Epidemic spreading driven by biased random walks

Cunlai Pu, Siyuan Li, Jian Yang

Nanjing University of Science and Technology, Nanjing 210094, China

Abstract

Random walk is one of the basic mechanisms of many network-related applications. In this paper, we study the dynamics of epidemic spreading driven by biased random walks in complex networks. In our epidemic model, infected nodes send out infection packets by biased random walks to their neighbor nodes, and this causes the infection of susceptible nodes that receive the packets. Infected nodes recover from the infection at a constant rate λ, and will not be infected again after recovery. We obtain the largest instantaneous number of infected nodes and the largest number of ever-infected nodes respectively, by tuning the parameter α of the biased random walks. Analytical and simulation results on model and real-world networks show that spread of the epidemic becomes intense and widespread with increase of either delivery capacity of infected nodes, average node degree, or homogeneity of node degree distribution.

Keywords: epidemic spreading, biased random walks, complex networks

1. Introduction

Unexpected epidemic outbreaks in biological systems[1, 2] and the spread of computer viruses in technology systems[3, 4, 5] result in a lot of death or great damage to related systems. The study of epidemiological models has a long history, especially in the field of social science[6, 7]. The SIR (susceptible-infected-removed) model and the SIS (susceptible-infected-susceptible) model are two representative models which capture the basic properties of epidemic spreading by defining some transitions among several

Email address: pucunlai@njust.edu.cn (Cunlai Pu)

disease states[8, 9]. In the SIR model, a susceptible individual becomes infected at some rate when the individual has contact with infected individuals. An infected individual recovers from the disease at a constant rate and is assumed to get a permanent immunity. Therefore, the SIR model always terminates when all infected individuals are recovered. However, in the SIS model there are only two states: susceptible and infected. A recovered individual can get infected again. If the fraction of infected individuals is large enough, the disease will spread indefinitely, otherwise it will die out after a period of time. In the past, epidemic models were discussed under the homogeneous mixing hypothesis[10], which assumes that at each time point an arbitrary individual has an equal opportunity to contact everyone in the population. Recently, results from the area of network science demonstrated that most real-world network systems have heterogeneous topological structures[11, 12, 13], and this caused many mathematicians and physicists to explore epidemic models on heterogeneous random networks[14, 15, 16, 17, 18, 19] by means of mean-field approximation[20, 21, 22], generating functions formalism[23], and percolation theory[24]. It was found that the epidemic threshold is absent for random networks with strongly heterogeneous degree distribution, which means an epidemic always has a finite probability to survive indefinitely[11, 20]. Most recently, much attention has been paid to the study of epidemic spreading in multiplex networks. For example, it was found that the critical point for the spread of a disease in a multiplex network is lower than the critical point in each isolated network[25, 26]. Also, many researchers obtained the interrelation between the spread of an epidemic and the information awareness of the epidemic on multiplex networks[19, 27].

Besides network structures, traffics in networks also have great impacts on epidemic spreading. For instance, in the Internet computer viruses are transmitted from a node to another with data packets. Without transmission of packets, viruses do not spread even if the two nodes are physically connected. In world city networks, air traffics speed the spread of diseases among different spatial areas. The combination of epidemic spreading and traffic dynamics was first considered in the metapopulation model[28] which characterizes the dynamics of systems composed of subpopulations. Then, Meloni et al[29] studied the impact of traffic dynamics on the spread of viruses in the Internet, in which information packets are transmitted with the shortest paths. Later, many mechanisms were proposed to suppress traffic-driven epidemic spreading, for instance controlling the traffic flow[30], the routing strategy[31, 32], the heterogeneous curing rate[33], and deleting some partic-

ular edges[34], etc.

Random walk is a basic mechanism related to many spreading processes[35, 36, 37, 38, 39, 40]. For example, a mobile phone virus may randomly dial some phone numbers from a directory. Some computer viruses propagate randomly by email or other online communication tools. Therefore, the role of random walks in epidemic spreading should be explored. We propose an epidemic model driven by biased random walks. In our model, infected nodes send out a constant number of infection packets to their neighbor nodes through biased random walks. Susceptible nodes are infected after receiving the infection packets, and then are removed from the set of infected nodes with a constant rate. We investigate the spreading properties and the optimal control parameters of our model, as well as the influence of network structures on our model.

2. SIR model

The SIR model is one of the traditional epidemic spreading models in literature. In the SIR model, there are three types of nodes: susceptible nodes, infected nodes and removed nodes. A susceptible node can contract the infection. An infected node was previously a susceptible node that got infected by the disease. A removed node is the one that recovered from the disease, and was removed from the set of infected nodes. Assuming that the numbers of susceptible, infected and removed nodes at time t are S, I and R respectively. There are three basic elements in the SIR model as follows[8]:

(1) $S + I + R = N$. N is the number of nodes in the network.
(2) An arbitrary infected node infects the susceptible nodes by the ratio β. The number of new infected nodes is $\beta * S * I$.
(3) The number of new removed nodes is proportional to the total number of infected nodes I, which is $\lambda * I$.

According to these three elements, the dynamics of the SIR model can be expressed as follows[8]:

$$\begin{cases} \frac{\mathrm{d}I}{\mathrm{d}t} &= \beta SI - \lambda I, \\ \frac{\mathrm{d}S}{\mathrm{d}t} &= -\beta SI, \\ \frac{\mathrm{d}R}{\mathrm{d}t} &= \lambda I. \end{cases} \tag{1}$$

When time t is large enough, all the infected nodes will become removed nodes, and the epidemic spreading stops.

The SIR model is based on the assumption that a node in the network contacts every other node with equal probability. However, in real situations, individuals often have heterogeneous numbers of contacts[11]. A few individuals have a larger number of contacts, and they will get more contacts according to the "rich-get-richer" mechanism, while most of the individuals have a few contacts.

3. Epidemic model driven by biased random walks

In real situation, epidemics are often spreading with other physical quantities on networks. In the Internet, a virus can't spread from a node to another node unless there is a transport of infection packets between the two nodes. Additionally, in city networks even roads within cities are physically well connected, an epidemic can't spread among cities unless infectious individuals travel among the cities. Therefore, based on the SIR model, we further consider the transport of infection packets in networks.

In our model, a randomly selected node is infected initially. In one time step, each of the infected nodes sends C infection packets independently to its neighbor nodes through biased random walks. An infection packet diffuses from one infected node to a neighbor node with a probability that is a function of the neighbor node's connectivity[41, 42], in contrast to the plain random walk where the neighbors of a node have all the same probability to be chosen (Fig.1 (a)). We assume that, at time t, the mth infected node, denoted by a, has D_m neighbor nodes. The degrees of a's neighbor nodes are $k_1, k_2, \cdots, k_{D_m}$ respectively. According to the biased random walk mechanism, the probability that a sends an infection packet to its ith neighbor node is as follows:

$$P_i = \frac{k_i^\alpha}{\sum_{j=1}^{D_m} k_j^\alpha}. \tag{2}$$

Where α is the control parameter of the biased random walk. When $\alpha = 0$, all the neighbor nodes have equal opportunity to receive an infection packet delivered from a. When $\alpha > 0$, nodes of larger degree have a larger probability to receive the packet. When $\alpha < 0$, nodes of smaller degree have a larger probability of receiving the packet. A susceptible node becomes an infected node when receive an infection packet (Fig. 1 (b)), and starts sending infection packets from next time step. An infected (or removed) node doesn't change its state when receive an infection packet, and will drop the infection

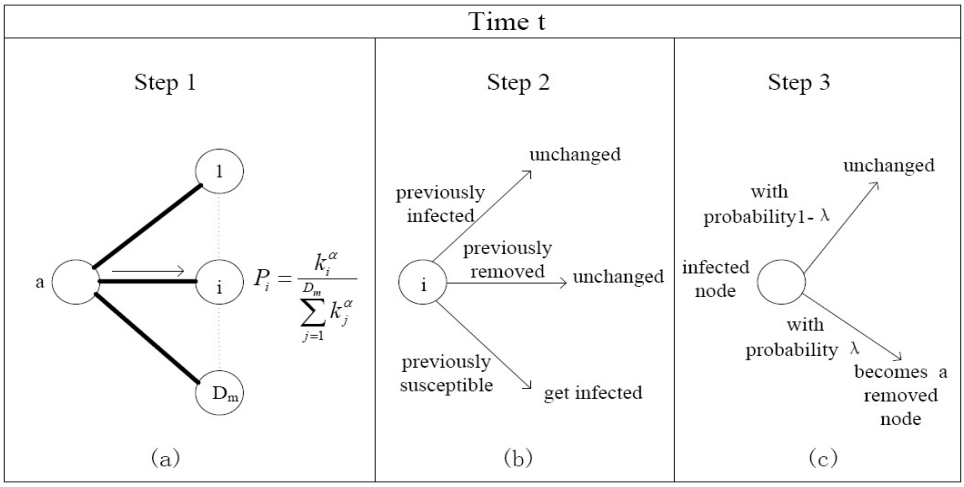

Figure 1: Illustration of our model.

packet. An infected node has a probability λ to become a removed node at time t (Fig. 1 (c)). The difference between our model and the original SIR model is that the spreading of epidemics in our model is driven by packets delivery, which is further based on the biased random walks.

3.1. Dynamics of our model

Since a sends out C infection packets, the probability that the ith neighbor node of a will not receive any infection packets from a is:

$$\bar{P}_i = (1 - \frac{k_i^\alpha}{\sum_{j=1}^{D_m} k_j^\alpha})^C. \tag{3}$$

Assuming X_i is a random variable. $X_i = 0$ represents the ith neighbor node of a didn't receive any infection packets sent from a. $X_i = 1$ means the ith neighbor node of a received at least one of the C infection packets sent from a. Then, the expected value of X_i is:

$$\mathrm{E}(X_i) = 1 - (1 - \frac{k_i^\alpha}{\sum_{j=1}^{D_m} k_j^\alpha})^C. \tag{4}$$

The average number of nodes that received infection packets sent from a is:

$$Y_m = \sum_{i=1}^{D_m} \mathrm{E}(X_i)$$

$$= D_m - \sum_{i=1}^{D_m}(1 - \frac{k_i^\alpha}{\sum_{j=1}^{D_m} k_j^\alpha})^C. \quad (5)$$

Where the sum is over all the D_m neighbor nodes of a. At time t, the mean number of nodes that received infection packets sent from an arbitrary infected node is as follows:

$$\begin{aligned} \langle Y \rangle &= \frac{1}{I}\sum_{m=1}^{I} Y_m \\ &= \frac{1}{I}\sum_{m=1}^{I}(D_m - \sum_{i=1}^{D_m}(1 - \frac{k_i^\alpha}{\sum_{j=1}^{D_m} k_j^\alpha})^C). \end{aligned} \quad (6)$$

Where I is the total number of infected nodes at time t. However, in the epidemic spreading process, the neighbor nodes of an infected node may not be susceptible nodes only. To precisely estimate the number of new infected nodes, we calculate the proportion of susceptible nodes in all the neighbors of infected nodes, which is as follows:

$$\mu = \frac{\sum_{m=1}^{I} n_m}{\sum_{m=1}^{I} D_m}. \quad (7)$$

Where n_m is the number of susceptible nodes among the neighbors of the mth infected node. According to Eq. 6 and Eq. 7, the total new infected nodes at time t is:

$$\begin{aligned} I_{new} &\approx \langle Y \rangle * \mu * I \\ &= \sum_{m=1}^{I}(D_m - \sum_{i=1}^{D_m}(1 - \frac{k_i^\alpha}{\sum_{j=1}^{D_m} k_j^\alpha})^C) * \frac{\sum_{m=1}^{I} n_m}{\sum_{m=1}^{I} D_m}. \end{aligned} \quad (8)$$

Combining Eq. 1 with Eq. 8, we get the dynamics equations of our model as follows:

$$\begin{cases} \frac{\mathrm{d}I}{\mathrm{d}t} &= \sum_{m=1}^{I}(D_m - \sum_{i=1}^{D_m}(1 - \frac{k_i^\alpha}{\sum_{j=1}^{D_m} k_j^\alpha})^C) * \frac{\sum_{m=1}^{I} n_m}{\sum_{m=1}^{I} D_m} - \lambda I, \\ \frac{\mathrm{d}S}{\mathrm{d}t} &= -\sum_{m=1}^{I}(D_m - \sum_{i=1}^{D_m}(1 - \frac{k_i^\alpha}{\sum_{j=1}^{D_m} k_j^\alpha})^C) * \frac{\sum_{m=1}^{I} n_m}{\sum_{m=1}^{I} D_m}, \\ \frac{\mathrm{d}R}{\mathrm{d}t} &= \lambda I. \end{cases} \quad (9)$$

We obtain the numbers of susceptible nodes $S(t)$, infected nodes $I(t)$ and

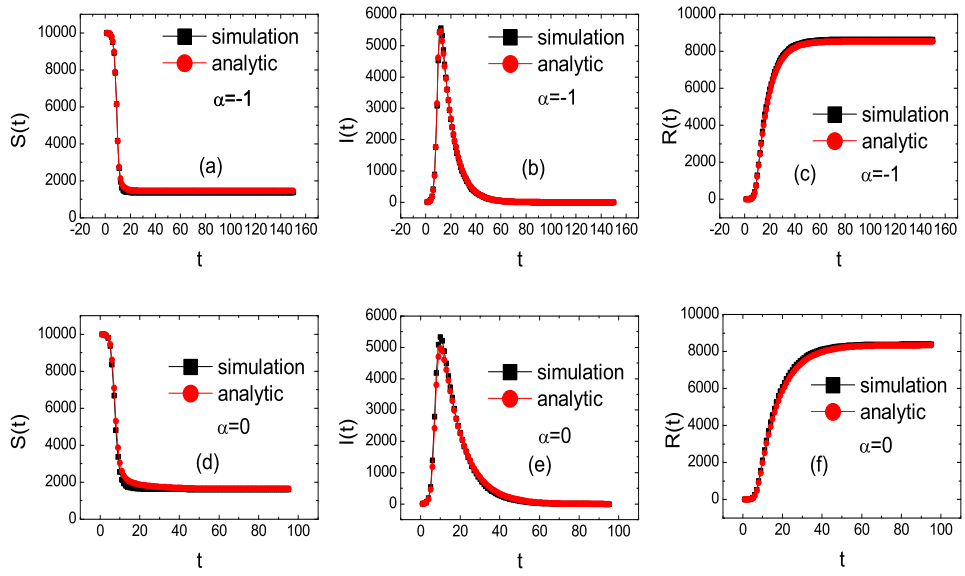

Figure 2: $S(t)$, $I(t)$, and $R(t)$ vs. t for $\alpha = 0$ and $\alpha = -1$. $C = 5$. $\lambda = 0.1$. The network is generated by the static model[43]. The parameters of the network are $N = 10000$, $\langle K \rangle = 5$, and $\gamma = 2.5$.

removed nodes $R(t)$ with increase of time t in a large scale-free network (see Appendix A). As shown in Fig. 2, $S(t)$ decreases abruptly, and then saturates with t. $I(t)$ first increases with t, then decreases with t, and finally saturates. The peak of $I(t)$, denoted as I_{peak}, corresponds to the maximum instantaneous size of infected nodes, which reflects how intense the spread of the epidemic is. $R(t)$ increases sharply, and then saturates with t, which is opposite to the trend of $S(t)$. When t is large enough, the epidemic spreading process stops, then the number of nodes that have ever been infected and finally recovered from the disease is denoted as R_{end}. R_{end} reflects how wide the spread of the epidemic is. The trends of the curves for biased random walks with $\alpha = -1$ are similar with that for simple random walks with $\alpha = 0$. Also, the simulation and analytical results obtained from Eq. 9 are consistent, as shown in Fig. 2.

3.2. Factors of our model

Delivery capacity C of infected nodes is a critical factor in our epidemic spreading model. The larger the delivery capacity is, the more susceptible nodes will likely be infected at each time step. According to Eq. 8, when

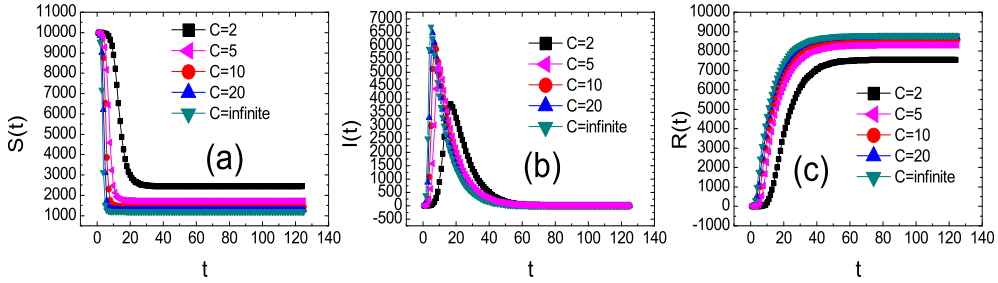

Figure 3: $S(t)$, $I(t)$, and $R(t)$ vs. t for different C. $\alpha = 0$. $\lambda = 0.1$. The network is generated by the static model with $N = 10000$, $\langle K \rangle = 5$, and $\gamma = 2.5$.

$C \to \infty$, $I_{new}(t) \to \sum_{m=1}^{I} n_m$. Then Eq. 9 is simplified as follows:

$$\begin{cases} \frac{\mathrm{d}I}{\mathrm{d}t} &= \sum_{m=1}^{I} n_m - \lambda I, \\ \frac{\mathrm{d}S}{\mathrm{d}t} &= -\sum_{m=1}^{I} n_m, \\ \frac{\mathrm{d}R}{\mathrm{d}t} &= \lambda I. \end{cases} \tag{10}$$

We show simulation results of $S(t)$, $I(t)$ and $R(t)$ for different value of C in Fig. 3. Clearly, the larger C, the faster $S(t)$, $I(t)$ and $R(t)$ converge, and the earlier the epidemic spreading stops. The larger C, the larger the population of nodes that has ever been infected, which is inferred from R_{end}. The larger C, the larger the maximum instantaneous number of infected nodes, which is indicated from I_{peak}.

The parameter α is another key factor in our epidemic model, which determines the probability that a node receives an infection packet. In Fig. 4, we see that $\alpha = -1$ corresponds to a larger I_{peak} and a larger R_{end} than $\alpha = 0$ and $\alpha = 1$. This indicates that random walks biased on small-degree nodes favors the epidemic spreading which holds true for the model network and the Epinions network, as shown in Fig.4. Furthermore, we obtain I_{peak} and R_{end} as a function of α by simulation, as shown in Fig. 5. Clearly, both I_{peak} and R_{end} increase with α first, and then decrease with α. There are optimal parameters α_{opt} that correspond to the maximum I_{peak} and the maximum R_{end} respectively. In our model, large-degree infected nodes facilitate the spread of the epidemic, since they connect a large number of nodes. However, large-degree removed nodes inhibit the spread of the epidemic, since they drop the infection packets passing by them. On the other hand, the large-degree infected nodes become removed nodes with a constant rate λ. Therefore, the epidemic should cautiously infect large-degree

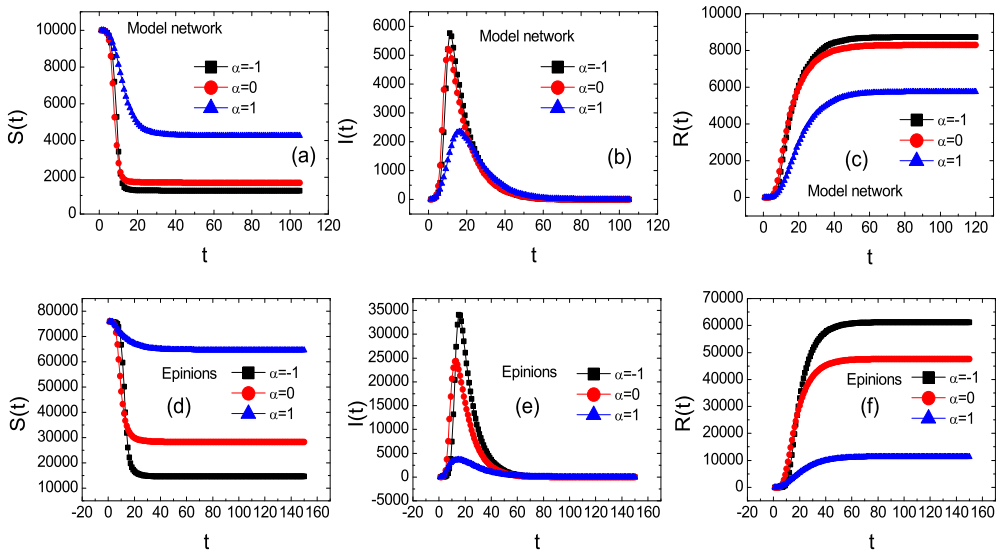

Figure 4: $S(t)$, $I(t)$, and $R(t)$ vs. t for different α. $C = 5$. $\lambda = 0.1$. The network for (a), (b) and (c) is generated by the static model with $N = 10000$, $\langle K \rangle = 5$, and $\gamma = 2.5$. The Epinions network for (d), (e) and (f) has 75, 879 nodes and 508, 960 edges.

nodes for spreading intensely and widely. We present the maximum I_{peak}, the maximum R_{end}, and the corresponding α_{opt} for some real-world networks[44, 45, 46, 47, 48, 49, 50, 51, 52, 53], as shown in Table 1. We see that generally the optimal parameters are slightly less than zero.

4. Impacts of network structures on our model

We investigate the influence of topological properties of complex networks, including average node degree and degree distribution, on our epidemic spreading model. We focus on the maximum instantaneous number of infected nodes I_{peak} and the number of nodes that have ever been infected R_{end}, as well as the related optimal parameters α_{opt}. According to Eq. 5, we have:

$$\begin{aligned} Y_m &= D_m - \sum_{i=1}^{D_m} (1 - \frac{k_i^\alpha}{\sum_{j=1}^{D_m} k_j^\alpha})^C \\ &= D_m - \sum_{i=1}^{D_m} (1 - C\frac{k_i^\alpha}{\sum_{j=1}^{D_m} k_j^\alpha} + \frac{C(C-1)}{2}(\frac{k_i^\alpha}{\sum_{j=1}^{D_m} k_j^\alpha})^2 - \cdots) \end{aligned}$$

Table 1: maximum I_{peak} and maximum R_{end} with corresponding α_{opt} for real-world networks (see Appendix B). $C = 5$. $\lambda = 0.1$.

NAME	NODES	EDGES	I_{peak}	α_{opt}	R_{end}	α_{opt}
Oregon-1[44]	10790	22469	2903.33	-0.4	6010.37	-0.8
Gnutella[45]	62586	147892	37833.43	-0.8	58155.71	-1.4
Epinions[46]	75879	508837	35274.13	-0.8	61560.15	-1.2
Wiki-Vote[47]	7115	103689	4294.22	-1	6623.23	-1.4
Yeast[48]	2361	7182	1314.21	-0.8	2214.63	-1.2
email-Enron[49]	36692	183831	12789.5	-0.8	24254.25	-1.4
Facebook[50]	4039	88234	2105.11	-0.8	3920.21	-0.4
Geom[51]	7343	11898	1778.5	-0.4	3202.5	-1
Political blogs[52]	1222	19021	813.26	-1.2	1192.14	-2.2
Power grid[53]	4941	6594	1278.36	0.4	4304.27	-0.2

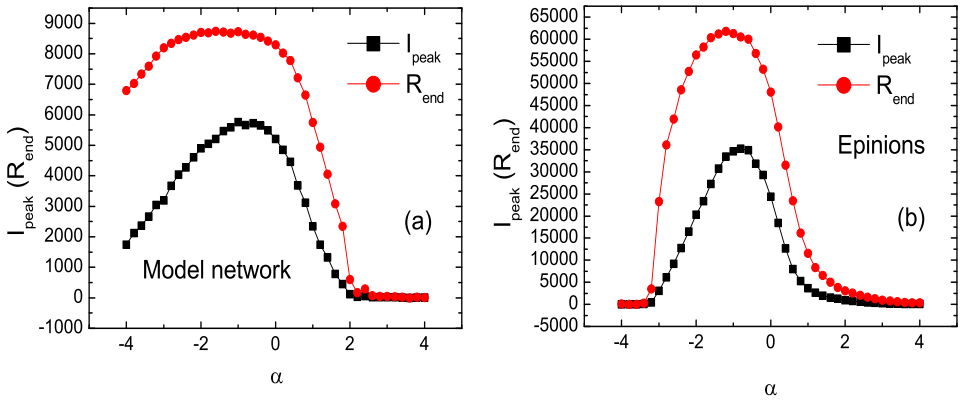

Figure 5: I_{peak} and R_{end} vs. α. $C = 5$. $\lambda = 0.1$. The networks are the same as in Fig. 4. The results are the average of 100 independent runs.

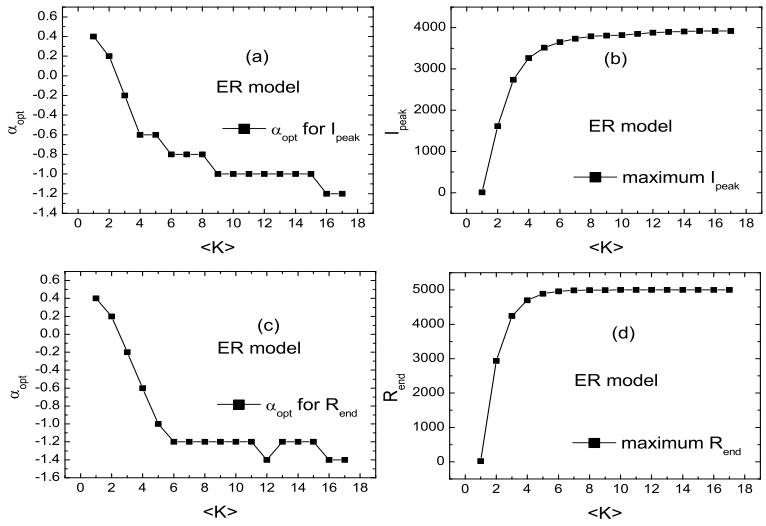

Figure 6: α_{opt} and the corresponding I_{peak} and R_{end} vs. $\langle K \rangle$ for random networks generated by the ER (Erdös-Rényi) model[54] with network size $N = 5000$. $C = 5$. $\lambda = 0.1$. The results are the average of 10^4 independent runs.

$$= \sum_{i=1}^{D_m} (C \frac{k_i^\alpha}{\sum_{j=1}^{D_m} k_j^\alpha} - \frac{C(C-1)}{2} (\frac{k_i^\alpha}{\sum_{j=1}^{D_m} k_j^\alpha})^2 + \cdots). \quad (11)$$

When $D_m \to \infty$, we get:

$$\begin{aligned} Y_m &\approx \sum_{i=1}^{D_m} \frac{C k_i^\alpha}{\sum_{j=1}^{D_m} k_j^\alpha} \\ &= C. \end{aligned} \quad (12)$$

Eq. 12 indicates that generally the number of nodes that an infected node can infect in one time step will increase with the number of neighbors it has, and tends to C. Also, for the entire epidemic spreading process, both the maximum I_{peak} and the maximum R_{end} increase substantially, then saturate with average degree $\langle K \rangle$, as shown in Fig. 6 and 7. Their corresponding optimal parameters α_{opt} are generally negative and decrease with $\langle K \rangle$. This indicates that when the networks become dense, random walks will be more biased towards small-degree nodes, which makes the epidemic spreading more intense and far-reaching. These results are consistent for both random networks (Fig. 6) and scale-free networks (Fig. 7). We also investigate the impact of degree distribution on epidemic spreading dynamics. In Fig. 8, both the maximum

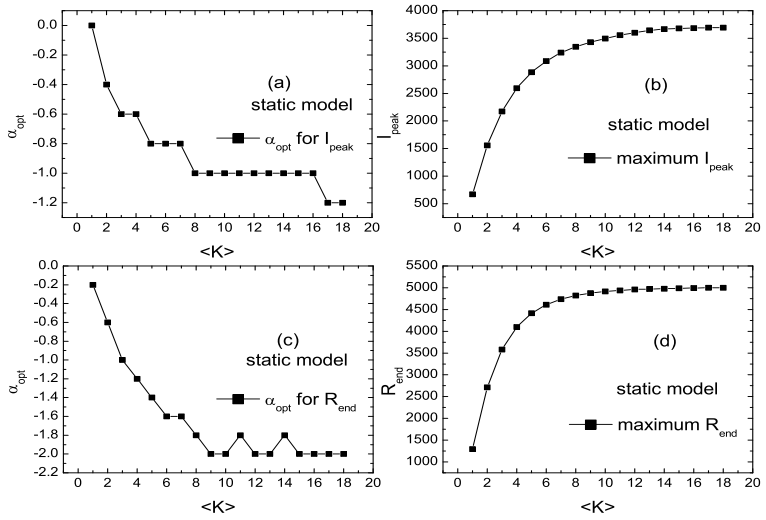

Figure 7: α_{opt} and the corresponding I_{peak} and R_{end} vs. $\langle K \rangle$ for scale-free networks generated by the static model with $N = 5000$ and $\gamma = 2.5$. $C = 5$. $\lambda = 0.1$. The results are the average of 10^4 independent runs.

I_{peak} and the maximum R_{end} increase abruptly, then saturate with γ, and this means when the degree distribution becomes homogeneous, the spreading of the epidemic becomes more intense and widespread in the network. The optimal parameters α_{opt} for I_{peak} and R_{end} increase with γ. This indicates when the network structure becomes homogeneous, random walks are less biased towards small-degree nodes for a more intense and widespread spreading of epidemics. However, the fluctuations in the curves are unambiguous.

5. Conclusion

In summary, we investigate the dynamics of epidemic spreading which is based on biased random walks. In our model, an epidemic disease spreads by means of the transmission of infection packets. The infection packets are sent by the biased random walk mechanism, in which an infection packet spreads from one infected node to a neighbor with a probability that is a function of the neighbor's degree. We combine the basic SIR equations with the equations of biased random walks. Also, we conduct agent-based simulations on model networks and real-world networks. By tuning the parameter of biased random walks, we obtain the maximum instantaneous size of infected nodes, as well as the maximum size of ever-infected nodes for our model. We

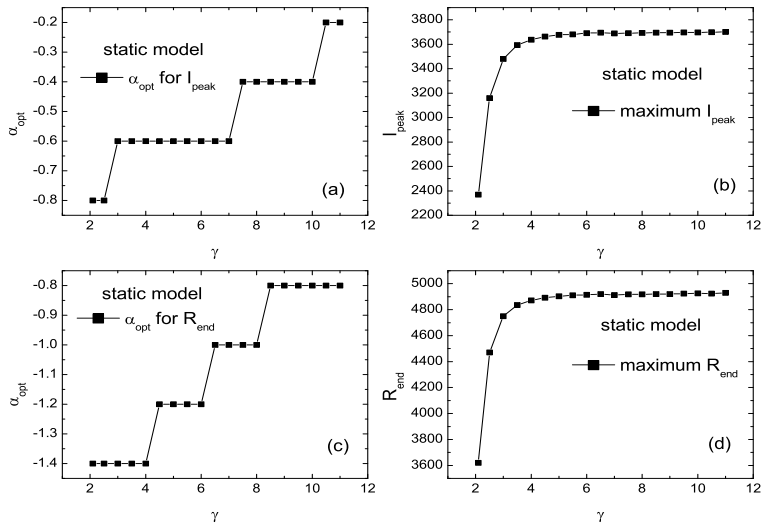

Figure 8: α_{opt} and the corresponding I_{peak} and R_{end} vs. power-law parameter γ. The networks are generated by the static model with $N = 5000$ and $\langle K \rangle = 5$. $C = 10$. $\lambda = 0.1$. The results are the average of 10^4 independent runs.

find that, generally the optimal parameters of the biased random walks are negative values, which means the epidemic is biased towards small-degree nodes for intense and widespread transmission. Furthermore, the analytical and simulation results demonstrate that epidemic spreading becomes fierce and widespread with the increase of either delivery capacity of infected nodes, average node degree, or homogeneity of the network.

It is worth mention that the biased random walks in our model are based on only degrees of the nearest neighbors. While the effect of biased random walks with more topological information on epidemic spreading still need to be explored. We only consider SIR spreading dynamics, while the dynamics of SI, SIS, SIRS, etc. are also interesting. We didn't discuss the immunization strategies in this paper, but we believe that our results give some clues for the immunization of the SIR epidemics.

Acknowledgments

The authors thank Ming Tang, Andrew Michaelson, Siyuan Zhou and the anonymous reviewers for their helpful comments. This work was supported by the Natural Science Foundation of China (Grant No. 61304154),

the Specialized Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20133219120032), and the Postdoctoral Science Foundation of China (Grant No. 2013M541673).

Appendix A.

We generate the scale-free networks by using the static model[43]. Initially, there are N nodes in the system indexed by an integer i ($i = 1, 2, \cdots, N$). We assign the weight $P_i = i^{\rho}$ to each node i, where ρ is a control parameter in $[0, 1)$. Then, we select two different nodes (i, j) with probabilities equal to the normalized weights, $p_i / \sum_k p_k$ and $p_j / \sum_k p_k$, respectively, and connect them with an edge unless they are already connected. This process repeated until mN edges are added in the system. For the generated scale-free network, the average node degree $\langle K \rangle$ is $2m$. The degree distribution follows the power law, $P_D(k) \sim k^{-\gamma}$, where γ is given by $\gamma = (1 + \rho)/\rho$. Since ρ is in $[0, 1)$, we get $2 < \gamma < \infty$.

Appendix B.

The datasets are downloaded from web sites, which are as follows:

1. Oregon-1[44]: autonomous system graph. Nodes represent autonomous systems. Two nodes are connected if their communication events appear in the BGP logs. `http://snap.stanford.edu/data/oregon1.html`.
2. Gnutella[45]: gnutella peer-to-peer file sharing network, in which nodes represent hosts in the Gnutella network topology and edges represent connections between the Gnutella hosts. `http://snap.stanford.edu/data/p2p-Gnutella31.html`.
3. Epinions[46]: who-trust-whom online social network of a general consumer review site Epinions.com, in which nodes are members of the site and edges represent the trust relationships among members. `http://snap.stanford.edu/data/soc-Epinions1.html`.
4. Wiki-Vote[47]: wikipedia vote network, in which nodes represent wikipedia users and a directed edge from node i to node j represents that user i voted on user j. `http://snap.stanford.edu/data/wiki-Vote.html`.
5. Yeast[48]: protien-protien interaction network in budding yeast. `http://math.nist.gov/~RPozo/complex_datasets.html`.

6. email-Enron[49]: Enron email network. Nodes of the network are email addresses. If an address i sent at least one email to address j, then there is an undirected edge from i to j. `http://snap.stanford.edu/data/email-Enron.html`.
7. Facebook[50]: ego networks collected from participants using the Facebook app. Nodes represent users and edges represent the friendships. `http://snap.stanford.edu/data/egonets-Facebook.html`.
8. Geom[51]: collaboration network in computational geometry. Nodes are the authors. Two authors are linked with an edge if they wrote a common work. `http://vlado.fmf.uni-lj.si/pub/networks/data/collab/geom.htm`.
9. Political blogs[52]: a directed network of hyperlinks between weblogs on US politics. `http://www-personal.umich.edu/~mejn/netdata/`.
10. Power grid[53]: an undirected network representing the topology of the Western States Power Grid of the United States. `http://www-personal.umich.edu/~mejn/netdata/`.

[1] D. G. Green, T. Bossomaier, eds, Complex systems: from biology to computation, IOS press, 1993.

[2] H. W. Hethcote, *SIAM Rev.* **42** (2000) 599.

[3] J. Balthrop, S. Forrest, M. E. J. Newman, et al, arXiv preprint cs/0407048, 2004.

[4] P. Wang, M. C. González, C. A. Hidalgo, et al, *Science* **324** (2009) 1071.

[5] A. Vespignani, *Nature Physics* **8** (2012) 32.

[6] F. Brauer, C. Castillo-Chavez, Mathematical models in population biology and epidemiology, Springer, 2011.

[7] L. F. Berkman, I. Kawachi, eds, Social epidemiology, Oxford University Press, 2000.

[8] N. T. J. Bailey, The Mathematical Theory of Infectious Diseases and its Applications, Griffin, London, 1975.

[9] R. M. May, Infectious diseases of humans: dynamics and control, Oxford University Press, 1995.

[10] R. M. Anderson, R. M. May, Infectious diseases of humans, Oxford: Oxford university press, 1991.

[11] M. Newman, Networks: an introduction, Oxford University Press, 2010.

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

[13] A. Barrat, M. Barthelemy, A. Vespignani, Dynamical processes on complex networks, Cambridge University Press, Cambridge, 2008.

[14] T. Zhou, Z. Q. Fu, B. H. Wang, *Progress in Natural Science* **16** (2006) 452.

[15] R. Yang, B. H. Wang, J. Ren, et al, *Phys. Lett. A* **364** (2007) 189.

[16] G. Yan, Z. Q. Fu, J. Ren, W. X. Wang, *Phys. Rev. E* **75** (2007) 016108.

[17] S. W. Chou, K. Wang, Q. Liu, et al, *Acta Phys. Sin* **61** (2012) 150201.

[18] Z. Ruan, M. Tang, Z. Liu, *Phys. Rev. E* **86** (2012) 036117.

[19] W. Wang, M. Tang, H. Yang, et al, *Sci. Rep.* **4** (2014) 5097.

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

[21] Z. Yang, T. Zhou, *Phys. Rev. E* **85** (2012) 056106.

[22] F. D. Sahneh, C. Scoglio, P. Van Mieghem, *IEEE/ACM Transactions on Networking (TON)* **21** (2013) 1609.

[23] M. E. J. Newman, *Phys. Rev. E* **66** (2002) 016128.

[24] R. Cohen, K. Erez, D. ben Avraham, et al, *Phys. Rev. Lett.* **85** (2000) 4626.

[25] A. Saumell-Mendiola, M. Á. Serrano, M. Boguñá, *Phys. Rev. E* **86** (2012) 026106.

[26] F. Sahneh, C. Scoglio, F. Chowdhury, *In: American Control Conference (ACC)* (2013) 2307.

[27] C. Granell, S. Gómez, A. Arenas, *Phys. Rev. Lett.* **111** (2013) 128701.

[28] V. Colizza, A. Vespignani, *Journal of Theoretical Biology* **251** (2008) 450.

[29] S. Meloni, A. Arenas, Y. Moreno, *PNAS* **106** (2009) 16897.

[30] P. Bajardi, C. Poletto, J. J. Ramasco, et al, *PLoS ONE* **6** (2011) e16591.

[31] H. X. Yang, W. X. Wang, Y. C. Lai, et al, *Phys. Rev. E* **84** (2011) 045101(R).

[32] H. X. Yang, Z. X. Wu, *J. Stat. Mech.* **3** (2014) P03018.

[33] C. Shen, H. Chen, Z. Hou, *Phys. Rev. E* **86** (2012) 036114.

[34] H. X. Yang, Z. X. Wu, B. H. Wang, *Phys. Rev. E* **87** (2013) 064801.

[35] L. Lovász, *Combinatorics: Paul Erdös is eighty* **2** (1993) 1.

[36] J. D. Noh, H. Rieger, *Phys. Rev. Lett.* **92** (2004) 118701.

[37] M. Bonaventura, V. Nicosia, V. Latora, *Phys. Rev. E* **89** (2014) 012803.

[38] Z. Zhang, T. Shan, G. Chen, *Phys. Rev. E* **87** (2013) 012112.

[39] Y. Lin, Z. Zhang, *Sci. Rep.* **4** (2014) 5635.

[40] R. Sinatra, J. Gómez-Gardeñes, R. Lambiotte, et al, *Phys. Rev. E* **83** (2011) 030103.

[41] V. Colizza, A. Barrat, M. Barthélemy, A. Vespignani, *PNAS*, **103** (2006) 2015.

[42] Z. Wu, L. A. Braunstein, V. Colizza, et al, *Phys. Rev. E* **74** (2006) 056104.

[43] K.-I. Goh, B. Kahng, D. Kim, *Phys. Rev. Lett.* **87** (2001) 278701.

[44] J. Leskovec, J. Kleinberg, C. Faloutsos, *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2005.

[45] M. Ripeanu, I. Foster, A. Iamnitchi, *IEEE Internet Computing Journal*, 2002.

[46] M. Richardson, R. Agrawal, P. Domingos, *ISWC*, 2003.

[47] J. Leskovec, D. Huttenlocher, J. Kleinberg, *CHI*, 2010.

[48] S. W. Sun, L. J. Ling, N. Zhang, et al, *Nucleic Acids Research* **31** (2003) 2443.

[49] J. Leskovec, K. Lang, A. Dasgupta, M. Mahoney, *Internet Mathematics* **6** (2009) 29.

[50] J. McAuley, J. Leskovec, *NIPS*, 2012.

[51] N. H. F. Beebe, http://www.math.utah.edu/ beebe/bibliographies.html, 2002.

[52] L. A. Adamic, N. Glance, *in Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem*, 2005.

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

[54] P. Erdös, A. Rényi, *Publ. Math. Inst. Hung. Acad. Sci.* **5** (1960) 17.