International Journal of Modern Physics C
Vol. 25, No. 5 (2014) 1440005 (10 pages)

DOI: 10.1142/S0129183114400051

Cascading failures in complex networks with community structure

Guoqiang Lin, Zengru Di and Ying Fan*

School of Systems Science
Beijing Normal University
Beijing 100875, P. R.China
**yfan@bnu.edu.cn*

Received 30 June 2013
Accepted 30 September 2013
Published 2 December 2013

Much empirical evidence shows that when attacked with cascading failures, scale-free or even random networks tend to collapse more extensively when the initially deleted node has higher betweenness. Meanwhile, in networks with strong community structure, high-betweenness nodes tend to be bridge nodes that link different communities, and the removal of such nodes will reduce only the connections among communities, leaving the networks fairly stable. Understanding what will affect cascading failures and how to protect or attack networks with strong community structure is therefore of interest. In this paper, we have constructed scale-free Community Networks (SFCN) and Random Community Networks (RCN). We applied these networks, along with the Lancichinett–Fortunato–Radicchi (LFR) benchmark, to the cascading-failure scenario to explore their vulnerability to attack and the relationship between cascading failures and the degree distribution and community structure of a network. The numerical results show that when the networks are of a power-law distribution, a stronger community structure will result in the failure of fewer nodes. In addition, the initial removal of the node with the highest betweenness will not lead to the worst cascading, i.e. the largest avalanche size. The Betweenness Overflow (BOF), an index that we developed, is an effective indicator of this tendency. The RCN, however, display a different result. In addition, the avalanche size of each node can be adopted as an index to evaluate the importance of the node.

Keywords: Cascading failures; community structure; power-law distribution; LFR benchmark.

PACS Nos.: 89.75.−k, 89.75.Da.

1. Introduction

The world in which we live is supported by large, complex networks, including technology networks, social networks, information networks and biological networks.[1,2] These networks all share the attributes of a small average distance between nodes and an organized distribution of links (or degree) per node.[3–5] Generally, the

*Corresponding author.

average distance will change very little if a randomly selected subset of the nodes is removed. However, when the intrinsic dynamics of the flows of physical quantities in the network is taken into account, the removal of nodes can have a much more devastating consequence.[6] Taking power grid as an example, each power station (node) handles maximum load of power. The removal of nodes changes the distribution of the flows over the network, which can trigger a cascade of overload failures.[7] The blackout that occurred on August 10, 1996, in the western United States power grid is a vivid example.[8] To investigate cascade-based attacks on complex networks, Motter and Lai introduced a model using the betweenness as the load of each node and assuming that the capacity of a node is proportional to its initial load (betweenness). They demonstrated that heterogeneity of a network makes it particularly vulnerable to attacks because a large-scale cascade may be triggered by disabling a single key node with an exceptionally large load.[6] Meanwhile, Newman *et al.* examined cascading failure in a simplified transmission system model and found two types of critical points respectively, characterized by transmission line flow limits and generator capability limits.[9] In 2008, Wang and Chen investigated cascading failure on weighted complex networks by adopting a local weighted flow redistribution rule and developed a tunable parameter to regulate the robustness of networks.[10]

However, the scale-free networks that Motter and Lai considered are randomly generated.[11] Because nodes link with each other randomly, there is little chance that these scale-free networks will have strong community structure. The following questions therefore remain: What will happen if the networks are of strong community structure? What will affect cascading failures in them? Is betweenness still a valuable index in these networks? Last but not least, it is of interest to explore what type of node will incite the worst avalanche size. In other words, we can evaluate the importance of each node by the cascading failures it causes.

2. Community Network Model and Cascading Model

2.1. *Random community networks*

First, we have built a type of graph with c ($c \geq 2$) communities. This model is constructed as follows: Assuming that the total number of nodes is n, the average degree is $\langle k \rangle$, and each community has $\frac{n}{c}$ nodes, then the total number of edges of the network E_{total} is $\frac{1}{2}n\langle k \rangle$. Then, we set

$$\frac{E_{\text{intra}}}{E_{\text{intra}} + E_{\text{inter}}} = \sigma,$$

where E_{intra} is the number of edges inside each community, while E_{inter} is the number of edges between every two communities. We then randomly connect nodes. Because all of the edges are randomly linked and the graph has a certain number of strong communities, we name it the Random Community Network (RCN). Figure 1 is a visualization of this model.

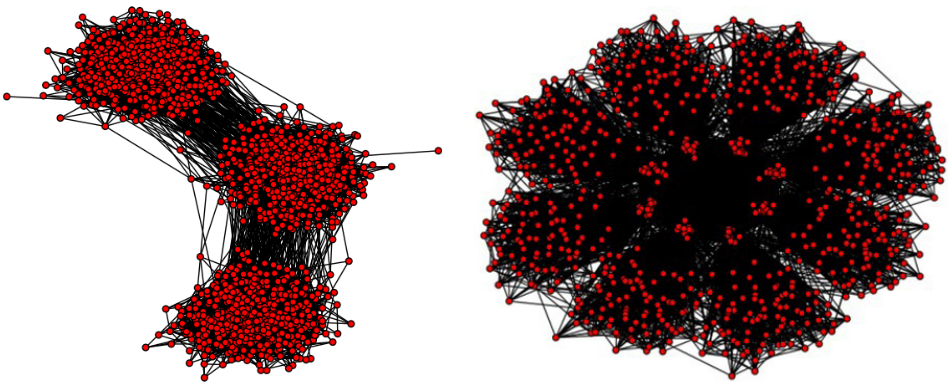


Fig. 1. (Color online) Graph(a) is an instance of Random Community Network, with $n = 1500, \langle k \rangle = 4, c = 3$ and $\sigma = 0.95$, such that $Q = 0.633$. Graph(b) is an instance of Scale-free Community Network, with $s = 12, m = 13$ and $c = 8$, such that $Q = 0.798$.

As first proposed by Newman and Girvan proposed and later developed by Kashtan and Alon,[12,13] the strength of community structure can be quantified by

$$Q = \Sigma_1^c \left(\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right), \tag{1}$$

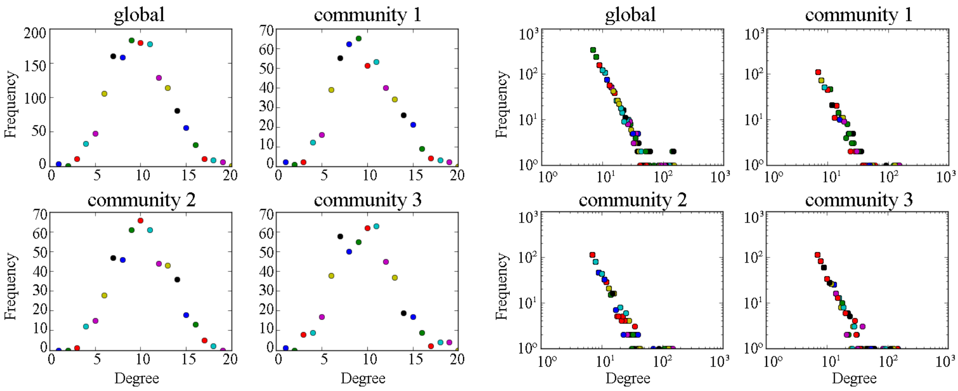
where c is the number of communities, L is the number of total edges in the network, l_s is the number of edges between nodes in community U_s and d_s is the sum of degree of nodes in community U_s. Then, we can obtain community strength of RCN Q_{RCN} as

$$Q_{\text{RCN}} = \sigma - \frac{(1+\sigma)^2}{4c}, \tag{2}$$

where $c \geq 2$. Because the nodes connect with each other randomly, no matter whether they are in the same community, the degree distribution of the whole graph and those of each individual community are Poisson distributions, as Fig. 2(a) shows.

2.2. *Scale-free community networks*

Second, we adopt the growth model proposed in Ref. 14 to generate a scale-free network with strong community structure. we name it the scale-free Community Network (SFCN). This model is constructed as follows: Initially, we start with c communities, denoted by $U_1, U_2, \ldots, U_c$, each with a small number (m_0) of nodes; to ensure network connectivity, the initial $m_0 c$ nodes link to one another. Then, at every time-step, we add a new node with $m(m < m_0)$ edges to each community. Among the m edges, s edges are to be linked to s different nodes in this community



(a) (b)

Fig. 2. (Color online) An illustration of the degree distributions of RCN (a) and SFCN (b) Each community or global graph is a power-law distribution or a Poisson distribution, respectively.

and $m - s$ edges linked to other $c - 1$ communities. Each link obeys the "preferential-attachment" rule,[5] where the probability P that the new node will connect to node i, if node i belongs to community U_l, depends on the degree (k_i) of node i, as follows:

$$P(k_i) = \frac{k_i}{\Sigma_{(j \in U_l)} k_j}.$$

If node i belongs to one of the other $c - 1$ communities, these communities are treated as a whole and the preferential-attachment mechanism is then applied in the same manner. Figure 1 is illustrates an example of the model. By changing the number of communities and the values of m and s, we can also generate Barabsi–Albert (BA) networks.

As we did for the RCN, we can obtain the community strength of the SFCN:

$$Q_{\mathrm{SFCN}} = \frac{s}{m} - \frac{1}{c}. \tag{3}$$

We can therefore adjust the strength of community structure by changing the values of s, m and c.

Furthermore, we can deduce the degree distribution of SFCN, which is as follows:

$$f(k) = \frac{1}{\theta} m^{\frac{1}{\theta}} k^{-(\frac{1}{\theta}+1)}, \tag{4}$$

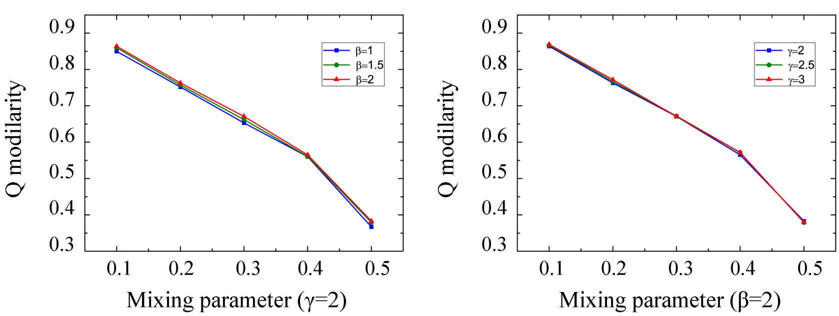
where θ is equal to $\frac{1}{2}(\delta + \frac{1-\delta}{c-1})$, $\delta = \frac{s}{m}$ and $c \geq 2$. Therefore, the degree distribution of the SFCN is a power-law with an exponent of the form $-(\frac{1}{\theta} + 1)$. Set $s = m$, then the exponent will be -3, the same exponent as the BA network. In Fig. 2(b), we illustrate the degree distribution of SFCN.

2.3. *LFR benchmark*

Because the community size of SFCN is homogeneous, we have explored the Lancichinetti–Fortuanto–Radicchi (LFR) benchmark,[15] which has not only a power-law degree distribution but also an adjustable strength of community structure with a heterogeneous community size that is also of a power-law distribution. We can generate an LFR benchmark of a certain strength of community structure by adjusting three tunable parameters: the degree-distribution exponent γ, the community-size-distribution exponent β and the mixing parameter that indicates the ratio of outside community edges to inside community edges. We have explored the relationship between the Q modularity and these three parameters. It is clear that the Q modularity decreases as the mixing parameter increases. When we set the degree-distribution exponent γ equal to 2, the Q modularity increases slightly as community-size-distribution exponent β increases with a fixed mixing parameter, as shown in Fig. 3(a). However, if we set the community-size-distribution exponent β to 2, the Q modularity shows no significant change as γ increases, as Fig. 3(b) shows. This result means that the degree distribution has little effect on the strength of community structure of a network as long as it follows a power-law distribution.

2.4. *Cascading model*

In Ref. 6, Motter and Lai proposed a cascading model. For a given network, assume that at each time-step one unit of the relevant quantity, which may be information or travel flow, is exchanged between every pair of nodes and travels along the shortest path linking them. The load at a node is the total number of shortest paths passing through the node, i.e. the betweenness of the node. In real-world networks, the capacity is the maximum load the node can handle, which is always limited by cost. The higher the budget is, to some extent, the larger the capacity is. Thus, the authors



(a) (b)

Fig. 3. (Color online) Relationship among Q modularity and γ, β and mixing parameter. In subgraph (a), the degree-distribution exponent γ is set to 2, while the community-size exponent β varies among 1, 1.5 and 2, denoted by the blue, green and red lines, respectively. In subgraph (b), we set β equal to 2, while γ ranges from 2 to 3. The figures are generated from 5 implementations of LFR networks.

assume that the capacity C_j of node j is proportional to its initial load L_j,

$$C_j = (1 + \alpha)L_j, \quad j = 1, 2, \ldots, N, \tag{5}$$

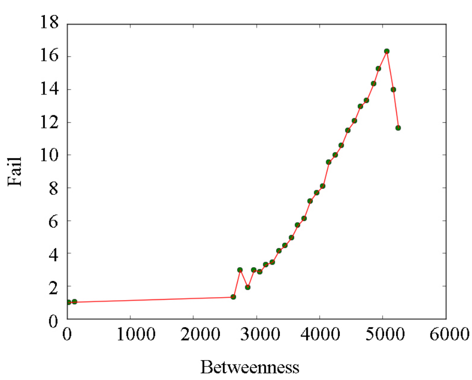
where the tolerance parameter α is tunable and always larger than zero, L_j is the initial betweenness of node j and N is the initial number of nodes. In our research, we set the tolerance parameter α permanently equal to 0.1.

3. Results and Analyses

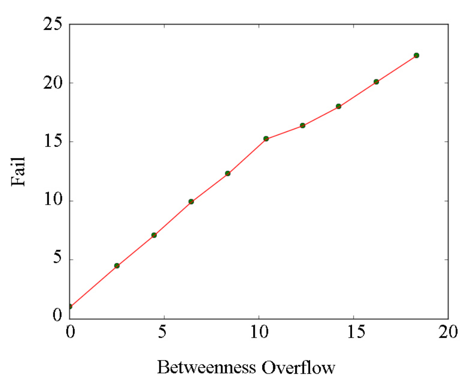
3.1. *Randomly generated scale-free networks versus SFCN*

For randomly generated scale-free networks with a given degree sequence that follows a power-law distribution,[11] we can remove the node with the highest betweenness to achieve the worst cascading failure,[6] i.e. to cause the failure of the largest number of nodes. Here, the betweenness well evaluates the importance of the node. A node with larger betweenness is more important than those of smaller betweenness.

However, randomly generated scale-free networks have weak community structure. To prove this claim, we generate 10 networks. Considered the complexity of the algorithm and computation time, each network has 1000 nodes and a power-law degree distribution, with an exponent of −3. We then use the Extremal Optimization (EO) method[16] to detect their network structures, and we find that the average Q modularity is 0.275. For a scale-free network with strong community structure ($Q = 0.798$ in our instance), we can see that the node that causes worst cascading failure is not the node of largest betweenness, as shown in Fig. 4(a).



(a)



(b)

Fig. 4. Scale-free networks with strong community structure. The total number of nodes is 800, $n = 12, m = 13, c = 8, \alpha = 0.1$ and $Q = 0.798$. In (a), the number of failed nodes decreases as the betweenness increases once the betweenness is higher than a certain level. In (b), there is a clear, consistent trend that the number of failed nodes increases as the BOF increases. The node of the highest BOF has the worst cascading failures. Figures are generated from 22 realizations of the network and we averaged betweenness with bin width equaling to 100 and BOF with that equaling to 2.

Comparing Figs. 4(a) and 4(b), it can be observed that betweenness is not as good an indicator of the severity of cascading failure that the removal of a node will cause as the Betweenness Overflow (BOF for short), an index we developed, which is the number of failed nodes in the first round of cascading. For example, if we attack node j, it directly leads to the failure of another 30 nodes. These 30 nodes will in turn cause a much greater number of nodes to cascade. The BOF of node j is thus 30. In Fig. 4(b), the tendency for the total number of failed nodes to increase as the BOF increase is clear. When we initially break down the node with the highest BOF, the network will cascade the most. Therefore, in SFCN, the BOF evaluates the importance of a node better than its betweenness.

It is also worth noting that, the worst cascading failure observed in SFCN has caused approximately 16 nodes of the 800 total nodes to fail. In contrast, for randomly generated scale-free networks, which have no strong community structure, the worst cascading failure has broken down more than 500 nodes of the same total number of nodes. So far, it seems safe to say that network structure (strong community structure) has an effect on cascading failures. To be specific, in scale-free networks, stronger community structure results in the failure of fewer nodes.

3.2. *RCN versus SFCN*

We now apply the RCN to the cascading-failure scenario. We generate a random network with strong community structure, where $n = 600, \langle k \rangle = 6, c = 3, \sigma = 0.9$ and $\alpha = 0.1$. Based on what we have previously observed, we expect the cascading failure to be small because of the strong community structure. However, almost half of the nodes (300 out of 600) broke down, as observed in Figs. 5(a) and 5(b). Therefore we can draw the conclusion that in addition to community structure, the degree distribution (e.g. Poisson distribution, power-law distribution, etc.) also has an effect on the cascade. As for the two indicators, i.e. betweenness and BOF, both fail to identify the worst cascading failures in this case.

Intrigued by the question of how strongly degree distribution affects the cascade in the RCN, we explored cascading failures in networks of different Q modularity, namely, different degree of community structure. We then checked the maximum and average cascading failures among all the nodes, respectively denoted as maxfail and avefail. We can see from Table 1 that as the Q modularity increases from 0.5 to 0.9 in a network in which the total number of nodes is 1000, the maxfail varies little with an average of approximately 625, and the avefail increases from 103 to 180. Therefore, it is rather safe to conclude that in RCN, community structure has little influence on cascading failures in comparison to the degree distribution. We similarly explored the SFCN. Interestingly, we found that in this case, the avefail is constantly 1, no matter what Q modularity is. Because most nodes in scale-free networks are leaf nodes, whose removal will not incite other nodes to fail, this behavior is not difficult to understand this phenomenon. When the maxfail of SFCN and RCN of the same Q modularity are compared, the latter is

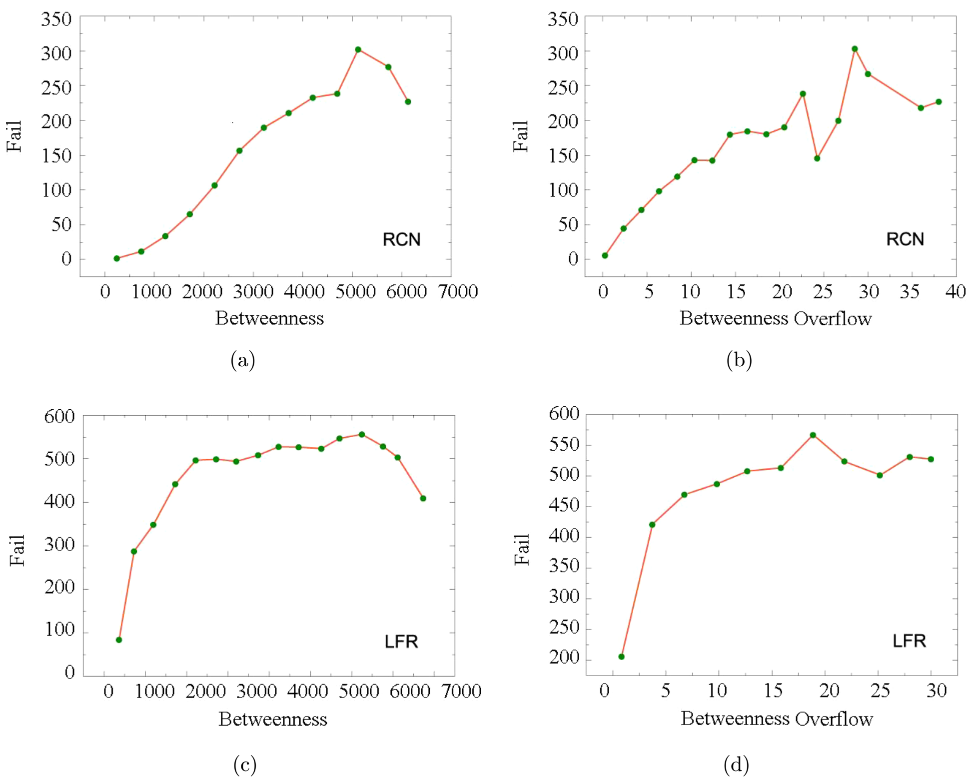


Fig. 5. Because the cascading failures of the RCN and the LFR are quite similar to one another, we present the outcomes together in a single graph. Subgraphs (a) and (b) are the results of random networks with strong community structure, where $n = 600, \langle k \rangle = 6, c = 3, \sigma = 0.9$ and $\alpha = 0.1$. (a) is the betweenness failure graph. The number of failed nodes initially increases with node betweenness, and it then decreases when node betweenness nears its maximum. (b) is the BOF failure graph. When BOF is approximately 25, the number of failed nodes changes rapidly. Initially removing the nodes with the largest BOF will not lead to the worst cascading, which indicates that the BOF index fails to reliably indicate the severity of cascading failure, unlike the BOF in the case of SFCN. Subgraphs (c) and (d) depict the cascade in the LFR benchmark with $\gamma = 3.14, \beta = 0.08, n = 1000$ and $Q = 8.7$. Approximately, six-tenths of the nodes failed. However, the largest betweenness and the largest BOF both failed to indicate the worst cascading, the same situation depicted in subgraphs (a) and (b). The figures are generated from 25 iterations. The bin width is 500 in (a) and (c) and 2 in (b) and (d).

Table 1. Maximum and average cascading failures of RCN and SFCN.

Maxfail	$Q = 0.5$	$Q = 0.6$	$Q = 0.7$	$Q = 0.8$
RCN	622	635	614	630
SFCN	18	20	18	24
Avefail	$Q = 0.5$	$Q = 0.6$	$Q = 0.7$	$Q = 0.8$
RCN	103	138	163	180
SFCN	1	1	1	1

almost 30 times larger than the former. Thus, we can infer that the degree distribution will strongly affect the cascade in networks.

3.3. *LFR benchmark versus SFCN*

The community size of SFCN is constrained to be homogenous, while networks in the real-world highly likely to have highly heterogeneous community-size distribution; for this reason, we have also explored the LFR benchmark.

As discussed in Sec. 2.3, the LFR benchmark has three tunable parameters: the degree-distribution exponent γ, the community-size-distribution exponent β and the mixing parameter. We have explored their relation to Q modularity previously. Here, we generate a specific LFR benchmark with $\gamma = 3.14, \beta = 0.08, n = 1000$ and $Q = 8.7$, which has the same degree-distribution exponent and the same outside–inside edges rate as the SFCN that we constructed before. We then apply this network to cascade scenario. The results are shown in Figs. 5(c) and 5(d). The number of failed nodes is far larger than the number of nodes that failed in the SFCN, indicating that the community-size distribution also has an impact on cascading failures. Again, betweenness and BOF both failed to identify the node that caused the worst cascading failure.

4. Conclusion

To summarize, we generated SFCN and RCN. Using these networks and the LFR benchmark,[15] we investigated the effects of the degree distribution, the community-structure strength and the community-size distribution on cascading failure,[6] and we have proposed an indicator, the BOF, to distinguish which node leads to the worst cascading failures. Our conclusions are as follows.

- In SFCN, the BOF can better evaluate the importance of a node than betweenness, which means that larger BOF for a node leads to worse cascading failures.
- The strength of network structure has an effect on the cascade. When networks are of a power-law degree distribution, a network with a stronger community structure is more resistant.
- The degree distribution (Poisson or power-law) affects the cascade in networks. When networks have the same strength of community structure, a network with a Poisson degree distribution is less robust to cascade.
- The community-size distribution has an effect on the cascade. When scale-free networks are of the same Q modularity, a network with a heterogeneous community-size distribution shows vulnerability to cascade.
- In LFR benchmarks and RCN, the BOF and betweenness both fail to identify the node that can lead to the worst cascade. In this situation, it is not appropriate to apply these indices to evaluate the importance of the nodes. Some other index is necessary, which is a topic that must be explored.

Acknowledgments

We thank Professor Wenxu Wang for his helpful discussion and instruction. We acknowledge the support from the National Natural Science Foundation of China under Grant No. 61174150, NCET-09-0228 and the Doctoral Fund of the Ministry of Education (20110003110027).

References

1. S. H. Strogatz, *Nature* **410**, 268 (2001).
2. R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
3. R. Albert, H. Jeong and A.-L. Barabási, *Nature* **406**, 378 (2000).
4. D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
5. A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
6. A. E. Motter and Y.-C. Lai, *Phys. Rev. E* **66**, 065102(R) (2002).
7. Y. Moreno, J. B. Gómez and A. F. Pacheco, *Europhys. Lett.* **58**, 630 (2002).
8. M. L. Sachtjen, B. A. Carreras and V. E. Lynch, *Phys. Rev. E* **61**, 4877 (2000).
9. B. A. Carreras, V. E. Lynch, I. Dobson and D. E. Newman, *Chaos* **12**, 985 (2002).
10. W. Wang and G. Chen, *Phys. Rev. E* **77**, 026101 (2008).
11. M. E. J. Newman, S. H. Strogatz and D. J. Watts, *Phys. Rev. E* **64**, 026118 (2001).
12. M. E. J. Newman and M. Girvan, *Phys. Rev. E* **69**, 026113 (2004).
13. N. Kashtan and U. Alon, *Proc. Natl. Acad. Sci. USA* **102**, 13773 (2005).
14. G. Yan, Z.-Q. Fu, J. Ren and W.-X. Wang, *Phys. Rev. E* **75**, 016108 (2007).
15. A. Lancichinetti, S. Fortunato and F. Radicchi, *Phys. Rev. E* **78**, 046110 (2008).
16. J. Duch and A. Arenas, *Phys. Rev. E* **72**, 027104 (2005).