

Coarse-Graining Method Based on Hierarchical Clustering on Complex Networks

Lin Liao, Zhen Jia*, Yang Deng

College of Science, Guilin University of Technology, Guilin, China

Email: *jjzzz0@163.com

How to cite this paper: Liao, L., Jia, Z. and Deng, Y. (2019) Coarse-Graining Method Based on Hierarchical Clustering on Complex Networks. *Communications and Network*, 11, 21-34.

<https://doi.org/10.4236/cn.2019.111003>

Received: January 18, 2019

Accepted: February 19, 2019

Published: February 22, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

With the rapid development of big data, the scale of realistic networks is increasing continually. In order to reduce the network scale, some coarse-graining methods are proposed to transform large-scale networks into mesoscale networks. In this paper, a new coarse-graining method based on hierarchical clustering (HCCG) on complex networks is proposed. The network nodes are grouped by using the hierarchical clustering method, then updating the weights of edges between clusters extract the coarse-grained networks. A large number of simulation experiments on several typical complex networks show that the HCCG method can effectively reduce the network scale, meanwhile maintaining the synchronizability of the original network well. Furthermore, this method is more suitable for these networks with obvious clustering structure, and we can choose freely the size of the coarse-grained networks in the proposed method.

Keywords

Complex Network, Synchronizability, Coarse-Graining Method, Hierarchical Clustering

1. Introduction

Our life is full of all kinds of networks. For example, metabolic networks, large-scale power networks, paper citation networks, global transportation networks, scientific research cooperation networks and so on. These networks have same analogous and obvious characteristics—massive nodes and complex interactions. Networks with complex topological properties are called complex networks [1]. A complex network is formed by abstracting the basic units of a complex system into nodes and the interactions between the units into edges. Complex networks are important tools for studying the relationship between the

structure and function of complex systems.

The existed research methods of complex networks are mainly designed for mesoscale networks. Coarse-graining technology is an effective way to study the large-size networks, which can reduce the complexity of networks by merging similar nodes. However, coarse-graining techniques go far beyond the requirement of clustering techniques, because coarse-graining techniques requires the coarse-grained networks to keep the initial network's topological properties or dynamic characteristics, such as the degree distribution, clustering coefficient, correlation properties [2], random walk properties [3], synchronizability [4]. In 2007, Gfeller *et al.*, proposed a spectral coarse-graining (SCG) algorithm, which is based on the eigenvalue spectrum of random walk probability matrix, the algorithm can maintain the random walk characteristics of original networks [5]. In 2008, Gfeller *et al.* further used SCG technology on the Laplace matrix of networks to extract coarse-grained networks, while maintaining the synchronizability of original networks [6]. In 2013, J Chen *et al.* [7] studied the coarse-graining method on the networks with obvious clustering structure in literature [5] [6]: through a large number of simulation experiments on WS small world networks, ER random networks, and BA scale-free networks, we can draw the conclusion that the clustering structure of networks can enhance the coarse-graining effect. In 2016, Shuang Xu *et al.* proposed a coarse-graining method based on K-means clustering which by redistributing the central cluster nodes of the network and clustering to reduce the network scale, the method can keep some statistical characteristics of initial networks in some extent, such as degree distribution, assortativity and so on [9]. In 2017, the ISCG [8] algorithm proposed by our research group Zhou *et al.*, the algorithm adopt split clustering method determining which nodes should be are grouped together, the algorithm can effectively keep the synchronizability of original networks, and the effect is better than SCG method in directed networks. Recently, our research team proposed a spectral coarse-graining method based on K-means clustering, by numerical simulations, they find that this method has better effect in preserving the synchronizability of original networks than SCG algorithm. Xingbo Liu proposed a cohesive hierarchical clustering algorithm based on variable merging threshold, and gave a reference formula for the distance threshold [10].

Coarse-graining technique is an important method for studying large-scale complex networks. However, almost every coarse-graining method has some inadequacies: for example, in the SCG method, computing the eigenvalues of network's Laplacian matrix will take a lot of time, so the method is difficult to be used in large-scale real networks, furthermore, the SCG method cannot accurately control the size of coarse-grained network; The K-means clustering coarse-graining method requires to define the objective function, and may arise the problem of trapping local minimum values or selecting initial points. This paper proposes a new coarse-graining method based on hierarchical clustering (HCCG) on complex networks. The distance and similarity of the coarse graining method are easy to define, and we can choose the size of the reduced network freely. More-

over, this method does not need to define the objective function and does not cause the problem of selecting the initial point. The new coarse-graining method can make up for some shortcomings of the above-mentioned methods. In the HCCG method, we use the hierarchical clustering method to cluster the network nodes, and update the weights of edges between clusters to extract the coarse-grained network. Furthermore, we apply the HCCG method to WS small world networks, ER random networks and BA scale-free networks. Simulation experiments show that the HCCG method can keep the synchronizability of the initial network well, and the method is more suitable for the networks with more obvious clustering structure.

The rest of the paper is organised as follows. In Section 2, the mathematical basis of the HCCG method is introduced. The steps of the HCCG method are presented in Section 3. In Section 4, a large number of numerical simulations on several typical networks verify the feasibility and effectiveness of the proposed method. Finally, conclusions and discussion are drawn in Section 5.

2. Mathematical Basis

2.1. Hierarchical Clustering Algorithm

Clustering is an unsupervised learning process. By clustering, the similarity is as large as possible in the same cluster and as small as possible in different clusters. The fundamental difference between sorting technique and clustering technique is that sorting technique must know the data characteristic on which it is based in advance, and the clustering technique is to find the data characteristics. Therefore, clustering analysis is usually used as a data preprocessing process in many fields, which is the basis for further data analysis and processing [11]. There are usually four clustering methods: partitioning methods, hierarchical methods, density-based methods, and grid-based methods [12]. In this paper, we use the hierarchical methods to cluster networks nodes.

Hierarchical clustering methods merge or separate data objects recursively until some termination condition met. According to the order of hierarchical decomposition, the method can be divided into bottom-up algorithm and top-down algorithm. This paper adopts the bottom-up method, which is a cohesive hierarchical clustering algorithm. The algorithm starts with every individual as a cluster, then searching for the nearest cluster to group into one cluster. After merging once, the total number of clusters is reduced by one, until the required number of clusters or the nearest threshold is reached. In this paper, Jaccard distance is used to calculate the distance between two different nodes. The methods of calculating the distance between two different clusters include Single Linkage, Complete Linkage, Average Linkage and so on. In this paper, we use the Average Linkage in the HCCG method.

$$dist_{avg}(C_i, C_j) = \frac{1}{n_i n_j} \sum_{p \in C_i, p' \in C_j} \{|p - p'|\} \quad (1)$$

where $|p - p'|$ is the distance between two nodes, i, j represent two different

nodes of the initial network, the coarse-grained network has \tilde{N} clusters, labeled with $C = \tilde{1}, \tilde{2}, \dots, \tilde{N}$, C_i is the label of the cluster of node i , $|C_i|$ is the cardinality of the set C_i , $n_i = |C_i|$ (n_i is the number of nodes in the cluster C_i).

2.2. Index of Network Synchronizability

Consider an unweighted and undirected network with N nodes, $A = (a_{ij})^{N \times N}$ is the adjacency matrix of the network, $(a_{ii}) = 0$ and $(a_{ij}) = (a_{ji}) = 1$ if node i connects to j ($i \neq j$), otherwise $(a_{ij}) = (a_{ji}) = 0$. $D = \text{diag}(k_1, k_2, \dots, k_N)$ is a diagonal matrix, k_i is the degree of i th node. $L = D - A$, $L = (L_{ij})_{N \times N} \in R^{N \times N}$ is the Laplacian matrix of the network. L satisfies the dissipative coupling condition: $\sum_{i=1}^N L_{ij} = 0$. Since the network is connected, L is a symmetric matrix with nonnegative eigenvalues: $0 = \lambda_1 < \lambda_2 \leq \dots \leq \lambda_N$.

Generally, according to the difference of synchronization regions, network systems can be divided into four types, each type corresponds to a kind of synchronization region: 1) Type-I networks correspond to bounded regions (α_1, α_2) ; 2) Type-II networks correspond to unbounded regions (α_1, ∞) ; 3) Type-III networks correspond to union of several bounded regions $\cup(\alpha_{i_1}, \alpha_{i_2})$; 4) Type-IV networks correspond to empty sets. Generally, the networks of the case (3) and (4) are difficult or impossible to achieve synchronization. Fortunately, most networks are in the case of (1) and (2). In Type-I networks, The synchronizability of networks can be characterized by the minimum non-zero eigenvalue λ_2 . The larger the value of λ_2 , the smaller the coupling strength is needed to achieve synchronization, the synchronizability of Type-I networks are stronger. In Type-II networks, the synchronizability of networks can be characterized by the ratio of $\tilde{\lambda}_2 / \tilde{\lambda}_N$. Only when the eigenvalues ratio $\tilde{\lambda}_2 / \tilde{\lambda}_N > \alpha_1 / \alpha_2$ are satisfied can the network achieve synchronization. So, the larger the value of $R = \tilde{\lambda}_2 / \tilde{\lambda}_N$, the stronger the synchronizability of Type-II networks. Therefore, if λ_2 or $\tilde{\lambda}_2 / \tilde{\lambda}_N$ is unchanged in the process of coarse-graining, we can deem the synchronizability of the network is maintained [13].

3. Hierarchical Clustering Coarse-Graining Scheme

There are three main steps in the HCCG method: the first step is calculating the distance between two nodes for obtaining the distance matrix of the initial network by using Jaccard distance, and constructing n single-member clusters; the second is calculating the distance between two different clusters, using the hierarchical clustering method to cluster the network nodes; the last step is updating the weight of the links between different clusters to extract the coarse-grained network. In this section, the HCCG scheme is introduced revolving around the above content.

3.1. Jaccard Distance Calculation

Supposing $\mathcal{G}(i) = \{j \mid a_{ij} \neq 0\}$ is the set of neighbor nodes of node i . $|\mathcal{G}(i)|$ is

the cardinality of the set $\mathcal{G}(i)$. $|\mathcal{G}(i)| = k_i$, k_i are the degrees of nodes i . The Jaccard coefficient is defined as follows:

$$J_{ij} = \frac{|\mathcal{G}(i) \cap \mathcal{G}(j)|}{|\mathcal{G}(i) \cup \mathcal{G}(j)|} = \frac{|\mathcal{G}(i) \cap \mathcal{G}(j)|}{|\mathcal{G}(i)| + |\mathcal{G}(j)| - |\mathcal{G}(i) \cap \mathcal{G}(j)|} \quad (2)$$

here, i, j are two different nodes. $\mathcal{G}(i) \cap \mathcal{G}(j)$ is the common neighbor nodes set of node i and j , $\mathcal{G}(i) \cup \mathcal{G}(j)$ is the union of the neighbor nodes of node i and j . When $|\mathcal{G}(i)| = 0$ and $|\mathcal{G}(j)| = 0$, $J_{ij} = 1$. Jaccard distance is an index related to Jaccard coefficient. The lower the nodes similarity, the larger the Jaccard distance. The Jaccard distance is defined as follows:

$$d_{ij} = 1 - J_{ij} = \frac{|\mathcal{G}(i) \cup \mathcal{G}(j)| - |\mathcal{G}(i) \cap \mathcal{G}(j)|}{|\mathcal{G}(i) \cup \mathcal{G}(j)|} \quad (3)$$

A toy network is used to illustrate how to calculate the distance between two different nodes, using Jaccard distance method obtain the **Table 1**.

In **Figure 1**, $\mathcal{G}(1) = \{3, 4\}$, $\mathcal{G}(2) = \{3, 4, 5\}$, $\mathcal{G}(3) = \{1, 2\}$, $\mathcal{G}(4) = \{1, 2\}$, $\mathcal{G}(5) = \{2, 6\}$, $\mathcal{G}(6) = \{5\}$. **Table 1** can be obtained based on Equation (3).

3.2. Node Clustering of Networks

The basic idea of hierarchical clustering method is to calculate the similarity between nodes by some similarity index, and to rank the nodes according to the similarity from high to low, then to merge the nodes step by step. The main steps are as follows:

- 1) Obtain the adjacency matrix A of the network;
- 2) Use Jaccard distance method calculating the distance between two different nodes to obtain the distance matrix of the network, construct n single-member clusters, the height of each cluster is 0;
- 3) Use Average Linkage method calculating the distance between two different clusters to search for the nearest clusters C_i, C_j , merging C_i, C_j , reducing the number of all clusters by 1, take the distance between the two clusters merged as the height of the upper layer;
- 4) Calculate the distance between the newly generated cluster and other clusters in this layer. If the termination condition is satisfied, the algorithm will end, otherwise it will turn to (3).

To better explain the steps of clustering, we use the example in **Figure 2**, a 6-node toy network is clustered into one cluster. The obtained hierarchical clustering dendrogram is shown in **Figure 3**.

In **Table 1**, the closest distance is $d_{3,4} = 0$ on the **Figure 2(a)**, so node 3 and node 4 are grouped together; based on Equation (3), we know $d_{1,2} = \frac{1}{3}$ is the minimum distance on the **Figure 2(b)**, so node 1 and node 2 are grouped together; according to Equation (1),

$$\text{dist}_{\text{avg}}(C_{1,2}, C_{3,4}) = \frac{1}{2 \times 2}(1+1+1+1) = 1, \quad \text{dist}_{\text{avg}}(C_{1,2}, C_5) = \frac{1}{2 \times 1}(1+1) = 1,$$

Table 1. Distance matrix of the toy network.

d_{ij}	1	2	3	4	5	6
1		$\frac{1}{3}$	1	1	1	1
2			1	1	1	$\frac{2}{3}$
3				0	$\frac{2}{3}$	1
4					$\frac{2}{3}$	1
5						1
6						

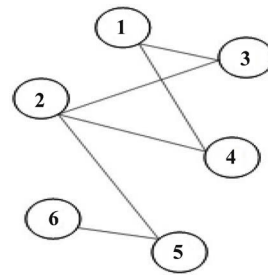


Figure 1. A toy network.

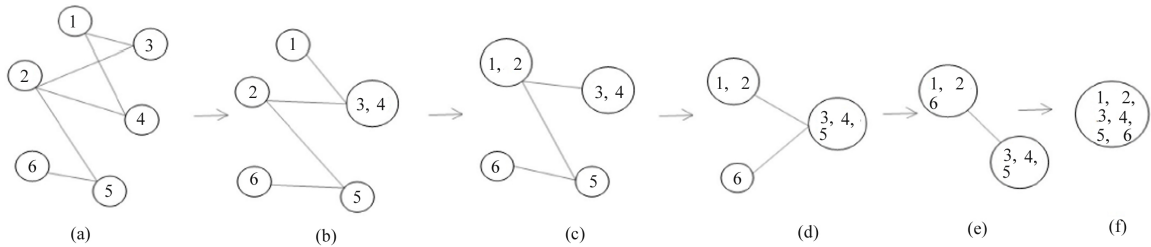


Figure 2. Node clustering of the toy network.

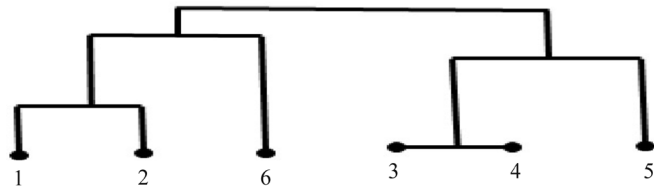


Figure 3. Hierarchical clustering dendrogram of the toy network.

$$dist_{avg}(C_{1,2}, C_6) = \frac{1}{2 \times 1} \left(1 + \frac{2}{3} \right) = \frac{5}{6}, \quad dist_{avg}(C_{3,4}, C_5) = \frac{1}{2 \times 1} \left(\frac{2}{3} + \frac{2}{3} \right) = \frac{2}{3},$$

$$dist_{avg}(C_{3,4}, C_6) = \frac{1}{2 \times 1} (1 + 1) = 1, \quad dist_{avg}(C_5, C_6) = \frac{1}{1 \times 1} (1 + 1) = 1,$$

the nearest distance is $dist_{avg}(C_{3,4}, C_5) = \frac{2}{3}$ on the **Figure 2(c)**, the cluster of $C_{3,4}$ and node 5 are grouped together; based on Equation (1),

$$dist_{avg}(C_{1,2}, C_{3,4,5}) = \frac{1}{2 \times 3} (1+1+1+1+1+1) = 1,$$

$$dist_{avg}(C_{1,2}, C_6) = \frac{1}{2 \times 1} \left(1 + \frac{2}{3}\right) = \frac{5}{6}, \quad dist_{avg}(C_{3,4,5}, C_6) = \frac{1}{3 \times 1} (1+1+1) = 1,$$

the nearest distance is $dist_{avg}(C_{1,2}, C_6) = \frac{5}{6}$ on the **Figure 2(d)**, so the cluster of $C_{1,2}$ and node 6 are grouped together; according to Equation (1), the nearest distance is

$$dist_{avg}(C_{1,2,6}, C_{3,4,5}) = \frac{1}{3 \times 3} (1+1+1+1+1+1+1+1+1) = 1$$

on the **Figure 2(e)**, all nodes are grouped together into one cluster.

3.3. Updating Edges of the Coarse-Grained Network

Clusters are obtained according to the clustering steps which in Section 3.2, then based on the Equation (4) update the edge weight between different clusters to extract the coarse-grained network. C_s, C_t are two different clusters, and the weight of C_s to C_t are defined as:

$$W_{C_s \rightarrow C_t} = \frac{\sum_{x \in C_s, y \in C_t} a_{xy}}{|C_t|} \quad s, t = 1, 2, \dots, \tilde{N}, s \neq t \quad (4)$$

In **Figure 4**, $C_{1,2} = \tilde{1}$, $C_{3,4,5} = \tilde{2}$, $C_6 = \tilde{3}$,

$$W_{\tilde{1} \rightarrow \tilde{2}} = \frac{\sum_{x \in \tilde{1}, y \in \tilde{2}} a_{xy}}{|\tilde{2}|} = \frac{1+1+1+1+1}{3} = \frac{5}{3}, \quad W_{\tilde{2} \rightarrow \tilde{1}} = \frac{\sum_{x \in \tilde{2}, y \in \tilde{1}} a_{xy}}{|\tilde{1}|} = \frac{1+1+1+1+1}{2} = \frac{5}{2},$$

$$W_{\tilde{2} \rightarrow \tilde{3}} = \frac{\sum_{x \in \tilde{2}, y \in \tilde{3}} a_{xy}}{|\tilde{3}|} = \frac{1}{1} = 1, \quad W_{\tilde{3} \rightarrow \tilde{2}} = \frac{\sum_{x \in \tilde{3}, y \in \tilde{2}} a_{xy}}{|\tilde{2}|} = \frac{1}{3} = \frac{1}{3}.$$

4. Application and Simulation

In this section, we apply the HCCG method to several typical complex networks. The several types of networks including: WS small world networks, ER random networks and BA scale-free networks. N is the size of initial networks; \tilde{N} is the size of coarse-grained networks, $\tilde{\lambda}_2$ is the minimum non-zero eigenvalue of the coarse-grained network, $\tilde{R} = \tilde{\lambda}_2 / \tilde{\lambda}_N$ is the eigenvalue ratio of the coarse-grained

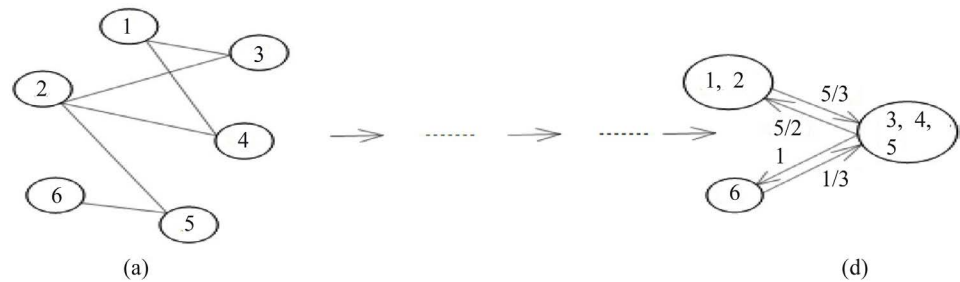


Figure 4. Updating the links weight between clusters.

network. The closer the values of $\tilde{\lambda}_2$ and λ_2 (or R and \tilde{R}), the better the effect of the HCCG method on keeping the synchronizability of original networks.

4.1. WS Small World Network

In 1998, Watts and Strogatz [14] [15] proposed the algorithm for the formation of small world networks and established WS model, found that most real networks have short average path lengths and high clustering coefficients. The WS network is based on the random edge rewiring of regular networks, which stipulated that there mustn't be self-ring and multiple edges. Random reconnection probability $p=0$ corresponds to complete regular networks and $p=1$ corresponds to complete random networks. By adjusting the value of p , the transition from regular networks to random networks can be realized. **Figures 5(a)-(c)** shows the evolution curves of $\tilde{\lambda}_2$ by using the HCCG method in WS small world networks with initial sizes $N=1000, 2000, 3000$ respectively, and **Figures 5(d)-(f)** shows the evolution curves of \tilde{R} by using the HCCG method in WS small world networks with initial sizes $N=1000, 2000, 3000$ respectively. For each cases, networks with average degrees $\langle k \rangle = 4, 6, 8$ and rewiring probabilities $p = 0.1, 0.2, 0.45$ are considered. Each simulation result was obtained by 5 independent repeated experiments.

As shown in **Figures 5(a)-(c)**, in Type-I networks, when $p = 0.1, 0.2$, $\langle k \rangle = 4, 6, 8$ the synchronizability of original networks can be maintained well in the coarse-graining process. However, with the increasing of p , the effect has not been so good. When the p value is fixed, the smaller the $\langle k \rangle$, the better the effect of the HCCG method. In the literature [16]: The clustering structure of networks contribute to improve the synchronicity of coarse-graining methods. The construction algorithm of WS small world model starts with the regular graph, at this time, on each nodes left and right sides is connected with $\langle k \rangle / 2$ adjacent nodes respectively, the clustering structure of the network is clear relatively, it is effective to keep the synchronizability of the original network by using the HCCG method; then the network is reconnected randomly with probability p . With the increasing of p and p , the clustering structure becomes weaker. The reason is that: $\langle k \rangle \times p$ is the number of the nodes randomly reconnected, when the network scale is fixed, the larger p and p are, the more random reconnected nodes, the weaker the clustering structure. Therefore, the smaller the $\langle k \rangle$ and p , the better the effect of the HCCG method.

As shown in **Figures 5(d)-(f)**, in Type-II networks, the evolution curves of \tilde{R} are consistent with that of $\tilde{\lambda}_2$. The reason is also consistent with in the case of Type-I networks. The difference is: when $\langle k \rangle$ and p are quite large, the effect of the HCCG method in Type-II networks is better than that in Type-I networks.

When p increases gradually, The minimum non-zero eigenvalue $\tilde{\lambda}_2$ and the maximum eigenvalue $\tilde{\lambda}_N$ increase simultaneously, $\tilde{\lambda}_2$ increases much faster than $\tilde{\lambda}_N$ which lead to the rise in $\tilde{R} = \tilde{\lambda}_2 / \tilde{\lambda}_N$ [17], so with p increases the synchronizability of the network is enhanced. However, with the increase of

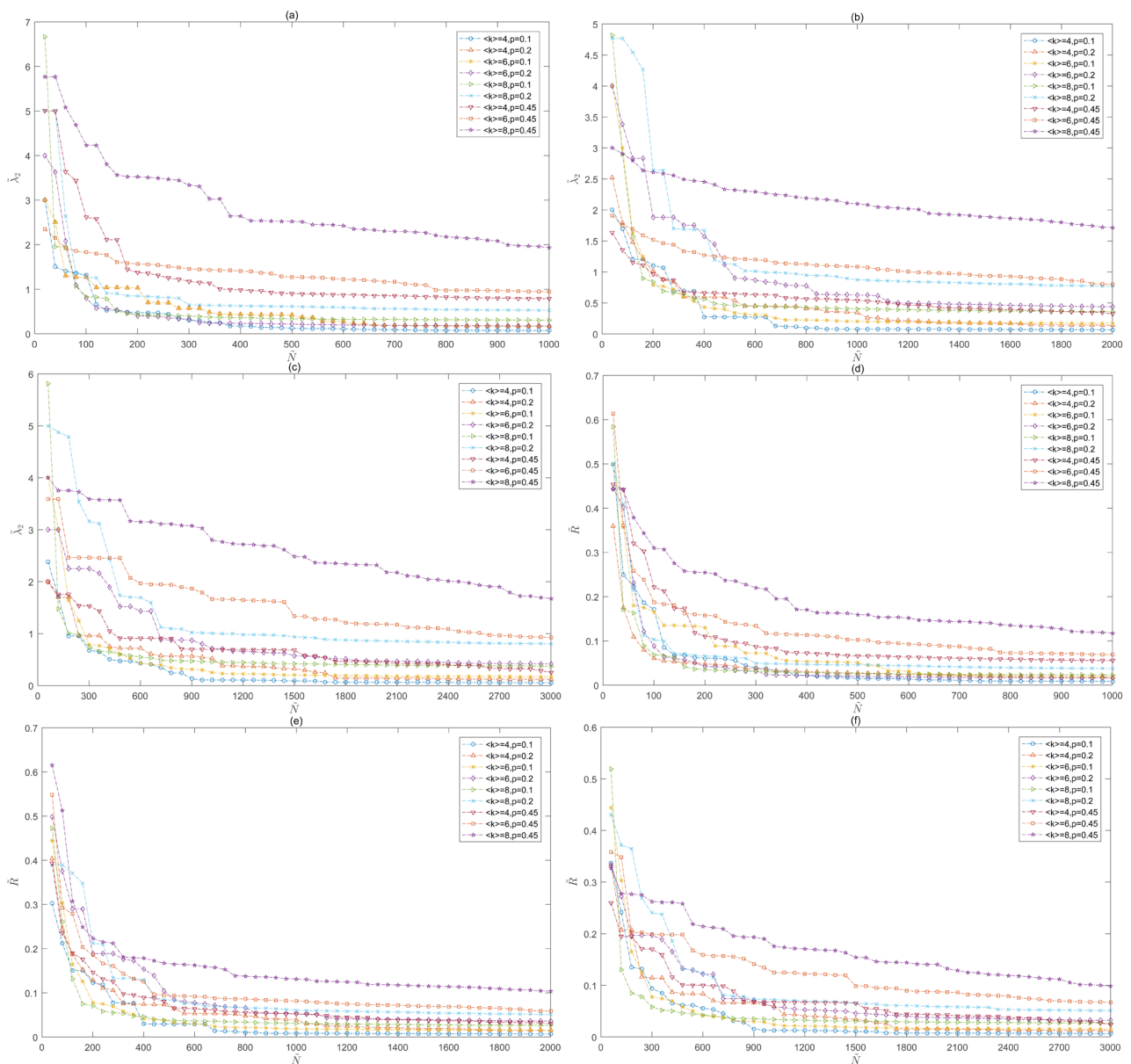


Figure 5. (a)-(c) Evolution of $\tilde{\lambda}_2$ by using the HCCG method in WS small world network; (d)-(f) Evolution of \tilde{R} by using the HCCG method in WS small world network.

$\langle k \rangle$ and p , the clustering structure of the network becomes weaker, the effect of maintaining original network's synchronizability becomes not so good. So the HCCG method is more suitable for the networks with obvious clustering structure.

4.2. ER Random Network

In 1959, Erdos and Renyi, two Hungarian mathematicians, proposed algorithm for the formation of random graphs [18]. The ER random graph $G(N, p)$ given N nodes firstly, the connecting probability between any two different points is p , ($0 < p < 1$). **Figures 6(a)-(c)** shows the evolution curves of $\tilde{\lambda}_2$ by using the

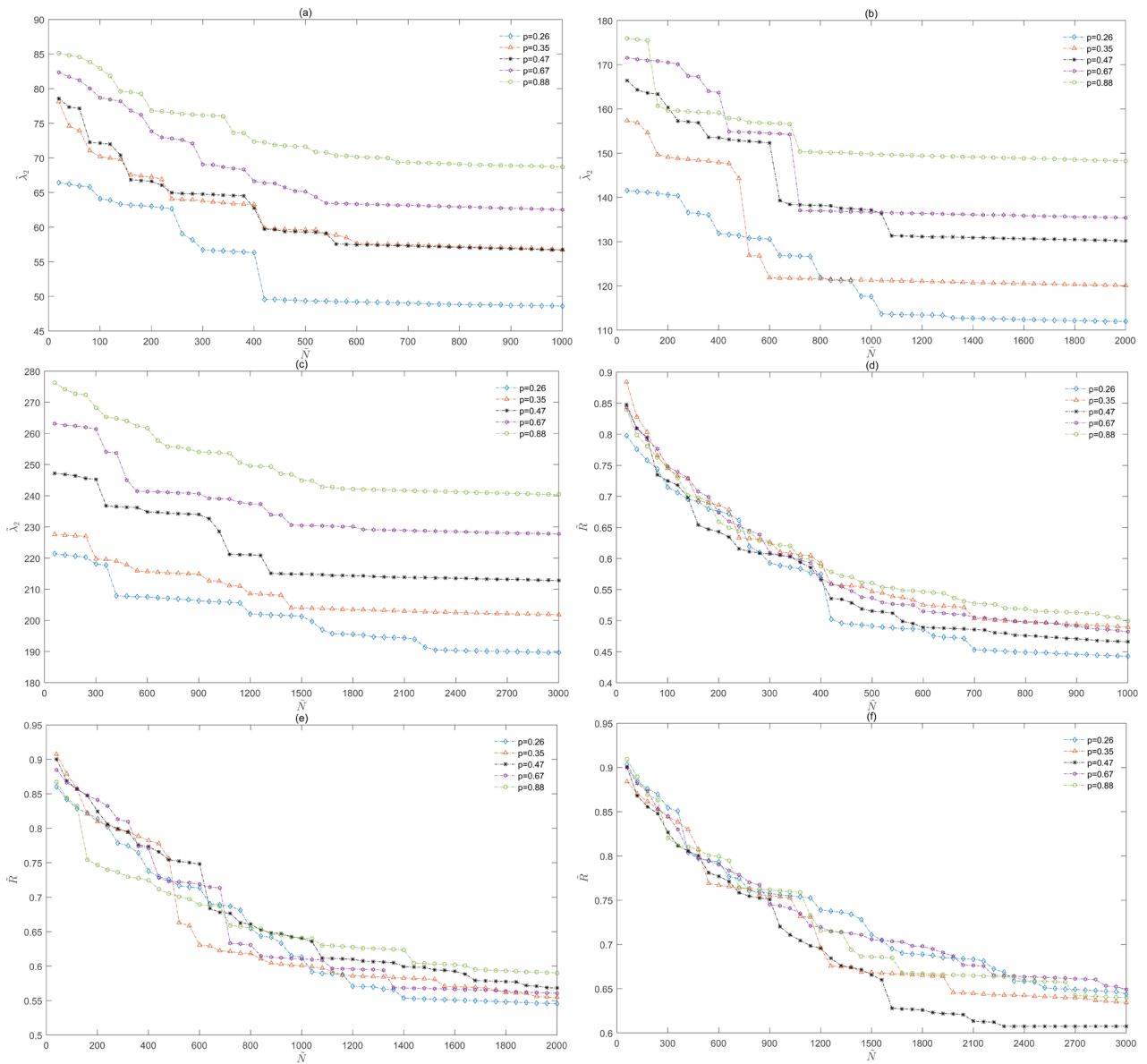


Figure 6. (a)-(c) Evolution of $\tilde{\lambda}_2$ by using the HCCG method in ER random network; (d)-(f) Evolution of \tilde{R} by using the HCCG method in ER random network.

HCCG method in ER random networks with initial sizes $N = 1000, 2000, 3000$ respectively, and **Figures 6(d)-(f)** shows the evolution curves of \tilde{R} by using the HCCG method in ER random networks with initial size $N = 1000, 2000, 3000$ respectively. For each case, the connecting probability is $p = 0.26, 0.35, 0.47, 0.67, 0.88$. Each simulation result was obtained by 5 independent repeated experiments.

The eigenvalue spectral distribution of ER random networks is symmetrical, and as the connection probability p increases, the range of eigenvalue distribution shrinks rapidly, at the same time, $\tilde{\lambda}_2$ and $\tilde{\lambda}_N$ increase. The main reason is that with p increases, the number of isolated clusters decreases, the maximum degree d_{\max} increases, so $\tilde{\lambda}_N$ is increased. However, the growth rate of $\tilde{\lambda}_2$ is

faster than that of $\tilde{\lambda}_N$, which leads to the increase of $\tilde{R} = \tilde{\lambda}_2 / \tilde{\lambda}_N$. Therefore, from a view of statistical perspective, the synchronizability is gradually increased as the connection probability p increases [19]. **Figures 6(a)-(c)**, in Type-I networks, when the size of the coarse-grained network is larger than or equal to that of original network's 60 percent, although the difference of the value $\tilde{\lambda}_2$ is quite big in the case of different p , the method can maintain the synchronizability of the original network well. We can see that the larger p , the larger the value of $\tilde{\lambda}_2$. So the larger p , the stronger the synchronizability of the network. The minimum size (that can maintain the synchronizability of original network) in the case of $p = 0.88$ is significantly smaller than that in the case of $p = 0.26$. Therefore, the larger p , the better the effect of the HCCG method, which is consistent with the reference [19]. As shown in **Figures 6(d)-(f)**, in Type-II networks, the HCCG method also can maintain the synchronizability of the original network well. So, the HCCG method is suitable for ER networks.

4.3. BA Scale-Free Network

In 1999, Barabási and Albert established BA model to explain the scale-free property of complex networks [19] [20]. They believed that most complex systems in the real world evolve dynamically, and these complex systems are open and self-organizing. Scale-free phenomena in real networks comes from two important factors: Growth pattern and Priority connection mode. Another an important discovery in network science research is that the degree distribution of networks can be described by appropriate power-law form. The degrees of nodes in these networks have no obvious characteristic length, so these networks are called scale-free networks. **Figures 7(a)-(c)** shows the evolution curves of $\tilde{\lambda}_2$ by using the HCCG method in BA scale-free networks with initial sizes $N = 1000, 2000, 3000$ respectively, and **Figures 7(d)-(f)** shows the evolution curves of \tilde{R} by using the HCCG method in BA scale-free networks with initial sizes $N = 1000, 2000, 3000$ respectively. The larger the power-law exponents γ of the network, the weaker the heterogeneity. In the real complex networks, γ mostly range among 2 and 3. So we main consider the networks with power-law exponents $\gamma = 2.2, 2.4, 2.6, 2.8, 3.0$, Each simulation result was obtained by 5 independent repeated experiments.

Figures 7(a)-(c), In the case of Type-I networks, when the scale of coarse-grained network is larger than or equal to 60 percent of the original scale, the HCCG method can maintain the synchronizability of the original network well; **Figures 7(d)-(f)**, In the case of Type-II networks, when the scale of coarse-grained networks is larger than 70 percent of the original scale, the HCCG method can maintain the synchronizability of the original network well, but subsequently, there is a sharp increasing in \tilde{R} , at this time the average degree of the corresponding network increases suddenly. This is consistent with the conclusion in the literature [21]: in Type-II networks, the synchronizability of the network is enhanced with the increasing of average degree; in Type-I

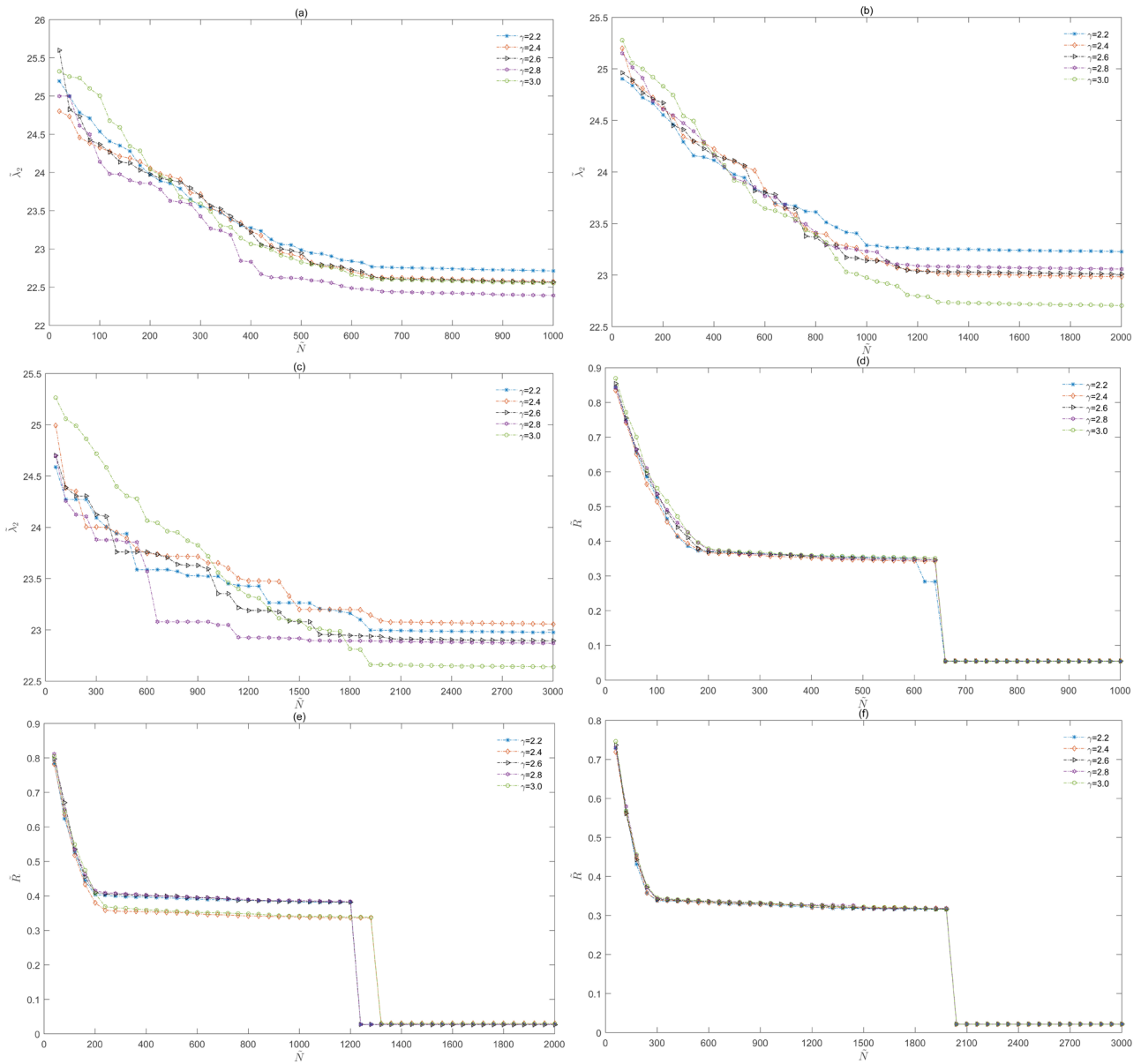


Figure 7. (a)-(c) Evolution of $\tilde{\lambda}_2$ by using the HCCG method in BA scale-free network; (d)-(f) Evolution of \tilde{R} by using the HCCG method in BA scale-free network.

networks, with the increasing of average degree, the synchronizability of the network basically unchanged. ER random networks have some emergence properties: for any given connecting probability p , either almost every graph $G(N, p)$ has a certain nature Q , or almost every graph $G(N, p)$ does not have this property Q . The suddenly increasing of \tilde{R} whether mean that arising certain emergence properties in BA scale-free networks, the problem has yet to be studied.

5. Conclusion and Discussion

Coarse-graining technique is an important method for studying large-scale

complex networks. In the HCCG method, the distance and similarity are defined simply. Jaccard distance is used to obtain the distance matrix of the network. Average Linkage method is used to find the nearest clusters. We use the hierarchical clustering method to cluster the network nodes, and update the weights of edges between clusters to extract the coarse-grained network. Massive simulation experiments show the HCCG method can keep the synchronizability of the original network well, and the method more suitable for the network with obvious clustering structure. Furthermore, the size of coarse-grained networks can be chosen freely in the HCCG method. However, the suddenly increasing of \tilde{R} on BA scale-free network whether mean that certain emergence properties arising, the problem has yet to be studied. In addition, greedy algorithm is used in hierarchical clustering, thence the clustering result is local optimum, may not be global optimum, the problem can be solved by adding random effects, which is also the direction of our future researches.

Acknowledgements

This project was supported by National Natural Science Foundation of China (Nos. 61563013, 61663006) and the Natural Science Foundation of Guangxi (No. 2018GXNSFAA138095).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Jin, Y.P. (2016) Coarsening and Anomaly Detection of Complex Networks. (In Chinese)
- [2] Kim, B.J. (2004) Geographical Coarse Graining of Complex Networks. *Physical Review Letters*, **93**, 168701. <https://doi.org/10.1103/PhysRevLett.93.168701>
- [3] Weiss, G.H. and Havlin, S. (1986) Some Properties of a Random Walk on a Comb Structure. *Physica A Statistical Mechanics and Its Applications*, **134**, 474-482. [https://doi.org/10.1016/0378-4371\(86\)90060-9](https://doi.org/10.1016/0378-4371(86)90060-9)
- [4] Duan, Z., Chen, G. and Huang, L. (2007) Complex Network Synchronizability: Analysis and Control. *Physical Review E Statistical Nonlinear and Soft Matter Physics*, **76**, 056103. <https://doi.org/10.1103/PhysRevE.76.056103>
- [5] Gfeller, D. and Rios, P.D.L. (2007) Spectral Coarse Graining of Complex Networks. *Physical Review Letters*, **99**, 038701. <https://doi.org/10.1103/PhysRevLett.99.038701>
- [6] Gfeller, D. and Rios, P.D.L. (2008) Spectral Coarse Graining and Synchronization in Oscillator Networks. *Physical Review Letters*, **100**, 174104. <https://doi.org/10.1103/PhysRevLett.100.174104>
- [7] Zhou, J., Jia, Z. and Li, K.Z. (2017) Improved Algorithm of Spectral Coarse Graining Method of Complex Network. *Acta Physica Sinica*.
- [8] Chen, J., Lu, J.A., Lu, X.F., Wu, X.Q. and Chen, G.R. (2013) Spectral Coarse Graining of Complex Clustered Networks. *Communications in Nonlinear Science and Numerical Simulation*, **18**, 3036-3045. <https://doi.org/10.1016/j.cnsns.2013.03.020>

-
- [9] Shuang, X. and Pei, W. (2016) Coarse Graining of Complex Networks: A k-Means Clustering Approach. *Control and Decision Conference*.
- [10] Liu, X.B. (2008) Research on the Clustering Algorithm of Cohesive Hierarchy. *Scientific and Technological Information: Scientific Teaching and Research*, **11**, 202. (In Chinese)
- [11] Duan, M.X. (2009) Research and Application of Hierarchical Clustering Algorithms. Central South University. (In Chinese)
- [12] Zhang, J.H. (2007) K-Means Clustering Algorithms Research and Application. Wuhan University of Technology, Wuhan. (In Chinese)
- [13] Anonymous (2012) Introduction to Network Science. (In Chinese)
- [14] Watts, D.J. (1999) Networks, Dynamics, and the Small-World Phenomenon. *American Journal of Sociology*, **105**, 1-10. <https://doi.org/10.1086/210318>
- [15] Watts, D.J. (1999) Small Worlds: The Dynamics of Networks between Order and Randomness. Princeton University Press, Princeton.
- [16] Chen, J., Lu, J.A., Lu, X.F., Wu, X.Q. and Chen, G.R. (2013) Spectral Coarse Graining of Complex Clustered Networks. *Communications in Nonlinear Science and Numerical Simulation*, **18**, 3036-3045. <https://doi.org/10.1016/j.cnsns.2013.03.020>
- [17] Chen, J. and Lu, J. (2012) Medium-Scale Study of Complex Networks Uncovers the Process of Network Synchronization. *Journal of University of Electronic Science and Technology*, **41**. (In Chinese)
- [18] Erdos, P. and Rnyi, A. (1960) On the Evolution of Random Graphs. *Publication of the Mathematical Institute of the Hungarian Academy Offences*, **38**, 17-61.
- [19] Barabasi, A.L. and Albert, R. (1999) Emergence of Scaling in Random Networks. *Science*, **286**, 509-512.
- [20] Barabasi, A.L., Albert, R. and Jeong, H. (1999) Mean-Field Theory for Scale-Free Random Networks. *Physica A*, **272**, 173-187. [https://doi.org/10.1016/S0378-4371\(99\)00291-5](https://doi.org/10.1016/S0378-4371(99)00291-5)
- [21] Li, Z., Liu, J., Jiang, P., et al. (2008) Effect of Average Degree on Synchronization Ability of Complex Networks. *Guangxi Physics*, No. 1, 18-20 (In Chinese)