

A Preliminary Study on Effects of Community Structures on Epidemic Spreading and Detection in Complex Networks

Yi Yu, Gaoxi Xiao

Abstract—Community structures widely exist in almost all real-life networks. Extensive researches have been carried out on detecting community structures in complex networks. However, many aspects of how community structures may affect the dynamics and properties of complex networks still remain unclear. In this work, we examine the impacts of community structures on the epidemic spreading and detection in complex networks. Extensive simulation results show that community structures may not help decrease the infection size at steady state, yet they could indeed help slow down the infection spreading. Also, networks with strong community structures may expect to have a smaller average infection size when equipped with a number of sparsely deployed monitors.

Keywords—Complex network, epidemic spreading, infection size, infection monitoring.

I. INTRODUCTION

COMPLEX network has been a hot topic for more than a decade. A lot of real-life systems can be conveniently modeled as networks and then be effectively studied. Such systems include computer networks [1], Internet autonomous systems [2], human sexual contacts [3], World Wide Web [4], and many more. It is well known that the structures of the networks may have paramount impacts on the dynamics and properties of the systems. For example, one of the most significant discoveries is that many real-life systems can be presented into *scale-free* networks with a power-law nodal degree distribution [5], which leads to some critical properties of such systems including strong fault tolerance [6] and fragility under intentional attacks [7], [8], etc.

An important topic which has attracted extensive research interests recently is epidemic spreading and control in complex networks, for which the network structures also play a critical role. For example, it is found that scale-free networks are generally speaking in absence of an epidemic threshold [9] and one of the most effective immunization methods, known as targeted immunization, is to immunize the high-degree *hub* nodes [10], even when such immunization is not perfect [11], etc.

Community structures, which reflect some important properties of real-life systems such as social structures with dense internal connections and sparse external connections due to geographic distances and obstacles, emergence of clustering

in online social media etc., have been attracting increasing research interests. The existing results however have been rather heavily focusing on the construction and detection of various community structures in complex networks. Specifically, the pioneering work by Newman [12]-[14] introduced the *modularity score* for measuring the intensity of community structures. He also proposed algorithms for maximizing the modularity score by grouping the nodes according to the eigenvalues of the adjacency matrix of the network. Other relevant works include algorithms for adjusting modularity level without changing any nodal degree [15], detecting the modularity in bipartite graph [16], methods for defining and detecting overlapping communities in large-scale networks [17], and methods for detecting community structures in directed networks [18] etc.

Despite of all these existing results, the effects of community structures on dynamics and properties of epidemic spreading in complex networks, such as the effects on infection size and transmission speed, remain rather unclear. In [19], by conducting theoretical analyses and numerical simulations on a simple random network model with adjustable modularity level, it is shown that when the modularity score increases, which denotes an increasing number of intra-community links and a decreasing number of inter-community links, the epidemic threshold becomes lower. However, it does not discuss on the effects of community structures on infection size when an outbreak does happen. In [20], based on a different network model and by adopting the SI spreading dynamics model where infected nodes always remain infectious, it is shown that, when an outbreak does happen, the epidemic spreading is slower in networks with stronger community structures.

In this work, we carry out extensive numerical simulations to examine on two aspects of the effects of community structures on epidemic spreading and control in complex networks. First, we examine how the infection size changes over time and whether the final steady-state infection size changes with community structures; second, by adopting the monitoring allocation method as proposed in [21], we exam, when there is a certain number of monitors deployed in the network for detecting epidemic spreading, whether and how community structures affect the average infection size until the infection spreading is detected and stopped.

The rest of this paper is organized as follow. In Section II, we define models and assumptions adopted in this work. Simulation results and discussions are presented in Section III. Finally, Section IV concludes this paper.

Yi Yu and Gaoxi Xiao are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (e-mail: yyu6@ntu.edu.sg, egxxiao@ntu.edu.sg).

II. MODELS

A. Random Network with Community Structures

In this paper, we adopt the method proposed in [19] to generate random networks with community structures. Specifically, the main idea of constructing a network with community structures is to generate inter-community links at a lower probability than that for generating intra-community links. The construction process can be briefly described as composing of the following three steps [19]:

Step 1. Randomly divide N node into M communities with

$$n_i \text{ nodes in } i^{\text{th}} \text{ group such that } \sum_{i=1}^M n_i = N.$$

Step 2. Within each group, let every pair of nodes be connected with a probability p .

Step 3. For each pair of nodes in different groups, connect them with a lower probability q . Denote $\sigma = p/q$.

In this case we are able to calculate the overall number of links in the network as:

$$V = \sum_{i=1}^M \frac{1}{2} n_i (n_i - 1) p + \sum_{i < j} n_i n_j q$$

For networks with very strong community structures and equal size of each community, we have $\sigma \gg 1$ and the total number of edges therefore can be approximated as

$$V \approx \frac{N^2 p}{2M} - \frac{Np}{2}$$

In the rest part of this paper, we use this function to generate network with community structures. The number of edges in the network can be controlled by adjusting p .

To make comparisons between the networks with community structures versus their counterparts with the same nodal degree by without community structure, we generate their counterparts using the Monte-Carlo Switching algorithm [22]. The method is to randomly choose two links, denoted as (A-B, C-D), and rewire the links into (A-D, B-C) if, and only if, there is no multiple links of self-links generated. By carrying out such rewiring operation for a large enough times, the network will closely resemble a random network with no community structures. In our simulations, we rewire 100V times where V denotes the number of nodes in the network. Such a large number of rewiring operations is sufficient to basically eliminate the community structures in the networks [22].

B. Scale-Free Network with Community Structures

In scale-free networks, the nodal degrees follow a power-law distribution, i.e.,

$$P(k) \sim k^{-\gamma}$$

where $P(k)$ denotes the portion of nodes with degree k , and γ is the exponent which typically has a value within the range of

[2], [3] in real-life systems. The well-known BA model [5] which generates a scale-free network by growth and preferential attachment has an exponent value of $\gamma = 3$.

An algorithm is proposed in [20] to generate scale-free networks with community structures, basically still by adopting growth and preferential attachment. Specifically, when a new node is added to the network, it is randomly assigned to a certain community. First it is connected to a fixed number of existing nodes within the same community by intra-community links. Then it has a certain given probability to be connected to a fixed number of nodes outside its own community by inter-community links. For both the intra- and inter-community connections, the existing nodes to be connected are randomly selected in probabilities proportional to their degrees, i.e., by using the preferential attachment method. Specifically, the construction process can be described as follows:

Initialization: Build an initial network with M communities and m_0 nodes in each community, where each community is connected into a complete graph (i.e., every node is connected to every other node in the same community). For every pair of communities, randomly choose a single node in each community and connect them with an inter-community link. Keep separate records of every node's intra- and inter-community degrees respectively.

Evolving: When a new node is added to the network, randomly choose a community for it. Connect this new node with m existing nodes in the same community by intra-community preferential attachment, i.e., the probability that the new node is connected to an existing node is proportional to the intra-community degree of the existing node. Each new node also has a probability σ to have n inter-community links. When the new node indeed can have inter-community links, the probability that an existing node in another community is connected to the new node is proportional to the inter-community degree of the existing node.

This above procedure is repeated until the network size is large enough.

The network generated in this way has a power law degree distribution of [20]:

$$P(k) = \frac{2(m + \sigma n)^2 t}{M m_0 + t} k^{-3}$$

Once again, the Monte-Carlo Switching algorithm [22] is applied to generate the counterpart network with the same nodal degree but without community structures.

C. Epidemic Spreading Models

In this work, we adopt the well-known Susceptible-Infected-Recovered (SIR) model for evaluating infection size and spreading speed, and Susceptible-Infected (SI) model [23] for evaluating the average infection size in networks equipped with sparsely deployed monitors. Specifically, for the SIR model, we assume that time is slotted and any susceptible node adjacent to an infected node has a probability λ of being

infected at the beginning of each time step. At the same time slot, each infected node has a probability of μ to recover. A recovered node is no longer infectious and will never get infected again. We also assume that the epidemic starts from a single infection source. For the SI model, it differs from the SIR model by not allowing any node to recover. In early stage of strong infections when recovery hardly starts yet, the SI model closely resembles the real-life cases.

III. THE EFFECTS OF THE COMMUNITY STRUCTURES

In this section, we compare epidemic spreading in networks with and without community structures, respectively. For the random network model, we generate 10 networks, each of which has 1000 nodes, 10000 edges, 10 communities and a community parameter of $\sigma = p/q = 1000$. In each generated network, we choose 1000 different infection sources and average the results of these 1000 realizations. The spreading rate is set as $\lambda/\mu = 0.1$. For the scale-free network model, we generate 10 networks, each of which has 1000 nodes, 5000 edges, 10 communities and a community parameter value of $\sigma = 0.01$. Again, for each network generated, we test on 1000 different infection sources and calculate the average results of the 1000 realizations. The spreading rate is set as $\lambda/\mu = 0.4$. The final results presented in this paper, unless otherwise specified, are the average results in all the generated networks.

A. Infection Speed and Overall Infection Size

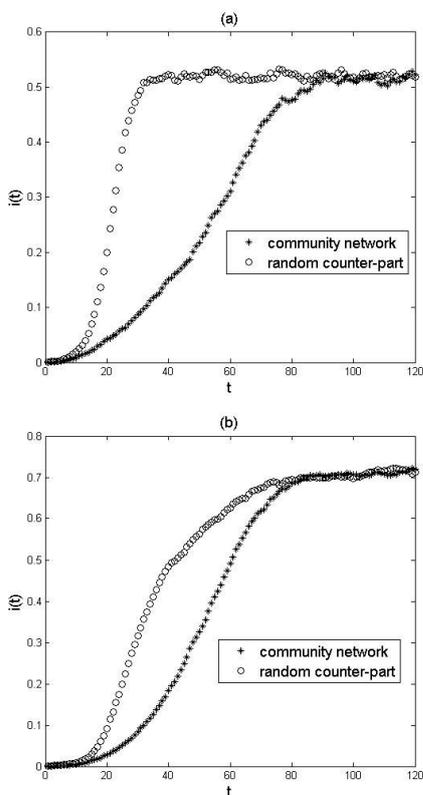


Fig. 1 Dynamics of overall infection size along time: (a) random network model; (b) scale-free network

Fig. 1 illustrates the dynamics of the overall infection size along time. The SIR model is adopted. We can see that in both networks, the existence of strong community structures helps slow down the infection spreading. The steady-state infection size, however, is hardly affected by the existence of the community structures. Nodal degrees of the network largely decide the overall infection size.

We then proceed to examine on the effects of changing the number of communities in the networks. Specifically, we test different cases with 2, 6 and 10 communities in the random networks and 5, 10 and 15 communities in the scale-free networks. For each number of communities, we generate 5 random networks. All the other parameters remain unchanged as stated at the beginning of Section III.

The simulation results are presented in Fig. 2. We can observe that in random networks, the infection spreading becomes slower with an increasing number of communities. This can be easily understood: the strong community structures make it difficult for the infection spreading to enter into a new community. In scale-free networks, however, the effects of having different number of communities become much less significant and decisive. This can be explained by the existence of the hub nodes in the scale-free network, which helps distribute the virus quickly despite of the gaps between different communities.

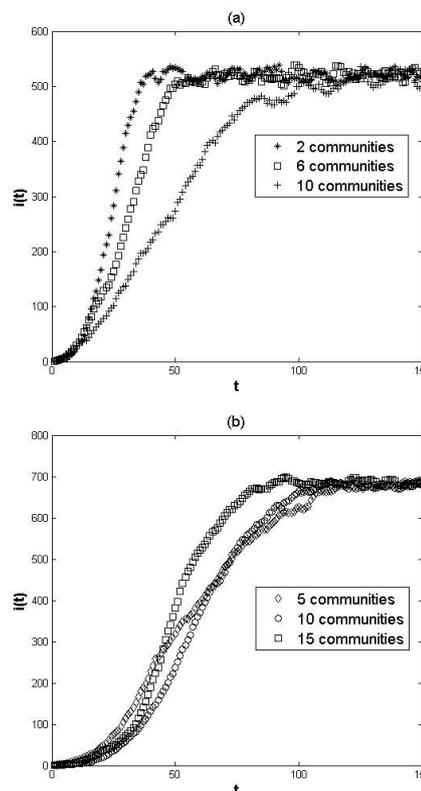


Fig. 2 Epidemic spreading in networks with different number of communities: (a) random networks; (b) scale-free networks

B. Infection Size with Network Monitors

Finally we test on the effects of community structures on the

average infection size in networks with sparsely deployed monitors. As aforementioned, since the main concern for infection detection is to find and stop strong infections at their early stage, we adopt the SI model where infected nodes never get recovered. As that in [21], we assume that once the infection reaches any of the monitors, it will be detected and stopped immediately

We test on both the random and scale-free network models. For random network model we generate 5 networks, each of which has 1000 nodes and 10000 edges with 10 communities and a modularity parameter $\sigma = p/q = 1000$. For scale-free network models we also generate 5 networks, each of which has 1000 nodes and 5000 edges with 10 communities and a modularity parameter of $\sigma = 0.01$. For each network we deploy up to 30 monitors in it using the algorithm proposed in [21] and randomly choose 1000 different infection sources to start the virus spreading. The spreading rate is $p = 50\%$. Once again we compare the cases in networks with and without community structures respectively. The average infection sizes are presented in Fig. 3.

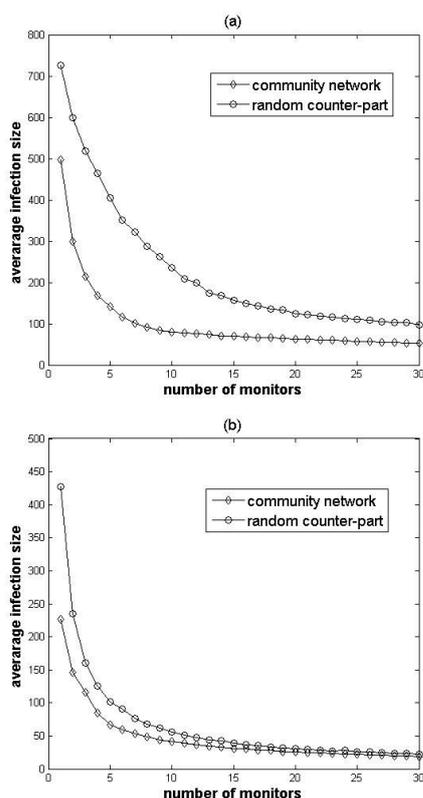


Fig. 3 Average infection size in networks equipped with a given number of monitors: (a) random networks; (b) scale-free networks

As we can observe, the existence of strong community structures does help reduce the average infection sizes in networks with a given number of monitors for infection detection. And the effects are much more significant in random networks than those in scale-free networks. Such observations can be explained: when there exist strong community structures

in networks, by using the algorithms in [21], monitors tend to be allocated on the “gateways” of the communities (i.e., the nodes with a large number of inter-community links) and high-degree hub nodes. Such allocations of monitors help reduce the infection to a fraction of a community before it is detected and stopped. This explains the significantly reduced average infection sizes in random networks. In scale-free networks, however, the existence of hub nodes plays a more significant role in epidemic spreading than the existence of communities. Since in both the community-structured networks and their counterpart community-less networks, the greedy algorithm in [21] shall make sure that monitors are allocated on high-degree hub nodes, the effects of the existence of community structures become much less impactful. Controlling the epidemic spreading in scale-free networks basically remains as a challenge, while the existence of community structures generally speaking may not help too much.

IV. CONCLUSION

In this paper, we evaluated the effects of community structures on epidemic spreading and control in complex networks. Specifically, by conducting extensive simulations, we studied on the effects of community structures on overall infection size, spreading speed and the average infection size in networks deployed with monitors, respectively. It was found that, generally speaking, the existence of community structures does not help reduce the overall infection size yet it may indeed slow down the spreading speed and hence helps buy more time before the vaccination could be available. In networks with monitors for infection detection, the existence of strong community structures helps reduce the average infection size since the infection sourced from a community has a very low chance to penetrate into any other community. Another important observation is that the effects of the strong community structures on epidemic spreading and control tend to be much more limited in scale-free networks than those in random networks. The existence of high-degree hub nodes persistently plays the most impactful role in epidemic spreading and control in scale-free networks, with or without community structures.

ACKNOWLEDGMENT

This work is partially supported by Ministry of Education (MOE), Singapore, under research grant RG 69/12.

REFERENCES

- [1] C. Qian, C. Hyunseok, R. Govindan, and S. Jamin, "The origin of power laws in Internet topologies revisited," *Proc. IEEE Infocom*, vol. 2 pp. 608-617, 2002.
- [2] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," *Proc. ACM SIGKDD*, 2005.
- [3] F. Liljeros, C. R. Edling, L. A. Amaral, H. E. Stanley, and Y. Aberg, "The web of human sexual contacts," *Nature*, vol. 411, pp. 907-908, 2001.
- [4] A.-L. Barabási, R. Albert, and H. Jeong, "Scale-free characteristics of random networks: the topology of the world-wide web," *Physica A*, vol. 281, pp. 69-77, 2000.
- [5] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509-512, October 15, 1999.

- [6] R. Cohen, K. Erez, D. ben-Avraham and S. Havlin, "Resilience of the Internet to random breakdowns". *Phys. Rev. Lett.*, vol. 85, pp. 4626–4628, 2000.
- [7] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, "Breakdown of the Internet under intentional attack". *Phys. Rev. Lett.* vol. 86, pp. 3682–3685, 2001.
- [8] S. Xiao, G. Xiao, and T. H. Cheng, "Tolerance of intentional attacks in complex communications networks," *IEEE Communications Magazine*, vol. 46, no. 1, pp. 146-152, Jan. 2008.
- [9] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, pp. 3200-3203, 2001.
- [10] R. Pastor-Satorras and A. Vespignani, "Immunization of complex networks," *Physical Review E*, vol. 65, p. 036104, 2002.
- [11] Y. Wang, G. Xiao, J. Hu, T. H. Cheng, and L. Wang, "Imperfect targeted immunization in scale-free networks," *Physica A*, vol. 388, no. 12, pp. 2535-46, 2009.
- [12] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, p. 026113, 2004.
- [13] M. E. J. Newman, "Detecting community structure in networks," *European Physical Journal B*, vol. 38, pp. 321-330, 2004/03/01 2004.
- [14] M. E. J. Newman, "Fast algorithm for detecting community structure in networks," *Physical Review E*, vol. 69, p. 066133, 2004.
- [15] J. Zhou, G. Xiao, L. Wong, X. Fu, S. Ma, and T. H. Cheng, "Generation of arbitrary two-point correlated directed networks with given modularity," *Physics Letters A*, vol. 374, pp. 3129-3135, 2010.
- [16] M. J. Barber, "Modularity and community detection in bipartite networks," *Physical Review E*, vol. 76, p. 066102, 2007.
- [17] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, pp. 814-818, 2005.
- [18] E. A. Leicht and M. E. J. Newman, "Community structure in directed networks," *Physical Review Letters*, vol. 100, p. 118703, 2008.
- [19] Z. Liu and H. Bambi, "Epidemic spreading in community networks," *EPL*, vol. 72, p. 315, 2005.
- [20] W. Huang and C. Li, "Epidemic spreading in scale-free networks with community structure," *Journal of Statistical Mechanics*, vol. 2007, p. P01014, 2007.
- [21] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," *Proc. ACM SIGKDD*, 2007.
- [22] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon, "On the uniform generation of random graphs with prescribed degree sequences," arXiv:cond-mat/0312028v2, 2004.
- [23] N. T. J., Bailey *The mathematical theory of infectious diseases and its applications*, volume 66, London, UK: Griffin, 1975.