Ranking of Nodal Infection Probability in Susceptible-Infected-Susceptible Epidemic

Bo Qu[1], **Cong Li**[2,*], **Piet Van Mieghem**[1], **and Huijuan Wang**[1]

[1]Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, 2628CD, The Netherlands
[2]Adaptive Networks and Control Lab, Department of Electronic Engineering; and Research Center of Smart Networks and Systems, School of Information Science and Engineering, Fudan University, Shanghai 200433, China
*cong_li@fudan.edu.cn

ABSTRACT

The prevalence, which is the average fraction of infected nodes, has been studied to evaluate the robustness of a network subject to the spread of epidemics. We explore the vulnerability (infection probability) of each node in the metastable state with a given effective infection rate τ. Specifically, we investigate the ranking of the nodal vulnerability subject to a susceptible-infected-susceptible epidemic, motivated by the fact that the ranking can be crucial for a network operator to assess which nodes are more vulnerable. Via both theoretical and numerical approaches, we unveil that the ranking of nodal vulnerability tends to change more significantly as τ varies when τ is smaller or in Barabási-Albert than Erdős-Rényi random graphs.

Introduction

The continuous outbreaks of epidemic diseases in a population and viruses in computer networks[1–4] motivate the study of epidemic spreading on a network. The Susceptible-Infected-Susceptible (SIS) epidemic process[5–12] has been widely studied as a model of virus spread on a network. In the SIS model, a node is either infected or susceptible at any time t. Each infected node infects each of its susceptible neighbors with an infection rate β. Each infected node recovers with a recovery rate δ. Both infection and recovery processes are independent Poisson processes and the ratio $\tau = \beta/\delta$ is the effective infection rate. There is an epidemic threshold τ_c and above the threshold $\tau > \tau_c$ a nonzero fraction of nodes is infected in the metastable state. The infection probability $v_{k\infty}(\tau)$ of a node k in the metastable state at a given effective infection rate τ indicates the vulnerability of node k to the virus, and the prevalence, which equals the average fraction $y_\infty(\tau)$ of infected nodes reflects the global vulnerability of the network.

Researchers have mainly concentrated on the average fraction y_∞ of infected nodes in the metastable state to estimate the vulnerability of a network against a certain epidemic or virus. Great effort has been devoted to understand how the network topology influences the vulnerability and the epidemic threshold[6,13–15]. When the effective infection rate is just above the epidemic threshold [16, p. 469]. In this case, it is found [16, p. 469] that, the metastable-state infection probability vector $V_\infty = [v_{1\infty} v_{2\infty} \cdot \cdot \cdot v_{N\infty}]^T$), obtained by the N-Intertwined Mean-Field Approximation (NIMFA) of SIS model is proportional to the principal eigenvector x_1 of the adjacency matrix A. In this work, we aim to explore the nodal infection probability in a systematic way, in different network topologies and when the effective infection rate τ varies. As a starting point, we investigate the ranking of nodal infection probabilities, which crucially informs a network operator which nodes are more vulnerable or require protection. Interestingly, we find that the ranking of the nodal infection probability changes as the effective infection rate τ varies. The observation points out that we cannot find a topological feature of a node i to represent the vulnerability of node i to an SIS epidemic, because the rankings in vulnerability of nodes in a network may be different when the effective infection rate τ varies, whereas a topological feature of node i remains the same. Our observation explains the finding of Hebert-Dufresne et al.[17] that different nodal features (such as degree, betweenness, etc.) should be used to select the nodes to immunize in different scenarios (based on different infection rates, link densities, etc.), i.e. different nodes should be immunized at different infection rates. In this paper, we explore two questions: (I) which network topology changes the ranking of nodal infection probabilities more significantly and (II) in which effective infection rate range, does the increment of the effective infection rate lead to a more significant change in the ranking for a given network topology?

We first assume that, for an arbitrary pair of nodes, the trajectory $v_{k\infty}(\tau)$ and $v_{m\infty}(\tau)$ as a function of the effective infection rate τ cross at most once in any interval (τ_0, τ_1). We call this assumption the one-crossing assumption and Section "Discussion about the one-crossing assumption" of the supplementary information shows that the assumption is reasonably good. Then

the rankings of the vulnerabilities $v_{k\infty}(\tau)$ and $v_{m\infty}(\tau)$ change or equivalently the trajectories $v_{k\infty}(\tau)$ and $v_{m\infty}(\tau)$ cross if $(v_{k\infty}(\tau_0)-v_{m\infty}(\tau_0))(v_{k\infty}(\tau_1)-v_{m\infty}(\tau_1))<0$, when the effective infection rate τ changes from τ_0 to τ_1. To estimate the maximal change in the ranking of nodal infection probabilities in a network, we consider the total number of crossings between the trajectories of the infection probabilities of all the nodes in a network, when the effective infection rate τ changes from just above the epidemic threshold to a large value, above which the ranking remains the same. The total number of crossings is a simple and straightforward measure of the changes in the ranking of nodal infection probabilities. (We also briefly discuss how the correlation of the ranking of nodal infection probabilities changes as the effective infection rate increases in Section "The Spearman rank correlation ρ as a function of α" of the supplementary information.) A higher total number of crossings may lead to a more complicated protection policy for a network operator. Given a network, we find a lower bound of the total number of crossings, which can be computed from the topology properties, in particular, from the degree vector and the principal eigenvector of the adjacency matrix. The lower bound is roughly proportional to, thus an accurate indicator of, the total number of crossings for an arbitrary network. Hence, these two topological features (i.e. the degree vector and the principal eigenvector of the adjacency matrix) could indeed characterize to what extent the ranking of nodal vulnerabilities would change on a network. Since the lower bound is computationally simple, it can be used to compare the total number of crossings for different network topologies. This result explains why the total number of crossings tends to be larger in networks with a smaller average degree if the degree distribution is given or with a larger degree variance if the average degree is given. Regarding to question (II), we analytically derive the number of crossings when the effective infection rate τ_0 increases with a small value ε, given the infection probability vector $V_\infty(\tau_0)$ at the effective infection rate τ_0. This theoretical result, validated by numerical results, explains the reason why the crossings are more likely to occur when the effective infection rate τ is smaller.

Results

We first introduce in detail how to count or quantify the changes of the nodal ranking of the infection probability. Afterwards, we investigate the changes in the ranking (I) in different topologies when the effective infection rate τ increases from just above the epidemic threshold to a large value, above which the ranking remains the same, and (II) when the effective infection rate varies in different ranges.

The counting of the nodal ranking changes

To explore the changes of the nodal ranking of the infection probability, we investigate in a given network how many crossings, denoted by χ, between the trajectory $v_{k\infty}(\tau)$ and $v_{m\infty}(\tau)$ for all pairs of nodes can occur in the effective infection rate interval (τ_0,τ_1), where $\tau_0>\tau_c^{(1)}$, and where $\tau_c^{(1)}=\frac{1}{\lambda_1}$ (λ_1 is the largest eigenvalue of the adjacency matrix) is the NIMFA epidemic threshold: the epidemic dies out if the effective infection rate $\tau<\tau_c^{(1)}$. (More details on $\tau_c^{(1)}$ are introduced in Section Methods.) In Fig. 1, we illustrate the trajectories $v_{k\infty}(\tau)$ of 10 nodes, randomly selected from a real-world network called Roget ($N=994$ nodes, average degree $E[D]=7.32$ and detailed in Section "Real-world graphs" of the supplementary information). For example, the vulnerability of the node corresponding to the red dash line in Fig. 1 changes dramatically from the medium vulnerability when $\tau=0.12$ to the high vulnerability when $\tau=0.24$. Network operators should be alert to such a change of nodal vulnerabilities. The trajectories $v_{k\infty}(\tau)$ of other groups of nodes in Roget are shown and discussed in the first section of the supplementary information.

For a graph with N nodes, the maximum possible number of crossings is $\frac{N(N-1)}{2}$ under the one-crossing assumption. To count the number of crossings in the interval (τ_0,τ_1), we define an $N\times N$ matrix F with elements f_{ij}:

$$f_{ij}(V_\infty(\tau_0),V_\infty(\tau_1))=(v_{i\infty}(\tau_0)-v_{j\infty}(\tau_0))(v_{i\infty}(\tau_1)-v_{j\infty}(\tau_1))$$

Since $f_{ii}=0$, the matrix F has a zero diagonal just as the adjacency matrix A. A negative matrix element $f_{ij}<0$ means that there is a crossing between the trajectory $v_{i\infty}(\tau)$ and $v_{j\infty}(\tau)$ in the interval (τ_0,τ_1). The number of crossings in the interval (τ_0,τ_1) of the effective infection rate then equals

$$\chi(\tau_0,\tau_1)=\sum_{i=1}^{N}\sum_{j=1}^{i-1}1_{f_{ij}(V_\infty(\tau_0),V_\infty(\tau_1))<0} \tag{1}$$

where $1_{\{x\}}$ is the indicator function: $1_{\{x\}}=1$ if the event or condition $\{x\}$ is true, else $1_{\{x\}}=0$. Specifically, if all nodal degrees are the same in a random graph, the nodal ranking in any interval of τ does not change, since the infection probability of every node[6] equals the average fraction of infected nodes for any effective infection rate τ. In this work, we focus on the NIMFA nodal infection probability in the meta-stable state which is obtained by solving (11), hence the initial conditions (such as how many nodes are initially infected) are not necessary.

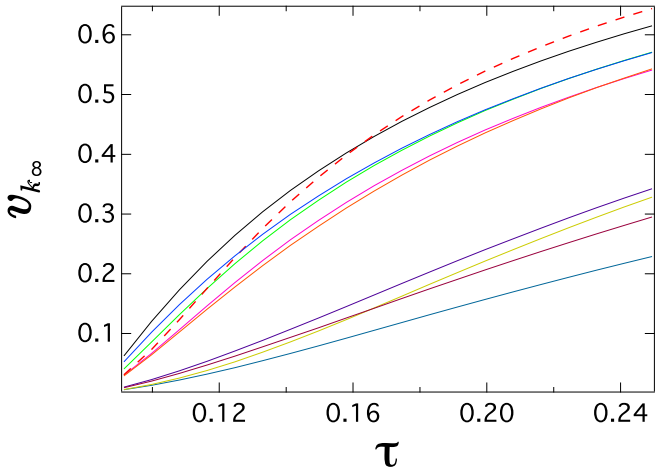


Figure 1. The meta-stable infection probability $v_{k\infty}$ as a function of the effective infection rate τ for 10 random nodes in a real-world network called Roget (details in Section "Real-world graph" of the supplementary information). The meta-stable infection probability $v_{k\infty}$ is obtained by solving (11).

We can compute the SIS metastable viral infection probability $v_{k\infty}$ of any node k both by the N-Intertwined Mean-Field Approximation (NIMFA)[6,18] and by simulations[8] of the SIS continuous-time Markov process. We then further compare the number of crossings χ as a function of the increment in the effective infection rate τ over different ranges, obtained by NIMFA and the continuous-time simulations of the SIS model. As shown in Section "The comparison between NIFMA and the continuous-time simulation" of the supplementary information, the number of crossings obtained from NIMFA is relatively close to that from the simulations, so we compute the number χ of crossings mainly by NIMFA due to its computational efficiency. However, NIMFA may not be accurate when the effective infection rate is close to the epidemic threshold[8]. Hence, the number of crossings obtained by NIMFA and simulations may be different from each other when the effective infection rate is close to the epidemic threshold as shown in Section "The comparison between NIFMA and the continuous-time simulation" of the supplementary information.

The total number of crossings in different topologies

We explore the total number of crossings in different graph topologies $\chi(\tau_c^{(1)}+\varepsilon,\tau_u)$ when the effective infection rate τ changes from just above the epidemic threshold, i.e. $\tau_c^{(1)}+\varepsilon$, to a large value τ_u, above which the ranking of the nodal infection probability hardly changes. In Section "Methods – The derivation of the lower bound χ_l" , we prove that there exists a value of τ, above which the ranking of the nodal infection probabilities does not change. We derive a lower bound of the total number of crossings and show that the lower bound is actually an accurate indicator of the total number of crossings in different types of graphs.

As shown in Methods, we derive a lower bound χ_l of the total number of crossings in a given graph:

$$\chi_l=\sum_{i=1}^{N}\sum_{j=1}^{i-1}1_{f_{ij}(x_1,d)<0}\leq\chi(\tau_c^{(1)}+\varepsilon,\tau_u) \tag{2}$$

where x_1 is the principal eigenvector of the adjacency matrix A, belonging to the largest eigenvalue λ_1 and d is the degree vector of the given graph.

With the one-crossing assumption, we can compute $\chi(\tau_c^{(1)}+\varepsilon,\tau_u)$ from the infection probability vector $V_\infty(\tau_c^{(1)}+\varepsilon)$ and $V_\infty(\tau_u)$. However, we have to select a proper value of τ_u which is large enough and practical. We set the value of τ_u as the minimum infection rate such that the average fraction of infected nodes $y_\infty(\tau_u)\geq 0.9$, since we find for most Erdős-Rényi (ER), Barabási-Albert (BA) random graphs and the aforementioned real-world network, that the rankings of the nodal degree and the infection probability are almost the same when the average fraction of infected nodes $y_\infty\geq 0.9$. We discuss how we select the value of τ_u in Section "The value of τ_u" of the supplementary information. The scatter plot of the lower bound χ_l vs $\chi(\tau_c^{(1)}+\varepsilon,\tau_u)$ is shown in Fig. 2 for different graphs including ER random graphs, BA random graphs and six graphs constructed from real-world datasets (as described in Section "Real-world graphs" of the supplementary information), and the

dash line in Fig. 2 is $\log \chi_l = \log \chi(\tau_c^{(1)} + \varepsilon, \tau_u) + \log 0.88$, equivalent to

$$\chi_l = 0.88\chi(\tau_c^{(1)} + \varepsilon, \tau_u) \tag{3}$$

We employ the average degree $E[D] = 8$, 10, 12, 14, 16, 18, 20, 40, 60, 80 for ER random graphs and $E[D] = 4$, 6, 8, 10, 12, 14, 16, 18, 20 for BA random graphs. Both ER and BA random graphs have the same size $N = 1000$. We confine ourselves to the connected graphs in this work. Hence, we employ the link density $p = \frac{\mathrm{E}[D]}{N-1}$ of ER random graphs, which is larger than the critical link density $p_c = \frac{\ln N}{N} \approx 0.007$ (equivalently the average degree $\mathrm{E}[D] > 7$), to ensure the connectivity. Fig. 2 and Equation (3) show that the lower bound χ_l is indeed always smaller than and approximately proportional to $\chi(\tau_c^{(1)} + \varepsilon, \tau_u)$. Hence, the lower bound χ_l is a computationally simple indication of the total number of changes in the ranking of the metastable state infection probability in a graph. Moreover, we find that for graphs generated by the same random graph model (ER or BA model), a graph with a small average degree tends to have a large number of crossings; given the average degree, a graph with a large degree variance tends to have more crossings. We can understand this observation as follows. The principal eigenvector component of any node i obeys the eigenvalue equation $(x_1)_i = \sum_{j=1}^{N} a_{ij}(x_1)_j$. The principal eigenvector is positively correlated with the degree vector[19]. Such correlation weakens if the principal eigenvector has a large variance, leading to a large χ_l. When the degree variance is large, the variance of the principal eigenvector tends to be large as well, contributing to a large χ_l. As more links are added to a network, the network becomes more homogeneous and the variance of the principal eigenvector decreases, resulting in a smaller χ_l, or equivalently less crossings.

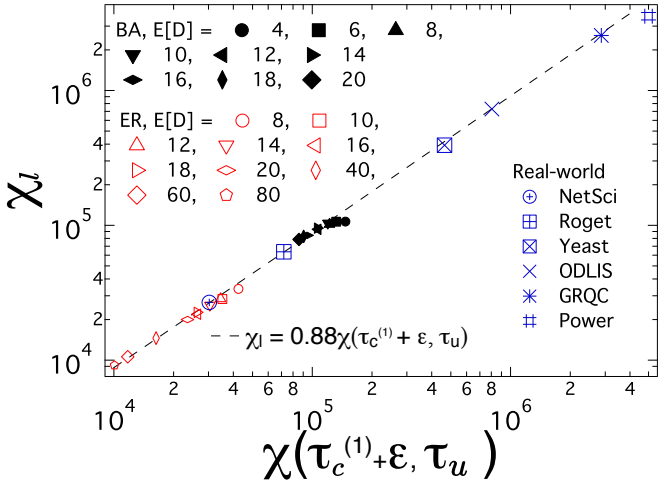


Figure 2. The lower bound χ_l versus the total number of crossings $\chi(\tau_c^{(1)} + \varepsilon, \tau_u)$ in ER random graphs (with the size $N = 1000$), BA random graphs (with the size $N = 1000$) and real-world networks (details in Section "Real-world graphs" of the supplementary information).

The number of crossings in different intervals of τ

As shown in (1), we can compute the number $\chi(\tau_0, \tau_1)$ of crossings in the given interval (τ_0, τ_1) based on the knowledge of the infection probability vectors $V_\infty(\tau_0)$ and $V_\infty(\tau_1)$ only. Here, we show that we can theoretically derive the number of crossings in a small interval $(\tau_0, \tau_0 + \Delta\tau)$ with the only knowledge of $V_\infty(\tau_0)$. Afterwards, we will validate this theory by numerical results, and illustrate in which ranges of the effective infection rate the number of crossings tends to be larger.

The crossings close to a given τ

For sufficiently small $\varepsilon = \Delta\tau > 0$, the Taylor expansion of the steady-state NIMFA infection probability $v_{k\infty}$ for any node k is

$$v_{k\infty}(\tau + \varepsilon) = \sum_{m=0}^{\infty} \frac{\varepsilon^m}{m!} \frac{\partial^m v_{k\infty}(\tau)}{\partial \tau^m} = v_{k\infty}(\tau) + \varepsilon \frac{\partial v_{k\infty}(\tau)}{\partial \tau} + \frac{\varepsilon^2}{2} \frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} + O\left(\varepsilon^3\right) \tag{4}$$

explicit up to order 2. In Section "Derivatives of $v_{i\infty}$ with respect to τ" of the supplementary information, we show the procedure to determine the m-th order derivative $v_{i\infty}(\tau)$ with respect to the effective infection rate τ for any node k.

If $v_{k\infty}(\tau) - v_{m\infty}(\tau) > 0$ and $\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau} > 0$, then $v_{k\infty}(\tau+\varepsilon) - v_{m\infty}(\tau+\varepsilon) > 0$ for sufficiently small $\varepsilon > 0$ and the ranking at $\tau+\varepsilon$ and at τ is unchanged. On the other hand, if $v_{k\infty}(\tau+\varepsilon) - v_{m\infty}(\tau+\varepsilon) = 0$, which implies, for sufficiently small $\varepsilon > 0$ (so that we can ignore the higher order terms in ε^m for $m > 1$ in (4)), that

$$v_{k\infty}(\tau) - v_{m\infty}(\tau) \approx -\varepsilon \left(\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau} \right)$$

In other words, given $v_{k\infty}(\tau)$ of all nodes at τ, then there can be a zero or crossing at $\tau + \varepsilon_{km}$, where

$$\varepsilon_{km} = -\frac{v_{k\infty}(\tau) - v_{m\infty}(\tau)}{\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}} \tag{5}$$

if ε_{km} is small compared to τ. This approach is actually known as the Newton-Raphson method and corresponds with the first term in the Lagrange series for the inverse function (see[20] in Page 304). A second order approximation, by ignoring terms of order $O\left(\varepsilon^3\right)$ in (4), equating $v_{k\infty}(\tau+\varepsilon) - v_{m\infty}(\tau+\varepsilon) = 0$ and solving for ε , yields

$$\varepsilon_{km} = \frac{-\left(\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau} \right) \pm \sqrt{\left(\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau} \right)^2 - 2\left(\frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} - \frac{\partial^2 v_{m\infty}(\tau)}{\partial \tau^2} \right)\left(v_{k\infty}(\tau) - v_{m\infty}(\tau) \right)}}{\left(\frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} - \frac{\partial^2 v_{m\infty}(\tau)}{\partial \tau^2} \right)} \tag{6}$$

which is expected to be more accurate, in spite of the higher computational complexity since now also the set of second order derivatives needs to be solved. We rewrite (6) as

$$\varepsilon_{km} = -\left(\frac{\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}}{\frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} - \frac{\partial^2 v_{m\infty}(\tau)}{\partial \tau^2}} \right) \left\{ 1 \pm \sqrt{1 - 2\left(\frac{\frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} - \frac{\partial^2 v_{m\infty}(\tau)}{\partial \tau^2}}{\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}} \right)\left(\frac{v_{k\infty}(\tau) - v_{m\infty}(\tau)}{\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}} \right)} \right\}$$

Using the generalized binomial expansion $(1+x)^\alpha = \sum_{k=0}^{\infty} \binom{\alpha}{k} z^k$, valid for any $|z| < 1$, up to first order yields

$$\varepsilon_{km} \simeq -\left(\frac{\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}}{\frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} - \frac{\partial^2 v_{m\infty}(\tau)}{\partial \tau^2}} \right) \left\{ 1 \pm \left[1 - \left(\frac{\frac{\partial^2 v_{k\infty}(\tau)}{\partial \tau^2} - \frac{\partial^2 v_{m\infty}(\tau)}{\partial \tau^2}}{\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}} \right)\left(\frac{v_{k\infty}(\tau) - v_{m\infty}(\tau)}{\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \frac{\partial v_{m\infty}(\tau)}{\partial \tau}} \right) \right] \right\}$$

After only retaining the root with the minus sign, we arrive again at (5), illustrating that (5) is accurate when (5) is as small as possible (so that higher order evaluations are not needed). The discriminant must be positive in order to obtain feasible ε_{km}. A positive discriminant is a condition for the existence of crossing in the interval $(\tau, \tau+\varepsilon)$. Hence, given an effective infection rate τ_0 and the corresponding infection probability vector $V_\infty(\tau_0)$, there is a crossing close to τ_0 between the trajectory $v_{k\infty}(\tau)$ and the trajectory $v_{m\infty}(\tau)$ at $\tau + \varepsilon_{km}$ if ε_{km} computed by (5) is positive and small enough.

Numerical results

In the following, we propose to normalize the effective infection rate by the NIMFA epidemic threshold: $\alpha = \frac{\tau}{\tau_c^{(1)}} \geq 1$, so that we can compare the number χ of crossings in different intervals of α in the same range $(1, \alpha_{\max})$ for different network topologies, i.e. different average degrees and different degree distributions. We explore the crossings of the infection probability trajectories when the effective infection rate varies over the range $(1, \alpha_{max})$. We divide the range $(1, \alpha_{\max})$ into intervals (α_{j-1}, α_j) where $j = 1, 2, ..., n$ is the index and $\alpha_n = \alpha_{max}$.

We aim to explore in which interval of the normalized effective infection rate α the crossings are more likely to appear. Hence, instead of directly exploring the number of crossings between the trajectory of every node in the whole interval $(1, \alpha_{\max})$ of the effective infection rate α, we investigate the number $\chi(\alpha_{j-1}, \alpha_j)$ of crossings in (1) in each small interval (α_{j-1}, α_j). We denote $\alpha_0 = 1$ (since the effective infection rate below the epidemic threshold corresponds to the all-healthy state), $\alpha_n = \alpha_{\max}$ and $\alpha_j = \alpha_0 + j\Delta\alpha$, where $\Delta\alpha = (\alpha_{\max} - 1)/n$ is the length of each interval. We will study how the number of crossings changes at different regions of the effective infection rate τ or scaled α. The infection probability $v_{k\infty}(\alpha)$ at any given value of the normalized effective infection rate α is computed by solving the NIMFA equation (11). On one hand, we can further compute the number $\chi(\alpha_{j-1}, \alpha_j)$ of crossings between all node pairs within any interval (α_{j-1}, α_j) by employing our theoretical result (5). On the other hand, we can also numerically compute the number $\chi(\alpha_{j-1}, \alpha_j)$ by (1). We first compare the theoretical (5) and numerical (1) when the normalized effective infection rate α is not close to 1, i.e. when the effective infection rate τ is not close to the epidemic threshold τ_c; specifically, we start from $\alpha_0 = 2$ and $\alpha_j = \alpha_0 + j\Delta\alpha$, where

$\Delta\alpha = 1$. The main figures in Fig. 3 demonstrate that, for both ER and BA graphs, our theoretical result (5) agrees well with the numerical result (1) except for BA graphs in the interval $(2,3)$. The lower accuracy of our theoretical result for small α can be explained as follows. Compared to $\tau_{j-1} = \alpha_{j-1}\tau_c^{(1)}$, a small value of $(\alpha_j - \alpha_{j-1})\tau_c^{(1)}$ is required for the accuracy of the theoretical results (5), since ε in (5) is required to be small with respect to the given effective infection rate τ. Hence, when α_j is smaller, a smaller value of $(\alpha_j - \alpha_{j-1})\tau_c^{(1)}$ is needed for (5) to be accurate.

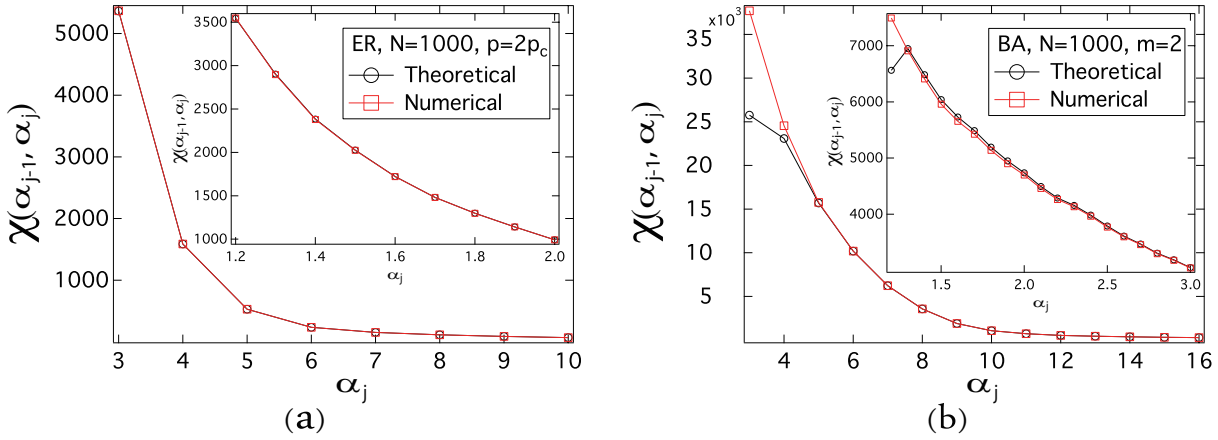


Figure 3. The number $\chi(\alpha_{j-1},\alpha_j)$ of crossings as a function of the normalized effective infection rate α_j. For ER graphs, we employ the link density $p = 2p_c$, thus the average degree E[D]=14, the size $N = 1000$ and the NIMFA epidemic threshold $\tau_c^{(1)} \approx 0.0673$. For BA graphs, we employ the number of newly added links in each step $m = 2$, thus the average degree $E[D] = 4$, the size $N = 1000$, and the NIMFA epidemic threshold $\tau_c^{(1)} \approx 0.0902$. The meta-stable infection probability $v_{k\infty}$ is obtained by solving (11) and the number $\chi(\alpha_{j-1},\alpha_j)$ of crossings is obtained by (1). The results are averaged over 10 realizations.

We further plot the number $\chi(\alpha_{j-1},\alpha_j)$ of crossings in the interval (α_{j-1},α_j) as a function of α_j, when the normalized effective infection rate α is close to 1 and the length of the interval is reduced to $\Delta\alpha = 0.1$. When the length of the interval, i.e. $\Delta\alpha$, is smaller, the theoretical (5) results are more consistent with the numerical (1) results for BA random graphs in the range of $\alpha \in (2,3)$ in the inset than in the main figure of Fig. 3(b). For both ER and BA graphs, the two methods agree with each other well when the intervals of α are small, even when the normalized effective infection rate α is close to 1 as shown in the insets of Fig. 3.

Physical explanation

Figure 3 shows that more crossings appear when the effective infection rate is smaller. In this section, we give a physical explanation of that observation.

At an effective infection rate τ or a normalized effective infection rate α, Equation (14) shows that the comparison of the infection probabilities $v_{k\infty}(\alpha)$ and $v_{m\infty}(\alpha)$ is actually equivalent to the comparison of the sum of the infection probabilities of their neighbors, i.e. $\sum_{j=1}^{N} a_{kj}v_{j\infty}(\alpha)$ and $\sum_{j=1}^{N} a_{mj}v_{j\infty}(\alpha)$. Without loss of generality, we assume that the degree d_k of node k is larger than the degree d_m of node m, i.e. $d_k > d_m$. As discussed in Methods, the infection probability $v_{k\infty}(\alpha) > v_{m\infty}(\alpha)$ if the effective infection rate is large enough. If there exists a value of α_1 at which $\sum_{j=1}^{N} a_{kj}v_{j\infty}(\alpha_1) < \sum_{j=1}^{N} a_{mj}v_{j\infty}(\alpha_1)$ while $d_k > d_m$, there must be a crossing between $v_{k\infty}(\alpha)$ and $v_{m\infty}(\alpha)$ in the interval (α_1,∞). If the infection probabilities $v_{j\infty}(\alpha)$ (where $j = 1,2,...,N$) of all nodes vary in a larger range with respect to the average infection probability $\frac{1}{N}\sum_{j=1}^{N} v_{j\infty}$, i.e. the average fraction y_∞ of infected nodes, then there may be a higher probability that $\sum_{j=1}^{N} a_{kj}v_{j\infty}(\alpha) < \sum_{j=1}^{N} a_{mj}v_{j\infty}(\alpha)$ and thus more crossings could be expected when the effective infection rate τ exceeds α_1. This hypothesis further motivates us to study the normalized standard deviation of the nodal infection probability:

$$\sigma^*(\alpha) = \frac{\sqrt{\sum_{i=1}^{N}(v_{i\infty}(\alpha) - y_\infty(\alpha))^2/N}}{y_\infty(\alpha)} \tag{7}$$

(where we define $\sigma^*(\alpha = 1) = \lim_{\alpha\downarrow 1}\sigma^*(\alpha)$) and explore whether a larger difference $|\sigma^*(\alpha_{j-1}) - \sigma^*(\alpha_j)|$ of σ^* would imply more crossings in the interval (α_{j-1},α_j).

The number $\chi(\alpha_{j-1},\alpha_j)$ of crossings as a function of the difference $\sigma^*(\alpha_{j-1}) - \sigma^*(\alpha_j)$ is shown in Fig. 4(a) for ER random graphs and in Fig. 4(b) for BA random graphs. For both ER and BA random graphs, the number $\chi(\alpha_{j-1},\alpha_j)$ of

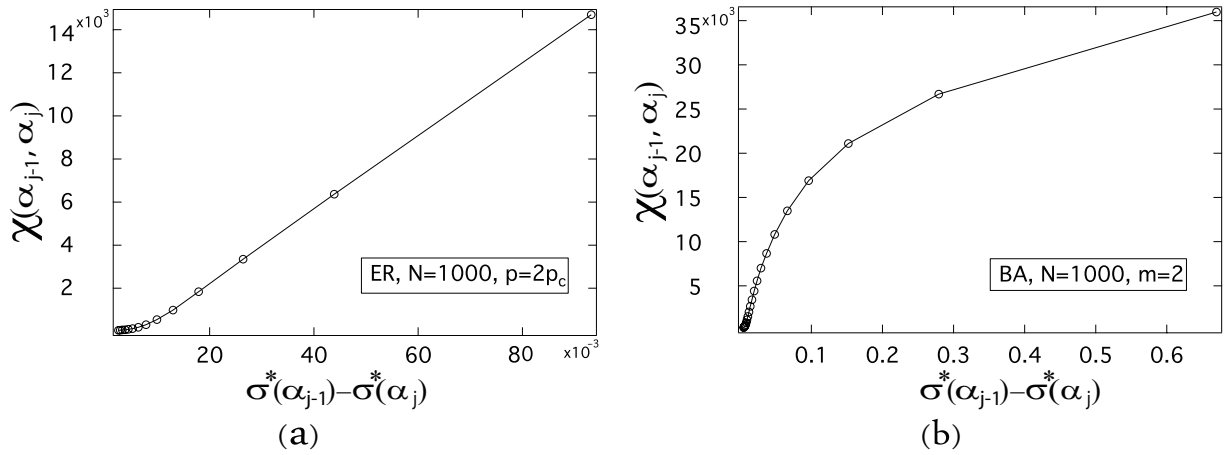


Figure 4. The number $\chi(\alpha_{j-1},\alpha_j)$ of crossings as a function of the difference $\sigma^*(\alpha_{j-1})-\sigma^*(\alpha_j)$ of the normalized standard deviation of the metastable infection probability. For ER graphs, we employ the link density $p=2p_c$, thus the average degree $\mathrm{E}[D]=14$, and the size $N=1000$ (the NIMFA epidemic threshold $\tau_c^{(1)}\approx 0.0673$). For BA graphs, we employ the minimum degree $m=2$, thus the average degree $\mathrm{E}[D]=4$, and the size $N=1000$ (the NIMFA epidemic threshold $\tau_c^{(1)}\approx 0.0902$). The meta-stable infection probability $v_{k\infty}$ is obtained by solving (11), the number $\chi(\alpha_{j-1},\alpha_j)$ of crossings is obtained by (1) and the value of $\sigma^*(\alpha)$ is obtained by (7). The results are averaged over 10 realizations.

crossings are positively correlated with the difference $\sigma^*(\alpha_{j-1})-\sigma^*(\alpha_j)$ in the interval (α_{j-1},α_j). We observe the same in ER and BA random graphs with various average degrees though not shown here. The numerical results support that more crossings tend to appear in an interval where the variable σ^* changes more.

We then further explore how the value of the variable $\sigma^*(\alpha)$ changes with the normalized effective infection rate α. We plot the variable σ^* as a function of the normalized effective infection rate α in Fig. 5(a) for ER random graphs and in Fig. 5(b) for BA random graphs with $N=1000$ and various average degrees, and find that for both types of random graphs the curves can be fitted by a power law function, i.e. σ^* is proportional to $\alpha^{-\gamma}$, especially when the average degree is not small. More figures and the curve fittings are shown in the last section of the supplementary information for both ER and BA random graphs.

Fig. 5 illustrates that the power law exponent γ of the fitting curves is close to 1 as the average degree $\mathrm{E}[D]$ increases for ER random graphs, and that is always approximately 1 for BA random graphs even though the average degree $\mathrm{E}[D]$ is small. Furthermore, the relationship between the variable σ^* and the normalized effective infection rate α follows a power law when the effective infection rate is much larger, as shown in Section "σ^* as a function of τ" of the supplementary information.

When α is large, we can theoretically prove the power-law relationship between the variable σ^* and the normalized effective infection rate α. By (12) and assuming a large enough effective infection rate, we obtain $v_{i\infty}(\tau)=1-\frac{1}{\tau d_i}+O(\tau^{-2})$ for node i and consequently $y_\infty(\tau)=1-\frac{1}{\tau}\mathrm{E}[\frac{1}{D}]+O(\tau^{-2})$, so that (7) becomes

$$\sigma^*=\frac{\sqrt{\mathrm{Var}[\frac{1}{D}]}}{\tau-\mathrm{E}[\frac{1}{D}]}+O(\tau^{-2}) \tag{8}$$

In a finite graph, $\mathrm{Var}[\frac{1}{D}]$ and $\mathrm{E}[\frac{1}{D}]$ are finite, hence σ^* is proportional to τ^{-1}. The NIMFA epidemic threshold $\tau_c^{(1)}$ is a constant for a given graph, and with $\alpha^{-1}=\tau^{-1}\tau_c^{(1)}$, we obtain that σ^* is proportional to α^{-1}. Although the power-law relationship between σ^* and α can be clearly observed in Fig. 5, the effective infection rate τ corresponding to the variable α in this figure may be smaller than 1 and the theoretical proof is only valid when the effective infection rate $\tau\gg 1$. Our result (8) is based on connected graphs, because the terms $\mathrm{E}[\frac{1}{D}]$ and $\mathrm{Var}[\frac{1}{D}]$ are undefined in unconnected graphs with isolated nodes.

The power-law decay of the variable σ^* with the effective infection rate τ explains why there are more crossings when the effective infection rate is smaller.

Validation on a real-world network

Finally, we validate our previous findings on the real-world network – Roget, detailed in Section "Real-world graphs" of the supplementary information. As shown in Fig. 6(a), the number $\chi(\alpha_{j-1},\alpha_j)$ of crossings at normalized effective infection rate

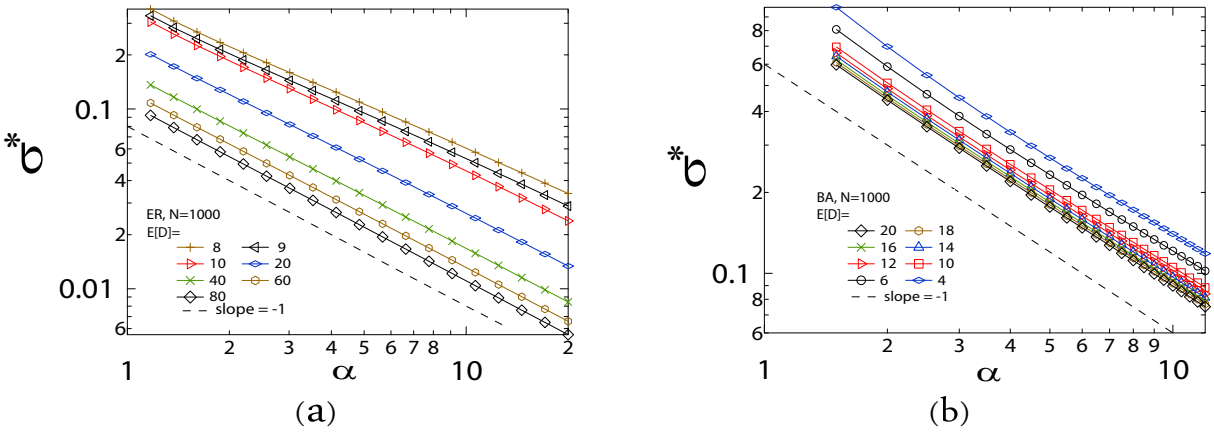


Figure 5. The normalized standard deviation σ^* of infection probabilities of all nodes as a function of α in (a) ER and (b)BA random graphs. The dash line is a power-law curve with the exponent $\gamma = -1$. The sizes of all random graphs are 1000 and the average degree $E[D]$ is shown in the figures. The meta-stable infection probability $v_{k\infty}(\alpha)$ is obtained by solving (11) and the value of $\sigma^*(\alpha)$ is obtained by (7). The NIMFA epidemic threshold $\tau_c^{(1)} \approx 0.1097$, 0.0993, 0.0902, 0.0476, 0.0244, 0.0164 and 0.0124 for ER random graphs with the average degree $E[D] = 8$, 9, 10, 20, 40, 60 and 80 respectively, and $\tau_c^{(1)} \approx 0.0902$, 0.0698, 0.0479, 0.0416, 0.0368, 0.0329, 0.0300 and 0.0274 for BA random graphs with the average degree $E[D] = 4$, 6, 10, 12, 15, 16, 18 and 20 respectively. The results are averaged over 10 realizations.

α interval obtained by theoretical and numerical methods are consistent with each other. The number of crossings decreases fast as α increases, similar to ER and BA models. The main figure of Fig. 6(b) shows that the number $\chi(\alpha_{j-1}, \alpha_j)$ of crossings increases with the difference $\sigma^*(\alpha_{j-1}) - \sigma^*(\alpha_j)$ in the interval (α_{j-1}, α_j). In the inset of Fig. 6(b), we observe the power-law relationship between the variable σ^* and the normalized effective infection rate α. All these findings are well in line with previous results on ER and BA random graphs.

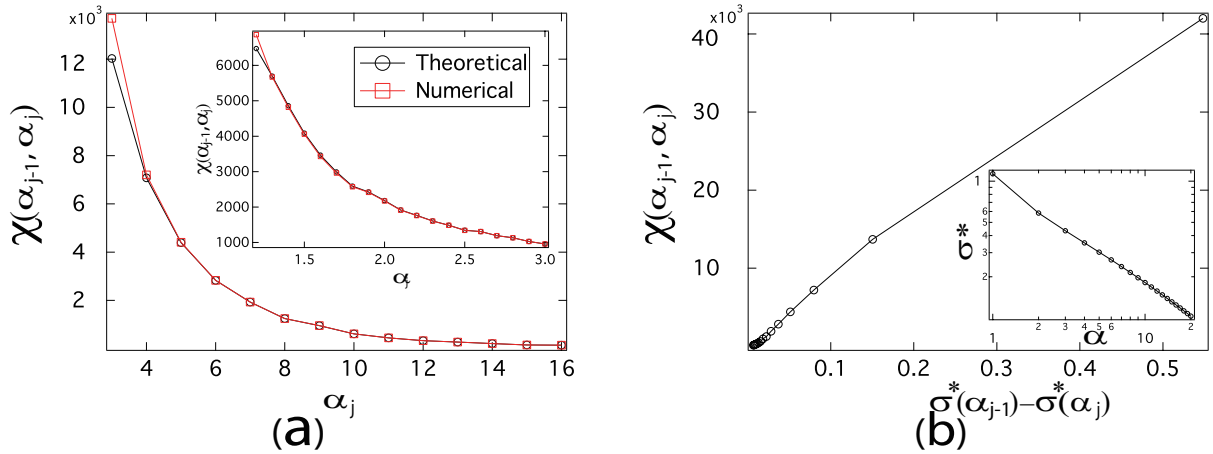


Figure 6. (a) The number $\chi(\alpha_{j-1}, \alpha_j)$ of crossings as a function of the normalized effective infection rate α_j. (b) Main figure: the number $\chi(\alpha_{j-1}, \alpha_j)$ of crossings as a function of the difference $\sigma^*(\alpha_{j-1}) - \sigma^*(\alpha_j)$ of the normalized standard deviation of the metastable infection probability; Inset: the normalized standard deviation σ^* of infection probabilities of all nodes as a function of α. The real-world network – Roget, detailed in Section "Real-world graphs" of the supplementary information, is employed. The meta-stable infection probability $v_{k\infty}$ is obtained by solving (11), the number $\chi(\alpha_{j-1}, \alpha_j)$ of crossings is obtained by (1) and the value of $\sigma^*(\alpha)$ is obtained by (7). The NIMFA epidemic threshold $\tau_c^{(1)} \approx 0.0831$.

Discussion

In the SIS model, the infection probability trajectory $v_{k\infty}(\tau)$ of node k and the infection probability trajectory $v_{m\infty}(\tau)$ of node m may cross if $(v_{k\infty}(\tau_0)-v_{m\infty}(\tau_0)(v_{k\infty}(\tau_1)-v_{m\infty}(\tau_1)<0$, when the effective infection rate τ varies from τ_0 to τ_1. The number $\chi_(\tau_0,\tau_1)$ of crossings of all node pairs within an interval (τ_0,τ_1) of the effective infection rate measures the change in the ranking of the nodal vulnerabilities when the effective infection rate changes from τ_0 to τ_1. We explore in what types of network topologies and in what ranges of the effective infection rates the crossings are more likely to appear. Theoretically, we find a lower bound χ_l in (2) of the total number of crossings in a graph. The lower bound χ_l only depends on topological features, i.e. the degree vector and principal eigenvector of the adjacency matrix. That lower bound χ_l is also shown to reflect the total number of crossings for a given graph. Moreover, we analytically predict the crossings close to an effective infection rate τ_0, given the infection probabilities of all nodes at the effective infection rate τ_0. This theory can be used to estimate the changes of the ranking of the nodal vulnerabilities if the effective infection rate τ slightly increases from its current value τ_0. We find that more crossings tend to appear when the effective infection rate is smaller. Our findings may help network operators to estimate how significant the ranking of nodal vulnerabilities may change for a given change of the effective infection rate on a given network.

This work inspires interesting further questions. For example, how much is the change in the value of the nodal infection probabilities when the trajectories of the nodal infection probability crossing? Can we use the changes in the ranking of nodal infection probabilities to more effectively select the nodes to immunize?

Methods

Network construction

The Erdős-Rényi (ER) random graph[21] is one of the most widely-used and well-studied models. In an ER random graph $G_p(N)$ with N nodes, each pair of nodes is connected with probability p independent from every other pair. The distribution of the degree of a random node is binomial: $\Pr[D=k]=\binom{N-1}{k}p^k(1-p)^{N-1-k}$ and the average degree $\mathrm{E}[D]=(N-1)p$. For large N and constant $\mathrm{E}[D]$, the degree distribution tends[16] to a Poisson distribution: $\Pr[D=k]=exp(-E[D])(\mathrm{E}[D])^k/k!$. Moreover, if the link density $p>p_c=\frac{\ln N}{N}$ for large N, the graph $G_p(N)$ is almost surely connected. We employ ER graphs with $p=2p_c$ (the average degree is approximately $\mathrm{E}[D]=14$) and $N=1000$ as an example in some discussions, but consider the ER graphs with various average degrees when needed.

Besides the ER random graph, the scale-free model is often used to capture the scale-free degree distribution of the real-world networks such as the Internet[22] and World Wide Web[23]. In this work, we consider the Barabási-Albert (BA) model[24], which begins with an initial connected network of m_0 nodes. At each step, a new node is connected to $m\leq m_0$ existing nodes. The probability that an existing node is chosen to be connected is proportional to the degree of the existing node. The degree distribution of BA random graphs[16] is $\Pr[D=k]=ck^{-3}$ for sufficiently large N, where $c=(\sum_{k=m}^{N-1}k^{-3})^{-1}$. The minimum degree of BA graphs is m, and we set $m_0=m+1$ to generate a BA graph with $N=1000$ nodes. Hence, the number of links is $L=\frac{m_0(m_0-1)}{2}+(N-m_0)m=(N-\frac{m_0}{2})m$ and the average degree is $\mathrm{E}[D]=\frac{2L}{N}=\frac{2N-m_0}{N}m$, thus approximately equals to $2m$. We employ the BA random graphs with $m=2$ (the average degree $\mathrm{E}[D]=4$) as an example in discussions and consider more average degrees when needed.

The N-Intertwined Mean-Field Approximation of the SIS model

The N-Intertwined Mean-Field Approximation (NIMFA) is one of the most accurate approximation of the SIS model that takes into account the influence of the network topology[6]. The single governing equation for a node i in the NIMFA is

$$\frac{\mathrm{d}v_i(t)}{\mathrm{d}t}=-\delta v_i(t)+\beta(1-v_i(t))\sum_{j=1}^{N}a_{ij}v_j(t) \tag{9}$$

where $v_i(t)$ is the infection probability of node i at time t, and the adjacency matrix element $a_{ij}=1$ or 0 denotes if there is a link or not between node i and node j. With $V(t)=[v_1(t)\ v_2(t)\ \cdots\ v_N(t)]^T$, the matrix evolution equation of NIFMA is

$$\frac{\mathrm{d}V(t)}{\mathrm{d}t}=(\beta\mathrm{diag}(1-v_i(t))A-\delta I)V(t) \tag{10}$$

where A is the $N\times N$ adjacency matrix of the network, I is the $N\times N$ identity matrix and $\mathrm{diag}(v_i(t))$ is the diagonal matrix with elements $v_1(t),v_2(t),....,v_N(t)$. In the steady state, defined by $\frac{\mathrm{d}V(t)}{\mathrm{d}t}=0$, or equivalently $\lim_{t\to\infty}v_i(t)=v_{i\infty}$ and $\lim_{t\to\infty}V(t)=V_\infty$, we have

$$(\tau\mathrm{diag}(1-v_{i\infty})A-I)V_\infty=0 \tag{11}$$

Given the network and the effective infection rate τ, we can numerically compute the infection probability $v_{i\infty}$ as a function of the effective infection rate τ for each node i by solving (11). The trivial, i.e. all-zero, solution indicates the absorbing state where all nodes are susceptible. The non-zero solution of V_∞ in (11), if exists, points to the existence of a metastable state with a non-zero fraction of infected nodes. Or else, the metastable state can be figured as 0 or not existing. In this paper, we are interested in actually the metastable state.

Furthermore, the NIMFA epidemic threshold $\tau_c^{(1)} = \frac{1}{\lambda_1}$, where λ_1 is the largest eigenvalue of the adjacency matrix A, is a lower bound of the exact epidemic threshold τ_c, i.e. $\tau_c^{(1)} < \tau_c$. The epidemic dies out if the effective infection rate $\tau < \tau_c^{(1)}$. Since the NIMFA is the main approach in this work, we also employ the NIMFA epidemic threshold $\tau_c^{(1)}$. The Laurent series of the steady-state infection probability is given by[16,25]

$$v_{i\infty}(\tau) = 1 + \sum_{m=1}^{\infty} \eta_m(i)\, \tau^{-m} \tag{12}$$

possesses the coefficients $\eta_1(i) = -\frac{1}{d_i}$ and

$$\eta_2(i) = \frac{1}{d_i^2}\left(1 - \sum_{j=1}^{N} \frac{a_{ij}}{d_j}\right) \tag{13}$$

and for $m \geq 2$, the coefficients obey the recursion

$$\eta_{m+1}(i) = -\frac{1}{d_i}\eta_m(i)\left(1 - \sum_{j=1}^{N} \frac{a_{ij}}{d_j}\right) - \frac{1}{d_i}\sum_{k=2}^{m} \eta_{m+1-k}(i) \sum_{j=1}^{N} a_{ij}\eta_k(j)$$

The derivation of the lower bound χ_l

As shown in [16, p. 469] when the effective infection rate $\tau = \tau_c^{(1)} + \varepsilon$ is just above the NIMFA epidemic threshold $\tau_c^{(1)} = \frac{1}{\lambda_1}$, the vector V_∞ with the NIMFA metastable-state infection probabilities is proportional to the principal eigenvector x_1 of the adjacency matrix A. In particular, $v_{k\infty} = \varepsilon (x_1)_k$, where $\varepsilon > 0$ and $(x_1)_k$ is the k-th component corresponding to node k of the principal eigenvector x_1 of the adjacency matrix A, belonging to the largest eigenvalue λ_1. The Perron-Frobenius Theorem[20] states that all vector components of x_1 are non-negative, and even positive if the graph G is connected. Hence, when the effective infection rate is just above the epidemic threshold, the ranking of the infection probability $v_{i\infty}(\tau_c^{(1)} + \varepsilon)$ is the same as the ranking of the component of the principal eigenvector $(x_1)_i$, i.e. $f_{km}(V_\infty(\tau_c^{(1)} + \varepsilon), x_1) = 0$ for any k and m.

On the other hand, the NIMFA steady-state infection probability for node k is given by[18], [16, p. 464] and expressed as

$$v_{k\infty}(\tau) = 1 - \frac{1}{1 + \tau \sum_{j=1}^{N} a_{kj} v_{j\infty}(\tau)} \tag{14}$$

from which we obtain

$$v_{k\infty}(\tau) - v_{m\infty}(\tau) = \tau\,(1 - v_{k\infty}(\tau))\,(1 - v_{m\infty}(\tau)) \sum_{j=1}^{N} \left(a_{kj} - a_{mj}\right) v_{j\infty}(\tau)$$

The sign of $v_{k\infty}(\tau) - v_{m\infty}(\tau)$ thus equals to the sign of $\sum_{j=1}^{N} \left(a_{kj} - a_{mj}\right) v_{j\infty}(\tau)$. Common neighbors of node m and k do not play a role in the sign change of $v_{k\infty}(\tau) - v_{m\infty}(\tau)$. (The common neighbors of node m and k are the set of nodes $\left\{j \in \mathcal{N} : a_{mj} = a_{kj}\right\}$.) Moreover, if the number of non-common neighbors is 1 (or 0), then there is no change in the sign of $v_{k\infty}(\tau) - v_{m\infty}(\tau)$ while the effective infection rate τ varies. Since the minimum infection probability $v_{\min}(\tau) > 0$ for $\tau > \tau_c^{(1)}$ as shown in [16, Lemma 17.4.2 on p. 464], the following bounds apply

$$d_k v_{\min}(\tau) - d_m v_{\max}(\tau) \leq \sum_{j=1}^{N} \left(a_{kj} - a_{mj}\right) v_{j\infty}(\tau) \leq d_k v_{\max}(\tau) - d_m v_{\min}(\tau)$$

where $v_{\max}(\tau)$ and $v_{\min}(\tau)$ are the maximum and minimum infection probability respectively and d_k is the degree of node k, so that the condition $v_{k\infty}(\tau) - v_{m\infty}(\tau) > 0$ at τ is surely satisfied if $d_k - d_m \frac{v_{\max}(\tau)}{v_{\min}(\tau)} > 0$. Using $v_{\max}(\tau) \leq 1 - \frac{1}{1+\tau d_{\max}}$ and

$v_{\min}(\tau) \geq 1 - \frac{1}{\tau d_{\min}}$ in [16, p. 464-465], we arrive at the conservative bound for the condition $v_{k\infty}(\tau) - v_{m\infty}(\tau) > 0$ at τ,

$$d_k > d_m \frac{\tau^2}{\left(\tau - \frac{1}{d_{\min}}\right)\left(\tau + \frac{1}{d_{\max}}\right)}$$

Hence, for large τ, the comparison between $v_{k\infty}(\tau)$ and $v_{m\infty}(\tau)$ reduces to a comparison in the nodal degree: if $d_k > d_m$, then $v_{k\infty}(\tau) > v_{m\infty}(\tau)$. This conclusion implies that there exists an effective infection rate τ_u, above which the ranking of the metastable-state infection probability is the same as the ranking of the nodal degree, i.e. $f_{km}(V_\infty(\tau), d) = 0$ for any k and m (where d is the degree vector), if $\tau \geq \tau_u$.

The above discussion suggests that the number $\chi(\tau_c^{(1)} + \varepsilon, \tau_u)$ of crossings in the interval $(\tau_c^{(1)} + \varepsilon, \tau_u)$ is the total number of crossings which a graph can possess. With the one-crossing assumption, we have

$$\chi(\tau_c^{(1)} + \varepsilon, \tau_u) = \sum_{i=1}^{N} \sum_{j=1}^{i-1} 1_{f_{ij}(V_\infty(\tau_c^{(1)}+\varepsilon), V_\infty(\tau_u))<0} \geq \sum_{i=1}^{N} \sum_{j=1}^{i-1} 1_{f_{ij}(x_1,d)<0} \tag{15}$$

Since only the crossings between two nodes with different degrees are considered in $\sum_{i=1}^{N} \sum_{j=1}^{i-1} 1_{f_{ij}(x_1,d)<0}$, we obtain a lower bound of the total number $\chi(\tau_c^{(1)} + \varepsilon, \tau_u)$ of crossings. In order to simplify the notation, we denote the lower bound of the total number of crossings by $\chi_l = \sum_{i=1}^{N} \sum_{j=1}^{i-1} 1_{f_{ij}(x_1,d)<0}$.

References

1. Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000).
2. Kephart, J. O. & White, S. R. Directed-graph epidemiological models of computer viruses. In *Research in Security and Privacy, 1991. Proceedings., 1991 IEEE Computer Society Symposium on*, 343–359 (IEEE, 1991).
3. Garetto, M., Gong, W. & Towsley, D. Modeling malware spreading dynamics. In *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 3, 1869–1879 (IEEE, 2003).
4. Ganesh, A., Massoulié, L. & Towsley, D. The effect of network topology on the spread of epidemics. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 2, 1455–1466 (IEEE, 2005).
5. Daley, D. J., Gani, J. & Gani, J. M. *Epidemic modelling: an introduction*, vol. 15 (Cambridge University Press, 2001).
6. Van Mieghem, P. The N-intertwined SIS epidemic network model. *Computing* **93**, 147–169 (2011).
7. Cator, E. & Van Mieghem, P. Susceptible-infected-susceptible epidemics on the complete graph and the star graph: Exact analysis. *Phys. Rev. E* **87**, 012811 (2013).
8. Li, C., van de Bovenkamp, R. & Van Mieghem, P. Susceptible-infected-susceptible model: A comparison of n-intertwined and heterogeneous mean-field approximations. *Phys. Rev. E* **86**, 026116 (2012).
9. Boccara, N. & Cheong, K. Critical behaviour of a probabilistic automata network SIS model for the spread of an infectious disease in a population of moving individuals. *J Phys A-Math Gen* **26**, 3707 (1993).
10. Shi, H., Duan, Z. & Chen, G. An SIS model with infective medium on complex networks. *Physica A* **387**, 2133–2144 (2008).
11. Wang, H. *et al.* Effect of the interconnected network structure on the epidemic threshold. *Physical Review E* **88**, 022801 (2013).
12. Li, D., Qin, P., Wang, H., Liu, C. & Jiang, Y. Epidemics on interconnected lattices. *EPL (Europhysics Letters)* **105**, 68004 (2014). URL http://stacks.iop.org/0295-5075/105/i=6/a=68004.
13. Li, C., Wang, H. & Van Mieghem, P. Epidemic threshold in directed networks. *Physical Review E* **88**, 062802 (2013).
14. Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *arXiv preprint arXiv:1408.2701* (2014).
15. Yang, Z. & Zhou, T. Epidemic spreading in weighted networks: an edge-based mean-field solution. *Physical Review E* **85**, 056106 (2012).
16. Van Mieghem, P. *Performance analysis of communications networks and systems* (Cambridge University Press, 2014).

17. Hébert-Dufresne, L., Allard, A., Young, J.-G. & Dubé, L. J. Global efficiency of local immunization on complex networks. *Scientific Reports* **3** (2013).

18. Van Mieghem, P., Omic, J. & Kooij, R. Virus spread in networks. *Networking, IEEE/ACM Transactions on* **17**, 1–14 (2009).

19. Li, C., Li, Q., Van Mieghem, P., Stanley, H. E. & Wang, H. Correlation between centrality metrics and their application to the opinion model. *The European Physical Journal B* **88**, 1–13 (2015).

20. Van Mieghem, P. *Graph spectra for complex networks* (Cambridge University Press, 2010).

21. Erdős, P. & Rényi, A. On random graphs i. *Publ. Math. Debrecen* **6**, 290–297 (1959).

22. Caldarelli, G., Marchetti, R. & Pietronero, L. The fractal properties of internet. *EPL (Europhysics Letters)* **52**, 386 (2000).

23. Albert, R., Jeong, H. & Barabási, A.-L. Internet: Diameter of the world-wide web. *Nature* **401**, 130–131 (1999).

24. Barabási, A.-L., Albert, R. & Jeong, H. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A* **281**, 69–77 (2000).

25. Van Mieghem, P. The viral conductance of a network. *Computer Communications* **35**, 1494–1506 (2012).

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China (No.61603097) and Natural Science Foundation of Shanghai (No.16ZR1446400)

Author contributions statement

P.V.M., B.Q. C.L. and H.W. planned the study; P.V.M., B.Q. and H.W. did the theoretical derivation; B.Q., C.L and H.W. performed the experiments,analyzed the data and prepared the figures; All authors wrote the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.

Supplementary Information: The nodal infection probability in Susceptible-Infected-Susceptible epidemic spreading

Bo Qu[1], Cong Li[2,*], Piet van Mieghem[1], and Huijuan Wang[1]

[1]Delft University of Technology, Delft, 2624CZ, The Netherlands
[2]Fudan University, Shanghai, postcode, China
[*]cong_li@fudan.edu.cn

June 26, 2017

The crossing behavior of the trajectories $v_{k\infty}$

In this section, we use a real-world network as an example to illustrate the crossing behavior by plotting the infection probability $v_{k\infty}$ as a function of the effective infection rate τ for a small number e.g. 10 nodes. The real-world network is called Roget (detailed in Section Real-world graphs), with 994 nodes and the average degree $\mathrm{E}[D] = 7.32$. If we plot all values of the infection probability $v_{k\infty}$ as a function of the effective infection rate τ for a network with hundreds of nodes, it would be difficult to tell which two curves actually cross. Hence, we sample 10 nodes, but according to different strategies to illustrate the crossing behavior. In Fig. S1(a), S1(b), S1(c) and S1(d), 10 nodes are randomly selected from all nodes; in Fig. S1(e), S1(f) and S1(g), 10 nodes are random selected from the nodes with degree $d = 4$, 5 and 6 respectively; Thus, the 10 nodes selected have the same degree in each of these three figures; in Fig. S1(h), the top 10 nodes with largest degrees are selected. We find that the crossing of a pair of nodes is indeed significant with respect to the value of their infection probabilities, when the nodes have quite different degrees, as shown in Fig. S1(a), S1(b) and S1(c) where the nodes are selected randomly. The crossing is less significant when the nodes have similar degrees as shown in Fig. S1(e), S1(f), S1(g) and S1(h). Since most real-world networks have a heavy tail degree distribution, significant crossing/change in infection probability for pairs of nodes is expected when the infection probability varies.

Discussion about the one-crossing assumption

We assume that the two trajectories $v_{k\infty}(\alpha)$ and $v_{m\infty}(\alpha)$ crosses at most once as the effective infection rate α changes. Although our theoretical result about the lower bound of the total number of crossings does not depend on this assumption, our method to compute the number of crossings does depend on such an assumption. Hence, we discuss whether the assumption is reasonably good.

Our simulation results so far show that more than one crossing seldom happen. For example, in a real-world network – Roget, only three pairs of nodes have two crossings in their infection probability trajectories in the infection-rate intervals we observed.

When we count numerically the number of crossings between two infection probability curves when α is changed from 1 to any large value α_{max}, we divide the interval $(1, \alpha_{max})$ into a number of m bins. If $m = 1$, we could find maximally one crossing by comparing the infection probability of the two nodes at $\alpha = 1$ and at $\alpha = \alpha_{max}$ respectively. As the number of bins increases, we may have the possibility to discover the multiple crossings if they exist. Hence, we explore further whether we observe few node pairs whose infection probability curves cross twice is due to the the fact that the bin size we chose is not small enough. Would it be possible that actually two crossings exist within

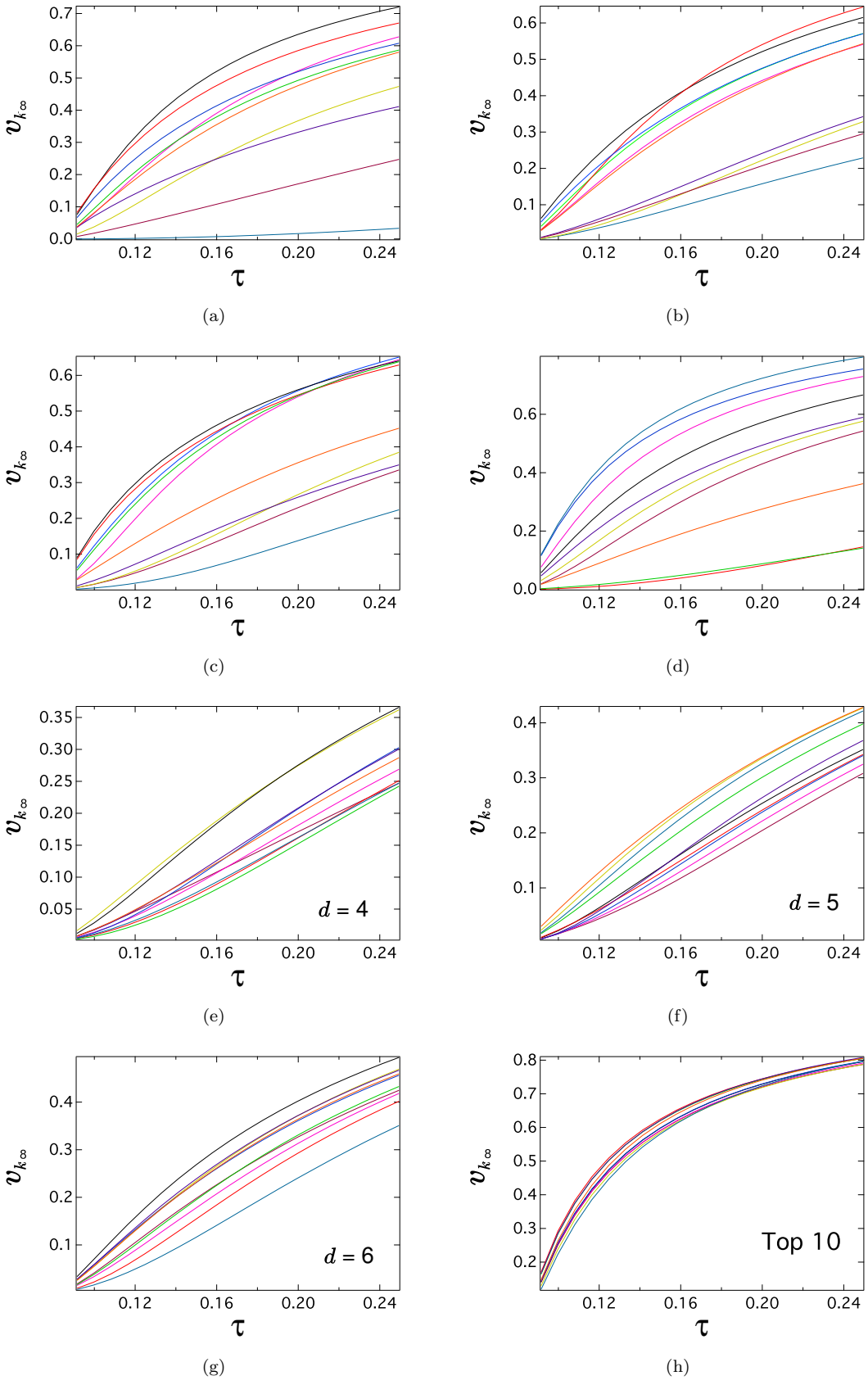
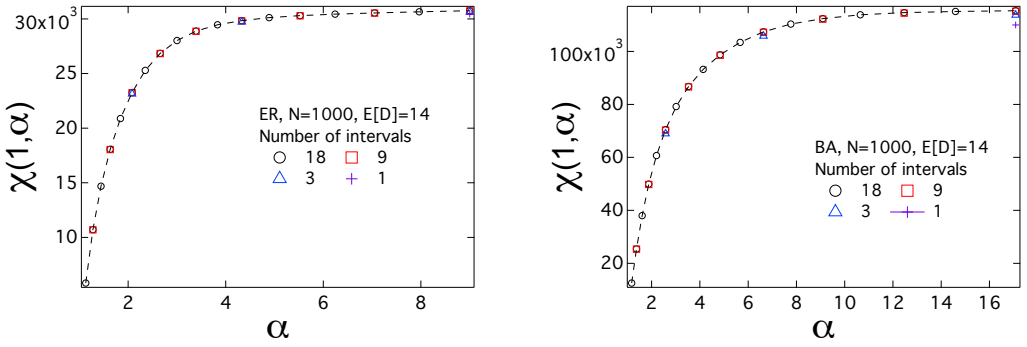


Figure S1: $v_{k\infty}$ as a function of τ for a real-world network.

the same bin which would not be observable if we don't split the bin into smaller ones. Hence, we gradually increase the number of bins to explore whether we could find more crossings. As shown in Fig. S2, we employ ER and BA random graphs with the average degree $\mathrm{E}[D] = 14$ as the examples to show how the number of crossings change when the interval is divided into small ones. We plot the number $\chi(1, \alpha)$ of crossings as a function of the normalized effective infection rate α. We do not observe evident increase of the number of crossings (taking all node pairs into account) as the number of bins increases.

Finally, the bin size should not be too small either. As the bin size becomes small, the change of infection probability for each node when the infection rate is changed from α to $\alpha+\epsilon$ is small. In this case, the precision of the numerical solution to compute the infection probability of each node using NIMFA may not be able to distinguish the ranking change of two nodes if their infection probabilities are close. The seemingly two crossings of a node pair may be due to the limited precision of our numerical solution when the bin size is too small.



(a) (b)

Figure S2: The number $\chi(1, \alpha)$ of crossings as a function of the normalized effective infection rate α for (a) ER random graphs and (b) BA random graphs with the same average degree $E[D] = 14$.

Derivatives of $v_{i\infty}$ with respect to τ

Only for vectors, we use the notation in which a function of a vector is equal to that function applied to the vector components; thus $f(R) = (f(r_1), f(r_2), \ldots, f(r_N))$. Obviously, this convention does not apply to matrices, where the matrix $f(A)$ is different than the matrix with elements $f(a_{ij})$.

Theorem 1 *Let V_∞ be the $N \times 1$ vector with k-th component $v_{k\infty}$, which obeys the NIMFA steady-state equation. Then, all higher order derivative vectors $\frac{\partial^m V_\infty}{\partial \tau^m}$ obey the linear equation*

$$\mathcal{Q}\left(\frac{1}{\tau\left(1-v_{i\infty}\right)^2}\right)\frac{\partial^m V_\infty}{\partial \tau^m} = R_m \tag{1}$$

where $\mathcal{Q}(q_i) = diag(q_i) - A$ is the generalized Laplacian and where the right-hand side vector R_m depends on the previously computed vectors $\left(V_\infty, \frac{\partial V_\infty}{\partial \tau}, \ldots, \frac{\partial^{m-1} V_\infty}{\partial \tau^{m-1}}\right)$. In addition, the generalized Laplacian matrix $\mathcal{Q}\left(\frac{1}{\tau(1-v_{i\infty})^2}\right)$ and its inverse are positive definite matrices.

Proof: Following the approach in [1], the i-th component of the governing steady-state equation [1, (17.45) on p.466]

$$\frac{1}{\tau}\mathrm{diag}\left(\frac{1}{1-v_{i\infty}}\right)V_\infty = AV_\infty \tag{2}$$

written as a generalized Laplacian $\mathcal{Q}(q_i) = \mathrm{diag}(q_i) - A$,

$$\mathcal{Q}\left(\frac{1}{\tau\left(1-v_{i\infty}\right)}\right)V_\infty = 0$$

is

$$\frac{v_{i\infty}}{1-v_{i\infty}} = \tau\sum_{j=1}^{N} a_{ij}v_{j\infty}$$

With $\frac{v_{i\infty}}{1-v_{i\infty}} = \frac{1}{1-v_{i\infty}} - 1$ and $\sum_{j=1}^{N} a_{kj}v_{j\infty} = \frac{v_{k\infty}}{\tau(1-v_{k\infty})}$, differentiation with respect to τ yields

$$\frac{1}{\left(1-v_{k\infty}\right)^2}\frac{\partial v_{k\infty}(\tau)}{\partial \tau} - \tau\sum_{j=1}^{N} a_{kj}\frac{\partial v_{j\infty}(\tau)}{\partial \tau} = \sum_{j=1}^{N} a_{kj}v_{j\infty} = \frac{v_{k\infty}}{\tau\left(1-v_{k\infty}\right)} \tag{3}$$

In matrix form, with the definition [1, p. 472] of the generalized Laplacian $\mathcal{Q}(q_i) = \mathrm{diag}(q_i) - A$, the vector with the derivatives obeys

$$\mathcal{Q}\left(\frac{1}{\tau\left(1-v_{i\infty}\right)^2}\right)\frac{\partial V_\infty}{\partial \tau} = \frac{1}{\tau^2}\mathrm{diag}\left(\frac{1}{1-v_{i\infty}}\right)V_\infty = \frac{1}{\tau^2}\frac{V_\infty}{1-V_\infty} \tag{4}$$

In [2], we have shown that $\mathcal{Q}\left(\frac{1}{\tau(1-v_{i\infty})^2}\right)$ is positive definite (as well as its inverse),

$$\frac{\partial V_\infty}{\partial \tau} = \frac{1}{\tau^2}\mathcal{Q}^{-1}\left(\frac{1}{\tau\left(1-v_{i\infty}\right)^2}\right)\mathrm{diag}\left(\frac{1}{1-v_{i\infty}}\right)V_\infty$$

from which

$$\frac{\partial v_{k\infty}(\tau)}{\partial \tau} = \frac{1}{\tau^2}\sum_{j=1}^{N}\left(\mathcal{Q}^{-1}\left(\frac{1}{\tau\left(1-v_{i\infty}\right)^2}\right)\right)_{kj}\frac{v_{j\infty}}{1-v_{j\infty}} \geq 0$$

Hence, given the knowledge of the vector V_∞ at the effective infection rate τ, the solution of a linear set returns the components of the vector $\frac{\partial V_\infty}{\partial \tau}$.

After a second differentiation with respect to τ of (3) and some manipulations, we have

$$\frac{1}{\tau\left(1-v_{k\infty}\right)^2}\frac{\partial^2 v_{k\infty}\left(\tau\right)}{\partial\tau^2}-\sum_{j=1}^{N}a_{kj}\frac{\partial^2 v_{j\infty}\left(\tau\right)}{\partial\tau^2}=\frac{2}{\tau}\sum_{j=1}^{N}a_{kj}\frac{\partial v_{j\infty}\left(\tau\right)}{\partial\tau}-\frac{2}{\tau\left(1-v_{k\infty}\right)^3}\left(\frac{\partial v_{k\infty}\left(\tau\right)}{\partial\tau}\right)^2$$

In matrix form, we obtain

$$\mathcal{Q}\left(\frac{1}{\tau\left(1-v_{i\infty}\right)^2}\right)\frac{\partial^2 V_\infty}{\partial\tau^2}=\frac{2}{\tau}A\frac{\partial V_\infty}{\partial\tau}-\frac{2}{\tau}\mathrm{diag}\left(\frac{1}{\left(1-v_{k\infty}\right)^3}\right)\left(\frac{\partial V_{k\infty}\left(\tau\right)}{\partial\tau}\right)^2$$

We can avoid the matrix computation $A\frac{\partial V_\infty}{\partial\tau}$, because (4) supplies us with

$$\begin{aligned}A\frac{\partial V_\infty}{\partial\tau}&=\mathrm{diag}\left(\frac{1}{\tau\left(1-v_{i\infty}\right)^2}\right)\frac{\partial V_\infty}{\partial\tau}-\frac{1}{\tau^2}\mathrm{diag}\left(\frac{1}{1-v_{i\infty}}\right)V_\infty\\&=\frac{\frac{\partial V_\infty}{\partial\tau}}{\tau\left(1-V_\infty\right)^2}-\frac{1}{\tau^2}\frac{V_\infty}{1-V_\infty}\end{aligned}$$

while the NIMFA matrix equation (2) shows that

$$A\frac{\partial V_\infty}{\partial\tau}=\frac{d}{d\tau}\frac{1}{\tau}\frac{V_\infty}{1-V_\infty}=\frac{d}{d\tau}\frac{1}{\tau}\left(\frac{1}{1-V_\infty}-u\right)$$

where $u=(1,1,\ldots,1)$ is the all-one vector. Hence,

$$\mathcal{Q}\left(\frac{1}{\tau\left(1-v_{i\infty}\right)^2}\right)\frac{\partial^2 V_\infty}{\partial\tau^2}=R_2 \tag{5}$$

with

$$R_2=\frac{2}{\tau^2}\mathrm{diag}\left(\frac{1}{\left(1-v_{i\infty}\right)^2}\right)\frac{\partial V_\infty}{\partial\tau}-\frac{2}{\tau}\mathrm{diag}\left(\frac{1}{\left(1-v_{k\infty}\right)^3}\right)\left(\frac{\partial V_\infty\left(\tau\right)}{\partial\tau}\right)^2-\frac{2}{\tau^3}\mathrm{diag}\left(\frac{1}{1-v_{i\infty}}\right)V_\infty$$

or

$$R_2=\frac{2}{\tau}\left\{\frac{d}{d\tau}\frac{1}{\tau}\left(\frac{1}{1-V_\infty}-u\right)-\frac{\left(\frac{\partial V_\infty(\tau)}{\partial\tau}\right)^2}{\left(1-V_\infty\right)^3}\right\}$$

which is a same matrix equation as in (4), but a different right-hand side vector, which can only be determined, after solving (4).

A next differentiation with respect to τ of

$$\frac{1}{\left(1-v_{k\infty}\right)^2}\frac{\partial^2 v_{k\infty}\left(\tau\right)}{\partial\tau^2}-\tau\sum_{j=1}^{N}a_{kj}\frac{\partial^2 v_{j\infty}\left(\tau\right)}{\partial\tau^2}=2\sum_{j=1}^{N}a_{kj}\frac{\partial v_{j\infty}\left(\tau\right)}{\partial\tau}-\frac{2}{\left(1-v_{k\infty}\right)^3}\left(\frac{\partial v_{k\infty}\left(\tau\right)}{\partial\tau}\right)^2$$

shows that the left-hand side L and the right-hand side R derivatives are

$$L=\frac{1}{\left(1-v_{k\infty}\right)^2}\frac{\partial^3 v_{k\infty}\left(\tau\right)}{\partial\tau^3}+\frac{2}{\left(1-v_{k\infty}\right)^3}\frac{\partial^2 v_{k\infty}\left(\tau\right)}{\partial\tau^2}\frac{\partial v_{j\infty}\left(\tau\right)}{\partial\tau}-\sum_{j=1}^{N}a_{kj}\frac{\partial^2 v_{j\infty}\left(\tau\right)}{\partial\tau^2}-\tau\sum_{j=1}^{N}a_{kj}\frac{\partial^3 v_{j\infty}\left(\tau\right)}{\partial\tau^3}$$

$$R=2\sum_{j=1}^{N}a_{kj}\frac{\partial^2 v_{j\infty}\left(\tau\right)}{\partial\tau^2}-\frac{3!}{\left(1-v_{k\infty}\right)^4}\left(\frac{\partial v_{k\infty}\left(\tau\right)}{\partial\tau}\right)^3-\frac{4}{\left(1-v_{k\infty}\right)^3}\frac{\partial^2 v_{k\infty}\left(\tau\right)}{\partial\tau^2}\frac{\partial v_{k\infty}\left(\tau\right)}{\partial\tau}$$

Again rewritten as

$$\frac{1}{\tau\left(1-v_{k\infty}\right)^{2}}\frac{\partial^{3}v_{k\infty}\left(\tau\right)}{\partial\tau^{3}}-\sum_{j=1}^{N}a_{kj}\frac{\partial^{3}v_{j\infty}\left(\tau\right)}{\partial\tau^{3}}=\frac{3}{\tau}\sum_{j=1}^{N}a_{kj}\frac{\partial^{2}v_{j\infty}\left(\tau\right)}{\partial\tau^{2}}-\frac{3!}{\tau\left(1-v_{k\infty}\right)^{4}}\left(\frac{\partial v_{k\infty}\left(\tau\right)}{\partial\tau}\right)^{3}$$
$$-\frac{6}{\tau\left(1-v_{k\infty}\right)^{3}}\frac{\partial^{2}v_{k\infty}\left(\tau\right)}{\partial\tau^{2}}\frac{\partial v_{k\infty}\left(\tau\right)}{\partial\tau}$$

leads to the matrix form

$$\mathcal{Q}\left(\frac{1}{\tau\left(1-v_{i\infty}\right)^{2}}\right)\frac{\partial^{3}V_{\infty}}{\partial\tau^{3}}=\frac{3}{\tau}A\frac{\partial^{2}V_{\infty}}{\partial\tau^{2}}-\frac{6}{\tau}\mathrm{diag}\left(\frac{1}{\left(1-v_{k\infty}\right)^{4}}\right)\left(\frac{\partial V_{k\infty}\left(\tau\right)}{\partial\tau}\right)^{3}$$
$$-\frac{6}{\tau}\mathrm{diag}\left(\frac{1}{\left(1-v_{k\infty}\right)^{3}}\right)\frac{\partial^{2}V_{k\infty}\left(\tau\right)}{\partial\tau^{2}}\frac{\partial V_{k\infty}\left(\tau\right)}{\partial\tau}$$

Introducing $A\frac{\partial^{2}V_{\infty}}{\partial\tau^{2}}$ from (5) as

$$A\frac{\partial^{2}V_{\infty}}{\partial\tau^{2}}=\mathrm{diag}\left(\frac{1}{\tau\left(1-v_{i\infty}\right)^{2}}\right)\frac{\partial^{2}V_{\infty}}{\partial\tau^{2}}-\frac{2}{\tau^{2}}\mathrm{diag}\left(\frac{1}{\left(1-v_{i\infty}\right)^{2}}\right)\frac{\partial V_{\infty}}{\partial\tau}+$$
$$\frac{2}{\tau}\mathrm{diag}\left(\frac{1}{\left(1-v_{k\infty}\right)^{3}}\right)\left(\frac{\partial V_{k\infty}\left(\tau\right)}{\partial\tau}\right)^{2}+\frac{2}{\tau^{3}}\mathrm{diag}\left(\frac{1}{1-v_{i\infty}}\right)V_{\infty}$$

yields

$$\mathcal{Q}\left(\frac{1}{\tau\left(1-v_{i\infty}\right)^{2}}\right)\frac{\partial^{2}V_{\infty}}{\partial\tau^{2}}=R_{3}$$

where

$$R_{3}=\frac{6}{\tau^{4}}\frac{V_{\infty}}{1-V_{\infty}}-\frac{6}{\tau^{3}}\frac{\frac{\partial V_{\infty}}{\partial\tau}}{\left(1-V_{\infty}\right)^{2}}+\frac{3}{\tau^{2}}\frac{\frac{\partial^{2}V_{\infty}}{\partial\tau^{2}}}{\left(1-V_{\infty}\right)^{2}}+\frac{6}{\tau^{2}}\frac{\left(\frac{\partial V_{\infty}(\tau)}{\partial\tau}\right)^{2}}{\left(1-V_{\infty}\right)^{3}}$$
$$-\frac{6}{\tau}\frac{\left(\frac{\partial V_{\infty}(\tau)}{\partial\tau}\right)^{3}}{\left(1-V_{\infty}\right)^{4}}-\frac{6}{\tau}\frac{\frac{\partial^{2}V_{k\infty}(\tau)}{\partial\tau^{2}}\frac{\partial V_{k\infty}(\tau)}{\partial\tau}}{\left(1-V_{\infty}\right)^{3}}$$

The computation illustrates the general structure (1) and demonstrates the Theorem. □

From a numerical point of view, the non-linear NIMFA steady-state matrix equation (2) only needs be solved once for a particular value of τ so that the vector $V_{\infty}\left(\tau\right)$ is known, as well as the generalized Laplacian $\mathcal{Q}\left(\frac{1}{\tau\left(1-v_{i\infty}\right)^{2}}\right)$. The Taylor expansion

$$V_{\infty}\left(\tau+\Delta\tau\right)=\sum_{m=0}^{\infty}\frac{\left(\Delta\tau\right)^{m}}{m!}\frac{\partial^{m}V_{\infty}\left(\tau\right)}{\partial\tau^{m}} \tag{6}$$

specifies the NIMFA infection probability vector $V_{\infty}\left(\tau+\Delta\tau\right)$ at another effective infection rate $\tau+\Delta\tau$, provided that the Taylor series converges at $\tau+\Delta\tau$. As mentioned earlier in [3], unfortunately, the convergence radius of the series in (6) is difficult to determine in general. The left-hand side positive definite matrix $\mathcal{Q}\left(\frac{1}{\tau\left(1-v_{i\infty}\right)^{2}}\right)$ in (1) is the same for all orders $m\geq1$ and can be inverted if a high precision and many terms in the Taylor series (6) are required.

The value of τ_u

We define $\chi_D(\tau) = \sum_{i=1}^{N} \sum_{j=1}^{i-1} 1_{f_{ij}(V(\tau),D)<0}$, then the larger χ_D is the higher the difference between the rankings of the infection probability at τ and the nodal degree is. As shown in S3, we plot $\chi_D(\tau)$ as a function of the average fraction y_∞ of infected nodes for ER and BA random graphs with the average degree $E[D] = 14$ as an example. We find that for both graphs $\chi_D \approx 0$ when the average fraction y_∞ of infected nodes is above 0.9, which suggests that we can employ the value of τ_u so that $y_\infty(\tau_u) = 0.9$. We have also done such tests on all the other networks in this paper and obtain the same conclusion. Hence, we employ the value τ_u, leading to $y_\infty(\tau_u) = 0.9$, for all networks in this paper.

This choice of 0.9, though not necessarily optimal, is supported by the following aspects. Practically, we would like to choose $y_\infty(\tau_u)$ as large as possible so that real-world prevalence levels are covered. Since real-world prevalence seldom reaches 0.9, $y_\infty(\tau_u) = 0.9$ is large enough. Also, we would like to choose $y_\infty(\tau_u)$ as large as possible so that $\chi(\tau_c^{(1)} + \epsilon, \tau_u)$ well counts the total number of crossings. Moreover, $y_\infty(\tau_u)$ should not be too large because the infection probability of the nodes are very close to each other when the prevalence is high, and the precision of numerical solution to compute the infection probability per node is not sufficient to distinguish nor to rank the nodes according to their infection probabilities. Furthermore, we observed that the crossing seldom happens when $y_\infty(\tau) > 0.9$ in all the networks generated by the two network models as well as in real-world networks. This is due to the fact that the number of crossings decreases as τ increases, as observed and discussed in the paper.

To compute the value of τ_u which leads to a high prevalence (0.9), we can employ the Laurent series of the steady-state infection probability [3, 1]:

$$v_{i\infty}(\tau) = 1 + \sum_{m=1}^{\infty} \eta_m(i)\, \tau^{-m} \tag{7}$$

where the coefficient $\eta_1(i) = -\frac{1}{d_i}$ and

$$\eta_2(i) = \frac{1}{d_i^2}\left(1 - \sum_{j=1}^{N} \frac{a_{ij}}{d_j}\right) \tag{8}$$

and for $m \geq 2$, the coefficients obey the recursion

$$\eta_{m+1}(i) = -\frac{1}{d_i}\eta_m(i)\left(1 - \sum_{j=1}^{N} \frac{a_{ij}}{d_j}\right) - \frac{1}{d_i}\sum_{k=2}^{m} \eta_{m+1-k}(i) \sum_{j=1}^{N} a_{ij}\eta_k(j)$$

Considering a large value of τ_u

$$v_{i\infty} = 1 - \frac{1}{\tau_u d_i} + O(\tau^{-2})$$

and

$$y_\infty = 1 - \frac{1}{\tau_u}\mathrm{E}[\frac{1}{D}] + O(\tau^{-2})$$

then, ignoring the second order condition $O(\tau^{-2})$,

$$\tau_u \approx \frac{1 - y_\infty}{\mathrm{E}[\frac{1}{D}]}$$

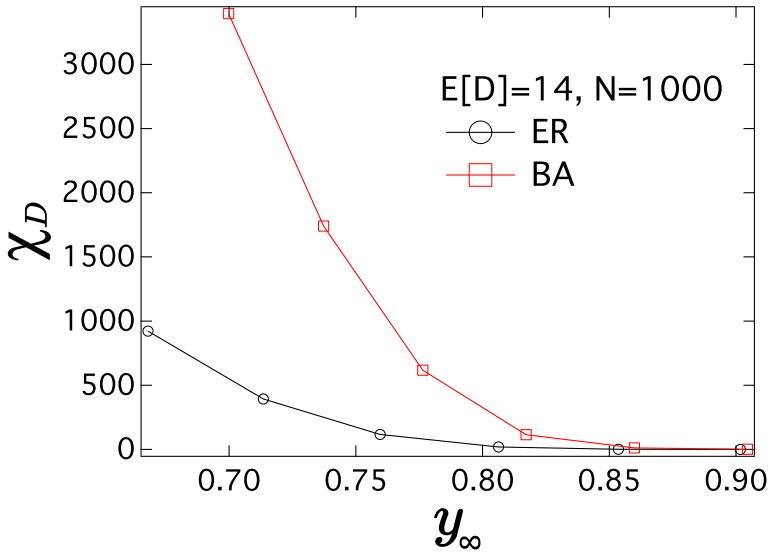


Figure S3: The plot of χ_D as a function of the average fraction of infected nodes. The results are averaged over 10 realizations

.

Real-world graphs

We use 6 connected and undirected graphs from real-world datasets. Some graphs are originally directed and may not be connected. We use the largest component of the unconnected graphs and change the directed graphs to undirected. The description of the 6 graphs are as follows:

1. GRQC: Arxiv GR-QC (General Relativity and Quantum Cosmology) collaboration network is from the e-print arXiv and covers scientific collaborations between authors papers submitted to General Relativity and Quantum Cosmology category. If an author i coauthored a paper with author j, there is link between i and j. The data covers papers in the period from January 1993 to April 2003.

2. NetSci: A coauthorship network of scientists working on network theory and experiment. The network was compiled from the bibliographies of two review articles on networks.

3. ODLIS: The network is based on the ODLIS: Online Dictionary of Library and Information Science (December 2000). The nodes are the terms in ODLIS and there is link between two terms if one is used to describe another one.

4. Roget: The network contains cross-references in Roget's Thesaurus, 1879. Each node of the graph corresponds to one of the categories in the 1879 edition of Peter Mark Roget's Thesaurus of English Words and Phrases. There is a link between two categories if one is the reference of the other. (See http://vlado.fmf.uni-lj.si/pub/networks/data/dic/roget/Roget.htm)

5. Power: The network represents the topology of the Western States Power Grid of the United States.

6. Yeast: The protein-protein interaction network in budding yeast. There is link between protein i and protein j if they have the interaction.

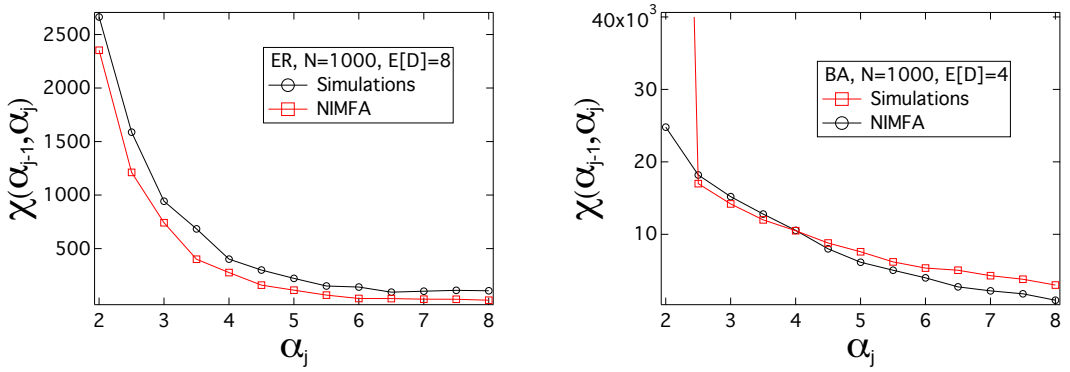
In Table S3, we list the size N, the average degree $E[D]$, the degree variance $Var[D]$ and the normalized degree variance $Var^*[D]$ of the 6 graphs.

Table S3: The real-world graph used in this paper.

	GRQC	NetSci	ODLIS	Roget	Power	Yeast
N	4158	379	2898	994	4941	2224
$E[D]$	6.46	4.82	11.30	7.32	2.67	5.94
$Var[D]$	74.42	15.46	679.61	23.66	3.21	63.70

The comparison between NIFMA and the continuous-time simulation

We compare the number of crossings obtained by NIMFA and the simulations of the exact SIS model. We show two examples of the comparison in Fig. S4. Because the NIMFA epidemic threshold $\tau_c^{(1)}$ is actually the lower bound of the real epidemic threshold, i.e. $\tau_c^{(1)} < \tau_c$, and to determine the value of $\alpha_c = \frac{\tau_c}{\tau_c^{(1)}} > 1$ for different topology is difficult, we start the comparison from $\alpha = 2$ (attempting to exclude the crossings near the epidemic threshold). Fig. S4 shows that the results of the simulation and NIMFA agree with each other quite well for both networks when α is not large. When α is large, i.e. the infection probability of each node is high and close to each other, there might be some crossings caused by the limited precision of the numerical NIMFA solution or the simulations. Because the precision of the numerical solution is higher than that of the simulation of the exact SIS model, the number χ of crossings obtained from the simulation tends to be larger than that from NIMFA if the effective infection rate α is large.



(a) (b)

Figure S4: The comparison of the number χ of crossings between NIMFA and the simulation of the exact SIS model for (a) an ER random graph with $N = 1000$ and $E[D] = 8$; (b) a BA random graph with $N = 1000$ and $E[D] = 4$. The linear sampling is employed with the step $\Delta\alpha = 0.5$.

The normalized standard deviation σ^* of the steady-state infection probability

ER random graphs with $N = 1000$

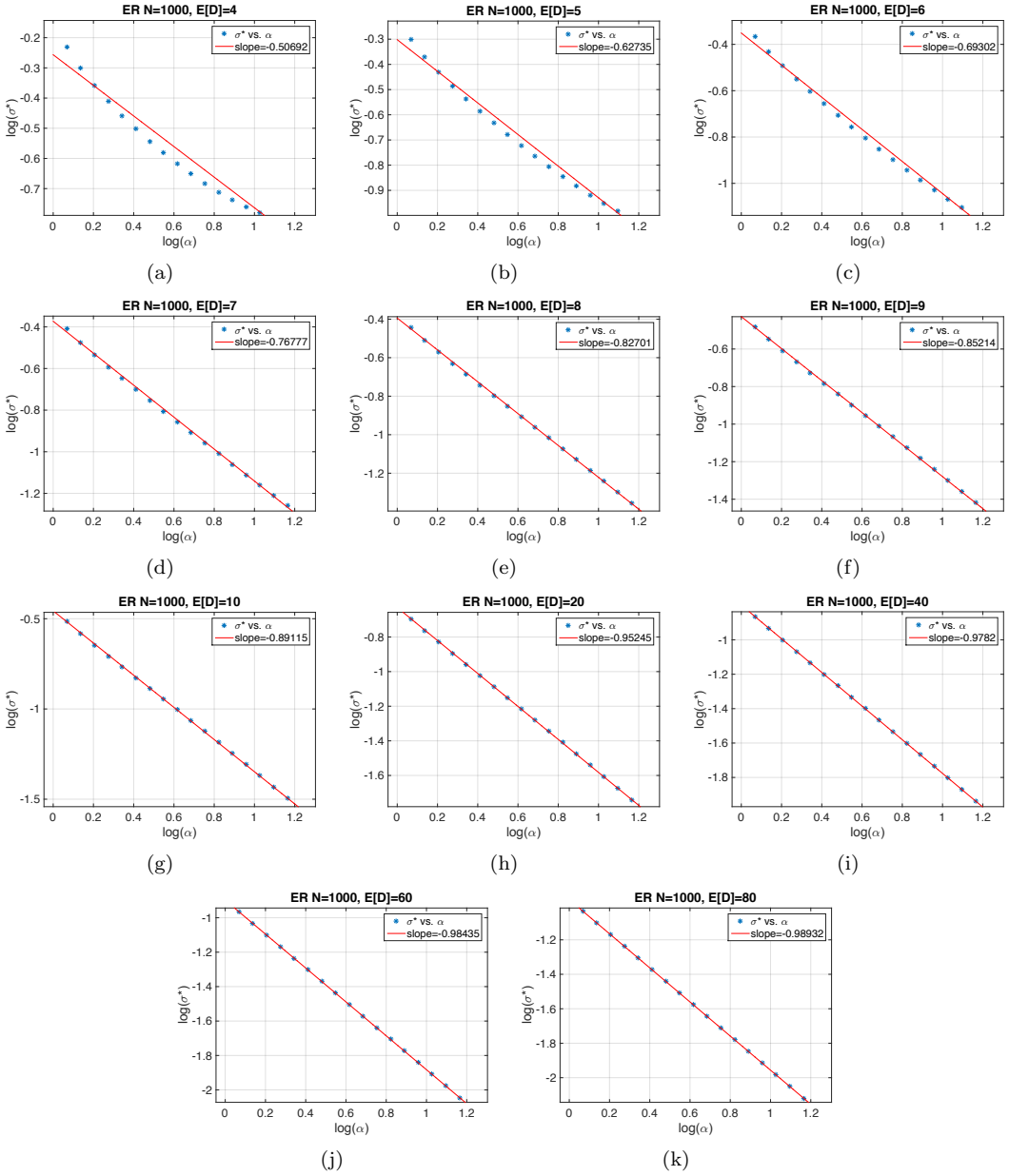


Figure S5: σ^* as a function of α for ER random graphs and the corresponding fitting curve.

BA random graphs with $N = 1000$

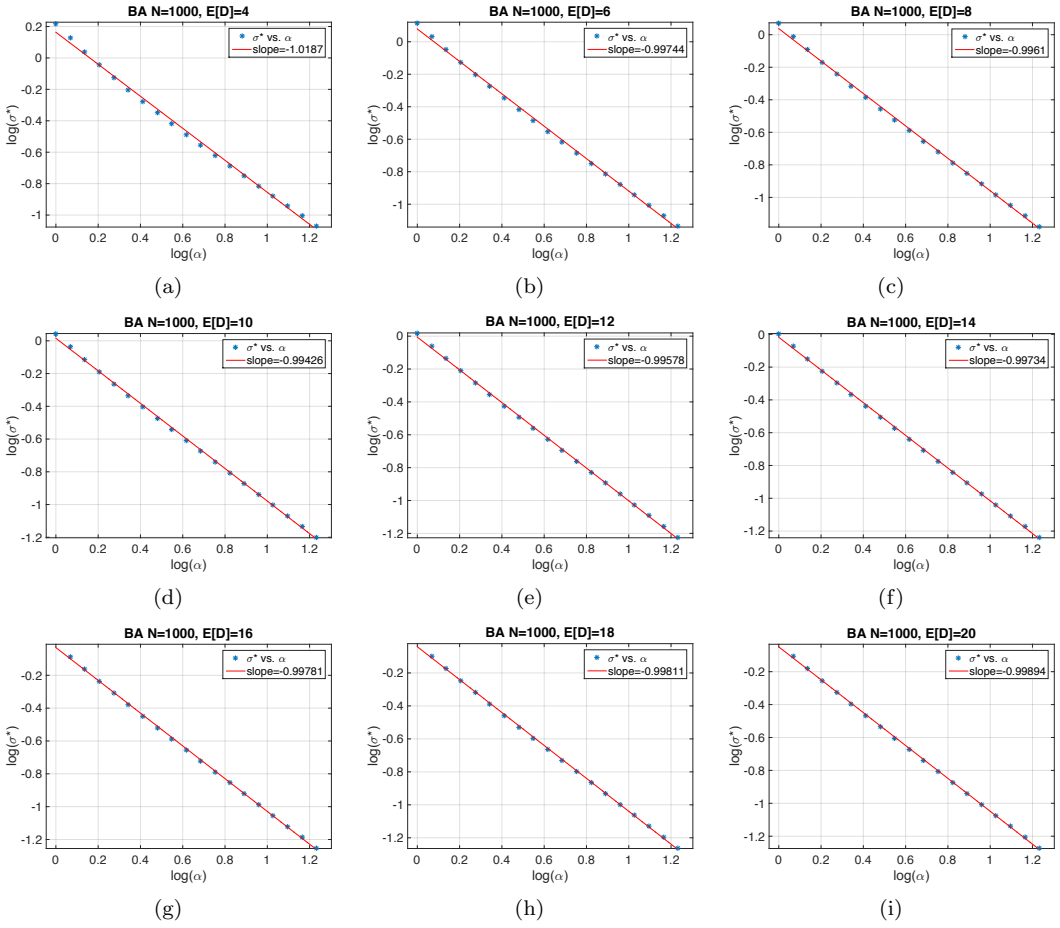
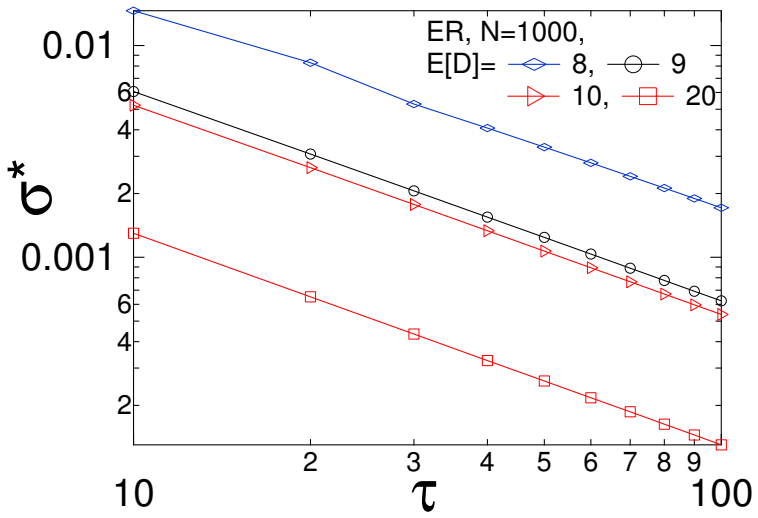


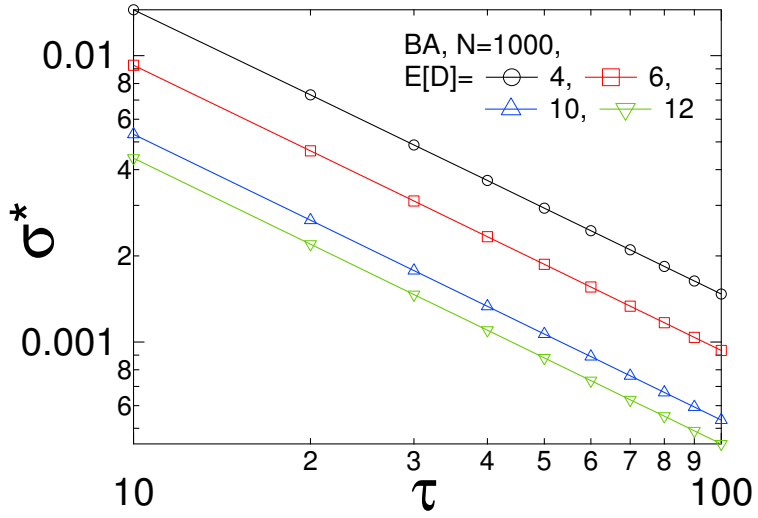
Figure S6: σ^* as a function of α for BA random graphs and the corresponding fitting curve.

σ^* as a function of τ

Here we show the relationship between σ^* and τ when $\tau \gg 1$.



(a)

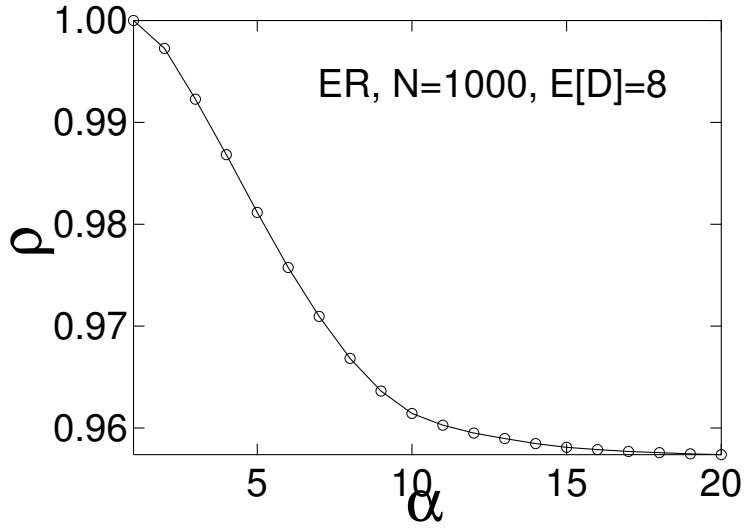


(b)

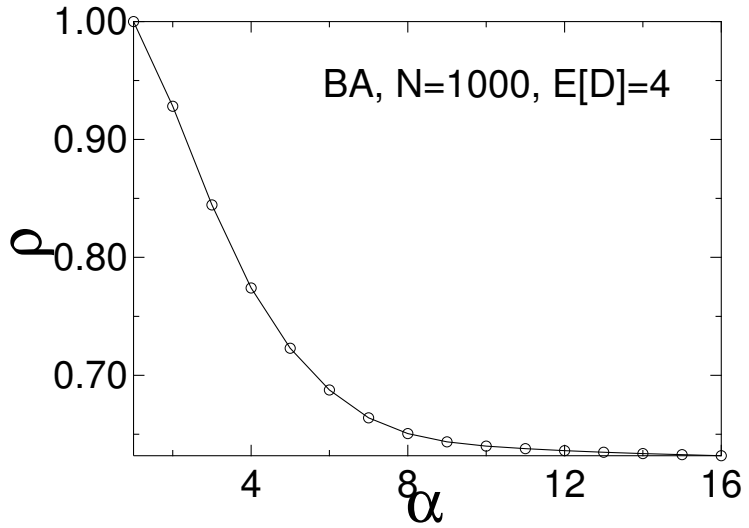
Figure S7: σ^* as a function of τ for (a) ER random graphs with the size $N = 1000$ and (b) BA random graphs with the size $N = 1000$.

The Spearman rank correlation ρ as a function of α

In Fig. S8, we choose two graphs, an ER random graph with $\mathrm{E}[D] = 8$ and a BA random graph with $\mathrm{E}[D] = 4$, as the example. We plot Spearman Rank Correlation between $V_\infty(\tau_c^{(1)} + \epsilon)$ and $V_\infty(\alpha\tau_c^{(1)})$. We find that the rank correlation decreases fast when the effective infection rate is small. Moreover, there tend to be a few nodes drastically changing ranks in BA random graphs but not in ER graphs.



(a)



(b)

Figure S8: The Spearman rank correlation between $V_\infty(\tau_c^{(1)} + \epsilon)$ and $V_\infty(\alpha\tau_c^{(1)})$ as a function of α for (a) ER random graphs with the size $N = 1000$ and the average degree $\mathrm{E}[D] = 8$ and (b) BA random graphs with the size $N = 1000$ and the average degree $\mathrm{E}[D] = 4$.

References

[1] Piet Van Mieghem. *Performance analysis of communications networks and systems*. Cambridge University Press, 2014.

[2] Piet Van Mieghem and Jasmina Omic. In-homogeneous virus spread in networks. *arXiv preprint arXiv:1306.2588*, 2013.

[3] Piet Van Mieghem. The viral conductance of a network. *Computer Communications*, 35(12):1494–1506, 2012.