;N}^2(\tau), \quad (\text{A9})$$

from which we can deduce that  $\sigma_y^2(0) = \frac{1}{N} \frac{\varepsilon^*}{(1+\varepsilon^*)^2}$ .

## 2. Scaling of the epidemic threshold

Let us consider the ratio  $\frac{\pi_{j+1}}{\pi_j} = \frac{(\tau j + \varepsilon^*)(N-j)}{j+1}$  of the steady-state probabilities in (A2). The ratio is maximal for

$$j^* = \sqrt{(N+1) \left(1 - \frac{\varepsilon^*}{\tau}\right)} - 1$$

and

$$\left(\frac{\pi_{j+1}}{\pi_j}\right)_{\max} = \left[ (N+2) - \frac{\varepsilon^*}{\tau} - 2\sqrt{(N+1) \left(1 - \frac{\varepsilon^*}{\tau}\right)} \right] \tau,$$

while the requirement  $\frac{\pi_{j+1}}{\pi_j} = 1$  yields

$$j_{\pm} = \frac{1}{2} \left( N - \frac{\varepsilon^* + 1}{\tau} \right) \pm \frac{1}{2} \sqrt{\left( N - \frac{\varepsilon^* + 1}{\tau} \right)^2 + \frac{4(N\varepsilon^* - 1)}{\tau}}.$$

If  $\varepsilon^* > \frac{1}{N}$ , there is only one (non-negative) index for  $j$  at which  $\frac{\pi_{j+1}}{\pi_j} = 1$ . Thus,  $\frac{\pi_{j+1}}{\pi_j} < 1$ , implying that  $\pi_j$  decreases with  $j$ , when  $j > j_+ \approx N - \frac{\varepsilon^* + 1}{\tau}$ . In other words,  $\pi_j$  increases with  $j$  when  $j < j_+$ .

The more interesting case appears if  $\varepsilon^* < \frac{1}{N}$ , then  $j_- \approx \frac{1-N\varepsilon^*}{\tau(N-\frac{\varepsilon^*+1}{\tau})}$  and  $\pi_j$  decreases with  $j$  when  $j < j_-$  and  $j > j_+$ . Now, when  $j_- = j_+$  there is only one value for which  $\frac{\pi_{j+1}}{\pi_j} = 1$  and this is the maximum value  $\left(\frac{\pi_{j+1}}{\pi_j}\right)_{\max}$ . This means that  $\pi_j$  decreases with  $j$  nearly everywhere, except in a small region around  $j = j_+ = j_-$ . We can consider this region as the onset of the epidemic which may define the epidemic threshold of the SIS epidemics (for negligibly small  $\varepsilon^*$ ). Hence, if we choose  $\left(\frac{\pi_{j+1}}{\pi_j}\right)_{\max} = 1$  and let  $\varepsilon^* = 0$ , then we arrive at the scaling for this defined epidemic threshold

$$\tau_c^* = \frac{1}{(N+2) - 2\sqrt{(N+1)}} = \frac{1}{N} \left[ 1 + \frac{2}{\sqrt{N}} + O\left(\frac{1}{N}\right) \right]. \quad (\text{A10})$$

We observe that  $\tau_c^* > \tau_c^{(1)} = \frac{1}{N-1} = \frac{1}{N} [1 + \frac{1}{N} + O(\frac{1}{N^2})]$ , in line with the fact that the  $N$ -intertwined mean-field approximation upper bounds the viral node probability [14].

## 3. Asymptotics for $\pi_j$

Since  $\prod_{j=1}^i (N-j) = N^i \exp[\sum_{j=1}^i \log(1 - \frac{j}{N})]$  and

$$\begin{aligned} \prod_{j=1}^i (N-j) &= N^i \exp \left[ -\sum_{j=1}^i \frac{j}{N} + O(i^3 N^{-2}) \right] \\ &= N^i \exp \left[ -\frac{i^2}{2N} + O(iN^{-1}) + O(i^3 N^{-2}) \right], \end{aligned}$$

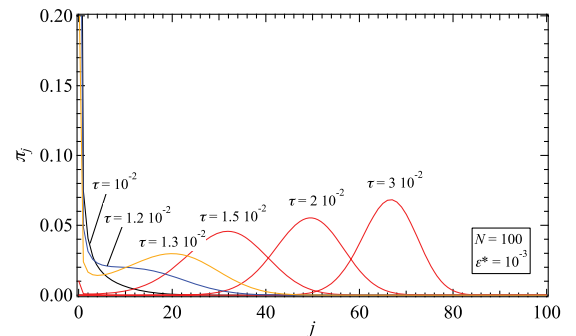


FIG. 8. (Color online) The steady-state probability  $\pi_j$  computed from (A3) versus  $j$  in  $K_{100}$  with  $\varepsilon^* = 10^{-3}$ , for various values of  $\tau$  around the defined epidemic threshold  $\tau_c \simeq 1.2 \times 10^{-2}$  in (A10).

we have that

$$\begin{aligned}\pi_j &= \frac{\pi_0}{j!} \varepsilon^* \tau^{j-1} \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau} + 1\right)} N \prod_{m=1}^{j-1} (N - m) \\ &= \frac{\pi_0}{j!} \frac{\varepsilon^*}{\tau} \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{\Gamma\left(\frac{\varepsilon^*}{\tau} + 1\right)} (\tau N)^j \\ &\quad \times \exp\left[-\frac{j^2}{2N} + O(jN^{-1}) + O(j^3N^{-2})\right] \\ &= \pi_0 \frac{\varepsilon^*}{\tau} \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{j! \Gamma\left(\frac{\varepsilon^*}{\tau} + 1\right)} \exp\left[-\frac{j^2}{2N} + j \log(\tau N)\right] \\ &\quad \times (1 + O(jN^{-1}) + j^3N^{-2})\end{aligned}$$

and

$$\begin{aligned}\pi_j &= \pi_0 \frac{\varepsilon^*}{\tau} \frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{j! \Gamma\left(\frac{\varepsilon^*}{\tau} + 1\right)} e^{\frac{N}{2}(\log \tau N)^2} \exp\left[-\frac{1}{2} \frac{(j - N \log \tau N)^2}{N}\right] \\ &\quad \times [1 + O(jN^{-1}) + j^3N^{-2}].\end{aligned}$$

For large  $N$  and small  $\frac{\varepsilon^*}{\tau}$  so that  $\frac{\Gamma\left(\frac{\varepsilon^*}{\tau} + j\right)}{j! \Gamma\left(\frac{\varepsilon^*}{\tau} + 1\right)} \approx \frac{1}{j!}$ , the appropriate scaling for  $\tau$  is found by requiring that  $N(\log \tau N)^2 = C^2$ , where  $C$  is independent of  $N$ . Hence,

$$\tau = \frac{1}{N} \exp\left(\frac{C}{\sqrt{N}}\right) = \frac{1}{N} \left\{1 + \frac{C}{\sqrt{N}} + O(N^{-1})\right\}$$

and

$$\begin{aligned}\pi_j &\simeq \pi_0 \frac{\varepsilon^*}{j\tau} e^{\frac{1}{2}C^2} \exp\left[-\frac{1}{2} \frac{(j - C\sqrt{N})^2}{N}\right] \\ &\quad \times [1 + O(jN^{-1}) + j^3N^{-2}],\end{aligned}$$

illustrating that, asymptotically,  $\pi_j$  resembles approximately a Gaussian with mean  $C\sqrt{N}$  and standard deviation  $\sigma = \sqrt{N}$ , in agreement with Fig. 8.

- 
- [1] A. L. Hill, D. G. Rand, M. A. Nowak, and N. A. Christakis, *Proc. Roy. Soc. B* **277**, 3827 (2010).
- [2] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications*, 2nd ed. (Charlin Griffin & Company, London, 1975).
- [3] J. O. Kephart and S. R. White, in *Proc. 1991 IEEE Computer Society Symposium on Research in Security and Privacy* (May 1991), pp. 343–359.
- [4] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).
- [5] D. J. Daley and J. Gani, *Epidemic Modelling: An Introduction* (Cambridge University Press, Cambridge, 1999).
- [6] C. Castellano and R. Pastor-Satorras, *Phys. Rev. Lett.* **105**, 218701 (2010).
- [7] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008).
- [8] M. Youssef, R. Kooij, and C. Scoglio, *J. Comput. Sci.* **2**, 286 (2011).
- [9] Y. Chen, G. Paul, S. Havlin, F. Liljeros, and H. E. Stanley, *Phys. Rev. Lett.* **101**, 058701 (2008).
- [10] P. Van Mieghem, D. Stevanović, F. A. Kuipers, C. Li, R. van de Bovenkamp, D. Liu, and H. Wang, *Phys. Rev. E* **84**, 016101 (2011).
- [11] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, in *22nd International Symposium on Reliable Distributed Systems (SRDS'03); IEEE Computer* (October 2003), pp. 25–34.
- [12] P. Van Mieghem, J. Omic, and R. E. Kooij, *IEEE/ACM Trans. Netw.* **17**, 1 (2009).
- [13] P. Van Mieghem, *Europhys. Lett.* **97**, 48004 (2012).
- [14] E. Cator and P. Van Mieghem, *Phys. Rev. E* **85**, 056111 (2012).
- [15] A. Ganesh, L. Massoulié, and D. Towsley, *IEEE INFOCOM*, Miami, pp. 1455–1466, 2005.
- [16] M. Draief and L. Massoulié, *Epidemics and Rumours in Complex Networks*, London Mathematical Society Lecture Note Series 369 (Cambridge University Press, Cambridge, 2010).
- [17] R. Durrett, *Proc. Natl. Acad. Sci. USA* **107**, 4491 (2010).
- [18] C. Li, R. van de Bovenkamp, and P. Van Mieghem, *Phys. Rev. E* (in preparation).
- [19] P. Bak, *How Nature Works: The Science of Self-organized Criticality* (Copernicus, Springer-Verlag, New York, 1996).
- [20] J. Omic, PhD. thesis, 2010, <http://repository.tudelft.nl/>.
- [21] P. Van Mieghem, *Performance Analysis of Communications Systems and Networks* (Cambridge University Press, Cambridge, 2006).
- [22] P. Van Mieghem, *Graph Spectra for Complex Networks* (Cambridge University Press, Cambridge, 2011).
- [23] P. Van Mieghem, *Computing* **93**, 147 (2011).