Enhancement of scale-free network attack tolerance

2010 Chinese Phys. B 19 110504

(http://iopscience.iop.org/1674-1056/19/11/110504)

Chin. Phys. B Vol. 19, No. 11 (2010) 110504

Rapid Communication

Enhancement of scale-free network attack tolerance*

Qu Ze-Hui(瞿泽辉)[a)b)†], Wang Pu(王 璞)[d)],
Song Chao-Ming(宋朝鸣)[b)c)], and Qin Zhi-Guang(秦志光)[a)]

[a)] *School of Computer Science and Engineering, University of Electric Science & Technology of China, Chengdu* 610054, *China*
[b)] *Center for Complex Network Research(CCNR), Department of Physics, Biology and Computer Science, Northeastern University, Boston, MA* 02115, *USA*
[c)] *Center for Cancer Systems Biology, Dana Farber Cancer Institute, Harvard University, Boston, Massachusetts* 02115, *USA*
[d)] *Department of Civil and Environmental Engineering, Massachusetts Institute of Technology Cambridge, MA* 02139, *USA*

(Received 6 July 2010; revised manuscript received 12 July 2010)

Despite the large size of most communication and transportation systems, there are short paths between nodes in these networks which guarantee the efficient information, data and passenger delivery; furthermore these networks have a surprising tolerance under random errors thanks to their inherent scale-free topology. However, their scale-free topology also makes them fragile under intentional attacks, leaving us a challenge on how to improve the network robustness against intentional attacks without losing their strong tolerance under random errors and high message and passenger delivering capacity. Here we propose two methods (SL method and SH method) to enhance scale-free network's tolerance under attack in different conditions.

Keywords: scale-free network, robustness spatial limited network, attack tolerance
PACC: 0520, 0250, 0547

1. Introduction

Scale- free topology is widely observed in many communication and transportation systems,[1−4] such as the Internet,[5,6] the World Wide Web (WWW),[7] the airline network[8] and the Wireless Sensor Network (WSN).[9] To maintain functionality, all these networks are characterized by the features of short path length[10,11] and high error tolerance. Not liking a random network, the node degree in these networks is not uniformly distributed but follows a power-law distribution,[11−19] implying that these networks are scale-invariant without a ‘typical’ node.[20] The scale-free property rooted in network inhomogeneous connectivity reduces the network attack survivability, making scale-free topology not a good candidate for communication and transportation systems when under an intentional attack.[21−24]

A challenge is left to us on how to improve the network robustness against intentional attacks without losing network short path length and robustness on random failures. In the communication network, we try to increase the network attack tolerance yet maintain fast and efficient delivery of messages. In the airline network, we hope to control the air traffic, avoiding the airline cancellations or delays caused by the problems happening at some airports. To explore the feasible and efficient methods for enhancing the attack tolerance for different kinds of networks, we propose two methods of the spatial limited network (where links are hard to re-link, i.e. the power grid network, the Internet router network) and the spatial unlimited network (i.e. the airline network and the Internet switcher network). We make the comparisons of the error and attack tolerance between the reconstructed networks and the original networks (the scale-free network constructed by the BA model, the Internet network and the airline network) and test the feasibility of our methodology.

The rest of the present paper is organized as follows. In Section 2, we introduce two methods of dealing with the two kinds of networks (the spatial limited

---

*Project supported in part by the China Scholarships Council (Grant No. 2007103794), the Defence Threat Reduction Agency Award HDTRA1-08-1-0027, the James S. McDonnell Foundation 21st Century Initiative in Studying Complex Systems, the National Science Foundation within the DDDAS (CNS-0540348), ITR (DMR-0426737) and IIS-0513650 programs, the US Office of Naval Research Award N00014-07-C, the National Natural Science Foundation of China (Grant Nos. 80678605 and 60903157), and the National High Technology Research and Development Program of China (Grant No. 2009AA01Z422).

†Corresponding author. E-mail: zehuiqu@163.com

http://www.iop.org/journals/cpb http://cpb.iphy.ac.cn

network and the spatial unlimited network). In Section 3, we provide a theoretical analysis on the average path length and the degree distribution for the reconstructed networks. In Section 4, we use our methods in BA model and two real networks to test their feasibilities. Conclusions are given in Section 5.

2. Methodology

The basic starting point of our methodology is to enhance the network attack tolerance by decreasing a Pc fraction of the hub degree without introducing a big increase to the network average path length l. P_c can be selected appropriately according to the real conditions. For the spatially unlimited network whose links can be easily re-linked (see switcher networks in Fig. 1(a)), the switch link (SL) method provides us an economic way to improve the network attack tolerance without adding new expensive nodes. In the SL method, we first address two links connecting a hub and then we keep the first link unchanged, disconnect the second link from the hub and connect it to the non-hub node connected by the first link (see Fig. 1(b)). For the spatially limited networks, the SL method is not economic and feasible because nodes are usually far from each other (see router networks in Fig. 1(a)). In this case we propose another method, i.e. split hub (SH) method. In the SH method, we replace the hub node by a 3-clique (see Fig. 1(b)). This method is suited for the networks where new nodes can be easily added. Figures 1(a) and 1(b) show the original Internet network and the reconstructed Internet network (which is composed of the switcher networks and the router networks) constructed using these two methods, respectively. We will introduce the details of these two methods in the following.

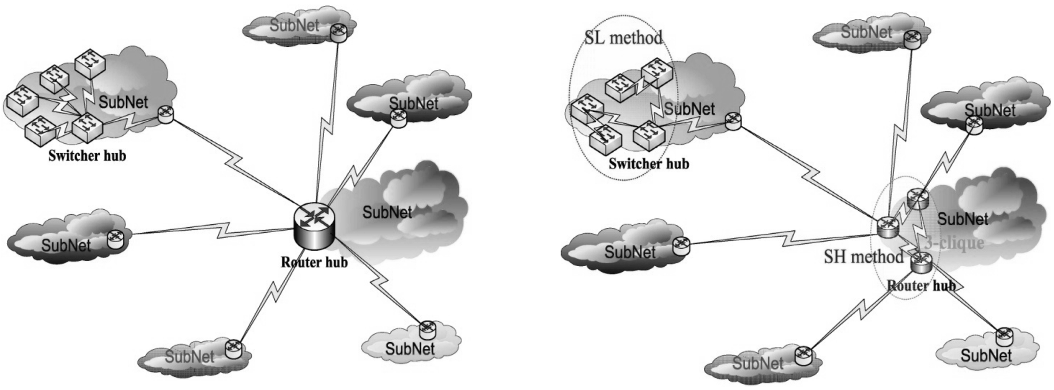

Fig. 1. SL and SH methods. In panel (a), the original Internet network is illustrated as the connections of the routers and some sub networks in which switchers are connected, the distances between the routers are not negligible. In the sub networks, there are many switchers close to each other. In panel (b), the reconstructed Internet network using the SL and SH methods. For the router hub, because the links between the routers are spatially limited we use the SH method. For the switchers in the sub networks, using the SL method is more appropriate, because switchers are near each other and the links between them are easily re-linked.

2.1. Switch link (SL) method

In this method, we define the top P_c fraction of the large degree nodes as hubs. For each hub, we can find two non-hub nodes connecting it. We keep the link connecting the first non-hub node and switch the link connecting the second non-hub node to the first non-hub node. We repeat this process until all the links connected to the hub nodes are addressed (If a hub has an odd degree, we do not deal with the last link). The detailed method is: (1) for each hub node, we select two non-hub nodes connecting it; (2) keep the link connecting the first non-hub node; (3) switch the link connecting the second non-hub node to the first non-hub node;. (4) repeat steps 2 and 3 until all the links from the hub have been considered; (5) repeat steps 1–4, until all the hub nodes are addressed.

2.2. Split hub (SH) method

In this method, we again define the top P_c fraction of nodes as hubs and replace the hub nodes by 3-cliques.[15] The detailed method is: (1) select one hub node; (2) replace the hub by a 3-clique; (3) connect the non-hub nodes which connect the original hub node randomly to the 3-clique nodes; (4) repeat steps

1–3 until all the hub nodes are addressed.

3. Analytical results

In this section, using the BA model as a sample, we analytically investigate the reconstructed network properties such as average path length l and degree distribution $P(k)$, demonstrating that our methods can preserve network small world property and scale-free topology.

3.1. Average path length l

After reconstructing the network, the average length l will change. We assume that a path with length l has l-1 nodes between the source and the destination and we also assume that $P_{\mathrm{h}l}$ ($P_{\mathrm{hl}} < 1$) is the fraction of hub nodes in the path. For the SL method, half of the non-hub nodes connecting the hubs are re-linked to the other half of the non-hub nodes, which makes the average path length have an increase of at most $\frac{1}{2}\left(l-1\right)P_{\mathrm{h}l}$, where it was not considered that re-linking could reduce the path length. Hence the range of the new average path length l' for the reconstructed network with using the SL method is $l' \geq l$ and $l' \leq \frac{1}{2}\left(l-1\right)P_{\mathrm{h}l} + l$. For the SH method, we use the 3-clique as the representation of stuff, the new average path length of the reconstructed networks lies between $l' \geq l$ and $l' \leq \frac{2}{3}\left(l-1\right)P_{\mathrm{h}l} + l$. These analytical results provide the lower and the upper bounds of l', demonstrating further that the network small world property will not change because l' and l have the same scales of magnitude.

3.2. Degree distribution $p(k)$

There are two categories of nodes: hub nodes and non-hub nodes. In the following, we theoretically calculate the new degree distribution $p'(k)$ in the reconstructed network to demonstrate that our SL and SH methods will not change the network scale-free property.

For the SL method from Ref. [16] the probability $p\left(k,i\right)$ of finding a link between two nodes with degrees k and i in the BA model is given by

$$p\left(k,i\right) = \frac{4\left(i-1\right)}{k\left(k+1\right)\left(k+i\right)\left(k+i+1\right)\left(k+i+2\right)} + \frac{12\left(i-1\right)}{k\left(k+i-1\right)\left(k+i\right)\left(k+i+1\right)\left(k+i+2\right)}, \tag{1}$$

where the nodes with degree larger than k_{c} are regarded as hub nodes.

In original network the probability of nodes with degree k is

$$p\left(k\right) = \alpha k^{-\gamma}.$$

The probability of nodes with degree k in the reconstructed network is given by

$$\begin{aligned} p'\left(k\right) = & \sum_{i=k_{\mathrm{c}}}^{\infty}\left(1-p\left(k,i\right)\right)p\left(k\right) \\ & + \frac{1}{2}\sum_{v=1}^{\infty}\sum_{i=k_{\mathrm{c}}}^{\infty}\left(p\left(v,i\right)p\left(v\right) - \frac{1}{2}p\left(k,i\right)p\left(k\right)\right) \\ & + \frac{1}{2}t\left(2k\right), \end{aligned} \tag{2}$$

where

$$t\left(2k\right) = \begin{cases} p\left(2k\right), & k \geq \frac{1}{2}k_{\mathrm{c}}, \\ 0, & k \leq \frac{1}{2}k_{\mathrm{c}}. \end{cases}$$

The first term on the right-hand side of Eq. (2) represents the new probability for the nodes without their connecting to the hubs, where the node degrees have no change. The second term represents the new probability for nodes with degree less or larger than k in the original network and with degree k after using the SL method. The case of $v = k$ has been calculated both in the first and second terms, so we substrate it from the third term. The last term represents the hub node whose degree k is larger than k_{c}, its degree in the reconstructed network is equal to half of its original degree.

For the SH method, the degree distribution is

$$p'\left(k\right) = \begin{cases} \frac{1}{3}p\left(3k\right), & k > k_{\mathrm{c}}, \\ p\left(k\right) + \frac{1}{3}p\left(3k\right), & \frac{1}{3}k_{\mathrm{c}} \leq k \leq k_{\mathrm{c}}, \\ p(k), & k < \frac{1}{3}k_{\mathrm{c}}, \end{cases} \tag{3}$$

where $p(k) = \alpha k^{-\gamma}$ and $p(3k) = \alpha(3k)^{-\gamma} = 3^{-\gamma}\alpha k^{-\gamma} = 3^{-\gamma}p(k)$.

There are three cases taken into account in Eq. (3): (1) in the reconstructed network, the hub nodes remain hub nodes; (2) hub nodes in the original network become non-hub nodes in the reconstructed network; (3) the degrees of non-hub nodes in the original network will not change.

We calculate the new degree distribution $p'(k)$ using different k_c values. For different k_c values the new degree distributions almost overlap with the original degree distribution, uncovering that our methodology will not change the network scale-free topology (see Fig. 2).

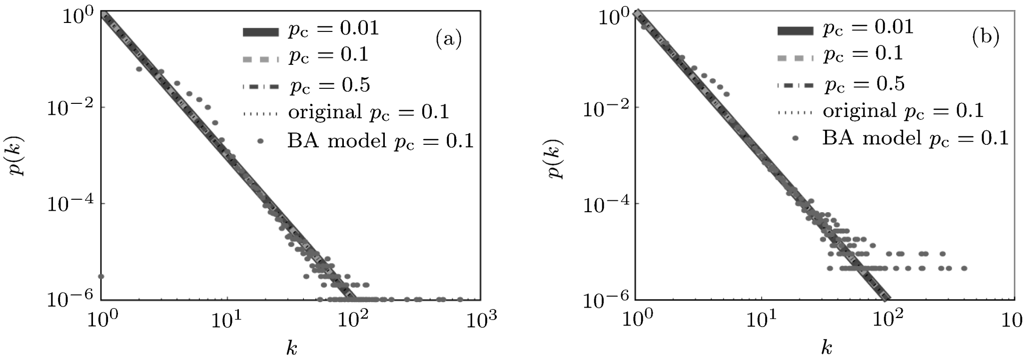

Fig. 2. Network degree distribution re-constructed by SL (a) and SH (b) methods, showing similar scaling laws to that of the original network constructed by BA model.

4. Experimental results and discussion

In this paper, we use a BA model network (20000 nodes and 101349 links), an AS level map of Internet (22963 nodes and 48436 links) network and an airline network (203 nodes and 4142 links) to test our methodology. The average path length l is an important parameter to evaluate the network efficiency for delivering message and passengers. Hence we look at the increase of l in the reconstructed networks. For the BA-model network, l increases less than 20% both for the SL method and SH method (see Fig. 3(a)). Figure 3(b) shows that l of the reconstructed airline network increases at most 48% for the SH method and 20% for the SL method (see Fig. 2(b)). The l for the reconstructed Internet network increases at most 22% for the SH method and 12% for the SL method (see Fig. 3(c)). All of these results show that our methods will not introduce big increase in the path length of the networks from technology to infrastructure field: the short path lengths of these networks are well maintained. Furthermore, the SL method introduces less increase in l in the airline network and the Internet network, hinting that the SL method is the preferential option if condition is permissible.

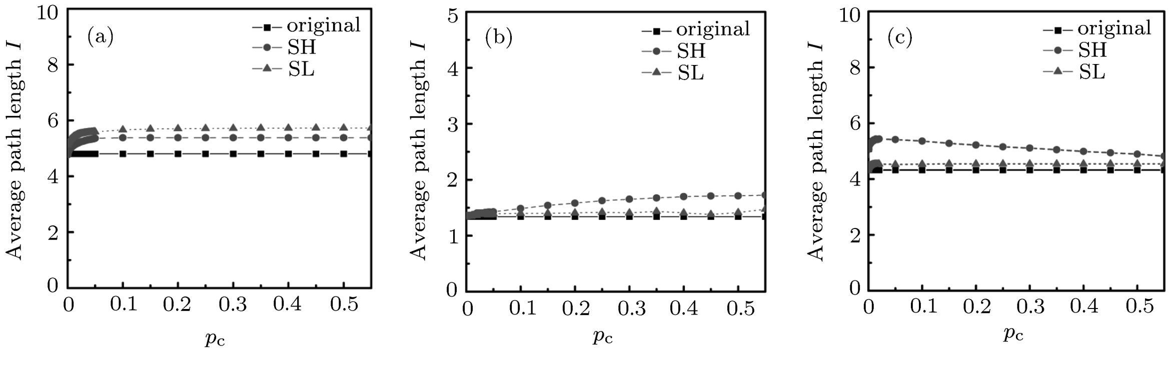

Fig. 3. Short path lengths for the reconstructed networks which are preserved. Panel (a) shows the average path length of the reconstructed BA-model network with 20000 nodes and average degree 10, panel (b) indicates the average path length of the reconstructed airline network and the original airline network with 203 nodes and an average degree 40, and panel (c) exhibits the average path length of the reconstructed Internet network with 22963 nodes and 48436 links.

The γ parameter is an important parameter to evaluate the network topology and structure. It gives the slope of the degree distribution. By testing this parameter, we find that the scale-free property is also well maintained using the SL and the SH methods. In Fig. 4, all of the dot lines show the ratio between γ value of the reconstructed network constructed using the SL method and that of the original network. The γ ratio increases at most 20% when P_c increases from 0 to 0.5, where P_c is the percentage of the most connected nodes.

For the SH method, all of the solid lines show that the γ ratio increases at most 10% (see dash linesin Fig. 3).

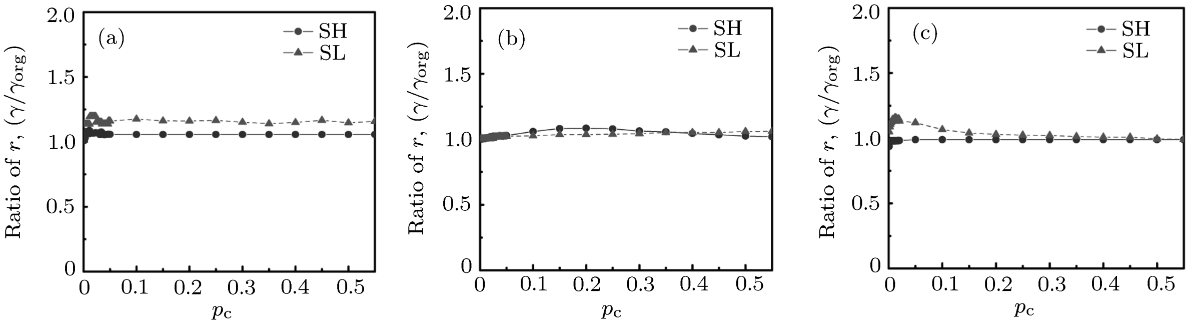

Fig. 4. Scale-free properties for the reconstructed networks, which are preserved. Panel (a) shows that ratio between γ value of the reconstructed BA-model network and that of the original BA-model network has a slight increase but smaller than 1.2, panel (b) for air line network and panel (c) for the Internet network show the same scenarios as panel (a).

Next we start to test the network error and attack tolerance. Here, we consider the top 1% nodes as hubs and perform the error and the attack experiments for all of the three networks mentioned above. To test their error tolerance, we randomly remove f fraction of the nodes, the giant component size S of each reconstructed network has a similar behaviour as that of the original networks as f increases. This implies that our methods keep the network surviving ability under random errors (see Fig. 5).

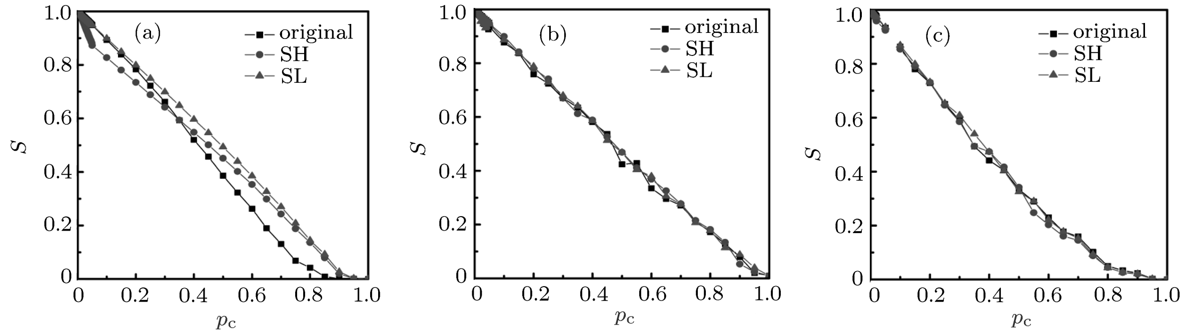

Fig. 5. Random error tolerances for the SL and SH methods, which are well maintained. Panel (a) shows sizes of giant component under random errors for BA-model network and the reconstructed BA model network using SL and SH methods, panel (b) for airline network and panel (c) for Internet network show the same scenarios as panel (a).

Next we perform the intentional attacks by removing f fraction of the most connected nodes. For the Internet network, we observe that the giant component in the reconstructed network can survive even when twice the high-degree nodes are removed compared as the original one (see Fig. 6(b)).

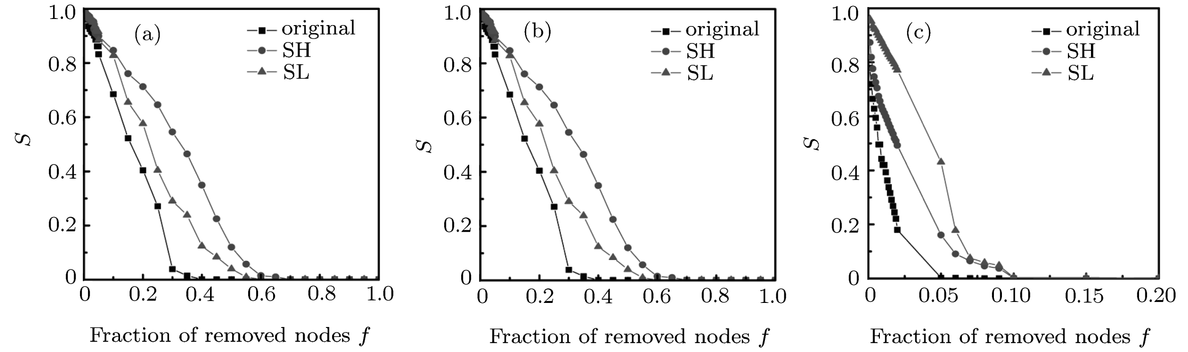

Fig. 6. Enhancement of attack tolerance for SL and SH methods. Panel (a) shows sizes of giant component after intentional attacks for the scale-free network constructed by BA-model and the networks after the structure optimization using our SL and SH methods. For the original airline network, the network fully falls apart at $f_{\rm c} = 0.4$, however $f_{\rm c} = 0.55$ and $f_{\rm c} = 0.65$ for the reconstructd network based on the SH and SL method, panel (b) for the air line network and panel (c) for the Internet network show the same scenarios as panel (a).

For the BA model network, there is still a giant component existing when more than 50% of most connected nodes are removed (see Fig. 6(a)). The reconstructed airline network also has larger $f_{\rm c}$ than the original network (see Fig. 6(b)). Hence the networks become much stronger and robust than the original ones under intentional attacks by our optimizing structure methods.

The underlying reason for these results can be explained into the increase of parameter γ for the scale-free networks. For the SL method, the degrees of selected hubs will become half of their original degrees after using the SL method, at the same time, the degrees of nodes having an edge connecting to the hub will increase, which results in the more even degree distribution. The degree distribution also becomes more average after reconstructing the network by the SH method. For a selected hub, its degree will decrease to $1/N$ of its original degree when the hub is replaced by the N-clique.

5. Conclusions

In this paper, we have proposed the SL and the SH methods to enhance the robustness of the scale-free networks under intentional attacks. We consider scale-free as a crucial topology related to the robustness of the networks and find that the networks re-constructed by using the SL or the SH method can maintain the network functionality and improve the network attack tolerance. The significant finding is that the robustness of the networks is improved greatly when the parameter γ in the scale-free networks changes a little. On the other hand, the average path lengths only have a small difference between the reconstructed networks and original networks. We also compare the advantages of SL and SH methods in different conditions. In the airline network and the Internet network, the SL method introduces a smaller increase in average path length; the SH method works on the spatially limited networks. We can conclude that the SL and the SH methods can be used to improve the attack tolerance and maintain the functionality for scale-free network by choosing an appropriate $P_{\rm c}$ value. Our modeling framework provides a fundamental tool for enhancing the security of different kinds of networks.

Acknowledgement

We thank Albert-László Barabási, Marta González, Dasun Wang, Yang Liu and James Bagrow for discussions and comments on the manuscript.

References

[1] Onnela J P, Saramaki J, Hyvonen J, Szabo G, Lazer D, Kaski D, Kertesz J and Barabási A L 2007 *Proceedings of the National Academy of Sciences* **104** 7332

[2] Wang P, Gonzalez M C, Hidalgo C A and Barabási A L 2009 *Science* **324** 1071

[3] Hu H, Myers S, Colizza V and Vespignani A 2009 *Proceedings of the National Academy of Sciences* **106** 1318

[4] Yook S H, Jeong H and Barabási A 2002 *Proceedings of the National Academy of Sciences* **99** 13382

[5] Park J and Newman M E J 2003 *Phys. Rev.* E **68** 026112

[6] Pastor-Satorras R, Vázquez A and Vespignani A 2001 *Phys. Rev. Lett.* **87** 258701

[7] Albert R, Jeong H and Barabási A L 1999 *Nature* **401** 130

[8] Colizza V, Barrat A, Barthélemy M and Vespignani A 2006 *Proceedings of the National Academy of Sciences* **103** 2015

[9] Li C J and Chen G R 2006 *Modelling of Weighted Evolving Networks with Community Structures* **370** 869

[10] Lewis F L *Wireless Sensor Networks. Smart Environments: Technologies, Protocols, and Applications* (New York: Wiley Interscience)

[11] Wu X J and Lu H T 2010 *Chin. Phys.* B **19** 070511

[12] Duncan W, Strogatz J and Steven H 1998 *Nature* **393** 440

[13] Albert R and Barabási A L 2002 *Rev. Mod. Phys.* **74** 47

[14] Barabási A L and Albert R 1999 *Science* **286** 509

[15] Wang X H, Jiao L C and Wu J S 2001 *Chin. Phys.* B **19** 020501

[16] Pu C L and Pei W J 2010 *Acta Phys. Sin.* **59** 3841 (in Chinese)

[17] Callaway D S, Newman M E J, Strogatz S H and Watts D H 2000 *Phys. Rev. Lett.* **85** 5468

[18] Palla G, Barabási A L and Vicsek T 2007 *Nature* **446** 664

[19] Guo J L and Wang L N 2007 *Acta Phys. Sin.* **56** 5635 (in Chinese)

[20] Krapivsky P L and Redner S 2001 *Phys. Rev.* E **63** 066123

[21] Li T, Pei W J and Wang S P 2009 *Acta Phys. Sin.* **58** 5903 (in Chinese)

[22] Li C J and Maini P K 2005 *J. Phys.* A **38** 9741

[23] Feng C F, Guan J Y, Wu Z X and Wang Y H 2010 *Chin. Phys.* B **19** 060203

[24] Albert R, Jeong H and Barabási A L 2000 *Nature* **406** 482