Robustness measurement of scale-free networks based on motif entropy

Yun-Yun Yang(杨云云), Biao Feng(冯彪), Liao Zhang(张辽), Shu-Hong Xue(薛舒红), Xin-Lin Xie(谢新林), and Jian-Rong Wang(王建荣)

What follows is a list of articles you may be interested in

Evolution of donations on scale-free networks during a COVID-19 breakout

Effects of the planarity and heterogeneity of networks on evolutionary two-player games

Structural and robustness properties of smart-city transportation networks

Co-evolution of the brand effect and competitiveness in evolving networks

Dynamical analysis of a sexually transmitted disease model on complex networks

----

Robustness measurement of scale-free networks based on motif entropy

Yun-Yun Yang(杨云云)[1,†], Biao Feng(冯彪)[1], Liao Zhang(张辽)[1],

Shu-Hong Xue(薛舒红)[1], Xin-Lin Xie(谢新林)[2], and Jian-Rong Wang(王建荣)[3]

[1]*College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, China*
[2]*Taiyuan University of Science and Technology, Taiyuan 030024, China*
[3]*School of Mathematical Sciences, Shanxi University, Taiyuan 030024, China*

As a classical complex network model, scale-free network is widely used and studied. And motifs, as a high-order subgraph structure, frequently appear in scale-free networks, and have a great influence on the structural integrity, functional integrity and dynamics of the networks. In order to overcome the shortcomings in the existing work on the robustness of complex networks, only nodes or edges are considered, while the defects of high-order structure in the network are ignored. From the perspective of network motif, we propose an entropy of node degree distribution based on motif to measure the robustness of scale-free networks under random attacks. The effectiveness and superiority of our method are verified and analyzed in the BA scale-free networks.

1. Introduction

Complex network is one of the methods to model and study the actual complex systems.[1] With the in-depth study of the related characteristics of complex networks, network robustness has gradually attracted the extensive attention of scholars. The so-called network robustness refers to the ability of the network to maintain the integrity and function of the original network structure after being subjected to random failures or deliberate attacks.

According to the current work on network robustness, the measurement of network robustness is mainly carried out around nodes. Albert *et al.*[2] investigated the random attacks and deliberate attacks on nodes in scale-free networks and random networks, and calculated the change of network diameter and the largest connected subgraph, and they confirmed the strong robustness of scale-free networks to random attacks and their vulnerability to deliberate attacks. Newman and Ghoshal[3] randomly removed nodes on real networks of different scales, and analyzed the robustness of the network by using the variation of two components in the network. Clemente and Cornaro[4] proposed to use the eigenvalue of the Laplace matrix of the network to calculate the Kirchhoff index,[5,6] then they calculated the effective resistance centrality based on nodes and edges respectively according to the Kirchhoff index, and they used the centrality value to measure the robustness of the network, and presented different calculation methods for different types of networks.

For different nodes, the coupling strength of the node it-self determines its influence on the network robustness, and nodes with large degrees are more sensitive to random attacks, showed the unity of dynamic robustness and structural robustness.[7] Kasthurirathna *et al.*[8] continuously attacked the scale-free network according to the node degree value, they used the average path length and clustering coefficient to measure the robustness. It was confirmed that the average path length is positively correlated with the network robustness, and the clustering coefficient is negatively correlated with the network robustness. Zhou *et al.*[9] used the Betti number to calculate the number of structural holes. Through random attacks and deliberate attacks on the network, it was found that the number of structural holes can change with the attack intensity. For power networks[10,11] and urban transportation networks,[12–15] they proved to have the characteristics of the scale-free network and small-world network, that is, they are robust under random attacks and weak under deliberate attacks. Other researches of network robustness include the using of node efficiency,[16] R-index,[17] fractal dimension,[18] natural connectivity[19] to measure the network robustness.

The research process of complex network itself is complex generally. But the research from the low-order perspective of nodes and edges turns relatively simple. The research from the high-order perspective can realize the transformation from low to high complexity, which is the general law of research problems. New rules and new properties can also be found with the help of high-order structure. For many actual complex networks, they are usually scale-free, and motifs appear frequently in the form of high-order structure. Compared

with nodes, motifs have the characteristics of large number, complex types and strong correlation, which play a decisive role in the overall structure and function of the network.[21] However, previous research work on network robustness ignored the existence of motifs. At the same time, in the actual complex systems, the probability of random failures is often higher than that of deliberate attacks. Therefore, we will study the robustness of scale-free networks from the perspective of motif, propose an entropy of node degree distribution based on motif to measure the robustness of different sizes of scale-free networks under random attacks, and verify its effectiveness and superiority through experimental analysis.

2. Theoretical basis

2.1. Scale-free networks

The scale-free network[22] is a classical complex network model, and its node degree distribution function obeys the power-law distribution, that is, $P(k) = Ck^{-\gamma}$, where C is the scale coefficient, k is the degree value of the node, γ is the exponent and usually $\gamma \in [2,3]$. In a scale-free network, the degree value of most nodes is very small, and a few hub nodes have large degree value.

In order to realize the construction of scale-free network, Barabási and Albert[23] proposed a simple and effective construction model — BA scale-free network, that is, on the basis of a randomly generated small network, BA scale-free network is generated according to the growth mechanism and preference connection mechanism. The growth mechanism refers to adding a new node to the original network at each time step, while the preference connection mechanism refers to that the newly added nodes tend to connect with the hub nodes in the original network. For a BA scale-free network with $N = 100$ and $M = 3$, the generation process is as follows: when a new node is added to the network, $M = 3$ edges will be added to the network until the scale of the network reaches $N = 100$ nodes.

2.2. Network motif

Motif is a kind of subgraph that appears frequently in the network. Motif plays a decisive role in the function and properties in the whole network. Identifying and analyzing motifs is helpful in revealing the dynamic and structural characteristics of the network. There have been many algorithms for recognizing and detecting the network motif. In this work, the software FANDOM[24] developed based on the ESU algorithm[25] is used to mine motifs. According to the number of nodes contained in the motif (n), the motif can be defined as an n-order motif. The specific structure of the possible 3rd motif and the 4th-order motif in the simple undirected network are shown in Figs. 1 and 2, respectively.

The most important statistical feature of the motif is the Z-score. The positive or negative Z-score can be used to deter-mine whether the motif appears in the network. The presence of the motif can be confirmed only when $Z > 0$. The value of Z can be calculated from the following formula:

$$Z = \frac{N_{\text{real}} - \langle N_{\text{rand}} \rangle}{\sigma_{\text{rand}}}, \tag{1}$$

where N_{real} represents the number of times of a specific motif appears in the actual network, $\langle N_{\text{rand}} \rangle$ represents the average number of times the same motif appears in a random network of the same size, and σ_{rand} represents the standard deviation of the number of times the motif appears in the random network.

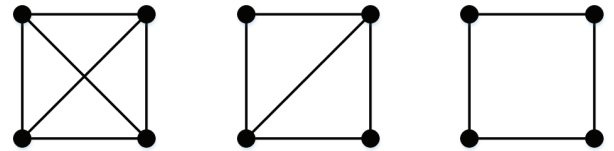**Fig. 1.** The third-order motifs of undirected network.

Fig. 2. The fourth-order motifs of undirected network.

3. Entropy of node degree distribution based on motif

Let the graph $G = (V, E)$ be an undirected and unweighted network, where $V = \{v_1, v_2, \ldots, v_N\}$ is a set of nodes in the network, N represents the total number of nodes, $E = \{e_1, e_2, \ldots, e_L\}$ refers to a set of edges in the network, and L denotes the total number of edges.

The classical node degree value refers to the total number of edges owned by the node, but this calculation method ignores the existence of motifs in the network. According to this, Han et al.[26] proposed a motif-based node degree value, which is specifically defined as follows: for a network G, let M be the motif of G, the total number of motif M composed of node i is defined as the motif-based node degree value of node i. In this work, the traditional node degree distribution value is extended to the node degree distribution value based on motif, and the entropy of node degree distribution based on motif, that is H_M, is further obtained, and its specific expression is

$$H_M = - \sum_{k_M=0}^{\max(k_M)} P_M(k_M) \cdot \log[P_M(k_M)], \tag{2}$$

where k_M is the node degree value based on the motif, $P_M(k_M)$ is the node degree distribution value based on the motif, that is, the node degree value based on the motif is equal to the probability of the occurrence of k_M. In the following, experimental results will show that the larger the value of H_M, the more robust the network is. In the process of calculation, it can be found that k_M does not change continuously, so it is not calculated for the case of $P_M(k_M) = 0$.

4. Experiment and analysis

In this work, all experimental networks are undirected and unweighted. In order to verify the effectiveness of our method, we will use H_M, and some classical robustness metrics, such as average edge betweenness B_G,[18] network efficiency E_G,[18] and network dispersion D_G,[27] to analyze the robustness of BA scale-free networks. At the same time, by analyzing the computational complexity of different indicators, the superiority of our method is highlighted. The 3rd-order motifs are used in the experiment unless otherwise specified.

Definition 1 Edge betweenness B_e [27] is one of the criteria to measure edge centrality based on the shortest path, and its calculation formula is as follows:

$$B_e = \sum_{\substack{j,k \in V \\ j \neq k}} \frac{L_{jk}(e)}{L_{jk}}, \tag{3}$$

where B_e represents the betweenness value of edge e, L_{jk} is the length of the shortest path from node j to node k, and $L_{jk}(e)$ refers to the length of the shortest path from node j to node k through edge e.

Definition 2 For the network G, the average edge betweenness B_G is one of the indicators to measure the robustness of the network from a local perspective, which is defined as follows:

$$B_G = \frac{1}{m} \sum_{e \in E} B_e, \tag{4}$$

where m is the total number of edges in network G, E is the set of all edges in network G, and the smaller the value of B_G, the stronger the robustness of the network is.

Definition 3 For network G, network efficiency E_G is one of the indicators to measure the robustness of the network from a global perspective, which can be obtained from

$$E_G = \frac{1}{n(n-1)} \sum_{\substack{j,k \in V \\ j \neq k}} L_{jk}, \tag{5}$$

where L_{jk} represents the length of the shortest path from node j to node k, n is the total number of edges in network G, V

refers to the set of all edges in network G, and the larger the value of E_G, the stronger the robustness of the network is.

Definition 4 For the network G, network dispersion D_G is one of the indicators to measure the robustness of the network from a local perspective, which is defined as follows:

$$D_G = 1 - 2 \frac{\sum\limits_{j>k} \frac{1}{L_{jk}}}{n(n-1)}, \tag{6}$$

where L_{jk} represents the length of the shortest path from node j to node k, $D_G \in [0,1]$, and the greater the value of D_G, the greater the dispersion of the network is and the worse the robustness of the network.

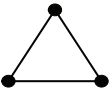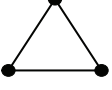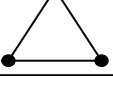### 4.1. Robustness of BA scale-free networks to random attacks

In this subsection, we will use different indicators to measure the robustness of BA scale-free networks under random attacks, and analyze the relationship between node degree distribution function based on motif, H_M, and network robustness.

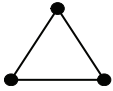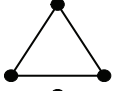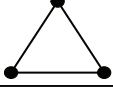#### 4.1.1. BA scale-free network robustness measurement

In experiment, nine different BA scale-free networks are generated according to the growth mechanism and preference connection mechanism. The calculation results of its basic attributes and robustness metrics H_M, B_G, and E_G are shown in Tables 1–3.

By analyzing the data in Tables 1–3, it can be found that for BA scale-free networks, when N is the same, the larger the value of M, the smaller the average edge betweenness is, the higher the network efficiency, and the greater the entropy of node degree distribution based on motif, that is, the stronger the robustness of the network is; when M is the same, the smaller the value of N, the smaller the average edge betweenness is, the higher the network efficiency, and the greater the entropy of node degree distribution based on motif, the stronger the robustness of the network.

Table 1. Robustness index of BA scale-free networks ($N = 100$).

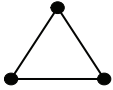| Network parameters	B_G	E_G	The 3rd-order motif	Z-score	H_M
$N = 100, M = 3$	43.4261	0.2158		24.8960	1.7415
$N = 100, M = 6$	18.0780	0.2640		37.5500	3.1540
$N = 100, M = 9$	11.4286	0.2867		35.7430	3.7778

Table 2. Robustness index of BA scale-free networks ($N = 300$).

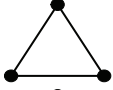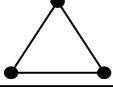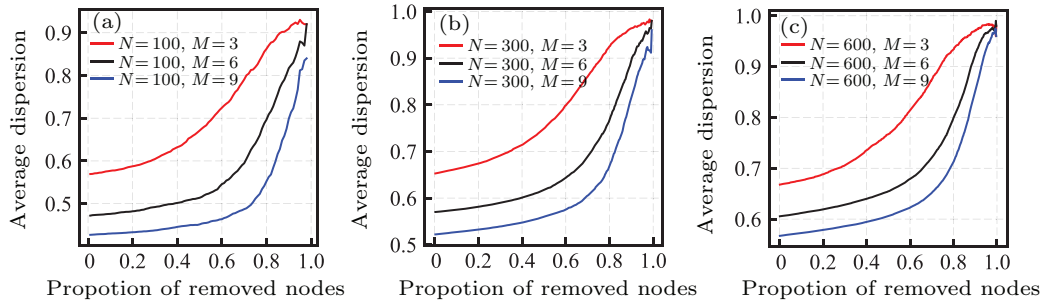| Network parameters	B_G	E_G	The 3rd-order motif	Z-score	H_M
$N = 300, M = 3$	156.6521	0.1737		66.4570	1.3542
$N = 300, M = 6$	63.6207	0.2150		65.5700	2.8258
$N = 300, M = 9$	38.5040	0.2390		61.1330	3.6219

Table 3. Robustness index of BA scale-free networks ($N = 600$).

Network parameters	B_G	E_G	The 3rd-order par motif	Z-score	H_M
$N = 600, M = 3$	321.5120	0.1661		73.1460	1.2226
$N = 600, M = 6$	135.3527	0.1973		72.1990	2.5155
$N = 600, M = 9$	83.1624	0.2165		67.9660	3.4017

Fig. 3. Dispersion curves of BA scale-free networks under random attacks (N is the same) for $N = 100$ (a), 300 (b), and 600 (c).

Fig. 4. Dispersion curves of BA scale-free networks under random attacks (M is the same) for $M = 3$ (a), 6 (b), and 9 (c).

In order to analyze the robustness of BA scale-free network under random attacks by using network dispersion, we use the method of randomly removing nodes to destroy the network, randomly removing one node each time, conducting 100 independent experiments on each network, and then calculating the average value of network dispersion obtained from 100 independent experiments. The dispersion curves are shown in Figs. 3 and 4.

In Figs. 3 and 4, the abscissa represents the proportion of removed nodes, and the ordinate represents the average value of network dispersion obtained from 100 independent experi-

ments. Since we remove nodes from the network randomly, when the number of removed nodes is large enough, there may be only one edge left in the network. Combining with formula (6), it can be seen that the network dispersion will become 0, which will reduce the average value of network dispersion obtained from 100 independent experiments to a certain extent, that is, the average dispersion of the network will be infinitely close to 1, but it may not be equal to 1.

In Fig. 3, from the overall change trend of the curve it follows that for the BA scale-free networks, when the network scale$N$ is the same, the larger the number of edges M in the

preferred connection, the smaller the dispersion in the whole process of random attacks is, that is, the stronger the robustness of the network is. At the same time, it can be found from Fig. 4 that for BA scale-free networks, if M in the preference connection is the same, the smaller the network scale N, the smaller the dispersion in the whole process of random attacks is, that is, the stronger the robustness of the network.

The experiment on BA scale-free networks shows that the measurement results of H_M, B_G, E_G, and D_G on network robustness are consistent with each other, which verifies the effectiveness of H_M in measuring network robustness. That is, in the BA scale-free network, H_M, the entropy of node degree distribution based on motif can be used as a new network robustness measurement index.

4.1.2. Analysis of cause for H_M change

It can be found from Eq. (2) that H_M is related to k_M and $P_M(k_M)$, the node degree distribution function based on motif is a comprehensive representation of k_M and $P_M(k_M)$. Therefore, for BA scale-free networks, we first study the node degree distribution function based on motif, and then further analyze the relationship between the node degree distribution function based on motif and H_M.

The node degree based on the motif is a generalization of the traditional node degree. In the BA scale-free networks, the traditional node degree distribution function obeys the power-law distribution. Therefore, it is of research significance to fit the node degree distribution function based on the motif according to the power-law distribution and judge whether it obeys the power-law distribution through the fitting results.

For the nine BA scale-free networks involved in Subsubsection 4.1.1, we will fit the traditional degree distribution function and the node degree distribution function based on the motif by power-law distribution. The curve fitting results and index calculation values are shown in Figs. 5 and 6.

By observing Figs. 5 and 6, it can be found that in the BA scale-free networks, like the traditional node degree distribution function, the node degree distribution function based on motif can be well fitted by power-law distribution. When N is the same and M increases gradually, the heavy-tailed distribution of the node degree distribution function based on the motif becomes more obvious, that is, the number of nodes with a larger motif degree becomes larger, resulting in larger H_M and stronger robustness. When M is the same and N increases gradually, although there is still a heavy-tailed distribution, the frequency of node degree value based on motif at a small value will gradually increase, that is, the number of nodes with a small motif degree value will increase, resulting in smaller H_M and weaker network robustness.

The statistical data shown in Table 4 can be obtained by combining the entropy of node degree distribution based on motif with results in Figs. 5 and 6.
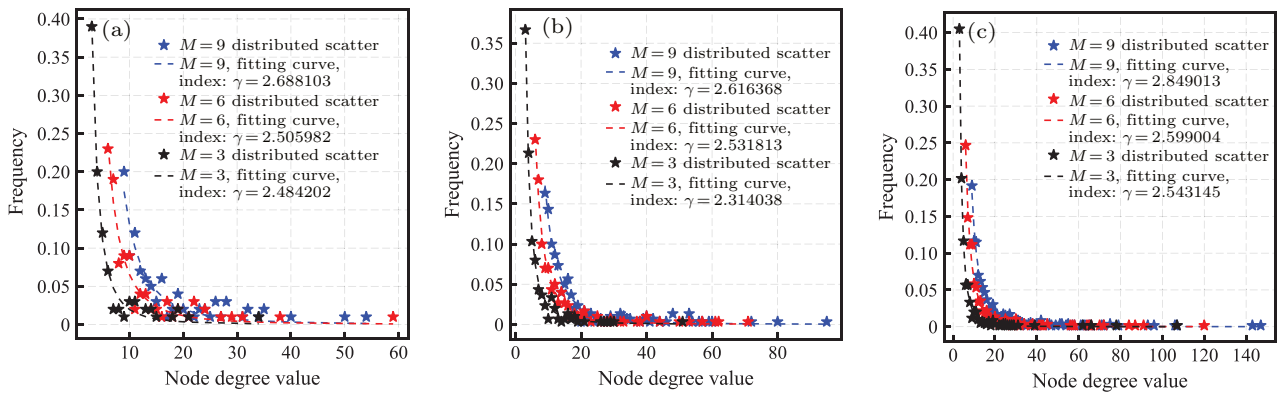

Fig. 5. Traditional degree distribution curve of BA scale-free networks for $N = 100$ (a), 300 (b), and 600 (c).
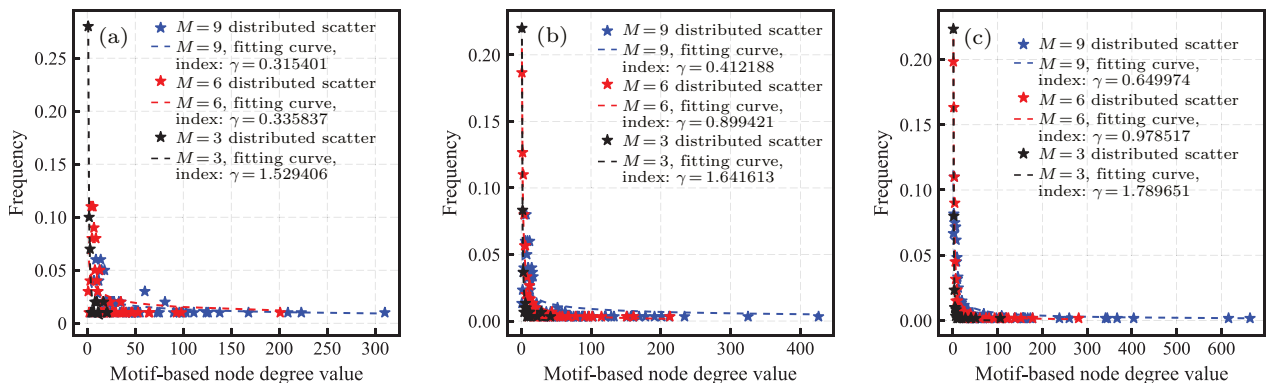

Fig. 6. Motif-based node degree distribution curve of BA scale-free networks, for $N = 100$ (a), 300 (b), and 600 (c).

Table 4. Entropy of node degree distribution based on motif and index γ.

Network parameters	H_M	Traditional node degree distribution function index γ_1	Motif-based node degree distribution function index γ_2
$N = 100, M = 3$	1.7415	2.4842	1.5294
$N = 100, M = 6$	3.1540	2.5060	0.3358
$N = 100, M = 9$	3.7778	2.6881	0.3154
$N = 300, M = 3$	1.3542	2.3140	1.6416
$N = 300, M = 6$	2.8258	2.5318	0.8994
$N = 300, M = 9$	3.6219	2.6164	0.4122
$N = 600, M = 3$	1.2226	2.5431	1.7897
$N = 600, M = 6$	2.5155	2.5990	0.9785
$N = 600, M = 9$	3.4017	2.8490	0.6500

It can be found from Table 4 that for the BA scale-free networks, when N is the same and M gradually increases, the trend of H_M, that is, the changing trend of network robustness is the same as that of γ_1, but opposite to γ_2. When M is the same and N increases gradually, the trend of H_M, that is, the changing trend of network robustness and γ_2 are opposite to each other.

4.2. Computational complexity analysis

In this subsection, we will calculate the time complexity of different indicators when measuring network robustness, and highlight the superiority of our method by comparing the time complexity.

When calculating H_M, k_M, and $P_M(k_M)$ will be involved, and the calculation of k_M turns the most complicated, with the time complexity being $O(N \cdot N_M)$, where N_M represents the number of the 3rd-order motifs in the network. For B_G, the computational complexity mainly focuses on B_e, the time complexity is $O(N^2 \cdot L)$, and the time complexity of E_G and D_G are $O(N^2)$ and $O(N^3)$ respectively.

In the BA scale-free networks, only when M is small, there is $O(N \cdot N_M) < O(N^2)$, that is, the computational complexity of H_M is less than that of E_G. However, E_G considers the robustness from a microscopic perspective by traversing nodes, H_M is based on motifs, and from a mesoscopic perspective, it also considers the coupling between the nodes themselves and the nodes, that is, although the time complexity of H_M is slightly higher than that of E_G, it considers the influence of more local information on the network robustness. For H_M, D_G, and B_G, in the BA scale-free networks, there is $N_M < N^2 < N \cdot L$, so $O(N \cdot N_M) < O(N^3) < O(N^2 \cdot L)$, namely, the computational complexity of H_M is lower than that of D_G and B_G, which highlights the superiority of H_M.

5. Conclusions and prospect

Motif, as a frequent interconnection pattern in the network, is used to measure the robustness of the network, and the decisive effect of the node itself and the interaction be-

tween the nodes on the robustness of the network is considered. Based on the 3rd-order motif, the entropy of node degree distribution is used to measure the robustness of the network under random attacks. At the same time, experiments on BA scale-free networks verify the effectiveness and superiority of our method, that is, the entropy of node degree distribution based on the motif can be used as a new index to measure network robustness, and the robustness of the network is directly proportional to it. In the future work, the robustness of directed networks and multilayer networks will be analyzed and measured based on motifs.

References

[1] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D U 2006 *Phys. Rep.-Rev. Sec. Phys. Lett.* **424** 175
[2] Albert R, Jeong H and Barabási A L 2000 *Nature* **406** 378
[3] Newman M and Ghoshal G 2008 *Phys. Rev. Lett.* **100** 138701
[4] Clemente G P and Cornaro A 2020 *Soft Comput.* **24** 13687
[5] Wang X R, Pournaras E, Kooij R E and Van Mieghem P 2014 *Eur. Phys. J. B* **87** 221
[6] Ellens W, Spieksma F M, Van Mieghem P, Jamakovic A and Kooij R E 2011 *Linear Alg. Appl.* **435** 2491
[7] He Z W, Liu S and Zhan M 2013 *Physica A* **392** 4181
[8] Kasthurirathna D, Piraveenan M and Thedchanamoorthy G 2013 *IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS),* April 16–19, 2013, Singapore, Singapore, p. 122
[9] Zhou A, Maletic S and Zhao Y 2018 *Physica A* **502** 459
[10] Liu D, Tse C K and Zhang X 2019 *IEEE Trans. Circuits Syst. II-Express Briefs* **66** 833
[11] Zhang X and Ise C K 2015 *IEEE International Symposium on Circuits and Systems (ISCAS),* May 24–27, 2015, Lisbon, Portugal, p. 2684
[12] Yang H H and An S 2021 *Alex. Eng. J.* **60** 2065
[13] Manzano M, Calle E, Torres-Padrosa V, Segovia J and Harle D 2014 *Comput. Netw.* **57** 3641
[14] Zou Z Y, Xiao Y and Gao J Z 2013 *Kybernetes* **42** 383

[15] Yang Y H, Liu, Y X, Zhou M X, Li F X and Sun C 2015 *Saf. Sci.* **79** 149

[16] Zhou X, Zhang F M, Zhou W P, Zou W and Yang F 2012 *Acta Phys. Sin.* **61** 190201 (in Chinese)

[17] Hong C, He N, Lordan O, Liang B Y and Yin N Y 2017 *Physica A* **478** 63

[18] Wu Y P, Chen Z L, Yao K, Zhao X D and Chen Y C 2019 *Fractals-Complex Geom. Patterns Scaling Nat. Soc.* **27**

[19] Wu J, Barahona M, Tan Y J and Deng H Z 2006 *IEEE Trans. Syst. Man Cybern. Paart A-Syst. Hum.* **41** 1244

[20] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D and Alon U 2002 *Science* **298** 824

[21] Onnela J P, Saramaki J, Kertesz J and Kaski K 2005 *Phys. Rev. E* **71** 065103

[22] Barabási A L and Bonabeau E 2003 *Sci. Am.* **288** 60

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

[24] Wernicke S and Rasche F 2006 *Bioinformatics* **22** 1152

[25] Wernicke S 2006 5*th International Workshop on Algorithms in Bioinformatics (WABI* 2005*),* October 3–6, 2005, Mallorca, Spain, p. 347

[26] Han H, Liu W L and Wu L Y 2013 *Acta Phys. Sin.* **62** 168904 (in Chinese)

[27] Borgatti S P 2003 *International Conference on Integration of Knowledge Intenzive Multi-Agent Systems,* September 30–October 4, 2003, Cambridge, Ma, p. 127