Susceptible-infected-susceptible model: A comparison of N-intertwined and heterogeneous mean-field approximations

Article *in* Physical Review E · August 2012

CITATIONS
8

READS
329

3 authors, including:

Cong Li
Fudan University
48 PUBLICATIONS **726** CITATIONS

SEE PROFILE

Susceptible-infected-susceptible model: A comparison of *N*-intertwined and heterogeneous mean-field approximations

Cong Li, Ruud van de Bovenkamp, and Piet Van Mieghem
Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands
(Received 20 March 2012; revised manuscript received 4 June 2012; published 28 August 2012; corrected 6 September 2012)

We introduce the ε-susceptible-infected-susceptible (SIS) spreading model, which is taken as a benchmark for the comparison between the *N*-intertwined approximation and the Pastor-Satorras and Vespignani heterogeneous mean-field (HMF) approximation of the SIS model. The *N*-intertwined approximation, the HMF approximation, and the ε-SIS spreading model are compared for different graph types. We focus on the epidemic threshold and the steady-state fraction of infected nodes in networks with different degree distributions. Overall, the *N*-intertwined approximation is superior to the HMF approximation. The *N*-intertwined approximation is exactly the same as the HMF approximation in regular graphs. However, for some special graph types, such as the square lattice graph and the path graph, the two mean-field approximations are both very different from the ε-SIS spreading model.

DOI: 10.1103/PhysRevE.86.026116 PACS number(s): 89.75.Hc, 89.75.Fb, 05.90.+m

I. INTRODUCTION

We consider the spread of a virus in an undirected graph $G(N,L)$, characterized by a symmetric adjacency matrix A. As a spreading model we use the susceptible-infected-susceptible [1] (SIS) epidemic process, which is described as follows. The arrival of an infection over a link and the curing of an infected node are assumed to be independent Poisson processes with rates β and δ, respectively. Only infected nodes can infect their healthy direct neighbors. The effective spreading rate is defined as $\tau = \frac{\beta}{\delta}$. The viral state of a node i at time t is specified by a Bernoulli random variable $X_i(t) \in \{0,1\}$, where $X_i(t) = 0$ refers to a healthy node and $X_i(t) = 1$ to an infected node. Every node i at time t is either infected, with probability $v_i(t) = \text{Prob}[X_i(t) = 1]$ or healthy (but susceptible) with probability $1 - v_i(t)$. Since an exact solution for any network has not been found yet, several approximations of SIS epidemics have been developed.

A fundamental question in the study of epidemics is whether a virus will spread through the entire network or will die out. The answer to this question is given by the epidemic threshold τ_c, which separates two different phases of the dynamic spreading process on a network: if the effective infection rate τ is above the threshold, the infection spreads and becomes persistent in time; if $\tau < \tau_c$, the infection dies out exponentially fast. Many authors (see [2–9]) mention the existence of an epidemic threshold τ_c. Here, we focus on the *steady-state* of two *mean-field approximations* of SIS epidemics: the *N*-intertwined approximation [10,11] and the Pastor-Satorras and Vespignani approximation [7]. A first-order mean-field epidemic threshold $\tau_c^{(1)} = \frac{1}{\lambda_1(A)}$, where $\lambda_1(A)$ is the largest eigenvalue of the adjacency matrix A, was first proposed by Wang *et al.* [9], and its existence rigorously proved by Van Mieghem *et al.* [10,11]; later it appeared in the physics community [12]. Van Mieghem *et al.* [10] also showed that this mean-field threshold lower-bounds the "in reality observed" epidemic threshold $\tau_c^{(1)} = \frac{1}{\lambda_1(A)} \leqslant \tau_c$. A more accurate lower bound (the second-order mean-field threshold) $\tau_c \geqslant \tau_c^{(2)} \geqslant \tau_c^{(1)}$ has been derived in [13]. Pastor-Satorras and Vespignani [7] proposed the heterogeneous mean-field HMF approximation, whose epidemic threshold [4,7] is given by $\tau_c^{\text{HMF}} = E[D]/E[D^2]$, where D is the degree of a randomly chosen node in G.

Here we present a detailed comparison of the two mean-field approximations. Usually, the quality of an approximation is assessed by two criteria: (1) which approximation is closer to the exact SIS model, and (2) which approximation's epidemic threshold is nearer to the epidemic threshold of the exact SIS model. A direct comparison to the SIS model is, however, not possible, because the steady state of the exact SIS model in a finite network is, as shown in [10], the overall-healthy state, which is equal to the absorbing state of the SIS Markov chain. The presence of an absorbing state is a major complication in the analysis of the SIS model. The steady state of both the above mean-field approximations corresponds, in fact, to the *metastable state* in the SIS model, which is not clearly defined for finite networks [10]. Therefore, we *define* here the metastable state of the SIS model via the steady state of the ε-SIS model for a prescribed value of ε. The ε-SIS process generalizes the SIS model by adding a nodal component to the infection. We assume that each node i can be infected spontaneously. The spontaneous infection process is a Poisson process with rate ε. Hence, besides receiving the infection over links from infected neighbors with rate β, the node i can also itself produce a virus with rate ε. All involved Poisson processes are independent. For $\varepsilon > 0$, the ε-SIS model has no absorbing state and Markov theory guarantees a unique steady state. When $\varepsilon = 0$, the ε-SIS model clearly reduces to the "classical" SIS model. Hence, for small values of $\varepsilon > 0$, the ε-SIS spreading model can be used to approximate the exact SIS model. Here, the ε-SIS spreading model with a small value of ε is used as a benchmark to compare the steady state of the *N*-intertwined approximation and the HMF approximation on different network types.

This paper is organized as follows. Section II overviews the *N*-intertwined approximation, the Pastor-Satorras and Vespignani HMF approximation, and the ε-SIS spreading model in detail. The steady state of infection in the ε-SIS model and these two approximations are described in Sec. III. Section IV compares the steady-state fraction of infected nodes in various

types of graphs: complete graphs, star graphs, Erdős-Rényi (ER) random graphs, small-world graphs, and Bárabasi-Albert graphs. An analytic comparison of the epidemic thresholds of the two mean-field approximations is shown in Sec. V. Conclusions are summarized in Sec. VI.

II. DESCRIPTION OF THE ε-SIS MODEL AND THE MEAN-FIELD APPROXIMATIONS

A. The ε-SIS spreading model

The ε-SIS spreading model was proposed recently by Hill *et al*. [14] in their analysis of emotions as a form of infection in a social contact network and earlier in [15] where ε is defined as the driving field conjugate to the density of infected nodes. Here, we will explain the simulation process, but defer to [16] for an analysis of the ε-SIS model.

In our simulations we take a nodal-central, event-driven approach. An event can either be the curing of a node or the spreading of the infection from one node to another. Events are stored in a timeline as tickets. A ticket contains, besides the time and the event type (spreading or curing), the owner of the ticket. The ticket owner is usually a node, but can also be the system to allow for scheduling of administrative tasks. Tickets are continuously taken from the timeline and passed on to the owner.

If the ticket owner is a node, the ticket indicates either a curing or a spreading event. In the case of a curing event, the node simply changes its state from infected to healthy; in the case of a spreading event, it will spread the infection to the neighbor mentioned in the ticket. If the neighbor was not already infected, it will now become infected and create one or more tickets.

A newly infected node will always create a ticket for its own curing event. According to continuous-time Markov theory (see [17]), the time between infection and curing is exponentially distributed with rate δ and is stored by the node for future reference. An infected node also generates spreading times at which it will spread the infection to its neighbors. The spreading times are again exponentially distributed but now with rate β. If the spreading time does not exceed the node's curing time, a ticket is created for the spreading event. All newly created tickets are stored in the timeline. Finally, the owner of the original ticket generates a new spreading time, which, if not exceeding its own curing time, creates a new spreading ticket for the same neighbor.

If the ticket is not owned by a node, it is a system ticket. System tickets are used to cause the spontaneous infections in nodes. Every node becomes infected spontaneously at a rate ε, but to minimize the number of tickets in the timeline, the system creates one spontaneous infection ticket at the time. The time between spontaneous infection tickets is exponentially distributed with rate $N\varepsilon$. When the system receives a spontaneous infection ticket, it selects a random node and tries to infect it. If the node is already infected, nothing will change, whereas a healthy node will become infected and create the tickets described above.

During the simulation, for each possible number of infected nodes (0 to N) how long the network was in a state with that many nodes infected is recorded. The average number of infected nodes during the simulation can be determined by multiplying the number of infected nodes by the fraction of time spent in that state, and sum over all the states.

B. The Pastor-Satorras and Vespignani HMF approximation

Pastor-Satorras and Vespignani [7] studied the susceptible-infected-susceptible epidemic on networks and proposed the heterogeneous mean-field approximation, in which the degree distribution plays an important role. Highly connected nodes are statistically significant and the strong fluctuations in the degree distribution cannot be neglected. Consider the relative density $\rho_k(t)$ of infected nodes with given degree k, i.e., the probability that a node with k links is infected. The fraction of infected nodes in a network is denoted by ρ. The dynamical mean-field reaction rate equation can be written as

$$\partial_t \rho_k(t) = -\delta\rho_k(t) + \beta k[1-\rho_k(t)]\Theta(\rho(t)).$$

$\Theta(\rho(t))$ is the probability that any given link points to an infected node. In steady state, $y_\infty = \lim_{t\to\infty}\rho(t)$ is a function of τ only, and as consequence, so is $\Theta(\rho(t))$. By imposing stationarity $[\partial_t\rho_k(t)=0]$, when $t\to\infty$, the relative density reduces to

$$\rho_k(\tau) = \frac{\tau k\Theta(\tau)}{1+k\tau\Theta(\tau)}, \tag{1}$$

where $\tau = \frac{\beta}{\delta}$ is the effective infection rate and

$$\Theta(\tau) = \frac{1}{E[D]}\sum_{k=1}^{N-1} k\,\mathrm{Prob}[D=k]\rho_k(\tau). \tag{2}$$

Here D is the degree of a randomly chosen node in the graph. Clearly, if $\tau=0$, then $\Theta(0)=0$. Substituting (1) into (2) leads to a self-consistent relation, from which $\Theta(\tau)$ can be determined as

$$\Theta(\tau) = \frac{\tau\Theta(\tau)}{E[D]}\sum_{k=1}^{N-1}\frac{k^2\mathrm{Prob}[D=k]}{1+k\tau\Theta(\tau)}. \tag{3}$$

Equation (3) has a trivial solution, $\Theta(\tau)=0$. For a nontrivial solution $\Theta(\tau)>0$ to exist, Eq. (3) must satisfy the following condition:

$$\frac{E[D]}{\tau} = \sum_{k=1}^{N-1}\frac{k^2\mathrm{Prob}[D=k]}{1+k\tau\Theta(\tau)}. \tag{4}$$

Next, we introduce the following expansion:

$$\frac{1}{1+k\tau\Theta(\tau)} = \sum_{j=0}^{\infty}(-1)^j[k\tau\Theta(\tau)]^j,$$

valid when $k\tau\Theta(\tau)<1$ for all k, and

$$\begin{aligned}\frac{E[D]}{\tau} &= \sum_{j=0}^{\infty}(-1)^j\left\{\sum_{k=1}^{N-1}\mathrm{Prob}[D=k]k^{j+2}\right\}\tau^j\Theta^j(\tau)\\ &= \sum_{j=0}^{\infty}(-1)^j E[D^{j+2}]\tau^j\Theta^j(\tau),\end{aligned}$$

where the latter series converges for $\Theta(\tau)<1/(D_{\max}\tau)$. Since $\tau=0$ leads to $\Theta(0)=0$, the nontrivial solution $\Theta(\tau)>0$

occurs when $\tau > \tau_c^{\mathrm{HMF}} \geqslant 0$ by the definition of the epidemic threshold. When $\Theta(\tau)$ is sufficiently small $[\Theta(\tau) < 1/(D_{\max}\tau)]$ and $\Theta(\tau) > 0$, we can write the above expansion up to first order as

$$\frac{E[D]}{\tau} = E[D^2] - \tau\Theta(\tau)E[D^3] + O(\Theta(\tau)^2), \tag{5}$$

in which $\tau\Theta(\tau)E[D^3] > 0$. Hence, when $\tau > \tau_c^{\mathrm{HMF}}$ but $\Theta(\tau)$ is small enough to ignore the second-order terms $O(\Theta(\tau)^2)$, we have from (5)

$$\frac{E[D]}{\tau} < E[D^2],$$

implying that for all $\tau > \tau_c^{\mathrm{HMF}}$, it holds that $\tau > \frac{E[D]}{E[D^2]}$. Thus, the epidemic threshold of the HMF approximation is

$$\tau_c^{\mathrm{HMF}} = \frac{E[D]}{E[D^2]}.$$

The same result was also deduced differently in [18]. For a regular graph [7] with degree r, $E[D^2] = E[D]^2 = r^2$, the epidemic threshold is $\tau_c^{\mathrm{HMF}} = \frac{1}{r} = \frac{1}{\lambda_1}$.

Finally, we can evaluate the fraction $y_\infty(\tau)$ of infected nodes using the relation

$$y_\infty(\tau) = \sum_{k=1}^{N-1} \mathrm{Prob}[D = k]\rho_k(\tau). \tag{6}$$

C. *N*-intertwined approximation

The HMF approximation considers the relative density $\rho_k(t)$ of infected nodes with given degree k during the epidemic process. However, the state of each node is not taken into account. The N-intertwined epidemic approximation [10,19] is derived by separately observing each node. Every node i at time t in the network is in one of two states: infected, with probability $\mathrm{Prob}[X_i(t) = 1]$, and healthy, with probability $\mathrm{Prob}[X_i(t) = 0]$. Since a node can only be in one of two states, $\mathrm{Prob}[X_i(t) = 0] + \mathrm{Prob}[X_i(t) = 1] = 1$. Since the curing and infection processes are Poisson processes, the whole epidemic process is a Markov process. If we apply Markov theory straightforwardly, the infinitesimal generator $Q_i(t)$ of this two-state continuous Markov chain is

$$Q_i(t) = \begin{bmatrix} -q_{1;i} & q_{1;i} \\ q_{2;i} & -q_{2;i} \end{bmatrix}$$

with $q_{2;i} = \delta$. Markov theory requires that the infinitesimal generator is a matrix whose elements are not random variables. However, this is not the case in our simple approximation: $q_{1;i}(t) = \beta \sum_{k=1}^{N} a_{ij} 1_{\{X_k(t)=1\}}$. Using a mean-field approximation [10] so that $E[q_{1;i}] = \beta \sum_{j=1}^{N} a_{ij}\mathrm{Prob}[X_j(t) = 1]$, the effective infinitesimal generator becomes

$$Q_i(t) = \begin{bmatrix} -E[q_{1;i}] & E[q_{1;i}] \\ \delta & -\delta \end{bmatrix}.$$

Then, in accordance with Markov theory in [17, Eqs. (10) and (11), p. 182], denoting $v_i(t) = \mathrm{Prob}[X_i(t) = 1]$ and $\mathrm{Prob}[X_i(t) = 0] = 1 - v_i(t)$, the set of nodes obey the differential equations

$$\begin{aligned}
\frac{dv_1(t)}{dt} &= \beta\sum_{j=1}^{N} a_{1j}v_j(t) - v_1(t)\left(\beta\sum_{j=1}^{N} a_{1j}v_j(t) + \delta\right), \\
\frac{dv_2(t)}{dt} &= \beta\sum_{j=1}^{N} a_{2j}v_j(t) - v_2(t)\left(\beta\sum_{j=1}^{N} a_{2j}v_j(t) + \delta\right), \\
&\vdots \\
\frac{dv_N(t)}{dt} &= \beta\sum_{j=1}^{N} a_{Nj}v_j(t) - v_N(t)\left(\beta\sum_{j=1}^{N} a_{Nj}v_j(t) + \delta\right),
\end{aligned}$$

written in matrix form as

$$\frac{dV(t)}{dt} = \beta AV(t) - \mathrm{diag}(v_i t)[\beta AV(t) + \delta u], \tag{7}$$

where the vector $V(t) = [v_1(t)\ v_2(t) \cdots v_N(t)]^T$. The average number of infected nodes in G is equal to $y(t) = u^T V(t)$, where u is the all-1 vector.

For the N-intertwined approximation, the largest eigenvalue λ_1 of the graph's adjacency matrix rigorously defines the first-order epidemic threshold $\tau_c^{(1)} = \frac{1}{\lambda_1}$. A second-order epidemic threshold $\tau_c^{(2)} \geqslant \tau_c^{(1)}$ is studied in [13] which also presents a different derivation of the N-intertwined equations. The threshold arises as a consequence of the mean-field approximation. A major property, proved in [10] as well as in [13], of the N-intertwined approximation is that $V_i(t) \geqslant V_i(t)|_{\mathrm{exact}}$. Hence, the N-intertwined approximation upper-bounds the SIS epidemics and, consequently, $\tau_c^{(1)} < \tau_c$.

III. THE STEADY-STATE INFECTION IN THE MODEL AND TWO APPROXIMATIONS

A. The ε-SIS spreading model

In this paper, we use the ε-SIS model as a benchmark to compare both mean-field approximations. Whereas the classical SIS model has an absorbing state, the ε-SIS model does not for $\varepsilon > 0$. The nonzero steady state of the ε-SIS model is reached as time progresses. We believe that the steady-state fraction of infected nodes in the ε-SIS model is the simplest and best way to determine the number of infected nodes in the metastable state of the SIS model. The metastable state of the classical SIS model, although easily recognized, is difficult to define precisely. One approach would be to run many independent instances of the virus spreading process, calculate the average number of infected nodes at sampled points in time, and look for a plateau. This will, however, lead to too low an average number of infected nodes as a function of time, as for smaller values of the effective spreading rate, many instances of the virus spreading process die out very quickly. These died-out instances have a large impact on the average number of infected nodes as a function of time. Since instances of the virus that die out quickly do not reach a metastable state they have to be filtered out, but that would require an assessment of how long a "reasonable" outbreak lasts. Such a reasonable outbreak will be dependent on the effective spreading rate and on the network topology, which makes it infeasible as a simulation method.

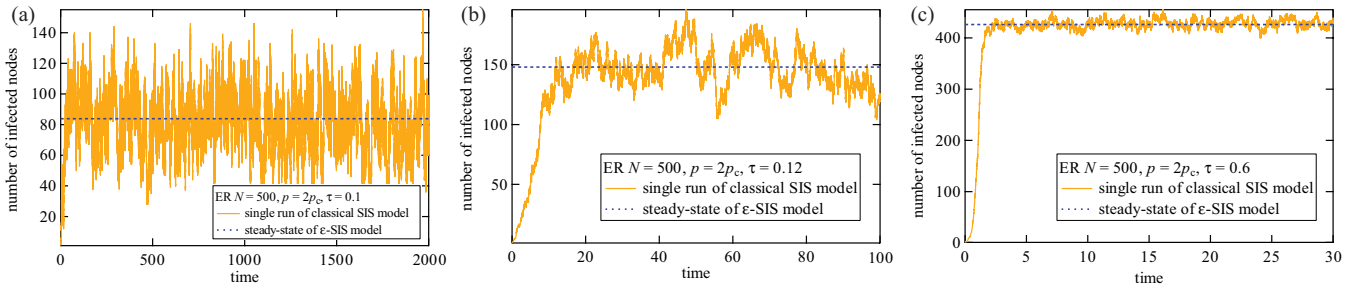

FIG. 1. (Color online) The steady state of the classical SIS model (solid yellow line) and the metastable state of the ε-SIS model (dashed blue line) in ER graphs ($\varepsilon = 10^{-3}$).

As the ε-SIS model has a well defined steady state, the steady-state number of infected nodes can be computed precisely. We start our simulations with no nodes infected and continue to run for a specified warm-up period. After the warm-up period, the measurement period starts during which we record the average number of infected nodes. For all simulations we have taken the warm-up and measurement period to be 10^7 time units and $\varepsilon = 10^{-3}$ time units. We have chosen the duration of 10^7 time units after careful experimentations. The accuracy of the ε-SIS simulations have been compared to the exact ε-SIS Markov chain (see [16]) for small ($N \leqslant 10$) networks, where more than three digits were accurate for all the considered τ ranges.

The steady-state number of infected nodes of the ε-SIS model will be close to the average number of infected nodes in the metastable state of the SIS model for small values of ε. In Fig. 1, we show a reasonable instance of a virus outbreak together with the steady-state number of infected nodes of the ε-SIS model. These examples illustrate that the steady-state average number of infected nodes of the ε-SIS model is precisely the line around which the number of infected nodes in the SIS model varies.

B. Pastor-Satorras and Vespignani HMF approximation

From (1) and (2), we obtain the set of nonlinear equations

$$\begin{aligned}
&\frac{\tau \sum_{k=1}^{N-1} k\text{Prob}[D=k]\rho_k}{E[D] + \tau \sum_{k=1}^{N-1} k\text{Prob}[D=k]\rho_k} - \rho_1 = 0,\\
&\frac{2\tau \sum_{k=1}^{N-1} k\text{Prob}[D=k]\rho_k}{E[D] + 2\tau \sum_{k=1}^{N-1} k\text{Prob}[D=k]\rho_k} - \rho_2 = 0,\\
&\qquad\qquad\vdots\\
&\frac{(N-1)\tau \sum_{k=1}^{N-1} k\text{Prob}[D=k]\rho_k}{E[D] + (N-1)\tau \sum_{k=1}^{N-1} k\text{Prob}[D=k]\rho_k} - \rho_{N-1} = 0. \qquad (8)
\end{aligned}$$

From the nonlinear set (8), the densities $\rho_1, \rho_2, \ldots, \rho_{N-1}$ can be calculated, and after using (6), we obtain the steady-state fraction $y_\infty(\tau)$ of infected nodes.

C. N-intertwined approximation

The steady-state of the N-intertwined approximation is obtained from (7), after letting $t \to \infty$ and $\lim_{t\to\infty} \frac{dv_j(t)}{dt} = 0$, as

$$\beta A V(t) - \text{diag}(v_i t)[\beta A V(t) + \delta u] = 0. \qquad (9)$$

Written as a nonlinear equation for a single node i, this leads to

$$v_{i\infty} = \frac{\beta \sum_{j=1}^{N} a_{ij} v_{j\infty}}{\beta \sum_{j=1}^{N} a_{ij} v_{j\infty} + \delta} = 1 - \frac{1}{1 + \tau \sum_{j=1}^{N} a_{ij} v_{j\infty}}. \qquad (10)$$

The steady-state fraction $y_\infty(\tau)$ of infected nodes can be calculated using (10).

For example, for the complete graphs K_N, when $t \to \infty$, $v_{i\infty} = y_\infty$, from which the fraction of infected nodes (10) reduces to

$$y_\infty = 1 - \frac{1}{1 + \tau(N-1)y_\infty}$$

or

$$y_\infty = 1 - \frac{1}{(N-1)\tau}, \qquad (11)$$

which is exactly the same as for the HMF approximation in (8) when $\rho_k = \rho_{N-1} = \rho = y_\infty$, as also illustrated in Fig. 5.

D. Asymptotics for large τ

We present the exact steady-state asymptotics of the epidemic for large τ. If τ is sufficiently large, the infection state $v_{j\infty} = \lim_{t\to\infty} \text{Prob}[X_j(t) = 1]$ of a node j with d_j neighbors tends to be independent of the viral state of its d_j neighbors, because the neighbors are with overwhelming probability infected. Hence, the nodal viral state of node j is not intertwined anymore with that of its neigbors, but independent, and is exceedingly well described by a two-state continuous Markov process with infection rate β and curing rate δ, where $v_{j\infty} = \frac{\beta d_j}{\delta + \beta d_j} = \frac{1}{1+1/\tau d_j} = \frac{1}{1+s/d_j}$ with $s = \frac{1}{\tau}$. The derivative for large τ or, equivalently, $s \to 0$, is

$$\left.\frac{dv_{j\infty}(s)}{ds}\right|_{s=0} = -\frac{1}{d_j}.$$

The average steady-state fraction of infected nodes is thus $y_\infty(s) = \frac{1}{N}\sum_{j=1}^{N} v_{j\infty}(s)$ and has a derivative at $s = 0$ equal to

$$\left.\frac{dy_\infty}{ds}\right|_{s=0} = -\frac{1}{N}\sum_{j=1}^{N}\frac{1}{d_j} = -E\left[\frac{1}{D}\right]$$

which is precisely equal to that computed in [20] for the N-intertwined mean-field approximation.

For the HMF approximation, we obtain, by substituting (1) into (6) and using the transform $s = \frac{1}{\tau}$,

$$y_{\infty;\mathrm{HMF}}(s) = \sum_{k=1}^{N-1} \frac{\mathrm{Prob}[D=k]}{1+\frac{s}{k\Theta(s^{-1})}},$$

from which, using $\lim_{s\to 0}\Theta(s^{-1}) = 1$,

$$\begin{aligned}\left.\frac{dy_{\infty;\mathrm{HMF}}(s)}{ds}\right|_{s=0} &= -\lim_{s\to 0}\sum_{k=1}^{N-1}\frac{\mathrm{Prob}[D=k]}{\left(1+\frac{s}{k\Theta(s^{-1})}\right)^2}\frac{1}{k} \\ &\quad\times\left(\frac{1}{\Theta(s^{-1})}-\frac{s}{\Theta^2(s^{-1})}\frac{d\Theta(s^{-1})}{ds}\right) \\ &= -\sum_{k=1}^{N-1}\frac{\mathrm{Prob}[D=k]}{k} = -E\left[\frac{1}{D}\right]\end{aligned}$$

because $\frac{d\Theta(s^{-1})}{ds}|_{s=0}$ is finite. Indeed, taking the derivative of the self-consistent relation (4),

$$E[D] = \sum_{k=1}^{N-1}\frac{k^2\mathrm{Prob}[D=k]}{s+k\Theta(s^{-1})},$$

yields

$$0 = \sum_{k=1}^{N-1}\frac{k^2\mathrm{Prob}[D=k]}{[s+k\Theta(s^{-1})]^2}\left(1-k\frac{d\Theta(s^{-1})}{ds}\right)$$

or

$$\frac{d\Theta(s^{-1})}{ds} = \frac{\sum_{k=1}^{N-1}\frac{k^2\mathrm{Prob}[D=k]}{[s+k\Theta(s^{-1})]^2}}{\sum_{k=1}^{N-1}\frac{k^3\mathrm{Prob}[D=k]}{[s+k\Theta(s^{-1})]^2}},$$

from which $\frac{d\Theta(s^{-1})}{ds}|_{s=0} = \frac{1}{\sum_{k=1}^{N-1}k\mathrm{Prob}[D=k]} = \frac{1}{E[D]}$.

Hence, both mean-field approximations return both $\lim_{s\to 0} y_\infty(s)$ and the derivative $\frac{dy_\infty(s)}{ds}|_{s=0}$ correctly in the large-τ regime.

IV. COMPARISON OF THE STEADY-STATE FRACTION $y_\infty(\tau)$ OF INFECTED NODES VERSUS τ

This section compares the ε-SIS model and the two approximations for different graph types. We take the following topologies into account: the bipartite graph, the star graph, the complete graph, the lattice graph, the path graph, the Erdős-Rényi random graph, the Bárabasi-Albert scale-free graph, and the small-world graph. The steady-state fraction $y_\infty(\tau)$ of infected nodes is calculated for increasing effective spreading rates τ and $\varepsilon = 10^{-3}$. The values of the N-intertwined approximation, the HMF approximation, and the simulations of the ε-SIS spreading model are shown in blue, red, and green lines, respectively. The different markers indicate the sizes of the graphs, e.g., circles in Fig. 2 indicate the results for graphs with $N = 10$ nodes.

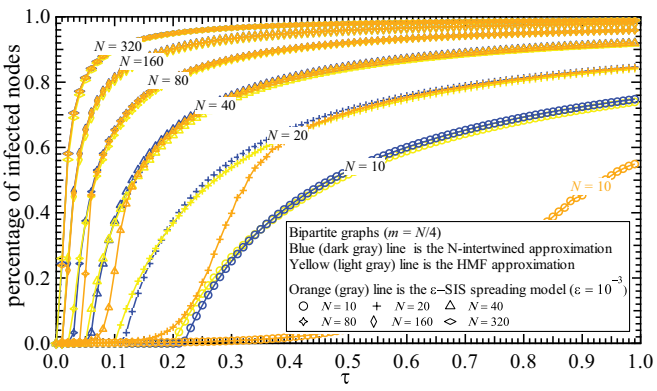

FIG. 2. (Color online) Comparison in bipartite networks.

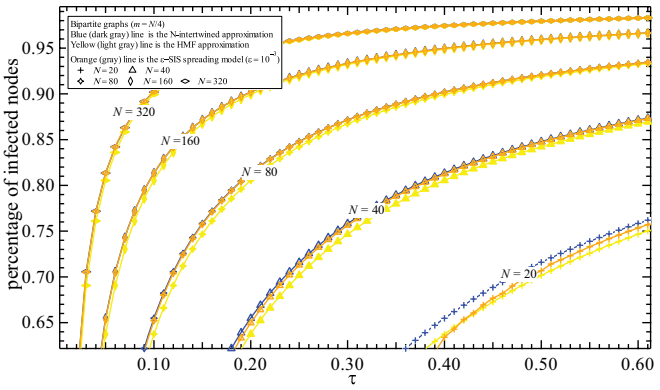

FIG. 3. (Color online) Zoom of the comparison in the bipartite graphs.

A. Complete bipartite graphs

A complete bipartite graph K_{M_1,M_2} consists of two disjoint sets S_1 and S_2 containing respectively M_1 and M_2 nodes. All nodes in S_1 are connected to all nodes in S_2, while nodes within a set do not connect. In this paper, we take $M_1 = N/4$ nodes, and $M_2 = 3N/4$ nodes. The steady-state fraction $y_\infty(\tau)$ of infected nodes as a function of τ are computed in bipartite graphs with $N = 10$, 20, 40, 80, 160, and 320 nodes. Figure 2 shows that the epidemic thresholds for the HMF approximation and the N-intertwined approximation are close to that of the ε-SIS spreading model ($\varepsilon = 10^{-3}$) in complete bipartite graphs. Since $\tau_c^{(1)}$ of the N-intertwined approximation is nearer to τ_c than τ_c^{HMF} of the HMF approximation, $\tau_c^{(1)}$ provides the better epidemic prediction for the SIS model in the complete bipartite graph K_{M_1,M_2}. Moreover, in [13] it is proved that $\tau_c \geqslant \tau_c^{(2)} \geqslant \tau_c^{(1)}$, which means that the second-order N-intertwined

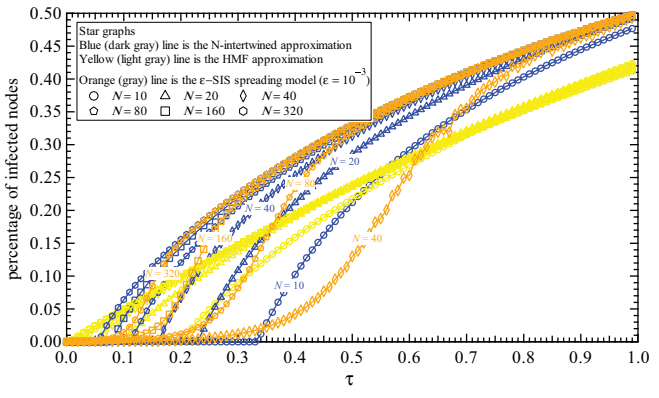

FIG. 4. (Color online) Comparison in star graphs.

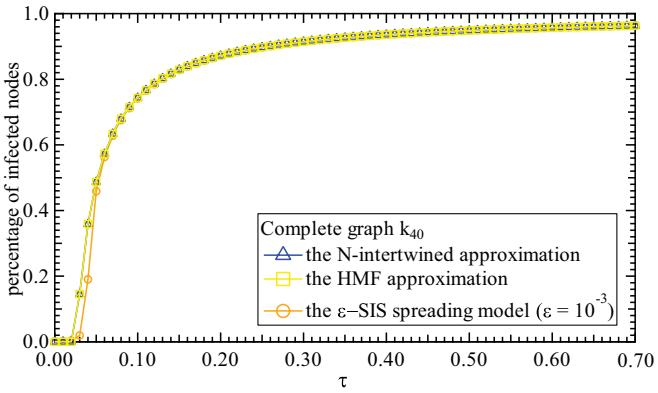

FIG. 5. (Color online) Comparison in complete graphs.

approximation is closer to the ε-SIS spreading model, and therefore the better approximation in bipartite graphs.

Three interesting results can be observed by zooming in on Fig. 2 as shown in Fig. 3. First, the N-intertwined approximation is an upper bound of the ε-SIS spreading model. Second, the difference between the N-intertwined approximation and the ε-SIS spreading model decreases with N. We observe that the N-intertwined approximation almost overlays the ε-SIS spreading model, when $N = 320$. Third, the HMF approximation is lower than the ε-SIS spreading model, showing that the HMF approximation is not upper-bounding the SIS model.

B. Star graphs

The star graph $K_{1,N-1}$ is a special bipartite graph where one of the disjoint sets contains only one node while the other set contains the rest of the nodes. The epidemic threshold for the first-order N-intertwined approximation equals $\tau_c^{(1)} = \frac{1}{\lambda_1}$. For any connected graph, the spectral radius is bounded [21] from above by $\lambda_1 \leqslant \sqrt{2L - N + 1}$, and equality is reached for the complete graph K_N and the star $K_{1,N-1}$. As a star graph contains $L = N - 1$ links, we obtain

$$\tau_c^{(1)} = \frac{1}{\sqrt{2L - N + 1}} = \frac{1}{\sqrt{N - 1}}. \tag{12}$$

The second-order mean-field threshold for the star was estimated in [13] to be $\tau_c^{(2)} = \frac{1}{\sqrt{0.53N - 1.3}}$, while exact computations indicate that $\tau_c = \frac{1}{\sqrt{N}}\sqrt{\frac{1}{2}\ln N + \ln\ln N + O(1)}$ for large N.

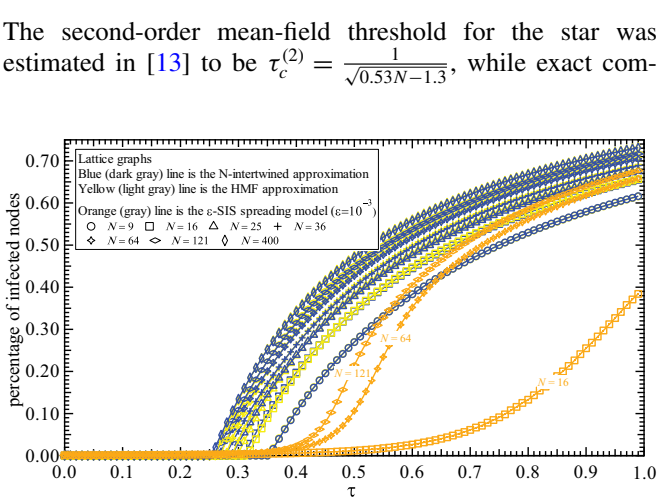

FIG. 6. (Color online) Comparison in lattice graphs.

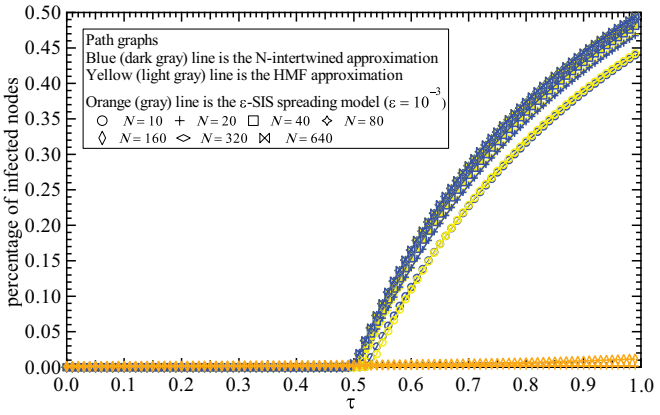

FIG. 7. (Color online) Comparison in path graphs.

Recall that the epidemic threshold of the HMF approximation is given by $\tau_c^{\mathrm{HMF}} = \frac{E[D]}{E[D^2]}$. For star graphs it holds that $E[D^2] = \frac{N^2 - N}{N}$ and $E[D] = \frac{2(N-1)}{N}$, so the HMF threshold reduces to

$$\tau_c^{\mathrm{HMF}} = \frac{2}{N}. \tag{13}$$

The equalities (12) and (13) indicate that, for $N > 2$, the epidemic threshold of the N-intertwined approximation is always larger than that of the HMF approximation in star graphs. Figure 4 shows the superiority of the N-intertwined approximation, especially when N is large. Nevertheless, the two epidemic thresholds are both quite far from the threshold of the ε-SIS spreading model ($\varepsilon = 10^{-3}$) in star graphs.

C. Complete graphs

The complete graph K_N is a graph in which every node pair is connected. For a complete graph $\tau_c^{\mathrm{HMF}} = \frac{E[D]}{E[D^2]} = \frac{N-1}{N(N-1)^2/N} = \frac{1}{N-1}$; at the same time $\lambda_1 = N - 1$. Hence, the epidemic threshold of the N-intertwined approximation $\tau_c^{(1)} = \frac{1}{\lambda_1}$ is equal to the threshold of the HMF approximation $\tau_c^{\mathrm{HMF}} = \frac{E[D]}{E[D^2]}$. For K_N, both approximations are very close

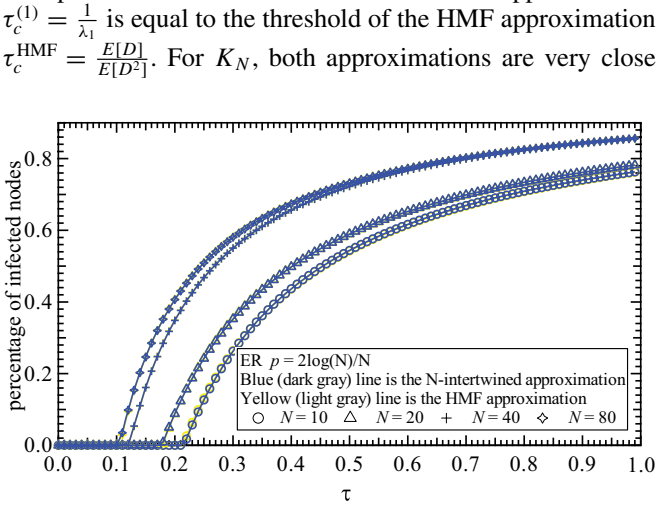

FIG. 8. (Color online) Comparison in Erdos-Renyi random graphs.

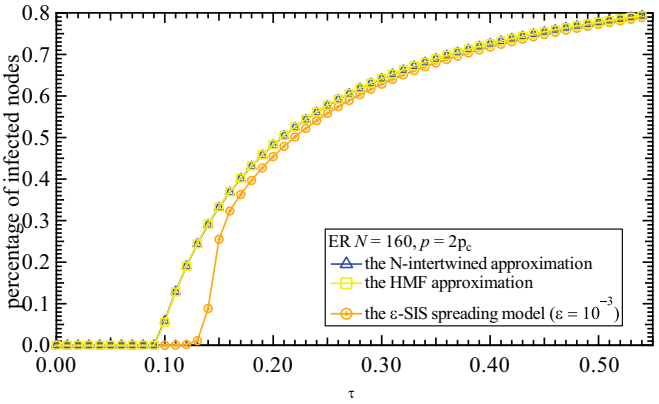

FIG. 9. (Color online) Comparation among the N-intertwined approximation, the Pastor-Satorras approximation, and the ε-SIS model in an ER network ($N = 160$).

to the ε-SIS spreading model ($\varepsilon = 10^{-3}$) (see Fig. 5). This is to be expected, since the mean-field approximation in the N-intertwined approximation is best for dense graphs, as explained in [10]. Moreover, for K_N, the steady-state equations (see Secs. III C and III B) in the N-intertwined and HMF approximations are the same. The steady-state fraction $y_\infty(\tau)$ of infected nodes as a function of τ has been deduced in (11).

D. Square lattice graphs

The square lattice graph is a two-dimensional grid. Ignoring the boundary nodes, the square lattice can be regarded as a regular graph, where all nodes have the same degree ($k_i = 4$). In this case, the equations of the N-intertwined approximation and the HMF approximation are almost the same, as verified from the simulations of the two approximations. Our simulations (see Fig. 6) show that the epidemic threshold of the ε-SIS spreading model ($\varepsilon = 10^{-3}$) decreases with the size N of the network. The HMF approximation performs slightly better than the N-intertwined approximation in approaching the ε-SIS spreading model in lattice graphs. The simulation illustrates that neither the N-intertwined approximation nor the HMF approximation predicts the epidemic threshold for epidemic processes in lattices. We remark that, in the related process of percolation, the critical probability [22–24] on the square lattice is equal to $1/2$.

E. Path graphs

The path graph is a example of a tree graph, in which every root node has only one branch and only the last root node is not branched at all. As shown in Fig. 7, the steady-state fractions $y_\infty(\tau)$ of infected nodes of the N-intertwined approximation and the HMF approximation are far from that of the ε-SIS spreading model ($\varepsilon = 10^{-3}$). The epidemic thresholds of the N-intertwined approximation and the HMF approximation are both near 0.5, since the average degree of the path graph is 2, ignoring boundary nodes. However, the steady-state fraction $y_\infty(\tau)$ of infected nodes of the ε-SIS spreading model increases very slowly with τ between $0 \leqslant \tau \leqslant 1$, and seems to always be around 0 in the range of network sizes that we considered.

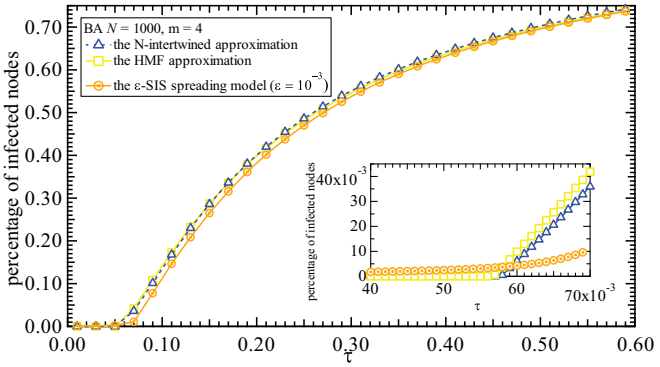

FIG. 10. (Color online) Comparison in Barabasi-Albert scale-free networks.

F. Erdős-Rényi random graphs

In this section we investigate the thresholds in Erdős-Rényi random graphs,[1] which have a binomial degree distribution [25]. An Erdős-Rényi random graph is connected with high probability, if $p > p_c \approx \frac{\ln N}{N}$ for large N, where p_c is the disconnectivity threshold. All the graphs in the simulations are generated with $p = 2p_c$, and checked for connectivity. Figure 8 shows that the steady-state fractions $y_\infty(\tau)$ of infected nodes of the N-intertwined approximation and the HMF approximation for ER graphs $N = 10$, 20, 40, and 80, are extremely close. However, they both differ from the epidemic threshold of the ε-SIS spreading model, especially when N is small. When N is large, the two approximations are close to the ε-SIS spreading model ($\varepsilon = 10^{-3}$) (see Fig. 9).

G. Bárabasi-Albert scale-free graphs

The Bárabasi-Albert (BA) graph[2] [26] is a characteristic model for complex networks because of its power-law degree distribution. Power-law degree distributions are widely, although approximately, observed in real-world complex networks. The steady-state fraction of infected nodes as a function of the effective spreading rate $y_\infty(\tau)$ is computed in a BA graph with $N = 1000$ and $m = 4$ and shown in Fig. 10. The N-intertwined approximation is close to the HMF approximation, but a little superior. This is to be expected, since the N-intertwined approximation is better than the HMF approximation in star graphs as explained in Sec. IV B, and the BA model can be regarded as a set of hubs with star graph features.

[1] An Erdős-Rényi random graph can be generated from a set of N nodes by randomly assigning a link with probability p to each pair of nodes.

[2] A Bárabasi-Albert graph starts with m nodes. At every time step, we add a new node with m links that connect the new node to m different nodes already present in the graph. The probability that a new node will be connected to node i in step t is proportional to the degree $d_i(t)$ of that node. This is referred to as preferential attachment.

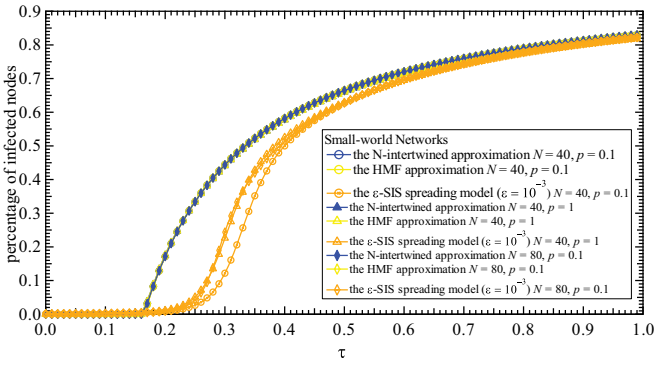

FIG. 11. (Color online) Comparison in WS small-world networks.

H. Watts-Strogatz small-world graphs

Watts-Strogatz (WS) small-world graphs[3] [27] have two main properties: a small average hop count $E[H]$, similar to Erdős-Rényi random graphs, and a high clustering coefficient C_G, similar to a ring lattice. The structural properties of small-world graphs have been found in various real-world networks, including social networks [28], neural networks [29], and biological oscillators [30]. In this paper, the WS graphs are generated with $N = 40$ and 80, $k_s = 6$, and $p = 0.1$ and 1. In Fig. 11 the steady-state fractions $y_\infty(\tau)$ of infected nodes, as predicted by the two approximations, are shown together with the ε-SIS simulations. The N-intertwined approximation and the HMF approximation are quite close to each other, but far away from the ε-SIS spreading model. The $\tau_c^{(1)} = \frac{1}{\lambda_1}$ and the $\tau_c^{\mathrm{HMF}} = \frac{E[D]}{E[D^2]}$ in small-world graphs are near to each other no matter what N and p are. This can be explained by observing that most nodes have the same degree in WS graphs, justifying the approximation of $E[D^2]$ by $E[D]^2$ and $\tau_c^{\mathrm{HMF}} = \frac{E[D]}{E[D^2]}$ by $\frac{1}{E[D]}$. Another consequence of the similar node degrees in WS graphs is that $E[D]$ is close to $D_{\max}$. Since λ_1 is bounded from below and above as $E[D] \leqslant \lambda_1 \leqslant D_{\max}$ ([21], article 43, p. 46 and article 48, p. 52), we can approximate λ_1 by $E[D]$ and $\tau_c^{(1)}$ by $\frac{1}{E[D]}$, just like τ_c^{HMF}.

V. ANALYTIC COMPARISON OF THE EPIDEMIC THRESHOLDS $\tau_c^{(1)}$ AND τ_c^{HMF}

In this section, we analyze the relation between the first-order epidemic threshold of N-intertwined approximation, $\tau_c^{(1)} = \frac{1}{\lambda_1}$, and the epidemic threshold of the HMF approximation, $\tau_c^{\mathrm{HMF}} = \frac{E[D]}{E[D^2]}$. From the comparison in Sec. IV, we find that the relation between the two epidemic thresholds strongly depends on the graph type. The two epidemic thresholds are equal to each other in regular graphs where each node has degree r increasing with N. Indeed, since $\lambda_1 = E[D] = r$ (see [21], article 43, p. 46]), and $\tau_c^{\mathrm{HMF}} = \frac{1}{r}$, we find that $\tau_c^{(1)} = \tau_c^{\mathrm{HMF}}$. There are graphs for which $\tau_c^{(1)} < \tau_c^{\mathrm{HMF}}$, while in most cases, our simulations in Figs. 2, 4, 8, and 10 demonstrate that $\tau_c^{(1)} > \tau_c^{\mathrm{HMF}}$.

Cases $\tau_c^{(1)} < \tau_c^{\mathrm{HMF}}$. The epidemic threshold τ_c^{HMF} is larger than the first-order threshold $\tau_c^{(1)} = \frac{1}{\lambda_1}$, when the assortativity[4] ρ_D is zero. Van Mieghem *et al.* [21,32] have reformulated the assortativity as follows:

$$\rho_D = \frac{N_1 N_3 - N_2^2}{N_1 \sum_{i=1}^{N} d_i^3 - N_2^2}, \tag{14}$$

where $N_k = u^T A^k u$ is the total number of walks with k hops. In [33], we have proved that $\lambda_1 \geqslant \frac{N_2}{N_1} = \frac{E[D^2]}{E[D]} = \frac{1}{\tau_c^{\mathrm{HMF}}}$, when $\rho_D = 0$.

Cases $\tau_c^{(1)} > \tau_c^{\mathrm{HMF}}$. Newman [31] pointed out that the assortativity ρ_D of the ER graph and the BA graph is zero when N is large. However, in most ER and BA graphs with finite size, the assortativity is only *approximately* zero. Our simulations in Figs. 8 and 10 show that $\tau_c^{\mathrm{HMF}} \leqslant \tau_c^{(1)}$ in ER and BA graphs, demonstrating that the precise $\rho_D = 0$ condition in (14) that led to $N_1 N_3 = N_2^2$ is not valid. Moreover, we have already proved that $\tau_c^{\mathrm{HMF}} \leqslant \tau_c^{(1)}$ in star graphs (see Sec. IV B).

It would be interesting to find all or the most prominent graph classes in which $\tau_c^{(1)} > \tau_c^{\mathrm{HMF}}$ and in which $\tau_c^{(1)} < \tau_c^{\mathrm{HMF}}$.

VI. CONCLUSION

Many approximations of the SIS model have been proposed to understand SIS epidemics. In this paper, we studied which mean-field approximation, the N-intertwined or the HMF, is better in approaching the SIS epidemic model. A direct comparison to the SIS model is, however, not possible, because the steady state of the exact SIS model in a finite network is the overall-healthy state. Although an infection in the SIS model will eventually die out, for high enough effective spreading rates the fraction of infected nodes as a function of time is metastable. We proposed to define the number of infected nodes in the metastable state of the SIS model via the number of infected nodes in the steady state of the ε-SIS model for a prescribed small value of ε. From the comparison between the N-intertwined and HMF approximations with the ε-SIS spreading model, we conclude that, overall, the N-intertwined approximation is better than the HMF approximation, except for square lattice graphs and path graphs. We have seen that the N-intertwined approximation can approach the ε-SIS epidemic model well in most graph types. The simulations show that the N-intertwined approximation almost overlaps with the ε-SIS spreading model, when the size of the network is large enough. While the HMF approximation is better than the N-intertwined approximation in the square lattice and path graphs, the difference between the two is small. Moreover, they are both far away from the ε-SIS spreading model. We also showed that

[3] A Watts-Strogatz small-world graph can be generated from a ring lattice with N nodes and k_s edges per node, by rewiring each link at random with probability p.

[4] The degree correlation, also called the assortativity ρ_D, is computed as the linear correlation coefficient of the degree of nodes connected by a link [31]. It describes the tendency of network nodes to connect preferentially to other nodes with either similar (when $\rho_D > 0$) or opposite (when $\rho_D < 0$) properties, i.e., degree.

the *N*-intertwined approximation and the HMF approximation are exactly the same in regular graphs with the degree of nodes increasing with N, such as complete graphs, and are similar in small-world graphs. In addition to our simulation results, we showed analytically the conditions under which the epidemic threshold of the *N*-intertwined approximation is larger than, smaller than, or equal to that of the HMF approximation.

[1] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1991).

[2] N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications*, 2nd ed. (Charlin Griffin & Company, London, 1975).

[3] A. Barrat, M. Bartelemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008).

[4] C. Castellano and R. Pastor-Satorras, Phys. Rev. Lett. **105**, 218701 (2010).

[5] D. J. Daley and J. Gani, *Epidemic Modeling: An Introduction* (Cambridge University Press, Cambridge, 1999).

[6] J. O. Kephart and S. R. White, in *Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy* (IEEE Computer Society Press, Oakland, California, 1991), pp. 343–359.

[7] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **63**, 066117 (2001).

[8] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).

[9] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, *IEEE Symposium on Reliable Distributed Systems* (IEEE Computer Society, Los Alamitos, CA, 2003).

[10] P. Van Mieghem, J. S. Omic, and R. E. Kooij, IEEE/ACM Trans. Netw. **17**, 1 (2009).

[11] P. Van Mieghem, Europhys. Lett. **97**, 48004 (2012).

[12] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, Europhys. Lett. **89**, 38009 (2010).

[13] E. Cator and P. Van Mieghem, Phys. Rev. E **85**, 056111 (2012).

[14] A. L. Hill, D. G. Rand, M. A. Nowak, and N. A. Christakis, Proc. R. Soc. B **277**, 3827 (2010).

[15] H. Hinrichsen, Adv. Phys. **49**, 815 (2000).

[16] P. Van Mieghem and E. Cator, Phys. Rev. E **86**, 016116 (2012).

[17] P. Van Mieghem, *Performance Analysis of Communications Systems and Networks* (Cambridge University Press, Cambridge, 2006).

[18] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **66**, 047104 (2002).

[19] P. Van Mieghem, Computing **93**, 147 (2011).

[20] P. Van Mieghem, Comput. Commun. **35**, 1494 (2012).

[21] P. Van Mieghem, *Graph Spectra for Complex Networks* (Cambridge University Press, Cambridge, 2011).

[22] H. Kesten, Commun. Math. Phys. **74**, 41 (1980).

[23] R. T. Smythe and J. C. Wierman, *First-Passage Percolation on the Square Lattice*, Lecture Notes in Mathematics, Vol. 671 (Springer, Berlin, 1978).

[24] D. J. A. Welsh, Sci. Prog. (Oxford) **64**, 65 (1977).

[25] P. Erdős and A. Rényi, Publ. Math. (Debrecen) **6**, 290 (1959).

[26] A.-L. Barabasi and R. Albert, Science **286**, 509 (1999).

[27] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).

[28] M. E. J. Newman and D. J. Watts, Phys. Lett. A **263**, 341 (1999).

[29] C. J. Stam, B. F. Jones, G. Nolte, M. Breakspear, and Ph. Scheltens, Cereb. Cortex **17**, 92 (2007).

[30] Y. Kuramoto, *Chemical Oscillations, Waves and Turbulence* (Springer, Berlin, 1984).

[31] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).

[32] P. Van Mieghem, H. Wang, X. Ge, S. Tang, and F. A. Kuipers, Eur. Phys. J. B **76**, 643 (2010).

[33] C. Li, H. Wang, W. de Haan, C. J. Stam, and P. Van Mieghem, J. Stat. Mech.: Theory Exp. (2011) P11018.