Changing motif distributions in complex networks by manipulating rich-club connections

Xiao-Ke Xu [a,b,*], Jie Zhang [c], Ping Li [a], Michael Small [a]

[a] *Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong, People's Republic of China*

[b] *School of Communication and Electronic Engineering, Qingdao Technological University, Qingdao 266520, People's Republic of China*

[c] *Centre for Computational Systems Biology, Fudan University, Shanghai 200433, People's Republic of China*

ARTICLE INFO

ABSTRACT

The role of rich-club connectivity is significant in the structural property and functional behavior of complex networks. In this study, we find whether or not a very small portion of rich nodes connected to each other can strongly affect the frequency of occurrence of basic building blocks (motifs) within a heterogeneous network. Conversely whether a homogeneous network has a rich-club or not generally has no significant effect on its structure. These findings open the possibility to optimize and control the structure of complex networks by manipulating rich-club connections. Furthermore, based on the subgraph ratio profile, we develop a more rigorous approach to judge whether a network has a rich-club or not. The new method does not calculate how many links there are among rich nodes but depends on how the links among rich nodes can affect the overall structure as well as the function of a given network.

1. Introduction

The motif, defined as a small connected subgraph that recurs in a graph, is the basic building block, or functional unit, of complex networks [1]. In real-world networks (e.g., gene regulatory networks), motifs represent the elementary interaction patterns between small groups of nodes, and the relative frequencies with which motifs appear represent different functions of the network [2–4]. Although it has been found that there is a topological relationship between the large-scale attributes (scale-free and hierarchical) and local interaction patterns (subgraph based) [5], it remains unclear for the accurate relationship between small functional units and other structural properties, such as rich-club connections, of complex networks. In our previous study we found that rich-club connections can dominate some global properties (e.g., assortativity and transitivity) of a network [6], which implies the possible relation between the rich-club property and the network's subgraph organization.

The rich-club property refers to the organization pattern of rich nodes [7], especially whether rich nodes tend to connect to one another, or with the remaining nodes [8–13]. Because rich nodes often play a central role in the static properties of, and dynamic processes on, complex networks [14–16], significant attention has been paid to the prominent effects of the richest elements [17] and the organization among them [6,18]. A systematic framework is needed to clearly understand the roles of rich nodes in different real-world networks with distinct degree distributions.

In this study, we find the influences of rich nodes and their organization pattern depends largely on the degree distributions of complex networks. Rich nodes are important in scale-free networks [19], because a power-law degree
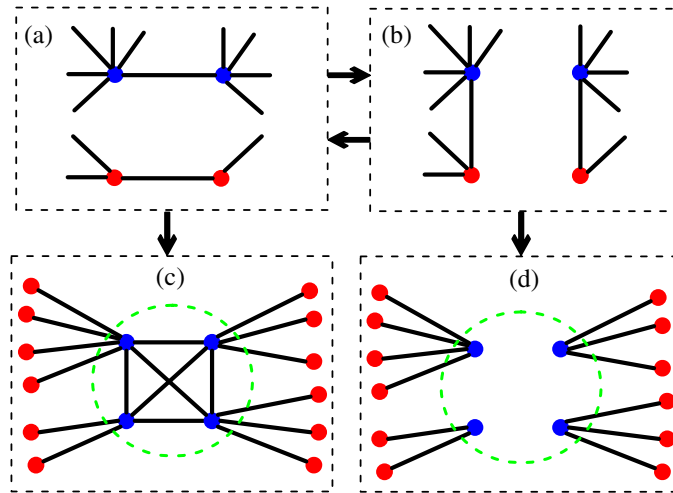
Fig. 1. (Color online) (a) and (b) are the two connection patterns for the four end nodes of a pair of links. (a) rich-club connection, where one link connects to the two rich nodes and the other link connects to the other low-degree nodes; (b) non-rich-club connection, where one link connects to one rich node and one low-degree node, and the other link connects to the two remaining nodes. Using the link rewiring algorithms, we can obtain (c) the network with a rich-club, or (d) the network without a rich-club.

distribution indicates that the majority of nodes participate in at most one or two motifs, while a few rich nodes take part in a very large number of small subgraphs. Manipulating a very small number of rich-club connections therefore can strongly affect the frequencies of the basic functional blocks (motifs) for a heterogeneous network. In comparison, for the network with a homogeneous degree distribution (e.g., the network of the US power grid), the links among rich nodes show a tiny effect on the whole network. The main reason behind this is that all nodes (including rich nodes) in a homogeneous network are engaged in only a few interactions, and there are no hubs linking to a significantly larger number of other nodes.

These results are helpful in understanding the origin of motifs and motif clusters in real-world complex networks, and the mechanisms by which small subgraphs aggregate into larger superstructures. Our finding has an important potential application: we can build a framework to optimize and control the functional behaviors of complex networks. In most cases we cannot regenerate or redesign a real-world network, but manipulating a small number of rich-club connections gives us a chance to optimize the structure of the network and control the relative frequencies of small functional units in a predictable manner.

Furthermore, although pioneering studies have developed a series of methods to judge whether a network has rich-club properties [11,9,13], these approaches are based on how many links there are among rich nodes instead of how these links affect the whole network. Based on subgraph ratio profile, the topological structure among rich nodes can be uncovered from the inspection of the basic functional units. In this study we develop a novel method to judge whether a network has a rich-club or not. The new method does not calculate how many links connect to rich nodes compared with its randomized version but rather it depends on how the organization pattern of rich nodes affects the appearance of different motifs.

Taken together, these findings indicate the strong ties between the local subgraphs and rich-club properties of complex networks, which complements our understanding of a network's topological and functional organization. Because each network can be characterized by a set of distinct types of subgraphs and rich-club connections are a significant property, our findings are expected to provide useful insights to design new models for characterizing real-world networks. Our work is a step in an ongoing effort to bridge the local topology of a network and its global statistical features.

2. Method

2.1. Link rewiring algorithms

Here we select the top 0.5% (selecting another small proportion, such as 0.3% or 0.8%, also gives the similar results) of the nodes with the highest degree as rich nodes in a network and manipulate the connections among them. We use link rewiring algorithms to generate the network with a rich-club and the network without a rich-club, respectively. The basic idea is very similar to the random rewiring method [20], while the main difference is that our new method only switches the links among rich nodes and a small number of low-degree nodes. First we make rich nodes fully connected to one another, so they form a completely connected rich-club. Secondly, we completely eradicate the edges among rich nodes, so that the network has no rich-club.

Now we specify the rewiring algorithms. First we make all rich nodes fully connected to generate a network with a distinct rich-club. If there is a link between two rich nodes, their structure remains unchanged (Fig. 1(a)). If there is no link between two rich nodes, we perform the operation from Fig. 1(b) to (a). That is, we select another two low-degree nodes that respectively connect to the two rich nodes while do not connect to each other (Fig. 1(b)). Then we cut the two links between the rich nodes and their low-degree neighbors, and connect the two rich nodes as well as the two low-degree
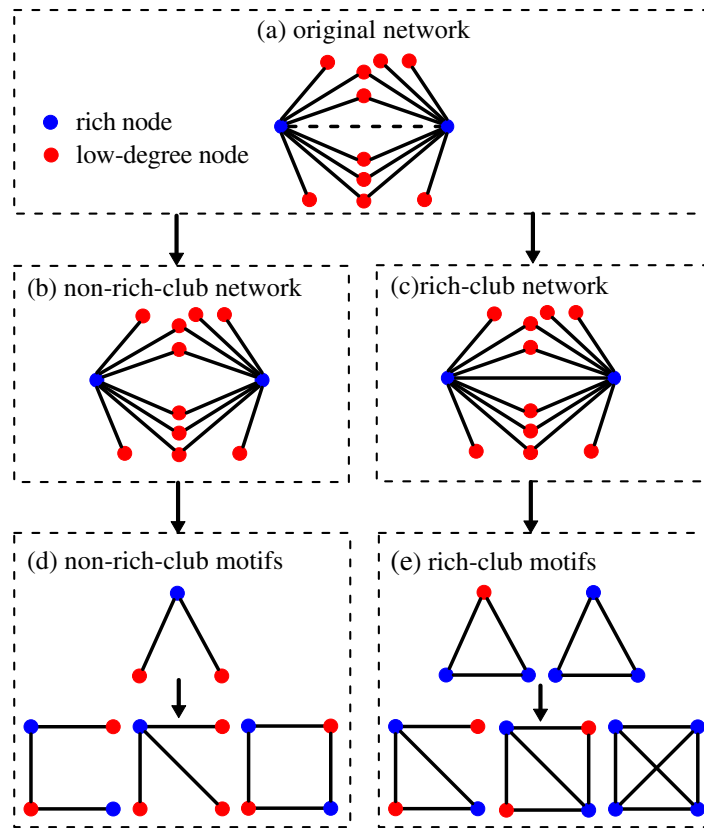
Fig. 2. (Color online) The demonstration for the aggregation of non-rich-club motifs when a network has no rich-club and the aggregation of rich-club motifs when a network has a rich-club.

nodes, respectively (Fig. 1(a)). After repeating this process until all rich nodes form a completely connected rich-club, we can obtain the network with a full-connected rich-club (Fig. 1(c)).

Secondly, we completely eradicate the edges among rich nodes, so that the network has no rich-club property. If there is no link between two rich nodes, we will do nothing (Fig. 1(b)). If there is a link between two rich nodes, we do the operation from Fig. 1(a) to (b). We randomly select another pair of low-degree nodes which connect to each other while we do not connect to either of the two rich nodes (Fig. 1(a)). Then we cut off the links both between the two low-degree nodes and between the two rich nodes respectively, and let each rich node connect to one low-degree node (Fig. 1(b)). Repeating the above process until the links among the whole rich nodes are completely eradicated, we will get a network without a rich-club property (Fig. 1(d)).

Because we use the rewiring method, the degree of every node in the original network remains exactly unchanged. For the topological structure of the original network, there is only small variation induced by manipulating rich-club connections, so we can monitor how the subgraph frequencies are affected by the rich-club property. Furthermore, we can compare the results of the subgraph ratio profile for the original network, the network with a rich-club, and the network without a rich-club, to a make more reliable inference on whether the original network has a rich-club property or not.

2.2. Motif clusters of rich nodes in non-rich-club and rich-club networks

Each network will be scanned for all possible n-node subgraphs (we choose $n = 4$). In a network with a skewed degree distribution, rich nodes have much higher degrees than the overwhelming majority, so whether they connect to each other to form a rich-club will strongly affect the frequencies of subgraphs. Actually, rich nodes can absorb a very large number of subgraphs and form a motif cluster. For example, triangles may not distribute uniformly within a scale-free network but tend to aggregate around the hubs, because a node with k links can carry up to $k(k-1)/2$ triangles [5]. The aggregation of motifs into motif clusters is important, because it implies that the potential functional properties of the large number of subgraphs also need to be evaluated at the level of subgraph clusters instead of being evaluated only at the level of a single subgraph.

Exploring rich-club connections provides a new way to evaluate the functional properties of abundant subgraphs at the level of subgraph clusters. A few rich nodes usually take part in a very large number of small subgraphs and they can form motif clusters in real-world complex networks. Actually, the organization of rich nodes can dominate the appearance of particular motifs prominently. In the non-rich-club network (Fig. 2(b)), rich nodes do not tend to connect to each other, so the non-rich-club subgraphs (Fig. 2(d)) will be more common. On the contrary, in the rich-club network (Fig. 2(c)), rich

Table 1

Statistics of six undirected networks: number of nodes n, average degree $\langle k \rangle$, the exponent of degree distribution if the distribution follows a power law: α (or "–" if not), structural cutoff degree $k_s = \sqrt{\langle k \rangle n}$ [21], maximal degree k_{max}. SW is the network generated by the small-world model [22], PG is the network of US power grid [19], BA is the network generated by the scale-free model [19], EPA is the network from the pages linking to www.epa.gov [23], PFP is the network generated by the model for the Internet topology [24] and AS is the network of the Internet topology at the level of autonomous systems [25]. The results for three models (SW, BA and PFP) are the average of 10 independent network realizations respectively.

Network	SW	PG	BA	EPA	PFP	AS
n	5000	4941	5000	4772	5000	5375
$\langle k \rangle$	6.0	2.7	6.0	3.7	6.0	3.9
α	–	–	3.0	2.0	2.2	2.2
k_{max}	16	19	219	175	1259	1193
k_s	173.2	115.4	173.2	132.9	173.2	144.8
k_{max}/k_s	0.09	0.16	1.26	1.32	7.26	8.24
Type	$k_{max} \ll k_s$		$k_{max} \approx k_s$		$k_{max} \gg k_s$	

nodes trend to connect to each other, so the network will demonstrate a larger number of the rich-club motifs (Fig. 2(e)). In the original network (Fig. 2(a)), rich nodes may or may not connect to each other. Comparing the relative prevalence of motifs in the above three networks, we can conclude whether the original network has a rich-club property.

It is obvious that by considering the subnetworks of rich nodes, the frequencies of the non-rich-club motifs and/or rich-club motifs are remarkably more than those of the randomized versions of the subnetworks. The inherent existence of two distinct classes of subgraphs (non-rich-club motifs and rich-club motifs) in a heterogeneous network demonstrates that, in contrast to the homogeneous network, the highly abundant motifs cannot exist in isolation but must naturally aggregate into subgraph clusters. Specifically, in the network with a rich-club, the neighbors of a highly connected node are linked to each other, therefore the chance that low-degree nodes participate in highly connected subgraphs is slim. In a homogeneous network, however, all nodes are engaged in only a few interactions and the appearance of motifs is the statistical average of the whole network, for there are no hubs linking to a significantly higher number of other nodes to form motif clusters.

3. Results

Table 1 lists the results of six undirected networks (including three real-world networks and three model networks) arranged with k_{max}/k_s increasing. The value of the structural cutoff degree k_s can be regarded as the first approximation of the maximum degree within a scale-free network [21]. Here k_{max}/k_s is a convenient index that can be used in complex networks with any degree distribution to show the proportion of links (or degrees) rich nodes possess in comparison with the remaining nodes in a network [6].

The low values of k_{max}/k_s for SW and PG mean that these two networks have a homogeneous degree distribution and the degrees of rich nodes are close to the majority of nodes. While a high value of k_{max}/k_s indicates that the network has a heterogeneous degree distribution and the degrees of a few rich nodes are far larger than the rest, like BA and EPA. Especially, PFP and AS not only have a power-law degree distribution, but also possess a few super-rich nodes [17] for $k_{max} \gg k_s$ in the two networks.

Although motifs are only local interaction patterns, the distribution of motifs can greatly reflect the topological properties of the networks [26]. Because undirected networks have only two types of triads (unclosed triple and triangle), we only analyze the profile of the six types of undirected connected tetrads (4-node motifs). The normalized Z scores of tetrads show a significant dependence on the network size, so we use the abundance of each subgraph i relative to random networks [4]:

$$\Delta_i = \frac{N\mathrm{real}_i - \langle N\mathrm{rand}_i \rangle}{N\mathrm{real}_i + \langle N\mathrm{rand}_i \rangle + \varepsilon}, \tag{1}$$

where $\varepsilon = 4$ ensures that $|\Delta_i|$ is not misleadingly large when the subgraph appears very few times in both the real and random networks. The Subgraph Ratio Profile (SRP) is the vector of $|\Delta_i|$ normalized to length 1:

$$\mathrm{SRP}_i = \Delta_i \Big/ \left(\sum \Delta_i^2 \right)^{1/2}. \tag{2}$$

Network motifs, which are patterns of interconnections occurring in complex networks are significantly higher than those for randomized networks [2]. The randomized network is generated by repeating the following process sufficient times: first randomly choose a pair of links and then swap two of their end nodes, which is equivalent to reconnecting the four end nodes using a wiring pattern chosen at random. The motif pattern reflects the local structural properties of complex networks and thus can be used to classify networks. If different types of networks share the similar result of SRP, these networks can be classified into the same "superfamily" [4,27,28]. The networks in the same triad superfamily share not only some particular types of motifs, but also very similar proportions of all types of subgraphs.

Here we show the SRP results for the original network, the network with a rich-club, and the network without a rich-club in Fig. 3. If the above three networks belong to the same superfamily, it means that the rich-club property has a weak effect on the original network, and this result shows the network is a homogeneous network. If the three networks belong to different superfamilies, it means that rich-club connections can strongly affect the structure and function of the original
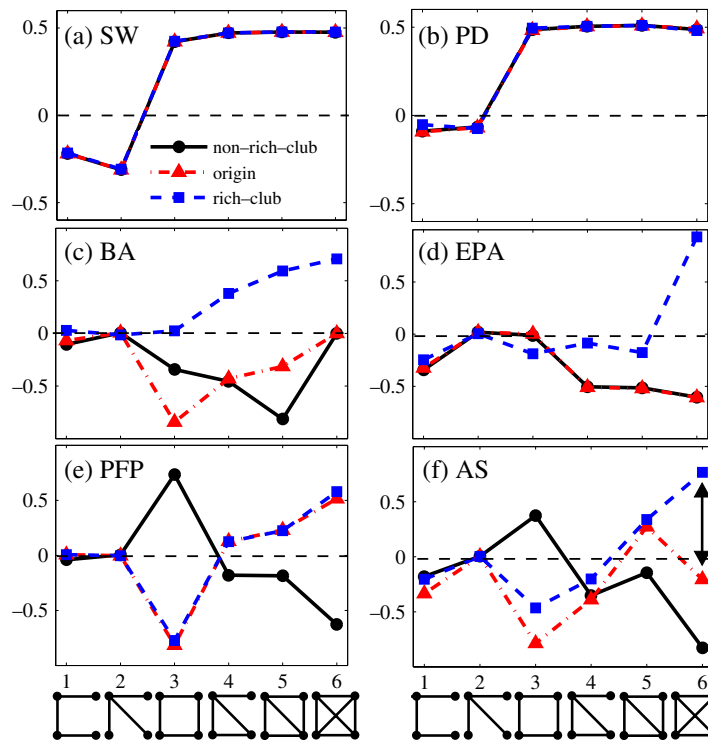
Fig. 3. (Color online) The subgraph ratio profile (SRP) for six undirected networks. SW, PG, BA, EPA, PFP and AS represent the same networks in Table 1.

network, and this result indicates that the network is heterogeneous. Furthermore, according to the fact that the original network belongs to the same superfamily as the network with a rich-club or the network without a rich-club, we can judge whether the original network has a rich-club property.

The networks of SW and PG have a homogeneous degree distribution, so rich nodes in the two networks are not significantly higher than others. Therefore, as has been shown in Fig. 3(a) and (b), whether the two networks have rich-club properties does not have any influence on SRP. Moreover, the original network, and the networks with and without a rich-club all belong to the same superfamily. The above results indicate whether a homogeneous network has a rich-club property is not very important, and rich-club connections cannot control the functions of such type of networks.

Because the networks of BA and EPA have a heterogeneous degree distribution, rich nodes possess many more links than the overwhelming majority. Therefore, whether the two networks have rich-club properties can greatly affect the result of SRP. As is shown in Fig. 3(c) and (d), the original network and the network with a rich-club do not belong to the same superfamily. Conversely, the original network and the network without a rich-club belong to the same superfamily, so neither BA nor EPA has the rich-club property.

For the networks of PFP and AS, they not only have a heterogeneous degree distribution but also have a few super-rich nodes, so whether the two networks have a rich-club can affect the result of SRP most significantly. The original network and the network without a rich-club do not belong to the same superfamily. For the original PFP and the network with a rich-club belonging to the same superfamily, PFP has the property of a rich-club as is shown in Fig. 3(e). The presence of a rich club on the Internet has been debated [8–11], and there is still no clear conclusion. Comparing with BA [19], Fitness-BA [29], and Inet-3.0 models [30], Zhou and Mondragón find that the Internet has a rich club [8]. However, Colizza et al. believe that a rich-club should be inferred by a comparison of the original network with its randomized counterparts [9], and using this method they conclude that the Internet has no rich-club [9]. Basically we can say that AS has a rich-club property, for the original AS has the very similar SRP to the network with a rich-club, except for the motif 6 (4-node clique) in Fig. 3(f). Yet the non-consistency of motif 6 for the original network and the network with a rich-club may help to explain the controversy of whether the Internet topology has a rich-club property.

4. Conclusions and discussions

In conclusion, we find that the influence of rich-club connections strongly depends on the degree distributions of complex networks. Our findings show that in a homogeneous network, whether the network has a rich-club or not is not very important for its structure. While rich-club connections in a heterogeneous network have a crucial role in the overall network structure, they can partially optimize and control the topological structure of the network.

Our new framework for measuring the subgraph ratio profile can provide a more impartial judgement on whether a network has a rich-club. Previous studies have placed more importance on determining whether the links among rich nodes appear more frequently in the original network compared with its randomized counterparts [9,10]. However, the actual

influence of rich-clubs in different degree distribution networks has not been thoroughly studied. In contrast, our approach is based specifically on the effect of the rich-club on the network structure.

We demonstrate that strong ties between the rich-club property and the local (subgraph-based) structure underscore the importance of understanding the properties of complex networks as fully integrated systems. Indeed, the abundance of some kinds of local interaction patterns reflects the rich-club property of a network, raising intriguing questions about the role of local events in shaping a network's overall behavior [5]. These results indicate that the analysis described here may have an impact on our understanding of other types of subgraphs (e.g., cliques [31] and cycles [32]) in complex networks.

Our results show the significance of the rich-club property and motif distributions in modeling and designing real-world networks [33]. An appropriate model should have a similar structure and function to the real-world network. To meet this demand the model can be designed from the basic motifs or the subgraph ratio profile, which can be easily controlled by the rich-club property. Following the framework in this work, we will contrive to bridge the gap between local topologies of a network and its global statistical features.

Acknowledgments

References

[1] A. Jamakovic, P. Mahadevan, A. Vahdat, M. Boguna, D. Krioukov, http://arxiv.org/abs/0908.1143.
[2] R. Milo, S. Shen-Orr, S. Itzkovic, N. Kashtan, D. Chklovskii, U. Alon, Science 298 (2002) 824.
[3] U. Alon, Nat. Rev. Genet. 8 (2007) 450.
[4] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, U. Alon, Science 303 (2004) 1538.
[5] A. Vázquez, R. Dobrin, D. Sergi, J.-P. Eckmann, Z.N. Oltvai, A.-L. Barabá, Proc. Natl. Acad. Sci. USA 101 (2004) 17940.
[6] X.-K. Xu, J. Zhang, M. Small, Phys. Rev. E 82 (2010) 046117.
[7] H. Zhao, Z.Y. Gao, Physica A 381 (2007) 473.
[8] S. Zhou, R.J. Mondragón, IEEE Commun. Lett. 8 (2004) 180.
[9] V. Colizza, A. Flammini, M.A. Serrano, A. Vespignani, Nat. Phys. 2 (2006) 110.
[10] L.A.N. Amaral, R. Guimera, Nat. Phys. 2 (2006) 75.
[11] S. Zhou, R.J. Mondragón, New J. Phys. 9 (2007) 173.
[12] J.J. McAuley, L.da F. Costa, T.S. Caetano, Appl. Phys. Lett. 91 (2007) 084103.
[13] Z.-Q. Jiang, W.-X. Zhou, New J. Phys. 10 (2008) 043002.
[14] R. Albert, H. Jeong, A.-L. Barabási, Nature (London) 406 (2000) 378.
[15] A.E. Motter, Y.-C. Lai, Phys. Rev. E 66 (2002) 065102(R).
[16] M. Small, C.K. Tse, D. Walker, Physica D 215 (2006) 146.
[17] X.-K. Xu, J. Zhang, J. Sun, M. Small, Phys. Rev. E 80 (2009) 056106.
[18] T. Opsahl, V. Colizza, P. Panzarasa, J.J. Ramasco, Phys. Rev. Lett. 101 (2008) 168702.
[19] A.-L. Barabási, R. Albert, Science 286 (1999) 509.
[20] S. Maslov, K. Sneppen, Science 296 (2002) 910.
[21] M. Boguñá, R. Pastor-Satorras, A. Vespignani, Eur. Phys. J. B 38 (2004) 205.
[22] D.J. Watts, S.H. Strogatz, Nature (London) 393 (1998) 440.
[23] http://vlado.fmf.uni-lj.si/pub/networks/data/.
[24] S. Zhou, R.J. Mondragón, Phys. Rev. E 70 (2004) 066108.
[25] http://pil.phys.uniroma1.it/~gcalda/cosinsite/.
[26] W.-J. Xie, W.-X. Zhou, http://arxiv.org/abs/1012.3850.
[27] X.-K. Xu, J. Zhang, M. Small, Proc. Natl. Acad. Sci. USA 105 (2008) 19601.
[28] C. Liu, W.-X. Zhou, J. Phys. A 43 (2010) 495005.
[29] G. Bianconi, A.L. Barabási, Europhys. Lett. 54 (2001) 436.
[30] J. Winick, S. Jamin, Univ. Michigan, Ann Arbor, Tech. Rep. UM-CSE-TR-456-02, 2002.
[31] G. Palla, I. Derényi, I. Farkas, T. Vicsek, Nature (London) 435 (2005) 814.
[32] G. Bianconi1, N. Gulbahce, A.E. Motter, Phys. Rev. Lett. 100 (2008) 118701.
[33] B. Karrer, M.E.J. Newman, Phys. Rev. E 82 (2010) 066118.