Numerical identification of epidemic thresholds for susceptible-infected-recovered model on finite-size networks

Panpan Shu,¹ Wei Wang,¹ Ming Tang,^{1,2,a)} and Younghae Do³

¹Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 610054, China

²State key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

³Department of Mathematics, Kyungpook National University, Daegu 702-701, South Korea

(Received 11 April 2015; accepted 26 May 2015; published online 4 June 2015)

Epidemic threshold has always been a very hot topic for studying epidemic dynamics on complex networks. The previous studies have provided different theoretical predictions of the epidemic threshold for the susceptible-infected-recovered (SIR) model, but the numerical verification of these theoretical predictions is still lacking. Considering that the large fluctuation of the outbreak size occurs near the epidemic threshold, we propose a novel numerical identification method of SIR epidemic threshold by analyzing the peak of the epidemic variability. Extensive experiments on synthetic and real-world networks demonstrate that the variability measure can successfully give the numerical threshold for the SIR model. The heterogeneous mean-field prediction agrees very well with the numerical threshold, except the case that the networks are disassortative, in which the quenched mean-field prediction is relatively close to the numerical threshold. Moreover, the numerical method presented is also suitable for the susceptible-infected-susceptible model. This work helps to verify the theoretical analysis of epidemic threshold and would promote further studies on the phase transition of epidemic dynamics. © 2015 AIP Publishing LLC.

[<http://dx.doi.org/10.1063/1.4922153>]

Epidemic threshold, which is one of the most important features of the epidemic dynamics, has attracted much attention recently. The existing studies have provided different theoretical predictions for epidemic threshold of the susceptible-infected-recovered (SIR) model on complex networks, while the numerical verification of these theoretical predictions is still lacking. As a result, it is very necessary to develop an effective numerical measure for identifying the SIR epidemic threshold. In this paper, the numerical identification of the SIR epidemic threshold is systematically studied. We present a numerical method by analyzing the peak of the epidemic variability to identify the epidemic threshold. To understand the effectiveness of the variability measure, the distribution of outbreaks sizes is investigated near the epidemic threshold on random regular networks. Based on the analysis of the cutoff hypothesis of the outbreak size distribution, we find that the variability measure can provide an excellent identification of the epidemic threshold. We further use the variability measure to verify the existing theoretical predictions on scale-free and real networks. The results show that the heterogeneous mean-field (HMF) prediction agrees very well with the numerical threshold, except the case that the networks are disassortative, in which the quenched mean-field (QMF) prediction is relatively close to the numerical threshold. The numerical method presented can effectively identify SIR epidemic thresholds on various networks, and could be extended to other dynamical processes such as information diffusion and behavior spreading. This work

provides us a deep understanding of epidemic threshold and would promote further studies on phase transition of epidemic dynamics.

I. INTRODUCTION

Models for disease propagation are at the core of our understanding about epidemic dynamics on complex networks.^{1,2} Two epidemic models of particular importance are the susceptible-infected-susceptible (SIS) and susceptible-infected-recovered (SIR) models.³ At each time step, an infected node can transmit a disease to each of its susceptible neighbors with probability β . At the same time, the infected nodes become susceptible again in the SIS model or recover in the SIR model with probability μ . In the SIS model, a critical value of the effective transmission rate $\lambda = \beta/\mu$ separates the absorbing phase with only healthy nodes from the active phase with a stationary density of infected nodes. Differently, no steady state is allowed in the SIR model, but a threshold still exists above which the final fraction of recovered nodes is finite.⁴

The traditional theoretical study on the epidemic threshold of the SIS model is based on the heterogeneous mean-field (HMF) theory, which means that all the nodes within a given degree are considered to be statistically equivalent.^{5,6} According to the HMF theory, the epidemic threshold of SIS model^{7,8} is given by $\lambda_c^{HMF} = \frac{\langle k \rangle}{\langle k^2 \rangle}$, where $\langle k \rangle$ and $\langle k^2 \rangle$ are the first and second moments of degree distribution $P(k)$,⁹ respectively. As the quenched structure of the network and dynamical correlations between the state of adjacent nodes is

^{a)}Electronic mail: tangminghuang521@hotmail.com

neglected in the HMF theory,¹⁰ researchers have proposed an important improvement over the HMF theory—quenched mean-field (QMF) theory. The QMF theory fully preserves the actual quenched structure of the network described as its adjacency matrix, and the epidemic threshold is predicted to be^{11–13}

$$\lambda_c^{QMF} = \frac{1}{\Lambda_N}, \quad (1)$$

where Λ_N is the maximum eigenvalue of the adjacency matrix of a given network. Considering that the existing theories more or less have some limitations (e.g., the HMF theory neglects the quenched structure of the network; QMF theory ignores dynamical correlations¹⁴), some numerical methods such as the finite-size scaling analysis,¹⁵ susceptibility,¹⁶ and lifetime measure¹⁷ have been proposed to check the accuracies of different theoretical predictions for the SIS model. Among these existing methods, the relatively common one for a network with finite size N is the susceptibility measure

$$\chi = N \frac{\langle \rho^2 \rangle - \langle \rho \rangle^2}{\langle \rho \rangle}, \quad (2)$$

where ρ denotes the outbreak size. In Ref. 18, the susceptibility measure has been shown to be very effective for identifying the SIS epidemic thresholds on various networks.

For another paradigmatic epidemic model, the SIR model, there have been a lot of theoretical studies on its epidemic threshold. The earliest theoretical study on the SIR epidemic threshold is based on the assumption of homogeneous mixing, showing that the SIR epidemic threshold is inversely proportional to the average connectivity $\langle k \rangle$.³ At the HMF level,¹⁹ the epidemic threshold of SIR model takes the value

$$\lambda_c^{HMF} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}. \quad (3)$$

On networks with power-law scaling $P(k) \sim k^{-\gamma}$,^{9,20} where γ is the degree exponent, the HMF approach predicts a vanishing threshold for scale-free networks with $\gamma \leq 3$ and the finite threshold for $\gamma > 3$.⁴ Mapping the SIR model to a bond percolation process,²¹ the epidemic threshold coincides with the result of Eq. (3) for a SIR model with unit infection time (e.g., $\mu = 1$), and when the infection times vary among infected nodes (e.g., $\mu < 1$), the epidemic threshold is given by⁴

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle - 2\langle k \rangle}. \quad (4)$$

According to the QMF theory,¹¹ the epidemic threshold of the SIR model has the same expression as Eq. (1). However, the QMF result is even qualitatively not correct, as the QMF predicts a vanishing threshold for power-law distributed networks with $\gamma > 3$ that is in conflict with the visually numerical results.²²

Although the numerical threshold of the SIS model has attracted much attention,^{15–18} the systematic study of the numerical identification of the SIR epidemic threshold is still

insufficient. It is well known that the outbreak size becomes finite above the threshold λ_c .²¹ However, as the value of λ increases, the outbreak size continuously changes from an infinitesimal fraction to a finite fraction in the SIR model,²³ and thus, it is difficult to determine the value of λ at which the outbreak size turns to be finite. To our knowledge, there has no effective numerical method for identifying the SIR epidemic threshold in previous studies. In this work, we perform extensive numerical simulations of the SIR model on networks with finite size, and present a numerical identification method by analyzing the peak of the epidemic variability^{24,25} (i.e., the maximal value of the epidemic variability) to identify the epidemic threshold. The effectiveness of the numerical measure is checked on random regular networks (RRN), where the HMF theory is exact. To get a deep understanding of the validity of the numerical method, we investigate the distribution of outbreaks sizes near the epidemic threshold. The robustness of the variability measure is confirmed by the analysis of cutoff hypothesis of the outbreak size distribution. We further employ the variability measure to verify the theoretical predictions on scale-free networks and real-world networks, where the results indicate that the HMF prediction agrees very well with the numerical threshold, except the case that the networks are disassortative, in which the QMF prediction is relatively close to the numerical threshold.

II. AN EFFECTIVE NUMERICAL IDENTIFICATION MEASURE

In this section, we give the detailed description of the simulation of SIR model, propose the numerical identification measure of the epidemic threshold, and make deep analysis of the effectiveness of the numerical measure. In the SIR model, each node can be one of three states which are the susceptible state, infected state, and recovered state, respectively. At the beginning, one node is randomly selected as the initial infected (i.e., seed), and all other nodes are susceptible. At each time step t , each susceptible node i becomes infected with probability $1 - (1 - \beta)^{n_i}$ if it has one or more infected neighbors, where n_i is the number of its infected neighbors. At the same time, all infected nodes recover (or die) at rate μ and the recovered nodes acquire permanent immunity. Time increases by $\Delta t = 1$, and the dynamical process terminates when all infected nodes are recovered. In this paper, μ is set to 1, unless otherwise specified.

A. Proposing the numerical identification measure

The susceptibility measure can not only identify an effective SIS epidemic threshold,¹⁸ but can also be used to determine the critical point of the percolation process.²⁶ Since the connection between SIR and bond percolation is made through the assimilation of the size of the percolating giant component with the final number of recovered individuals,²¹ we check the effectiveness of susceptibility measure χ for the SIR model on RRN, where all nodes have exactly the same degree k . On these networks, the HMF prediction $\lambda_c^{HMF} = 1/(k - 1)$ (Ref. 5) is accurate for the SIR model.

We compare the HMF prediction with the numerical threshold identified by the susceptibility measure λ_p^χ in Fig. 1(a), where the result shows that the numerical threshold of the SIR model identified by χ is significantly larger than $1/(k-1)$.

Considering that the fluctuation of the outbreak size is large near the epidemic threshold, we try to identify the epidemic threshold by the variability measure^{24,25}

$$\Delta = \frac{\sqrt{\langle \rho^2 \rangle - \langle \rho \rangle^2}}{\langle \rho \rangle}, \quad (5)$$

which is a standard measure to determine the critical point in equilibrium phase on magnetic system.²⁷ The inset of Fig. 1(a) shows that the variability Δ exhibits a peak over a wide range of λ , so we estimate the epidemic threshold from the position of the peak of the variability λ_p^Δ . On RRN with different values of k , we find that λ_p^Δ is always consistent with the HMF prediction. When the degree k is given, we further consider the relationship between the epidemic threshold and the network size N in Fig. 1(b), where the numerical thresholds λ_p^χ and λ_p^Δ do not change with N . λ_p^Δ is very close to the HMF prediction, while there is an obvious gap between λ_p^χ and HMF prediction. From the above, we know that the variability Δ can identify an effective SIR epidemic threshold, while the epidemic threshold identified by the susceptibility χ is overestimated on RRN. Thus, a new problem has arisen: Why the variability Δ performs well but the susceptibility χ goes awry for the SIR model?

FIG. 1. Comparison of theoretical thresholds with numerical thresholds on RRN. (a) The threshold λ_c vs. k , where N is set to 10^4 . (b) The threshold λ_c vs. N , where k is set to 10. “Squares,” “circles,” and “triangles” denote λ_p^χ , λ_p^Δ , and $1/(k-1)$, respectively. Inset: Susceptibility χ and variability Δ as a function of λ . The results are averaged over $10^2 \times 10^6$ independent realizations on 10^2 networks.

B. Analysis of the effectiveness of numerical identification measure

Next, we make an analysis of the effectiveness of the numerical identification measures above, by investigating the distribution of outbreak sizes which has the strong heterogeneity near the epidemic threshold. On RRN with $k = 10$, where the epidemic threshold $\lambda_c = 1/(k-1) = 1/9$, Fig. 2(a) shows the distribution of outbreak sizes near λ_c . The outbreak sizes follow approximately an exponential distribution at $\lambda = 0.1$ that is smaller than λ_c . At $\lambda = \lambda_c$, the outbreak sizes follow a power-law distribution $P(\rho) \sim \rho^\alpha$ with a cutoff at some values, where $\alpha \simeq -1.5$.^{28–30} Since the disease may die out quickly or infect a subset of nodes when $\lambda > \lambda_c$, the distribution of outbreak sizes is bimodal,^{31,32} with two peaks occurring at $\rho = 1/N$ and $\rho \simeq 0.2$ for $\lambda = 0.12$, respectively.

FIG. 2. Analysis of effectiveness and robustness of numerical methods on RRN. (a) Numerical distribution of outbreak sizes in SIR model for $\lambda = 0.10$ (circles), $\lambda = 1/9$ (triangles), and $\lambda = 0.12$ (squares), where blue solid, red short dash, and black dot lines represent the theoretical distributions given by Eq. (A4). (b) χ vs. λ , where only the small outbreak sizes with $\rho \leq r_c$ are considered to numerically count χ . (c) Δ vs. λ , where the lump is assumed to be located at r_c in the theoretical distribution diagram when $\lambda > \lambda_c$. In (b) and (c), “triangles,” “circles,” and “diamonds” denote cutoff values $r_c = 0.05, 0.2$ and 0.4 , respectively. The parameters are chosen as $N = 10^4$ and $k = 10$. The results are averaged over $10^2 \times 10^6$ independent realizations on 10^2 networks.

Moreover, the theoretical distribution of outbreak sizes (see Appendix for details) is compared with the results obtained by numerical simulations in Fig. 2(a), where the theoretical probability from Eq. (A4) is consistent with the numerical results for relatively small outbreak sizes ($\rho < 0.05$). At the epidemic threshold, the theoretical results also show that the outbreak sizes obey a power-law distribution with the exponent of about -1.5 . When $\lambda > \lambda_c$, some large outbreak sizes constitute a lump in the numerical scattergram, but the probability of large outbreak sizes cannot be solved from Eq. (A4). We thus speculate that the non-ignorable lump may affect the numerical identification of SIR epidemic threshold for the susceptibility measure.

To verify the speculation, Fig. 2(b) investigates the effectiveness of the susceptibility measures under some cutoff hypotheses. We set the cutoff value of the outbreak size as r_c , which means that only the outbreak sizes with $\rho \leq r_c$ are used to numerically count the susceptibility χ under the cutoff hypothesis. Three kinds of r_c are considered, where $r_c = 0.05$ corresponds to the maximum value of small outbreak size before the lump appears in the numerical scattergram, $r_c = 0.2$ means that the numerical scattergram consists of a part of the lump, and $r_c = 0.4$ means that there is a complete lump in the numerical scattergram. As shown in Fig. 2(b), the susceptibility measure can indeed give a quite effective estimate of the SIR epidemic threshold when the whole lump is ignored (i.e., $r_c = 0.05$). With the increase of r_c , the position of peak value of susceptibility χ gradually shifts to the right for large outbreak sizes are considered. This shows that the susceptibility χ loses its effectiveness in identifying the SIR epidemic threshold due to the existence of the lump.

In Fig. 2(c), the robustness of the variability Δ is further checked in theory. As the numerical distribution of the large outbreak sizes is concentrated, we assume that there is a lump located at r_c with $P(r_c) = 1 - \sum_{\rho < r_c} P(\rho)$ in the theoretical probability distribution diagram of outbreak sizes while $\lambda > \lambda_c$. Based on such theoretical distribution, we plot the variability measure as a function of λ for different values of r_c in Fig. 2(c). Since the variability Δ measures the heterogeneity of the outbreak size distribution, which is strongest at the epidemic threshold,^{28–30} the peak position of the variability measure does not change with the position of the lump.

From the above analysis, we can conclude that the variability Δ is effective in identifying the epidemic threshold of SIR model, while the bimodal distribution of outbreak sizes for $\lambda > \lambda_c$ leads to the overestimation of the SIR epidemic threshold when using the susceptibility χ .

III. VERIFICATION OF THE THEORETICAL PREDICTIONS ON SCALE-FREE AND REAL NETWORKS

We further verify the theoretical predictions on scale-free and real networks, by comparing them with the numerical thresholds from the variability Δ .

A. Scale-free networks

We build scale-free networks (SFN) with degree distribution $P(k) \sim k^{-\gamma}$ based on the configuration model.⁹ The

so-called structural cutoff $k_{\max} \sim N^{1/2}$ and natural cutoff $k_{\max} \sim N^{1/\gamma-1}$ (Ref. 33) are considered to constrain the maximum possible degree $k_{\max}$ on SFN. Differently, the degree-degree correlations vanish on scale-free networks with structural off, while the disassortative degree-degree correlations exist when $\gamma < 3$ for scale-free networks with natural cutoff,³³ because high degree vertices connect preferably to low degree ones in this case. We consider the SIR model on SFN with structural cutoff in Figs. 3(a) and 3(c), where the SIR epidemic threshold increases monotonically with the degree exponent γ , and the variation of epidemic threshold with network size N is approximately linear in logarithmic relationship.¹⁶ The HMF prediction λ_c^{HMF} is very close to the numerical threshold λ_p^Δ , while there is an obvious difference between the QMF prediction λ_c^{QMF} and λ_p^Δ .

The SFN with natural cutoff are considered in Figs. 3(b) and 3(d), where the variations of epidemic threshold with γ and N are similar to the results on SFN with structural cutoff. When $\gamma > 3$, the HMF prediction is close to the numerical threshold, while there is a gap between the QMF prediction and the numerical threshold. Since the disassortative degree-degree correlations exist for $\gamma < 3$, there is a slight difference between λ_c^{HMF} and λ_p^Δ . In Fig. 3(d), the distinction between λ_c^{HMF} and λ_p^Δ becomes large with the increase of N for $\gamma = 2.25$, while in such case the QMF prediction is always close to the numerical threshold since the principle eigenvector is delocalized when $2 < \gamma \leq 5/2$.³⁴ It can be seen from the above analysis that the numerical method presented provides quantitative indexes for the observations in Ref. 22, where the HMF theory is relatively accurate for the SIR model.

B. Real-world networks

To check the performance of the variability Δ on real-world networks,^{35–42} Fig. 4 depicts Δ as a function of λ for Hamsterster full network,³⁵ which contains friendships and family links between users of the website hamsterster.com, and Facebook (NIPS) network,⁴¹ which contains Facebook user-user friendships. The numerical results intuitively show that the variability Δ always reaches a maximum value near the value of λ above which the outbreak size ρ becomes finite. The theoretical predictions of the HMF theory and of the QMF theory are quite close to the numerical threshold identified by Δ on Hamsterster full network, but they become poor on Facebook (NIPS) network. Considering the difference in the behaviors of the theoretical predictions on the two networks above, the detailed comparisons between the numerical and theoretical thresholds on other real networks are presented in Table I. The results indicate that although the HMF prediction and the numerical threshold $\lambda_p^\Delta(\text{SIR})$ are nearly the same for assortative networks, there is an obvious difference between them for the networks showing significant disassortative mixing. The QMF prediction is relatively worse than the HMF prediction for assortative networks, but the former is close to $\lambda_p^\Delta(\text{SIR})$ for some disassortative networks (e.g., Router views,³⁷ CAIDI,³⁷ and email contacts⁴²). These findings numerically show the accuracies of existing theoretical predictions on real-world networks.

FIG. 3. Comparison of theoretical thresholds with numerical thresholds on SFN. The threshold λ_c vs. γ on SFN with structural cutoff (a) and natural cutoff (b), where N is set to 10^4 . λ_c vs. N on SFN with structural cutoff (c) and natural cutoff (d), where solid and empty symbols denote $\gamma = 2.25$ and 3.50 , respectively. “Squares,” “circles,” and “triangles” denote λ_c^{QMF} , λ_c^{HMF} , and λ_c^A , respectively. The results are averaged over $10^2 \times 10^6$ independent realizations on 10^2 networks.

IV. CONCLUSION AND DISCUSSION

In summary, we have studied the numerical identification of SIR epidemic threshold on complex networks with finite size. We have checked the effectiveness of the

susceptibility measure for SIR on RRN, where the HMF is exact. The results showed an obvious gap between the numerical threshold identified by the susceptibility χ and the HMF prediction. Then, we proposed the numerical identification method by analyzing the peak of the epidemic variability, and found that the numerical threshold identified by the variability measure Δ agrees very well with the HMF prediction on RRN.

In order to get a deep understanding of the effectiveness of the two numerical measures above, we have analyzed the distribution of outbreak sizes near the epidemic threshold λ_c . The outbreak sizes follow approximately an exponential distribution when $\lambda < \lambda_c$. At the epidemic threshold, the outbreak sizes follow a power-law distribution with the exponent of about -1.5 . When $\lambda > \lambda_c$, the numerical distribution of outbreak sizes is bimodal with two peaks occurring at $\rho = 1/N$ and $O(1)$, respectively. The probability of small outbreak sizes in theory is consistent with that obtained by numerical simulations, but the probability of large outbreak sizes that constitute a lump in the numerical scattergram cannot be obtained theoretically. Based on the analysis of the cutoff hypothesis of the outbreak size distribution, we found that the susceptibility measure can give a fairly effective SIR epidemic threshold when the lump is ignored. Since the variability measure reflects the heterogeneity of the outbreak size distribution, it is always effective in identifying the epidemic threshold, where the distribution of outbreak sizes has the very strong heterogeneity.

We further employed the variability measure to verify the theoretical predictions on scale-free and real networks. The HMF prediction is close to the numerical threshold on most of the networks, but on SFN with natural cutoff and degree exponent $\gamma < 5/2$, it becomes poor due to the

FIG. 4. Variability Δ and outbreak size ρ as a function of λ on Hamsterster full network (a) and Facebook (NIPS) network (b). “Triangleup” and “triangledown” denote $\lambda_c^{HMF} = \langle k \rangle / [\langle k^2 \rangle - \langle k \rangle]$ and $\lambda_c^{QMF} = 1/\Lambda_N$, respectively. The variability Δ for each λ is normalized by the maximal variability Δ_{max} . The results are averaged over 10^6 independent realizations on each network.

TABLE I. Characteristics and epidemic thresholds of real-world networks. N is the network size, k_{max} is the maximum degree, r is the correlation coefficient of the degrees, λ_c^{HMF} and λ_c^{QMF} are the HMF and QMF predictions for SIR, respectively, $\lambda_p^\Delta(\text{SIR})$ denotes the numerical threshold of SIR identified by Δ , and $\lambda_p^\Delta(\text{SIS})$ and $\lambda_p^\lambda(\text{SIS})$ represent the numerical thresholds of SIS identified by Δ and λ , respectively.

Network	N	k_{max}	r	λ_c^{HMF}	λ_c^{QMF}	$\lambda_p^\Delta(\text{SIR})$	$\lambda_p^\Delta(\text{SIS})$	$\lambda_p^\lambda(\text{SIS})$
Hamsterster full ³⁵	2000	273	0.023	0.023	0.020	0.023	0.025	0.025
Brightkite ³⁶	56739	1134	0.010	0.016	0.010	0.014	0.012	0.012
arXiv astro-ph ³⁷	17903	504	0.201	0.015	0.011	0.012	0.012	0.012
Pretty good privacy ³⁸	10680	206	0.239	0.056	0.024	0.053	0.033	0.033
US power grid ³⁹	4941	19	0.003	0.348	0.134	0.446	0.261	0.264
Euroroad ⁴⁰	1039	10	0.090	0.479	0.249	0.498	0.331	0.331
Facebook (NIPS) ⁴¹	2888	769	-0.668	0.004	0.036	0.075	0.079	0.077
Route views ³⁷	6474	1458	-0.182	0.006	0.022	0.037	0.034	0.036
CAIDA ³⁷	26475	2628	-0.195	0.004	0.014	0.019	0.019	0.019
email contacts ⁴²	12625	576	-0.387	0.009	0.02	0.027	0.024	0.025

existence of disassortative mixing. Similarly, the HMF prediction agrees well with the numerical method on real networks with assortative mixing, while it becomes very poor for disassortative networks, where the QMF prediction is relatively close to the numerical threshold. These findings provide quantitative indexes for the accuracies of existing theoretical predictions from the perspective of simulation.

As part of the discussion, we have considered the epidemic threshold for $\mu < 1$ in Fig. 5. The results on RRN and SFN all show that the numerical thresholds for $\mu = 0.1$ are a little larger than those for $\mu = 0.5$. As shown in the inset, $\mu \rightarrow 0$ leads to an epidemic threshold close to Eq. (4), while $\mu \rightarrow 1$ leads to an epidemic threshold close to Eq. (3). It should be pointed out that the numerical threshold of SIR model is inclined to the theoretical prediction $\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle + (\mu - 2)\langle k \rangle}$ from the edge-based compartmental theory^{43,44} when $0 < \mu < 1$. These findings could be complementary to some existing results.⁴

Moreover, we have tried applying the variability measure to the identification of the SIS epidemic threshold. As shown in Fig. 6 and Table I, the numerical threshold λ_p^Δ from the variability measure agrees very well with the threshold λ_p^λ identified by the susceptibility measure, whose validity for the SIS model has been confirmed in Ref. 18. This shows that the variability measure can also provide an effective estimate of the SIS epidemic threshold.

We have put forward a numerical method for identifying the epidemic threshold for SIR model, which is also suitable for the SIS model. This method can effectively identify epidemic thresholds on various networks, and could be extended to other dynamical processes such as information diffusion and behavior spreading. Further work should be done to check the effectiveness of this method on more complicated networks (e.g., temporal networks⁴⁵ and multilayer networks⁴⁶). Besides, the accurate analytic approximation of the epidemic threshold for general networks remains an

FIG. 5. Comparison of theoretical thresholds with numerical thresholds for $\mu < 1$. The threshold λ_c vs. k for $\mu = 0.1$ (a) and $\mu = 0.5$ (b) on RRN. The threshold λ_c vs. γ for $\mu = 0.1$ (c) and $\mu = 0.5$ (d) on SFN, where solid and empty symbols denote SFN with structural cutoff $k_{max} \sim N^{1/2}$ and natural cutoff $k_{max} \sim N^{1/(\gamma-1)}$, respectively. “Circles” and “squares” denote λ_c^{HMF} and λ_p^Δ , respectively. Inset: λ_c as a function of μ on RRN with $N = 10^4$ and $k = 6$. The results are averaged over $10^2 \times 10^6$ independent realizations on 10^2 networks.

FIG. 6. Comparison of theoretical thresholds with numerical thresholds for SIS model on RRN. (a) The threshold λ_c vs. k , where N is set to 10^4 . (b) The threshold λ_c vs. N , where k is set to 10. “Squares” and “circles” denote λ_p^Z and λ_p^A , respectively. The results are averaged over $10^2 \times 10^6$ independent realizations on 10^2 networks.

important problem. This work helps to verify theoretical analysis of epidemic threshold and would promote further studies on phase transition of epidemic dynamics.

ACKNOWLEDGMENTS

This work was partially supported by National Natural Science Foundation of China (Grant Nos. 11105025 and 91324002), China Postdoctoral Science Special Foundation (Grant No. 2012T50711), the Program of Outstanding Ph. D. Candidate in Academic Research by UESTC (Grant No. YXBSZC20131033), and Open Foundation of State key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications) (SKLNST-2013-1-18). Y. Do was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2013R1A1A2010067).

APPENDIX: THE THEORETICAL DISTRIBUTION OF SMALL OUTBREAK SIZES

For the case of the SIR model and similar models with no steady-state, the static properties (e.g., the final outbreak size and the epidemic threshold) of the epidemic outbreak can be mapped into a suitable bond percolation problem. In this framework, the distribution of occupied cluster sizes is related to the distribution of outbreak sizes. To get the distribution of small outbreak sizes in the SIR model with a fixed value of λ when recovery rate $\mu = 1$, we will present the derivation of the distribution of small occupied cluster sizes in bond percolation with bond occupation probability λ .²¹

After the percolation process on a general network with arbitrary degree distribution p_k , the average degree of the occupied network A_1 , which composes of vertices and occupied edges, is $\langle k_T \rangle = \lambda \langle k \rangle$, where $\langle k \rangle$ is the average degree of the original network A_0 . And, the size distribution of the small subgraphs of network A_1 is

$$\pi_s = \frac{\langle k_T \rangle}{(s-1)!} \left[\frac{d^{s-2}}{dz^{s-2}} [g_1(z)]^s \right]_{z=0}, \quad (\text{A1})$$

where s is the small subgraphs size and $g_1(z)$ is the generating function of the excess degree of network A_1 . In addition, the generating function of degree distribution of A_1 is

$$g_0(z) = \sum_{k=0}^{\infty} p_k (1 - \lambda + z\lambda)^k,$$

and we thus have

$$g_1(z) = \frac{g'_0(z)}{g'_0(1)}.$$

In a random regular network, which has an unique degree k with $p_k = 1$, we can easily obtain that

$$g_0(z) = [1 + (z-1)\lambda]^k, \quad (\text{A2})$$

and

$$g_1(z) = [1 + (z-1)\lambda]^{k-1}. \quad (\text{A3})$$

Substituting Eq. (A3) into Eq. (A1), we can obtain the distribution of small outbreak sizes

$$\pi_s = \frac{k\Gamma(a_2)}{\Gamma(a_0)\Gamma(a_1)} \lambda^{s-1} (1-\lambda)^{s(k-1)-(s-2)}, \quad (\text{A4})$$

where $\Gamma(x+1) = x!$, $a_0 = (s-2)$, $a_1 = s(k-1) - (s-1)$, and $a_2 = s(k-1) - 1$.

¹A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, New York, 2008).

²A. Vespignani, *Nat. Phys.* **8**, 32 (2012).

³R. M. Anderson and R. M. May, *Infectious Diseases in Humans* (Oxford University Press, Oxford, 1992).

⁴R. Pastor-Satorras, C. Castellano, P. V. Mieghem, and A. Vespignani, e-print [arXiv:1408.2701v1](https://arxiv.org/abs/1408.2701v1) (2014).

⁵S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Rev. Mod. Phys.* **80**, 1275 (2008).

⁶R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E* **63**, 066117 (2001).

⁷R. Pastor-Satorras and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).

⁸M. Boguñá and R. Pastor-Satorras, *Phys. Rev. E* **66**, 047104 (2002).

⁹M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).

¹⁰O. Givan, N. Schwartz, A. Cygelberg, and L. Stone, *J. Theor. Biol.* **288**, 21 (2011).

¹¹D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, *ACM Trans. Inf. Syst. Secur.* **10**, 1 (2008).

¹²P. Van Mieghem, J. Omic, and R. Kooij, *IEEE/ACM Trans. Network* **17**, 1 (2009).

¹³S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, *Europhys. Lett.* **89**, 38009 (2010).

¹⁴J. P. Gleeson, *Phys. Rev. Lett.* **107**, 068701 (2011).

¹⁵J. Marro and R. Dickman, *Nonequilibrium Phase Transitions in Lattice Models* (Cambridge University Press, Cambridge, 1999).

¹⁶K. Binder and D. W. Heermann, *Monte Carlo Simulation in Statistical Physics*, 5th ed. (Springer-Verlag, Berlin, 2010).

- ¹⁷M. Boguñá, C. Castellano, and R. Pastor-Satorras, *Phys. Rev. Lett.* **111**, 068701 (2013).
- ¹⁸S. C. Ferreira, C. Castellano, and R. Pastor-Satorras, *Phys. Rev. E* **86**, 041125 (2012).
- ¹⁹M. Barthélemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, *Phys. Rev. Lett.* **92**, 178701 (2004).
- ²⁰R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- ²¹M. E. J. Newman, *Phys. Rev. E* **66**, 016128 (2002).
- ²²C. Castellano and R. Pastor-Satorras, *Phys. Rev. Lett.* **105**, 218701 (2010).
- ²³C. Castellano and R. Pastor-Satorras, *Sci. Rep.* **2**, 371 (2012).
- ²⁴P. Crépey, F. P. Alvarez, and M. Barthélemy, *Phys. Rev. E* **73**, 046131 (2006).
- ²⁵P. Shu, M. Tang, K. Gong, and Y. Liu, *Chaos* **22**, 043124 (2012).
- ²⁶N. Bastas, K. Kosmidis, and P. Argyrakis, *Phys. Rev. E* **84**, 066112 (2011).
- ²⁷S. C. Ferreira, R. S. Ferreira, C. Castellano, and R. Pastor-Satorras, *Phys. Rev. E* **84**, 066102 (2011).
- ²⁸E. Ben-Naim and P. L. Krapivsky, *Phys. Rev. E* **69**, 050901(R) (2004).
- ²⁹E. Ben-Naim and P. L. Krapivsky, *Eur. Phys. J. B* **85**, 145 (2012).
- ³⁰D. A. Kessler and N. M. Shnerb, *Phys. Rev. E* **76**, 010901(R) (2007).
- ³¹D. H. Zanette, *Phys. Rev. E* **64**, 050901(R) (2001).
- ³²A. Khalique and P. Sen, *J. Phys. A: Math. Theor.* **46**, 095007 (2013).
- ³³M. Boguñá, R. Pastor-Satorras, and A. Vespignani, *Eur. Phys. J. B* **38**, 205 (2004).
- ³⁴A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, and J. F. F. Mendes, *Phys. Rev. Lett.* **109**, 128702 (2012).
- ³⁵J. Kunegis, see <http://konect.uni-koblenz.de/networks/petster-hamster> for Hamsterster full (accessed 5th March 2014).
- ³⁶E. Cho, S. A. Myers, and J. Leskovec, in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Diego, California, 2011), pp. 1082–1090.
- ³⁷J. Leskovec, J. Kleinberg, and C. Faloutsos, *ACM Trans. Knowl. Discovery Data* **1**, 1 (2007).
- ³⁸M. Boguñá, R. Pastor-Satorras, A. Diaz-Guilera, and A. Arenas, *Phys. Rev. E* **70**, 056122 (2004).
- ³⁹D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
- ⁴⁰L. Šubelj and M. Bajec, *Eur. Phys. J. B* **81**, 353 (2011).
- ⁴¹J. McAuley and J. Leskovec, in *Advances in Neural Information Processing Systems* (Lake Tahoe, Nevada, 2012), pp. 539–547.
- ⁴²M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse, *Nat. Phys.* **6**, 888 (2010).
- ⁴³J. C. Miller, *J. Math. Biol.* **62**, 349 (2011).
- ⁴⁴W. Wang, M. Tang, H. F. Zhang, H. Gao, Y. Do, and Z. H. Liu, *Phys. Rev. E* **90**, 042803 (2014).
- ⁴⁵P. Holme and J. Saramäki, *Phys. Rep.* **519**, 97 (2012).
- ⁴⁶S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadai, Z. Wang, and M. Zanin, *Phys. Rep.* **544**, 1 (2014).