An efficient method for restraining negative information cascades in online social networks

Yan Wang, Li Li, ZhaoHua Wang, XiaoHua Zheng
(School of Computer Science, Shaanxi Normal University, Xi'an 710119, China)
(E-mails: lili@snnu.edu.cn, wangy98@snnu.edu.cn)

***Abstract*—Information cascades is recognized as a major factor in disastrous social network phenomenons. The clustering of network community can block the cascade of negative information and users' behaviors in online social networks (OSNs). Confronting limited network resources, this paper takes the interaction relationship between network structure optimizing and neighbors' behaviors cascading as the breakthrough point, and study the restrain method of negative information cascades under emergency. This paper proposes a method for restraining the cascade of negative information, which is in order to improve the clustering of network communities by means of links that are logically removed (termed CLLR). By limited links with high betweenness being logically removed, CLLR method enhances the community density in a network, thereby effectively block information cascades. Experimental results show that CLLR method significantly improve the clustering of a network, efficiently block the speed and scope of information cascades in OSNs.**

Keywords-information cascades; the clustering of network community; links logically removed; restrain negative information

I. Introduction

Online social networks (OSNs) such as Twitter and Facebook have become an important part of people's daily lives. OSNs not only help people stay in touch with family and friends, but also keep abreast of breaking news and emerging contents. False or unverified information spreads just like accurate information on the web, thus possibly going viral and influencing the public opinion and its decisions [1-4]. Especially, information cascades are identified as a major factor in almost every disastrous social network phenomenon, ranging from viral marketing, rumor diffusion, cyber violence, and various negative information spread. During the COVID-19 outbreak in early 2020, various kinds of bad information on the Internet, such as "a certain place closed the city", "a certain place has a large-scale infection", etc., greatly misled the public and disrupted public order [5]. Therefore, effective methods for restraining negative information cascades are crucial for OSNs and it has been a hot topic in the last decades.

Vosoughi et al. pointed out that falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information; false news was more novel than true news, which suggests that people were more likely to share novel information [1]. Johnson et al. believe that policing within a single platform can make matters worse, and will eventually generate global 'dark pools' in which online hate will flourish [6]. In many cases, when the information provided to you by the people around you are more convincing than the information you get yourself, people tend to ignore their own information and choose to join the crowd. This situation is called information cascades [7]. Yang pointed out that in network emergencies, information cascades will have an impact on audience's cognitive bias [4]. The experiments designed by Anderson and Holt showed that the cascade is very easy to occur, and is also very fragile [8]. Therefore, it is of great significance to explore the characteristics of individual behaviors and reveal the influence of cascading effects on network information spread.

In a complex network, the network structure has an important impact on information spread. For example, the epidemic size decreases with the clustering [9]. The community structure in complex networks can help people better understand the function and organization of network systems and predict future trends [10]. In research on virus spread control, link reconnection strategy and link removal strategy are typical schemes to control virus propagation by changing the network structure [11-13]. Inspired by these studies, this paper restrains the negative information cascades by optimizing the network structure.

The main contributions of this paper are as follows:

- The clustering of network community not only reflects the closeness of network community structure, but also reflects the similar tendencies of users' behaviors.
- Homogeneity is often a barrier to information spread. Information cascades will stop when encounter a network community with relatively close internal connections.
- The method proposed in this paper effectively restrain negative information cascades by improving the clustering of network communities on the premise that the basic functions of the network are not affected.

II. Related Work

Currently, there are two kinds of methods for restraining negative information cascades in OSNs: the spread blocking method based on network structure optimization [12-17] and the spread suppression method based on network information competition [18-21]. The first method can quickly obtain a relatively stable control effect by directly blocking the critical path of information spread. However, in a real network, this method may make the network

978-1-6654-6131-3/22/$31.00 ©2022 IEEE

DOI 10.1109/NaNA56854.2022.00085

disconnected. The second method is to control indirectly, which is more suitable for online social networks.

Community structure generally exists in social networks. Compared with the sparse connections between communities, the connections between users within a community are denser. Members belong to the same community structure may hold similar views on the same matter. The clustering of network communities proposed in this paper emphasizes the similar tendencies of behaviors between users and their neighbors on the basis of the dense area structure. Homogeneity can often become an obstacle to spread: because of people tend to interact with their neighbors, and the spread of new things and new behaviors often comes from the "outside" world, so it is difficult for the spread of new things and new behaviors to enter a densely connected area. The waterfall will stop when it encounters a high-density community, which is the only reason why the waterfall stops spread [22]. That is to say, cascading and clustering are naturally opposites, and clustering can block cascading. Based on this conclusion, it can be seen that information cascades can be blocked by using the natural characteristics of the network structure. In addition, this conclusion also provides a theoretical basis for the intuitive understanding of blocking negative information cascades in the network.

Confronting limited network resources, this paper take the interaction relationship between network structure optimizing and neighbors' behaviors cascading as the breakthrough point, and study the restrain method of negative information cascades under emergency.

III. Clustering with Links Logically Removed Method

Confronting limited network resources, this paper take the interaction relationship between network structure optimizing and neighbors behavior cascading as the breakthrough point, and study the restrain method of negative information cascades under emergency. This paper proposes a method for restraining the cascade of negative information by improving the clustering with links logically removed (termed CLLR). By limited links with high betweenness being logically removed, CLLR method improves the clustering of the network communities in a network, thereby efficiently block information cascades.

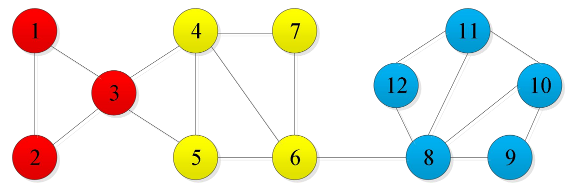
A. The Clustering

The community structure of the network is a dense connection branch in the whole network, which has the characteristics of dense interconnection between nodes in the same community and sparse interconnection between nodes in different communities. The clustering of network communities proposed in this paper emphasizes the similar tendencies of behaviors between users and their neighbors on the basis of the dense area structure.

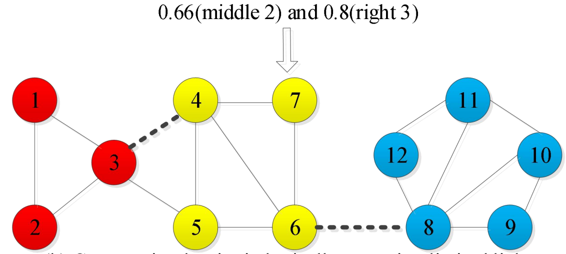
Definition 1 The clustering: The clustering of network communities not only reflects the closeness of the topology structure, but also reflects the similar tendencies of users' behaviors.

Definition 2 Community density: Community density can be used to quantify the clustering of network communities. The community density ρ means that each node in the community has at least a proportion ρ of neighbors with similar behaviors' tendencies in this community.

Let's take Fig.1 as an example to describe the change of community density. According to the definition of community density, there are three communities in Fig.1. The community 1 with a community density of 0.5 on the left, including nodes 1, 2 and 3. The middle community 2 with a community density of 0.66 includes nodes 4, 5, 6 and 7. The community 3 with a community density of 0.8 on the right, including nodes 8, 9, 10, 11 and 12.



(a) Community density in the initial graph is 0.5(left 1)、0.66(middle 2) and 0.8(right 3)



(b) Community density in logically removing limited links is 0.66(left 1)、0.66(middle 2) and 1.0(right 3)

Fig.1 An example of the change in community density

Densely connected regions and weak connections between regions have an important impact on information spread [7]. The links between different communities in the network are weak connections, which are generally have larger link betweenness. Link betweenness is often defined as the ratio of the number of paths passing through the link to the total number of shortest paths among all the shortest paths in the network [23]. If you consider the problem of graph partitioning, according to Girvan-Newman theory, it is an effective method to continuously remove links with high betweenness [24].

As shown in Fig.1, the connecting link (6-8) between nodes 6 and 8 is the link with the largest link betweenness. The community density of the three communities in Fig.1 changed after removing the two links (6-8) and (4-3) with the largest link betweenness. And the community density of community 1 and community 3 increased by 0.16 and 0.2, respectively. It can be seen that when temporarily removing or controlling links with large betweenness, it means removing the connection between the community structures. If there are no such connecting links between communities, the paths between many nodes will change, which will lead to longer propagation distances; at the same time, it can also be observed that removing links with high betweenness is an effective method to rapidly increase the community density of the network. In view of the fact that links with high betweenness are usually weak connections between different communities in the network. Removing the links with high betweenness means that the community density of the community where the two end nodes connected by the link are located is increased.

B. Information spread threshold

Threshold diffusion models reveal the advantage of weak connections [25], people we don't see often tend to form shortcuts in social networks. Taking Fig.1 as an example, the link(6-8) is the connecting link between community 2 and community 3, which is a weak connection. The message passing between different communities can be realized by the link (6-8).

The behavior of participating in the spread of false information has certain risks, and individual decision-making whether to participate in the spread of information should consider the risk and cost of participating in the behavior. This paper introduces the information spread threshold q to reflect the cost of participating in information spread. When the individual decides whether to participate in the information spread, it will judge according to the information spread threshold value and the tendencies of neighbors' behaviors, and use the tendencies of neighbors' behaviors with a high proportion to optimize its own decision-making.

Definition 3 Information spread threshold: The information spread threshold $q(q\in(0,1))$ is used to describe the cost of participating in information spread. The larger value of q, the greater harm of participating in information spread in the network, and the greater cost of individuals participating in information spread.

A lower value of q make the information spread faster to those areas in the current network that reject it. That is to say, the smaller q is, the lower proportion of neighbors participating in the information spread, which will prompt it choose to participate in the spread. Therefore, the lower information spread threshold q confirms the phenomenon that information with higher attention and lower threshold value is easier to spread on the network.

C. Clustering blocking cascading

Interactions between people tend to be localized, that is, people tend to interact with people who are close to each other, such as neighbors, friends, and colleagues. In a social network, if a user belongs to a community, most of the neighbors of the user also belong to the same community. When faced with a specific event, different users in the same community often hold the same or similar views and behaviors. It should be noted that at the edge of the community, different views on the same event are difficult to be absorbed by the community. People tend to interact with those around them, and it is difficult for external communication behavior to break into a tight community. Essentially, information cascades will stop when encounter a network community with relatively close internal connections [7].

D. CLLR algorithm

Considering the interaction between information cascades and topology evolution, by analyzing the interaction between network structure optimization and neighbors' behaviors cascading, the CLLR method is proposed to restrain negative information cascades. The basic idea of the CLLR method is improve the clustering of network communities by links being logically removed.

Given a network graph $G(N, E)$ with n nodes and m links. The basic steps of the CLLR algorithm are as follows:

Step 1. First, use the classical GN algorithm [24] to obtain the descending link remove order based on link betweenness in graph G;

Step 2. In the case of ensuring the network connectivity, use the BFS algorithm to determine the number $k1$ of removed links;

Step 3. Given the information spread threshold q, remove the links in sequence through the link betweenness in descending order, count the community density change of the largest community, and determine the number $k2$ of removed links;

Step 4. Then take the smaller value of $k1$ and $k2$ to determine the number k of removed links;

Step 5. Then logically remove a link that satisfies the condition according to the descending link remove order;

Step 6. Repeat step 5 until the number of removed links reaches k.

Specifically, the CLLR method has the following characteristics:

- Logically remove links. The CLLR method refers to temporarily removing links logically, and isn't equivalent to physically removing links. In the process of information spread, it may mean temporary isolation or other control measures.
- A limited number of links are removed. The number of links logically removed is jointly determined by the basic function of the network and the information spread threshold.
- Practical and effective method. The CLLR method does not completely cut off the connection between users; and links logically removed is easy to implement. This method efficiently block the speed and scope of information cascades.

IV. Simulation Results and Analysis

In this paper, the change of information spread threshold and community density are used to analyze and verify the feasibility and effectiveness of the CLLR method.

A. Dynamic SIS model

In the classic information spread SIS(Susceptible-Infected-Susceptible) model, there are only two states of nodes: susceptible state and infected state. In this paper, a dynamic SIS model based on network structure optimization and users decision-making is established by introducing dynamic information spread rate.

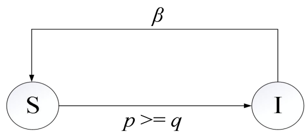


Fig.2 State transition graph in dynamic SIS model

In the dynamic SIS model, this paper focuses on how the users decide whether to participate in the information spread, that is, the transformation rule from the *S* state to the *I* state. From the concept of information spread threshold, participating in uncertain social movements has certain risks, and individual decision-making tends to be supported by a higher proportion of its neighbors. Parameter *q* represents the information spread threshold, which is used to describe the cost of users choose to participate in the information spread; parameter *p* represents the proportion of neighbors who choose to participate in the information spread; parameter β is the recovery probability that an individual changes from participating state to non-participating state, indicating that the individual withdraws from the information spread. The state transition relationship between the *S* state and the *I* state can be described as: when the *p* of the non-participating user is greater than or equal to the *q*, the *S* state of the user will be converted to the *I* state; otherwise, the *S* state is maintained; The *I* state is also converted to the *S* state with a probability of β.

In the dynamic SIS model, the information spread rate is not constant, and a user decides whether to participates in information spread is related to the information spread threshold and the similar tendencies of neighbors' behaviors.

B. Real network data set and parameter setting

The data set used in this experiment comes from the real network data set of the KONECT website, including Zachary network, Political Books network and Oz network. The topology characteristics of each network are shown in Table 1. Randomly select two initial spread nodes in the adjacent community in the given network, set the recovery rate β=0.20, and the information spread threshold q=0.20. Each curve value in the simulation experiment represents the average value of more than 100 runs.

Table 1 Network topology features of each network

Network	Node	Link	Average degree	Community
Zachary	34	78	4.59	2
Political Books	105	441	8.40	4
Oz	217	1839	16.95	6

C. Analysis of information spread threshold

In this paper, the information spread threshold *q* is used to describe the cost of participating in the information spread. The larger value of *q*, the greater cost of participating in the information spread, and the individual is more inclined to choose the behavior of participating neighbors with a high proportion.

Figs. 3-5 depict the changing trend of *I*(*t*) on different networks under different information threshold *q*. The *x*-axis represents the time, and the *y*-axis represents *I*(*t*) shows the ratio of the number of users that choose to participate in the information spread behavior to the total number of users in the network. The simulation experiment set 25 time steps.

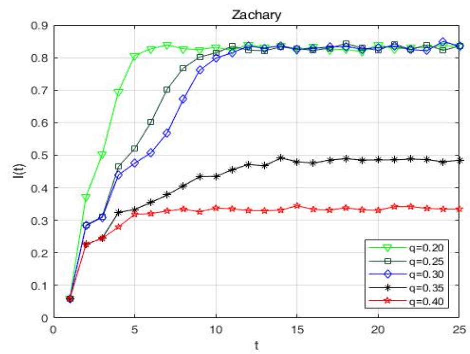


Fig.3 The changing trend of *I*(*t*) on Zachary network under different information spread threshold *q*

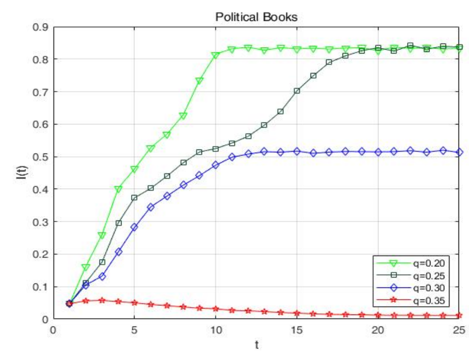


Fig.4 The changing trend of *I*(*t*) on Political Books network under different information spread threshold *q*

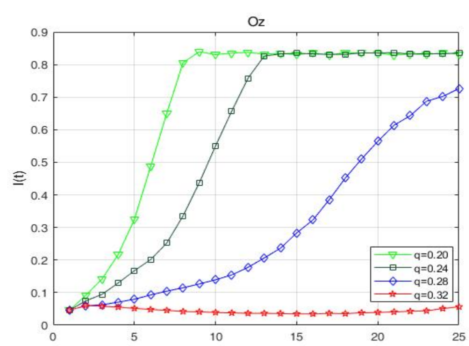


Fig.5 The changing trend of *I*(*t*) on Oz network under different information spread threshold *q*

It can be seen from the experimental results in Figs. 3-5 that: (1) When the information spread threshold is low, as long as a low proportion of neighbors choose to participate in the information spread, the individual will be followed, and the state of non-participating will become the state of participating. That is to say, the lower the information spread threshold, the lower the cost of participating in information spread, and the information is easier to spread in the network; (2) As the information spread threshold increases, the cost of participating in information spread is also increase, each user will tend to participate in the behaviors with more neighbors. In general, the larger the information spread threshold, the greater risk of participating in the information spread, individuals are more cautious when making choices.

D. Analysis of the change of community density

Table 2 shows the comparison of community density changes of each network with links logically removed based on the CLLR method. From the data in Table 2, it can be seen that the CLLR method can rapidly increase the community density in the whole network. Because clustering is an obstacle to cascading, the increase of network community density will block the spread of network information under a given information spread threshold.

Table 2 Comparison of community density changes of each network with links logically removed based on the CLLR method

Network		Comparison of community density changes with links logically removed					
Zachary		Community 1			Community 2		
	Initial community density	0.40			0.60		
	Reconfigured community density	0.80			0.75		
Political Books		Community 1		Community 2	Community 3		Community 4
	Initial community density	0.37		0.50	0.40		0.67
	Reconfigured community density	0.40		0.60	0.50		0.83
Oz		Community 1	Community 2	Community 3	Community 4	Community 5	Community 6
	Initial community density	0.25	0.25	0.21	0.17	0.38	0.35
	Reconfigured community density	0.31	0.38	0.33	0.46	0.50	0.33

E. Performance analysis based on simulation experiments

Figs. 6-8 depict the changing trend of $I(t)$ on different networks by different blocking methods with the same number of removed links. As we can seen from Figs. 6-8, the $I(t)$ curve under randomly remove links and degree-based remove links method shows that the nodes participating in information spread in the network have relatively. However, the $I(t)$ curve under the CLLR method shows that the proportion of nodes participating in information spread in the network is controlled within a small range, indicating that the CLLR method has significant advantages in blocking the speed and range of information spread.

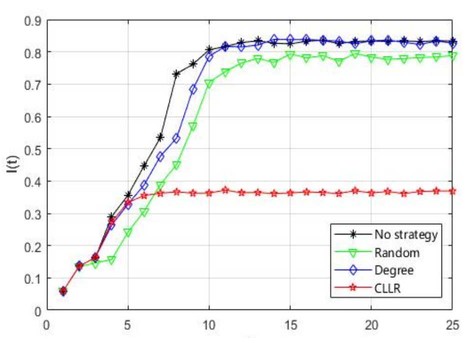


Fig.6 The changing trend of $I(t)$ on Zachary network by different restraining methods

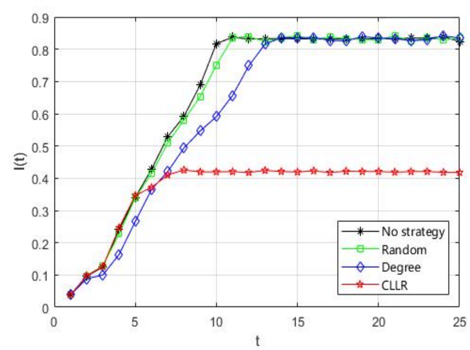


Fig.7 The changing trend of $I(t)$ on Political Books network by different restraining methods

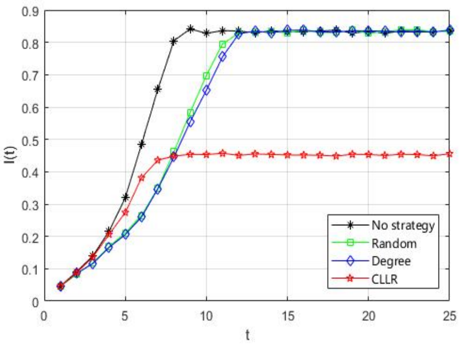


Fig.8 The changing trend of $I(t)$ on Oz network by different restraining methods

V. Conclusion

This paper is devoted to studying information cascades control in emergency response. Inspired by the clustering of network communities can block the cascade of negative information and users' behaviors in online social networks, this paper proposed a method for restraining negative information cascades, which is in order to improve the clustering of network communities by means of links that are logically removed (termed CLLR). This paper adopts the dynamic SIS model and validates the performance of the CLLR method with three real datasets. Experimental results show that the CLLR method proposed in this paper efficiently block the speed and scope of information cascades in OSNs, and the cost is small, and it is easy to implement.

VI. Acknowledgment

This work was supported by the National Natural Science Foundation of China (61303092), and the Shaanxi Province Natural Science Basic Research Foundation (2020JM-290, 2022JM-371).

References

[1] Vosoughi S, Roy D, Aral S. The spread of true and false news online [J]. Science, 2018, 359:1146–1151.

[2] Bondielli A, Marcelloni F. A survey on fake news and rumour detection techniques [J]. Information Sciences, 2019, 497:38-55.

[3] Zhang Z Y, Jing J C, Li F, et al. Survey on fake information detection, propagation and control in online social networks from the perspective of artificial intelligence [J]. Chinese Journal of Computers online, 2021, 44(11):2261-2282.

[4] Yang C Z. Research of effect of information cascade on cognitive bias in network emergency [J]. Journal of Intelligence, 2020, 39(2): 116-123, 160.

[5] Cai C C. Expert advice from public authorities on timely detection and disposal of bad information related to the epidemic situation legaldaily [EB/OL]. 2020.2.28.

[6] Johnson N F, Leahy R, Restrepo N J, et al. Hidden resilience and adaptive dynamics of the global online hate ecology [J]. Nature, 2019 (573): 261-265.

[7] David E, Jon K. Networks, crowds, and markets. Reasoning about a highly connected world [M]. Oxford city, Cambridge University Press, 2011.

[8] Anderson R L, Holt C A. Information cascades in the laboratory [J]. American Economic Review, 1997, 87(5): 847-862.

[9] Coupechoux E, Lelarge M. How clustering affects epidemics in random networks [J]. Advances in Applied Probability, 2014, 46(4): 985-1008.

[10] Yu E Y, Chen D B. Identifying critical edges in complex networks [J]. Scientific Reports, 2018,8:14469.

[11] Risau-Gusman S, Zanette D H. Contact switching as a control strategy for epidemic outbreaks [J]. Journal of Theoretical Biology, 2009, 257(1): 52-60.

[12] Yang X Y,Wu Y H and Zhang J H, "Analysis of rumor spreading with a temporal control strategy in multiplex networks" [J]. Journal of University of Electronic Science and Technology, 2020,49(04):511-518+554.

[13] Li L, Zhang R F, Du N N, et al. Virus propagation control strategy based on limited temporary links removed [J]. Journal of NanJing University (Natural Science), 2019, 55(4):651-659.

[14] Saha S, Adiga A, Prakash B A, et al. Approximation algorithms for reducing the spectral radius to control epidemic spread [C].//SIAM International Conference on Data Mining, (SDM), 2015: 568-576.

[15] Kimura M, Satio K, Motoda H. Blocking links to minimize contamination spread in social network [J]. ACM Transactions on Knowledge Discovery from Data. 2009,3(2):9.

[16] Khalil E B,Dilkina B, Song L. Scalable diffusion-aware optimization of network topology [J]. ACM Sigkdd International Conference on Knowledge Discovery and Data Mining ACM, 2014:1226-1235.

[17] Yao Z, Adiga A, Saha A, et al. Near-optimal algorithms for controlling propagation at group scale on networks [J]. IEEE Transactions on Knowledge and Data Engineering, 2016.28(12): 3339-3352.

[18] Budak C, Agrawal D, Abbadi A E. Limiting the spread of misinformation in social networks [C]. //20th international conference on World Wide Web,2011:665-674.

[19] He X R, Song G J, Chen W, et al. Influence blocking maximization in social networks under the competitive linear threshold model [C]. //SIAM International Conference on Data Mining (SDM12), 2012:463-474.

[20] Fan L D, Lu Z X, Wu W L, et al. Least cost rumor blocking in social networks [C]. //IEEE 2013 IEEE 33rd International Conference on Distributed Computing Systems (ICDCS), 2013: 540-549.

[21] Cao J X, Gao Q Q, Xia R Q, et al. Information propagation prediction and specific information suppression in social networks [J]. Journal of Computer Research and Development, 2021, 58(07):1490-1503.

[22] S. Morris. Contagion [J]. Review of Economic Studies, vol. 67: 57-78, 2000.

[23] Girvan M, Newman M E J. Community structure in social and biological networks [J]. Proceedings of the National Academy of Sciences, 2002, 99(12): 7821-7826.

[24] Damin C, Michael M. Complex contagions and the weakness of long ties [J]. American Journal of Sociology, 2007(113):702-734.

[25] Zhang G Q, Cheng S Q. Enhancing network capacity effects of edge-removal in small-world networks [J]. Chinese Science: Information Science, 2012, 42(2):151-160.