

Spreading Predictability in Complex Networks

Na Zhao

Yunnan University

Jian Wang

Kunming University of Science and Technology

Yong Yu

Yunnan University

Junyan Zhao

Beijing Special Vehicle Institute

Duanbing Chen (✉ dbchen@uestc.edu.cn)

University of Electronic Science and Technology of China

Research Article

Keywords: predicting infection scale, infectious diseases, synthetic and real networks

Posted Date: January 19th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-132928/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on July 12th, 2021. See the published version at <https://doi.org/10.1038/s41598-021-93611-z>.

Spreading predictability in complex networks

Na Zhao^{2,3,*}, Jian Wang^{4,*}, Yong Yu^{2,*}, Jun-Yan Zhao⁵, and Duan-Bing Chen^{1,6,7,**}

¹School of Computer Science & Engineering, University of Electronic Science and Technology of China, Chengdu 611731, P.R. China

²Key Laboratory in Software Engineering of Yunnan Province, School of Software, Yunnan University, Kunming 650504, P.R. China

³Electric Power Research Institute of Yunnan Power Grid, Kunming, 650200, P.R. China

⁴College of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650217, P.R. China

⁵Beijing Special Vehicle Institute, Beijing 100072, P.R. China

⁶The Research Base of Digital Culture and Media, Sichuan Provincial Key Research Base of Social Science, Chengdu 611731, P.R. China

⁷Union Big Data, Chengdu 610041, P.R. China

*These authors contributed equally to this work.

**Correspondence should be addressed to dbchen@uestc.edu.cn

ABSTRACT

Many state-of-the-art researches focus on predicting infection scale or threshold in infectious diseases or rumor and give the vaccination strategies correspondingly. In these works, most of them assume that the infected probability and initially infected individuals are known at the very beginning. Generally, infectious diseases or rumor has been spreading for some time when it is noticed. How to predict which individuals will be infected in the future only by knowing the current snapshot becomes a key issue in infectious diseases or rumor control. In this paper, a prediction model based on snapshot is presented to predict the potentially infected individuals in the future. Experimental results on synthetic and real networks demonstrate that the predicted infected individuals have rather consistency with the actual infected ones.

Introduction

Spreading dynamics is an important issue in spread and control^{1,2} of rumor^{3,4} and disease⁵⁻⁸, marketing⁹, recommending¹⁰⁻¹², source detecting^{13,14}, and many other interesting topics¹⁵⁻¹⁸. How to predict the infection rate¹⁹, infected scale^{20,21}, and even the infected nodes precisely has been gotten much attention in recent years.

Researchers have gotten many achievements on macro level of spread such as phase transition of spread²² and basic reproduction number²³. Up to now, many researches focus on estimating of infection scale. The simplest one is mean-field model, in which, the spread coverage can be predicted by using differential equations²⁰. Besides mean-field model, some more realistic models such as pair approximation²¹ and permutation entropy²⁴ are considered to predict the infection scale or infectious disease outbreaks. The main difference between mean-field and pair approximation is that the former(latter) approximates high-order moments in term of first (second) order ones. Researchers studied the predictability of a diverse collection of outbreaks and identified a fundamental entropy barrier for disease time series forecasting through adopting permutation entropy as a model independent measure of predictability²⁴. Funk et al²⁵ presented a stochastic semi-mechanistic model of infectious disease dynamics that was used in real time during the 2013–2016 West African Ebola epidemic to fit the simulated trajectories in the Ebola Forecasting Challenge, and to produce forecasts that were compared to following data points. Zhang et al²⁶ proposed a measurement to state the efforts of users on Twitter to get their information propagation. They found that small fraction of users with special performance on participation can gain great influence, while most other users play a role as middleware during the information propagation.

Up to now, most researches focus on macro level of spreading prediction. Besides analysis on macro level, we also should pay attention to the details of infected individuals so as to contain the spread of serious infectious diseases such as SARS²⁷, H7N7²⁸ and COVID-19²⁹. Chen et al. did some interesting works on this area¹⁹. They presented an iterative algorithm to estimate the infection rate of the spreading process and then to predict the spreading coverage from a given snapshot of the propagation. In this paper, we present a probability based prediction model to estimate the infection probability of a node from a snapshot, further, to determine the potentially infected nodes in the future.

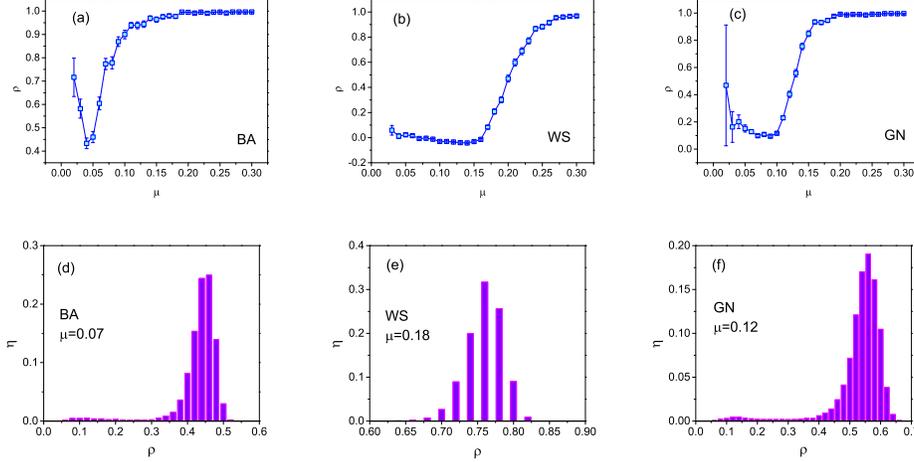


Figure 1. The correlation ρ under different infected rate μ on (a) BA, (b) WS and (c) GN networks. The distribution of correlation in (d) BA, (e) WS and (f) GN are shown. The network parameters are $N = 4000$, $\langle k \rangle = 10$, $p = 0.1$ for WS network, $N = 4000$, $\langle k \rangle = 10$ for BA network, and $N = 4000$, $\langle k \rangle = 10$, $\langle k_{in} \rangle = 7$ for GN network. The error bar in (a-c) and the distribution of correlation ρ in (d-f) are obtained by the results under 200 snapshots

Results and Discussion

We test our method on synthetic and real networks. Synthetic networks are Watts-Strogatz (WS) networks³⁰, Barabási-Albert (BA) networks³¹ and Given-Newman (GN) community networks³². Each synthetic network has 4000 nodes and each GN community network has 40 communities. We will discuss our model on three aspects: (1) the effect of infected rate, (2) the effect of network structure, and (3) the effect of stage of snapshot.

To simulate the spreading process on networks, we employ the Susceptible-Infected-Removed (SIR) model³³. In a network, we randomly select one node as the initial spreader. The information from this node will infect each of its susceptible neighbors with probability μ . After infecting neighbors, the node will immediately become recovered (i.e., the recovering probability is 1). The new infected nodes continue to infect their neighbors in next step. If it is not specially stated, we take the snapshot after five steps of spreading from the initial node as the known information.

The effect of infected rate

Fig. 1 shows the Pearson correlation ρ between the result of averaging on 200 simulations and that of probability prediction model under different infected rate μ on WS, BA and GN networks. Generally, the correlation get larger while μ getting larger. For large μ , e.g., $\mu = 0.3$, the correlation approach to 1 since most of nodes will be infected. From Fig. 1, it can be seen that there exists a transition point, in detail, the transition point at $\mu = 0.15$ for WS network (see Fig. 1(b)) and at $\mu = 0.1$ for GN network (see Fig. 1(c)). This can be explained as follows: the information almost do not diffuse if μ is small ($\mu < 0.15$ for WS networks and $\mu < 0.1$ for GN network), and the infected nodes are highly random for different simulations. It is noted that there hardly exist transition point in BA network. It can be explained as follows: the information will easily reach to the node with large degree regardless the location of initially infected node, eventually, reach to other nodes for its heterogenous structure. Interestingly, if μ is very small (e.g., $\mu = 0.02$), the correlation is getting large in BA network, as shown in Fig. 1(a). Actually, for very small μ , only a few snapshots in 200 simulations can be utilized to analyze the correlation ρ since spread stops in two or three steps in most simulations, the results have no statistical significance. Besides, the distribution of correlation ρ under the results of 200 independent runs are listed in Figs. 1(d-f). From these three subfigures, it can be seen that the distributions of correlation ρ of BA and GN networks are similar, while that of WS network are generally large comparing with BA and GN networks.

The effect of network structure

Fig. 2 shows the correlation for three types of networks with different structural parameters. For WS network, we study the effect of the rewiring parameter p on correlation. For BA network, we consider a variant form of it in which each new node u connects to an existing node v with probability $p_u = (k_u + B) / \sum_v (k_v + B)$ ^{34,35}. This modified model allows a selection of the exponent of the power-law scaling in the degree distribution $p(k) \sim k^{-\gamma}$ with $\gamma = 3 + B/m$ in the thermodynamic limit where m is the number of nodes should be connected when a new node is added and B is tunable parameter. With this network, we

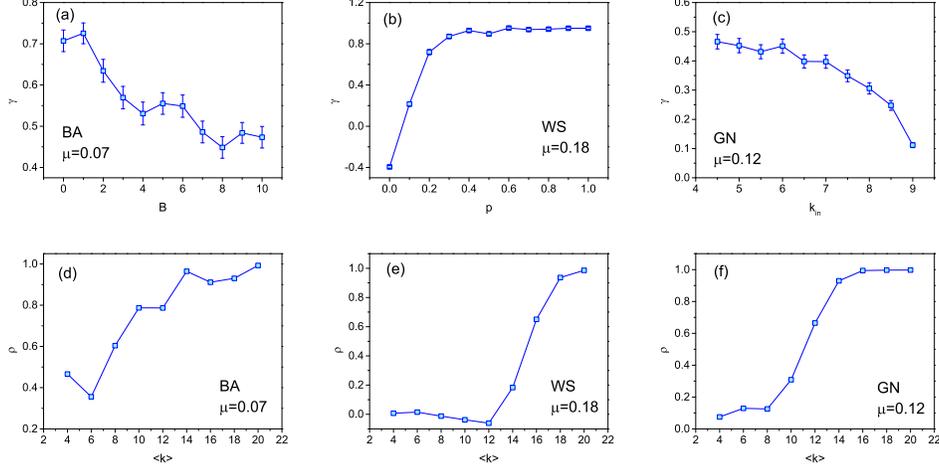


Figure 2. The correlation ρ for three types of networks with different structural parameters. In (a), B is a tunable parameter while generating network, (b) p is the rewiring probability, (c) $\langle k_{in} \rangle$ is the average internal degree, and (d-f) $\langle k \rangle$ is the average degree.

study the effect of B on correlation. For GN network, we study the effect of $\langle k_{in} \rangle$ on correlation, where $\langle k_{in} \rangle$ is the average internal degree of nodes in community. For a node u in community C , its internal degree k_u^{in} can be written as:

$$k_u^{in} = \sum_{u,v} \delta_{u,v}, \quad (1)$$

$\delta_{u,v} = 1$ if v is also in community C , otherwise $\delta_{u,v} = 0$. For standard BA network, i.e., $B = 0$, there are a few nodes with extremely large degree, the information can be spread out easily so long as it reaches to a node with large degree. So, it is relatively easy to predict which node will be infected in the future. As B increasing, the network evolves to random, a node getting infected or not will be hard to predict relatively, so the correlation decreases when B increases, as shown in Fig. 2(a). If rewiring probability $p < 0.2$, the information hardly diffuse to other nodes since the WS network is almost regular, so it is hard to predict the infected nodes. As rewiring probability p getting larger, the network getting more random, the information reaches to other nodes easily, consequently, it is easy to predict the infected nodes, as shown in Fig. 2(b). In GN network, if average internal degree $\langle k_{in} \rangle$ is larger, the community structure is clearer, correspondingly, the information is hard to escape the community boundary, and the correlation will getting worse, as shown in Fig. 2(c).

Besides the network parameter listed above, the density of network, i.e., average node degree $\langle k \rangle$, also affects the correlation, as shown in Fig. 2 (d) to Fig. 2(f). It can be seen that the correlation is small for small average node degree $\langle k \rangle$. Especially in WS and GN networks, for a large scope of average node degree ($\langle k \rangle < 12$ in WS and $\langle k \rangle < 8$ in GN), the correlation is extremely small, there exists an obvious transition points, as shown in Figs. 2(d) and (f).

The effect of stage of snapshot

We further analyze the correlation ρ under different stage of snapshot, as shown in Fig. 3. In Fig. 3, T is the spreading steps of snapshot. Generally, it is difficult to estimate the infected rate precisely if just the snapshot in the early stage is given since there is little usable information, so, it is hard to predict the infected nodes. As T increases, more information could be used, the correlation ρ are getting larger. In the late stage, many nodes of snapshot are infected or recovered, the left nodes are hard to be infected, so the correlation ρ are getting smaller, especially in BA network since most of all nodes are recovered.

Besides synthetic networks, we also analyze the correlation ρ on 11 real networks. The properties and analysis results on these real networks are shown in Table 1. From Table 1, it can be seen that the results are rather good, especial for the case of large number of infected nodes N_I of snapshot. It is noted that the correlation ρ and N_I have strong positive correlation. For networks Y2H and power, the correlation ρ is extremely low since N_I is very small. Actually, in these cases, there are few infected nodes in snapshot. Furthermore, the networks are very sparse, so, it is hard to predict the nodes being infected from snapshot in the future.

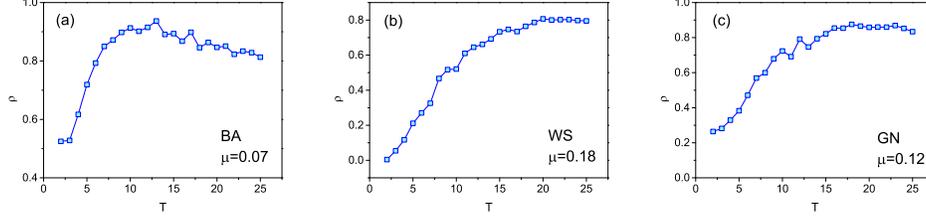


Figure 3. The correlation ρ under different stage of snapshot. Smaller T indicates earlier stage and larger T indicates latter stage.

Table 1. The properties and analyzing results on 11 real networks.

Networks	#Nodes	#Edges	ρ	N_I
cond-mat	39577	175693	0.9430	0.0152
astro-Ph	16046	121251	0.9426	0.0575
email	1133	5451	0.9860	0.0628
c.elegans	453	2025	0.9900	0.1143
ecoli	230	695	0.9558	0.0509
internet	22963	48436	0.9541	0.0625
PGP	10680	24316	0.8074	0.0069
TAP	1373	6833	0.5897	0.0101
HEP	7610	15751	0.5975	0.0016
Y2H	1846	2203	0.3214	0.0016
power	4941	6594	0.2762	0.0003

Methods

For a given snapshot, we use the IAIP¹⁹ method to estimate the infection probability. In IAIP model, we denote the number of infected nodes as N_I , the number of susceptible nodes as N_S and the number of recovered nodes as N_R . $N_S + N_I + N_R = N$ since we use SIR spreading model. The infection rate can be calculated by:

$$\mu = \frac{N_R + N_I}{\sum_{i \in R} (k_i - m_i)}, \quad (2)$$

where m_i is the number of infected nodes (both I and R nodes) among i 's neighbors when i tries to infect other nodes. It is noted that the exact value of m_i cannot be directly extracted from the snapshot. One can estimate m_i by its expected value

$$\bar{m}_i = \frac{\sum_{l=1}^{M_i} l \cdot \mu \cdot (1 - \mu)^{l-1}}{1 - (1 - \mu)^{M_i}}, \quad (3)$$

where M_i is the total number of I and R nodes among i 's neighbors in the observed snapshot.

On the basis of Eq. 2 and Eq. 3, μ and m_i are expected to respectively approach their true values.

For a given snapshot, a node u will be converted into infected one with a probability $P_u(t)$ at time t , we have,

$$P_u(t) = 1 - \prod_{v \in \Gamma_u} (1 - \mu P_v(t-1)), \quad (4)$$

where Γ_u is the neighbors of node u and infected probability μ can be estimated by IAIP model (Iterative Algorithm for estimating the Infection Probability)¹⁹. For node v in Eq. (4), it is reasonable to assume $P_v(t) = 1$ for infected node and $P_v(t) = 0$ for recovered node. Obviously, the initial condition is,

$$P_u(0) = \begin{cases} 0 & \text{if node } u \text{ is susceptible or recovered} \\ 1 & \text{if node } u \text{ is infected} \end{cases}, \quad (5)$$

By solving Eq. (4) under initial condition, the score $P_u(t)$ of node u will be converged to a unique steady state denoted by $P_u(t_c)$, where t_c is the convergence time. The final score $P_u = P_u(t_c)$ is the probability to be infected of susceptible node while spreading achieves steady state.

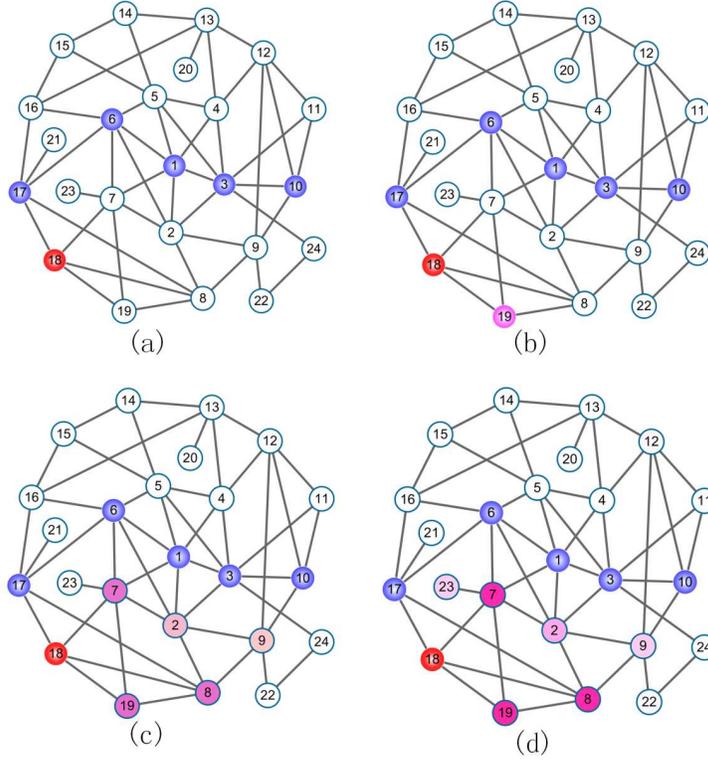


Figure 4. (Color online) A toy network with 24 nodes. (a) The snapshot includes 5 recovered nodes, i.e., 1, 3, 6, 10, 17, and 1 infected node, i.e., node 18, (b) a certain spreading simulation result from snapshot, only node 19 is infected when spreading achieves steady state, (c) average result on 10000 simulations from snapshot, and (d) result of probability prediction model from snapshot. In (c) and (d), the shades of nodes indicate the probability to be infected when spreading achieves steady state.

Fig. 4 is a toy network with 24 nodes. The snapshot includes 5 recovered nodes and 1 infected node, as shown in Fig. 4(a). A certain spreading simulation result, average result on 10000 simulations, and result of probability prediction model from snapshot are shown in Figs. 4(b), (c) and (d) respectively. From this toy network, it can be seen that the result obtained by the probability prediction model is coincident with that by the average over 10000 simulations very well, that is, nodes 7, 8, and 19 have high probability to be infected, nodes 2 and 9 have middle probability to be infected, while other nodes have relatively low probability to be infected, as shown in Fig. 4(c) and (d).

In order to evaluate the performance of the proposed model, we use Pearson correlation ρ between the result of averaging on N simulations and that of probability prediction model, that is:

$$\rho = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N \sum x_i^2 - (\sum x_i)^2} \sqrt{N \sum y_i^2 - (\sum y_i)^2}}, \quad (6)$$

where $\vec{p}_r = (x_1, x_2, \dots, x_N)$ and $\vec{p}_e = (y_1, y_2, \dots, y_N)$ are the vector of infected probability of nodes averaged by N simulations and obtained by probability prediction model respectively.

References

1. Iribarren, J. L. & Moro, E. Impact of human activity patterns on the dynamics of information diffusion. *Phys. Rev. Lett.* **103**, 038702 (2009).
2. Arruda, G. F., Petri, G., Rodrigues, F. A. & Moreno, Y. Impact of human activity patterns on the dynamics of information diffusion. *Phys. Rev. Res.* **2**, 013046 (2020).
3. Zhang, Y., Zhou, S., Zhang, Z., Guan, J. & S, Z. Rumor evolution in social networks. *Phys. Rev. E.* **87**, 032133 (2013).
4. Kwon, S., Cha, M. & K, J. Rumor detection over varying time windows. *PLoS ONE* **12**, e0168344 (2017).
5. Meloni, S. *et al.* Modeling human mobility responses to the large-scale spreading of infectious diseases. *Sci. Rep.* **1**, 62 (2011).
6. Goltsev, A. V., Dorogovtsev, S. N., Oliveira, J. G. & Mendes, J. F. F. Localization and spreading of diseases in complex networks. *Phys. Rev. Lett.* **109**, 128702 (2012).
7. Granell, C., Gómez, S. & Arenas, A. Competing spreading processes on multiplex networks: Awareness and epidemics. *Phys. Rev. E.* **90**, 012808 (2014).
8. Leventhal, G. E., Hill, A. L., Nowak, M. A. & Bonhoeffer, S. Authoritative sources in a hyperlinked environment. *Nat. Commun.* **8**, 61012 (2015).
9. Miquel-Romero, M. J. & Adame-Sánchez, C. Viral marketing through e-mail: the link consumer-company. *Manag. Decis.* **51**, 1970–1982 (2013).
10. Lü, L. *et al.* Recommender systems. *Phys. Rep.* **519**, 1–49 (2013).
11. Ren, X., Lü, L., Liu, R. & Zhang, J. Avoiding congestion in recommender systems. *New J. Phys.* **16**, 063057 (2014).
12. Chen, D. B., Zeng, A., Cimini, G. & Zhang, Y. C. Adaptive social recommendation in a multiple category landscape. *Eur. Phys. J. B* **86**, 61 (2013).
13. Pinto, P. C., Thiran, P. & Vetterli, M. Locating the source of diffusion in large-scale networks. *Phys. Rev. Lett.* **109**, 068702 (2012).
14. Shen, Z., Cao, S., Wang, W. X., Di, Z. & Stanley, H. E. Locating the source of diffusion in complex networks by time reversal backward spreading. *Phys. Rev. E.* **93**, 032301 (2016).
15. Seebens, H., Schwartz, N., Schupp, P. J. & Blasius, B. Similarity measures in scientometric research: The jaccard index versus salton's cosine formula. *Proc. Natl. Acad. Sci. USA.* **113**, 108–115 (2016).
16. Lü, L., Chen, D. B. & Zhou, T. The small world yields the most effective information spreading. *New J Phys.* **13**, 123005 (2011).
17. Cimini, G., Chen, D. B., Medo, M., Lü, L. & Zhang, Y. C. Enhancing topology adaptation in information sharing social networks. *Phys. Rev. E.* **85**, 046108 (2012).
18. Centola, D. The spread of behavior in an online social network experiment. *Sci.* **329**, 1174–1197 (2010).
19. Chen, D. B., Xiao, R. & Zeng, A. Predicting the evolution of spreading on complex networks. *Sci. Rep.* **4**, 6108 (2014).
20. Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200 (2001).
21. Mata, A. S., Ferreira, R. S. & Ferreira, S. C. Heterogeneous pair-approximation for the contact process on complex networks. *New J. Phys.* **16**, 053006 (2014).
22. Döbereiner, H. G., Dubin-Thaler, B., Giannone, G., Xenias, H. S. & M, P. Dynamic phase transitions in cell spreading. *Phys. Rev. Lett.* **93**, 108105 (2004).
23. Rodrigues, H. S., Monteiro, M. T. T., Torres, D. F. M. & Zinober, A. Dengue disease, basic reproduction number and control. *Int. J. Comput. Math.* **89**, 334–346 (2012).
24. Scarpino, S. V. & Petri, G. On the predictability of infectious disease outbreaks. *Nat. Commun.* **10**, 898 (2019).
25. Funk, S., Camacho, A., Kucharski, A. J., Eggo, R. M. & Edmunds, W. J. Real-time forecasting of infectious disease dynamics with a stochastic semi-mechanistic model. *Epidemics* **22**, 56–61 (2018).
26. Zhang, X., Han, D. D., Yang, R. & Zhang, Z. Users participation and social influence during information spreading on twitter. *PLoS One* **12**, e0183290 (2017).
27. Rota, P. A. *et al.* Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Sci.* **300**, 1394–1399 (2003).

28. Fouchier, R. A. M., Schneeberger, P. M., Frans, W., Rozendaal, F. W. & et al. Avian influenza a virus (h7n7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *Proc. Natl. Acad. Sci. USA*. **101**, 1356–1361 (2004).
29. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *Lancet* **395**, 689–697 (2020).
30. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nat.* **393**, 440–442 (1998).
31. Barabási, A. L. & Albert, R. Emergence of scaling in random networks. *Sci.* **286**, 509–512 (1999).
32. Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E*. **69**, 026113 (2004).
33. Anderson, R. M., May, R. M. & Anderson, B. *Infectious diseases of humans: dynamics and control* (Boston: Oxford Univ. Press, 1992).
34. Albert, R. & Barabási, A. L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
35. Dorogovtsev, S. N. & Mendes, J. F. F. Evolution of networks. *Adv. Phys.* **51**, 1079–1187 (2002).

Acknowledgements

This work was partially supported by the National Natural Science Foundation of China with Grant No 61673085, by the Science Strength Promotion Programme of UESTC under Grant No. Y03111023901014006 and by the National Key Research and Development Program of China under Grant No. 2018YFB2100100.

Author contributions statement

N. Z. and D.-B. C. designed the research and prepared all figures. N. Z., J. W. and Y. Y. performed the experiments and analyzed the data. All authors wrote the manuscript.

Additional information

Competing interests: The authors declare no competing interests.

Figures

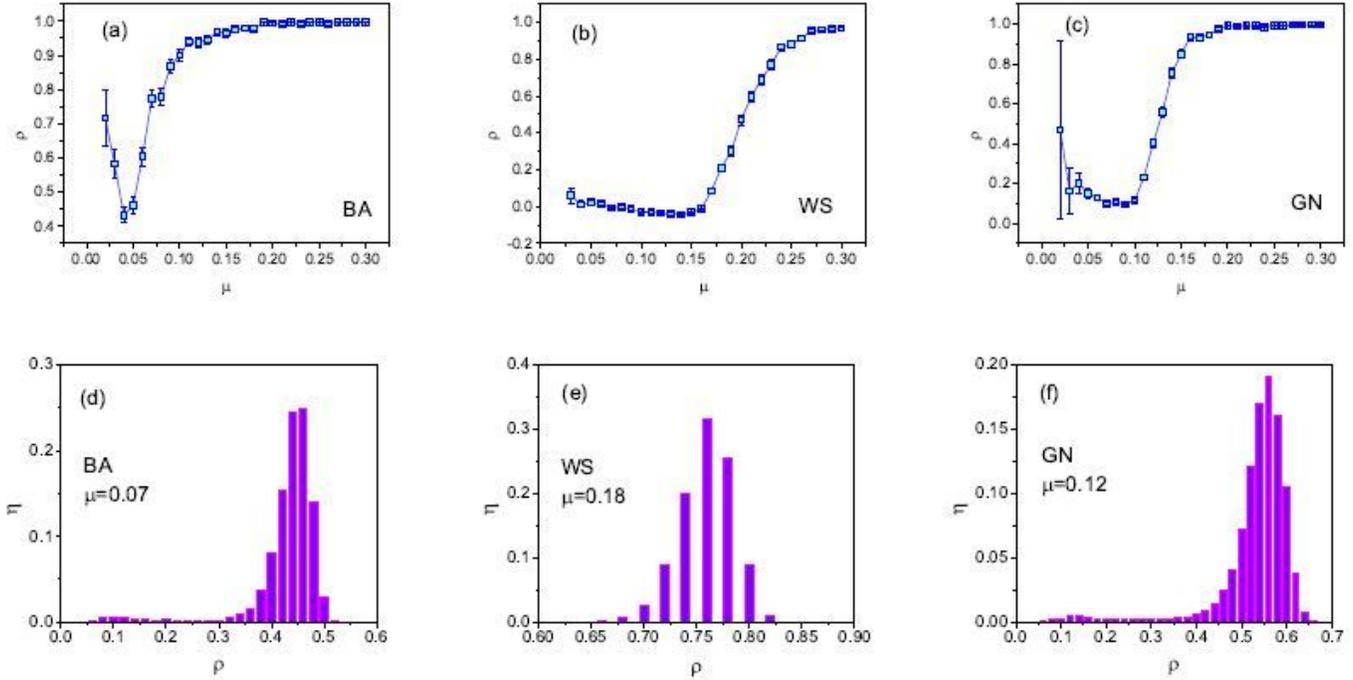


Figure 1

The correlation r under different infected rate μ on (a) BA, (b) WS and (c) GN networks. The distribution of correlation in (d) BA, (e) WS and (f) GN are shown. The network parameters are $N = 4000$; $\langle k \rangle = 10$; $p = 0:1$ for WS network, $N = 4000$; $\langle k \rangle = 10$ for BA network, and $N = 4000$; $\langle k \rangle = 10$; $\langle k_{in} \rangle = 7$ for GN network. The error bar in (a-c) and the distribution of correlation r in (d-f) are obtained by the results under 200 snapshots

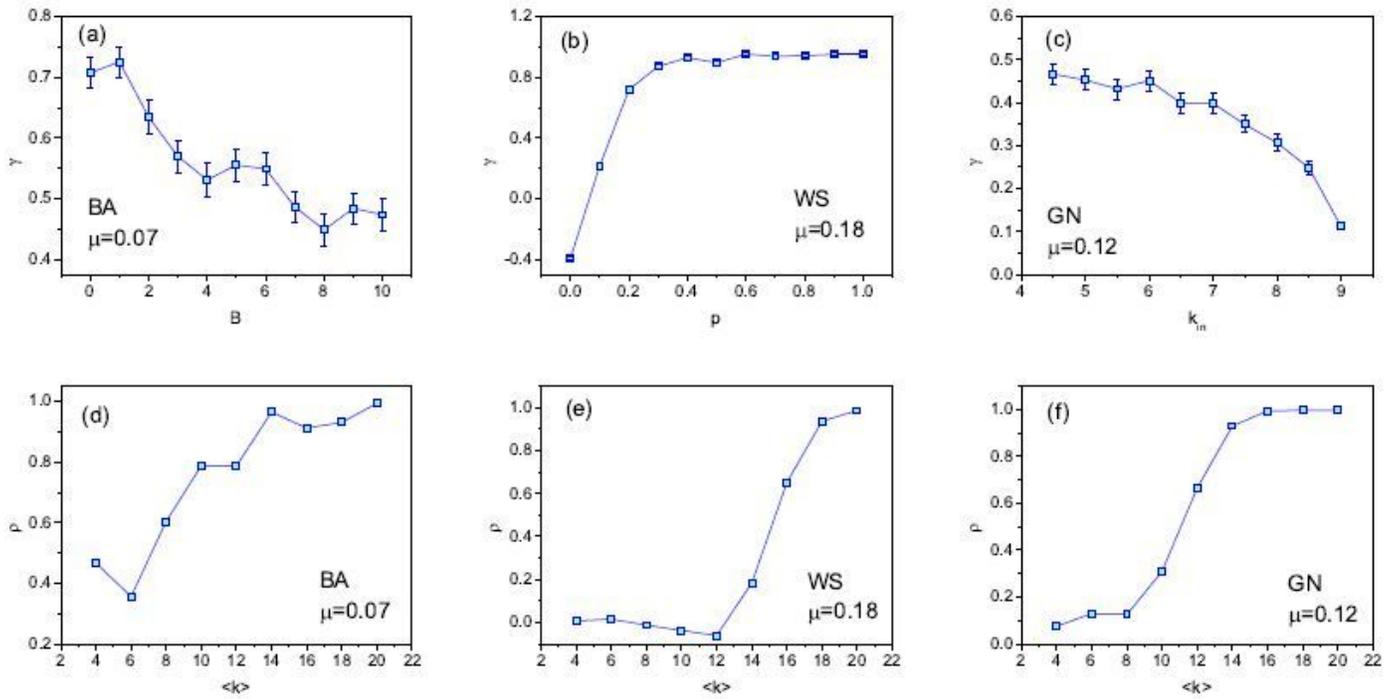


Figure 2

The correlation r for three types of networks with different structural parameters. In (a), B is a tunable parameter while generating network, (b) p is the rewiring probability, (c) $\langle k_{in} \rangle$ is the average internal degree, and (d-f) $\langle k \rangle$ is the average degree.

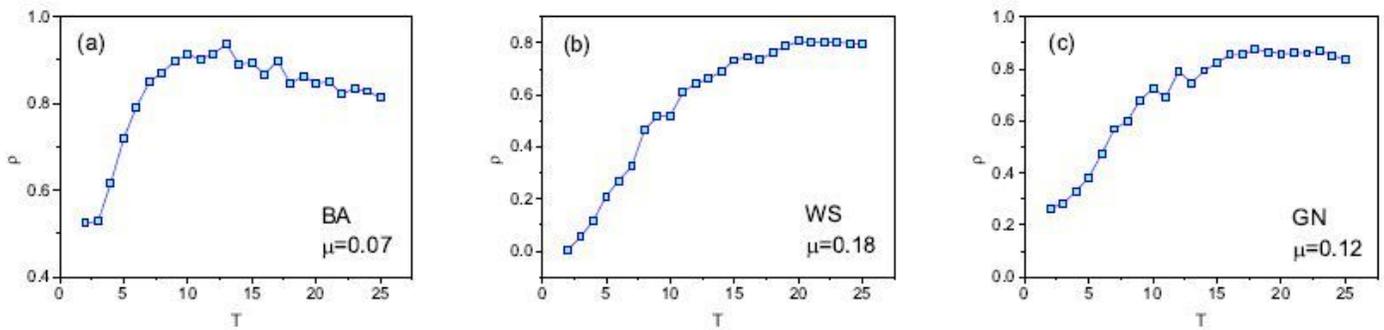


Figure 3

The correlation r under different stage of snapshot. Smaller T indicates earlier stage and larger T indicates latter stage.

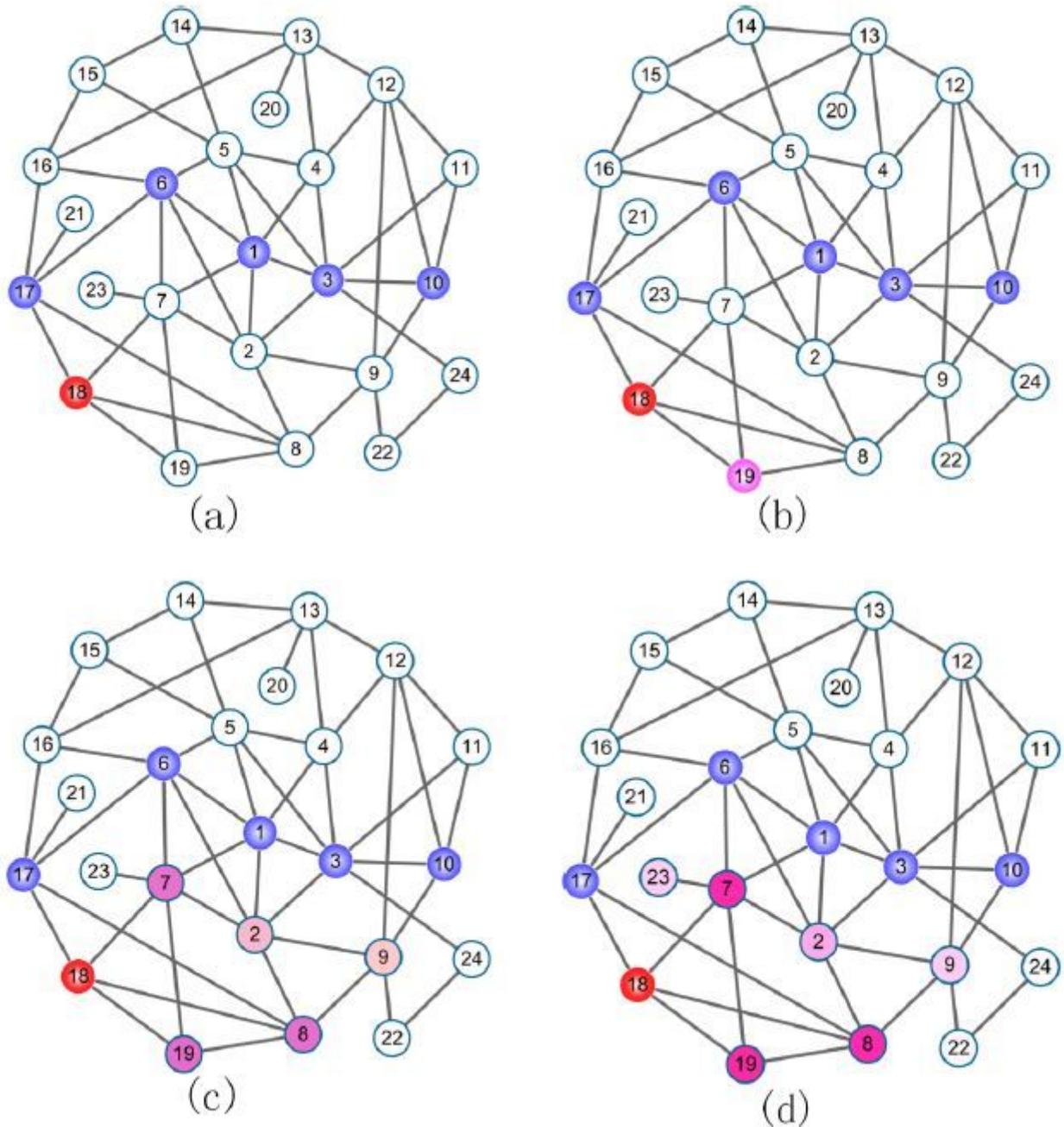


Figure 4

(Color online) A toy network with 24 nodes. (a) The snapshot includes 5 recovered nodes, i.e., 1, 3, 6, 10, 17, and 1 infected node, i.e., node 18, (b) a certain spreading simulation result from snapshot, only node 19 is infected when spreading achieves steady state, (c) average result on 10000 simulations from snapshot, and (d) result of probability prediction model from snapshot. In (c) and (d), the shades of nodes indicate the probability to be infected when spreading achieves steady state.