A Community Discovery Algorithm for Complex Networks

Lintao Lv¹, Jialin Wu¹ * and Hui Lv²

¹ XiJing University, Xi’an, Shanxi, 710123, China
² Chongqing Design Institute, Chongqing, 400015, China

*Corresponding author e-mail: 1211167803@qq.com

Abstract. Community structure is an important feature of complex networks. These community structures have the fractal characteristics, that is, there is a self similarity of statistical sense between the complex networks and their local. There have been more and more recent researches on communities’ discovery in complex network. However, most existing approaches require the complete information of entire network, which is impractical for some networks, e.g. the dynamical network and the network that is too large to get the whole information. Therefore, the study of community discovery in complex networks has rather important theoretical and practical value. Through the analysis and study of the complex network evolution models with renormalization and the community change of the complex network evolution, using the tool of adjusting scales as the renormalization process, a multi-scale network community detection algorithm based on fractal feature evolution was proposed. The purpose is to solve community discovery problems in dynamic complex networks, and the effectiveness of the proposed method is verified by real data sets. By comparing result of this paper with the previous methods on some real world networks, and experimental results verify the feasibility and accuracy.

Keywords: Complex Network, Renormalization, Complex Network Evolution Model, Multi-Scale, Community Discover

1 Introduction

There are various kinds of complex networks in the real world, such as social network, technology network, biological network, thesis co-authoring network, reference citation network. In order to further explore the structural characteristics of complex networks, understand and apply the functions of complex networks, the study of fruitful complex network community structure has been carried out, and a number of complex network community discovery methods have been proposed.

Literature [1] In order to effectively solve the problem of community structure in local network detection, an adaptive label propagation algorithm ALPA-S and an asynchronous adaptive label propagation algorithm ALPA-A based on local micro-structure maxima are proposed. The simulation results show that the asynchronous algorithm ALPA-A has certain advantages over the synchronous algorithm ALPA-S in the quality, adaptability and robustness of community detection. However, the ALPA-S algorithm does not consider the parallel problem, so this algorithm can not adapt to the community detection problem in large-scale and ultra-large-scale social networks. Literature [2] For
the traditional dynamic network community discovery method, the network at each time is detected by static algorithm, and then the relationship between communities is analyzed, which may lead to higher time overhead. According to the characteristics that the topological structure between adjacent times does not change much, a community discovery method considering the combination of community structure stability and incremental related nodes is proposed. Methods IPCSCDCA. It avoids re-partitioning the whole network, thus greatly reducing the time overhead of the algorithm. However, the paper still needs improvement in changing the node selection strategy. In addition, the parallel processing performance of the algorithm should be further improved based on cloud technology. For example, the condensation method, the splitting method, the optimization method and the simulation method[3]. Many results of them involve the analysis and theoretical research of complex networks based on fractal features, but they are only limited in the analysis and confirmation of the fractal characteristics of complex networks, and the practical and feasible complex network community discovery methods have not been proposed.

The essence of the complex network is a complex nonlinear system. Fractal characteristics can reveal the unity of order and disorder in nonlinear systems, and it is one of the characteristics of complex networks. In addition, community structure is another feature of complex networks. In many networks with community structure, the network with fractal features is not only stable, but also can effectively adjust the structure of the network through fractal features to meet the requirements of community discovery.

However, how to make use of the unity of certainty and randomness of nonlinear systems is very important to study the connection and influence between fractal characteristics and complex network community structure, and whether community discovery methods in complex networks are effective. Therefore, the application of fractal characteristics of complex networks in complex network dynamic community discovery not only has great theoretical and practical value, but also has a broad application prospect.

2 Complex network evolution model based on renormalization

2.1 Complex network evolution model

The evolution model of complex networks is one of the important theoretical foundations for building the complex network dynamic community discovery. The evolution of complex networks refers to the situation of complex networks changing with time, and the mechanisms and causes that lead to these changes can be analyzed [4]. The evolution of complex networks mainly includes the increase and extinction of nodes, the increase and disappearance of the edges, the reconnection of the edges, the change in the direction of the edges in the directed graph, and the change of the weight in the weighted graph. In a directed graph, it may also include the direction changes of edges; in a weighted graph, it may also include changes of weight values.

Many models are proposed for complex networks and their properties. The relationship between network evolution and its properties is explained from theory and practice. In the reference [5], Wilson et al. used the model of repainting edges to analyze the change from the rule network to the random network, and found the small world characteristics of the complex network. That is, because of the existence of a few short cuts, the characteristic path length of the network is smaller and the clustering coefficient is higher. Thus, the relationship between shortcut and small world is revealed. In the process of network growth, when a new node appears, it tends to connect the node which has large degree. As a result, these nodes with large degree will have more connections than those nodes with smaller degree, that is, the larger nodes will become larger, i.e. rich get richer. According to its own growth mechanism, the network will become a scale-free network. At the same time, the scale-free network formed by this way was vulnerable to deliberate attack on hub nodes.

In practical networks, the study of community changes with time was firstly proposed by Alexander et al. [6]. The data used is the citation network provided by the NEC CiteSeer database in reference [4] and the snapshot of complex networks at different times is mainly analyzed and studied.
By using a cohesive hierarchical cluster to find communities, the evolution of each community can be traced by analyzing the community structure of different snapshot networks, and the emergence of new communities is usually corresponding to the emergence of new research directions. They found that small communities were relatively stable while the changes of large communities were relatively violent. Mehdi Azaouzi et al. used the similar methods to study this problem [7]. Through the connection and change of the community structure at each time point, the community's key attributes of the complex network structure were found out with the change of time, and the behavior of the network was predicted and analyzed by them. The evolution model of the complex network is shown in Fig. 1.

![Fig. 1 The evolution of Complex Network](image)

### 2.2 Complex network evolution model based on renormalization

The literature published by Song et al. in the natural journal [8] pointed out that many complex networks had the self-similarity in reality, that is, the fractal characteristics. The concept of renormalization was used to study the evolution of fractal and non-fractal complex networks, that is, the evolution model of complex networks was simulated by the inverse process of renormalization, as shown in Fig. 2 (b).

The renormalization of complex networks is the initialization process of complex network evolution, and also the known data of the algorithm to be built in this paper. The renormalization of complex networks refers to the process of integrating the closer network nodes into a node. The method used by Song et al was the reverse weighting normalization process, which is similar to a growth process of a complex network, extending the node of the previous time to a set of points, as shown in Fig. 2 (c).
Suppose the community node is here, and the network is complex. Through the evolution of the fractal feature, the multi-scale network community discovery algorithm is proposed. This paper describes the main steps of this new algorithm as follows.

Step1. Input parameters. The algorithm has four parameters need to be input.

(1) The complex network at time $t$, i.e., $G(t) = (V(t), E(t))$.

(2) The community structure of complex network at time $t$, $|D_i(t)|, i = 1 \cdots K$, here $\bigcup_{i \in D_i(t)} V(t) = V(t)$, and $\forall i \neq j, D_i(t) \cap D_j(t) = \emptyset$.

(3) The minimum distance $d_n$ between nodes of a complex network in each community as a node.

(4) The changes of edges from time $t$ to time $t + 1$, here, $\Delta E^+$ and $\Delta E^-$ represent the new emergent edges and the death edges, respectively.

Step2. Update the topology structure of the network. That is to update the changes of edges of the complex network from time $t$ to time $t + 1$, i.e. update the topological structure from the complex network $G(t)$ to $G(t + 1)$, it will be used to calculate the distance.

Step3. Remove the edge changes that do not affect the community structure of complex networks. That is to remove the edge whose two ends are in the same community $D_i(t)$ from $\Delta E^+$. In the same way, the edges whose two ends are in two community $D_i(t)$ and $D_j(t)$ are removed from $\Delta E^-$, here $i \neq j$.

Step4. Adjusts the scale and achieve deep cohesion within the community. For the remaining elements in $\Delta E^-$ and $\Delta E^+$ and involving in the community $\{D_{ij}(t)\}, (p = 1 \cdots q)$, the renormalization is used to achieve the agglomeration of nodes and edges within each community. When one of the two cases is satisfied, the agglomeration is stopped. One is that the distance of each node in the agglomeration process is not less than $d_n$ for the first time; another is that all nodes in the community are agglomerated into one node. Here, the calculation method of distance is as follows. Suppose in a complex network $G = (V, E)$, $n_i$ and $n_j$ are the two point in $V$, and $(n_i, n_j) \in E$, then the distance between them is defined as following.

$$\|n_i - n_j\| = \frac{1}{d_i - d_j + 1} \cdot \min\{d_i, d_j\} \cdot w_{ij}$$

(1)

Here, $d_i$ and $d_j$ denote the degree of $n_i$ and $n_j$, respectively. $\min\{x, y\}$ denotes taking a smaller value, $w_{ij}$ is the weight value of the edge $(n_i, n_j)$.

The community involved in this step only need to consider the connection between its internal nodes, and does not need to consider the connection between communities. So computing for each

---

**Fig. 2** Evolution model of the renormalized complex network

3 Multi-scale network community discovery algorithm based on fractal feature evolution

Through the analysis of the fractal characteristics of complex networks, the evolution model and the impact of the real community, adopting the graph theory of undirected and unweighted values, a multi-scale network community discovery algorithm based on fractal feature evolution is proposed in this paper. The main steps of this new algorithm are described as follows.

Step1. Input parameters. The algorithm has four parameters need to be input.

(1) The complex network at time $t$, i.e. $G(t) = (V(t), E(t))$.

(2) The community structure of complex network at time $t$, $|D_i(t)|, i = 1 \cdots K$, here $\bigcup_{i \in D_i(t)} V(t) = V(t)$, and $\forall i \neq j, D_i(t) \cap D_j(t) = \emptyset$.

(3) The minimum distance $d_n$ between nodes of a complex network in each community as a node.

(4) The changes of edges from time $t$ to time $t + 1$, here, $\Delta E^+$ and $\Delta E^-$ represent the new emergent edges and the death edges, respectively.

Step2. Update the topology structure of the network. That is to update the changes of edges of the complex network from time $t$ to time $t + 1$, i.e. update the topological structure from the complex network $G(t)$ to $G(t + 1)$, it will be used to calculate the distance.

Step3. Remove the edge changes that do not affect the community structure of complex networks. That is to remove the edge whose two ends are in the same community $D_i(t)$ from $\Delta E^+$. In the same way, the edges whose two ends are in two community $D_i(t)$ and $D_j(t)$ are removed from $\Delta E^-$, here $i \neq j$.

Step4. Adjusts the scale and achieve deep cohesion within the community. For the remaining elements in $\Delta E^-$ and $\Delta E^+$ and involving in the community $\{D_{ij}(t)\}, (p = 1 \cdots q)$, the renormalization is used to achieve the agglomeration of nodes and edges within each community. When one of the two cases is satisfied, the agglomeration is stopped. One is that the distance of each node in the agglomeration process is not less than $d_n$ for the first time; another is that all nodes in the community are agglomerated into one node. Here, the calculation method of distance is as follows. Suppose in a complex network $G = (V, E)$, $n_i$ and $n_j$ are the two point in $V$, and $(n_i, n_j) \in E$, then the distance between them is defined as following.

$$\|n_i - n_j\| = \frac{1}{d_i - d_j + 1} \cdot \min\{d_i, d_j\} \cdot w_{ij}$$

(1)

Here, $d_i$ and $d_j$ denote the degree of $n_i$ and $n_j$, respectively. $\min\{x, y\}$ denotes taking a smaller value, $w_{ij}$ is the weight value of the edge $(n_i, n_j)$.

The community involved in this step only need to consider the connection between its internal nodes, and does not need to consider the connection between communities. So computing for each
community is independent, and parallel processing can be used to improve the processing speed. The result of this step is that the complex network can be condensed into a new network, which is composed of three elements, the undivided original community, the newly added node and the split community. The minimum distance of the network is \( d_a \), and these three elements are not necessarily all. If so, these three elements are represented as network nodes.

Step 5. Adjust the scale again. The community finding of renormalization of the Step 4 results is carried out until all the newly added original points are renormalized or properly structured.

Step 6. Results output. The output results of the algorithm include two aspects: (1) community structure of complex network at time \( t+1 \); (2) the estimated value of the correlation dimension of complex networks.

The estimation of correlation dimension can be obtained by the following methods and steps.

Step 1: The community structure \( \{D_l(t+1)\} (i=1 \cdots L) \) of complex network at time \( t+1 \), here \( \bigcup_{i=1}^{L} D_l(t+1) = V(t+1) \), for \( \forall i \neq j, D_l(t+1) \cap D_l(t+1) = \phi \). Among them, \( V(t+1) \) is formed by subtracting death points and adding new points from \( V(t) \).

Step 2: The value of a series of point pairs \( (C(d), d) \), and the fractal dimension \( D_c \) of the complex network estimated from this.

In summary, the results of multi-scale network community discovery algorithm based on fractal feature evolution include: (1) obtain the community structure of complex networks at time \( t+1 \); (2) obtain the minimum distance and the topological graphs of complex networks at time \( t+1 \). If the amount of edge change of the network \( \Delta E^+ \) and \( \Delta E^- \) can be obtained from time \( t+1 \) to \( t+2 \), one can continue to get the complex network community structure at time \( t+2 \). That is to say, the state of the next moment is determined by the state and amount of change at the previous moment, which is the evolution process of the complex network.

4 Comparison of experimental results

4.1 Test standard

The dataset used to test community discovery algorithms is a complex network in reality, and the community structure of these networks is known. Therefore, we only need to compare the community discovered by proposed algorithm with the actual community structure, and we can have a judgment on the feasibility and effectiveness of the algorithm. The adopted evaluation parameters are the common in pattern recognition and information retrieval, including the precision rate, the recall rate and the F-measure value. These parameters are often used to compare real results and experimental results, so as to quantify the accuracy of the experimental results. The following will be explained separately.

(1) The precision rate represents the correct ratio of node identification in the community discovered by the algorithm. Its mathematical model is as follows.

\[
\text{precision}(D_{\text{alg}}) = \frac{\text{number of correct node in } D_{\text{alg}}}{|D_{\text{alg}}|} \tag{2}
\]

Here \( D_{\text{alg}} \) denotes a community found by the algorithm, \( |D_{\text{alg}}| \) denotes the number of nodes in \( D_{\text{alg}} \).

(2) The recall rate represents the correct recognition rate of nodes in real communities, and it is defined as follows.

\[
\text{recall}(D_{\text{real}}) = \frac{\text{number of correct node in } D_{\text{real}}}{|D_{\text{real}}|} \tag{3}
\]
Here, $D_{\text{real}}$ represents a community that really exists in a complex network, $|D_{\text{real}}|$ denotes the number of nodes in $D_{\text{real}}$.

(3) The F-measure value is the harmonic mean of the precision rate and the recall rate. When $D_{\text{orig}}$ in equation (2) and $D_{\text{real}}$ in equation (3) represent the same community $D$, the definition of the F-measure value of the community $D$ is given by:

$$F(D) = \frac{2 \cdot \text{precision}(D) \cdot \text{recall}(D)}{\text{precision}(D) + \text{recall}(D)}$$ (4)

4.2 Experimental data set

The real data set selected in this paper is from the American National Collegiate Athletic Association (NCAA) rugby network, which is usually used to verify the performance and efficiency of the community discovery algorithm. The network of NCAA rugby is shown in Fig.3.

![NCAA rugby network](image)

**Fig. 3 NCAA rugby network**

It mainly includes the record of the rugby regular season of American College Sports League in the 2000 season in Fig.3. The nodes represent the university teams participating in each competition, they are expressed in the name of the University, and there are a total of 115 Universities. The edges between the nodes represent the regular matches between the school teams. There are 613 matches between these teams. In reality, these teams are divided into 11 different divisions, each consisting of 8 to 12 teams. There is more competition between teams in the same division than in different teams. The average number of games per team in the same division is seven, and the average number of each team in the different teams is four. But it is worth noting that the number of matches is uneven in different division. For teams belonging to different playing division, there is more competition between the close areas than the far away ones.

4.3 Experimental simulation and results analysis

In order to verify the incremental evolution of the algorithm proposed in this paper, the NACC rugby network is divided into two parts. The first part consists of 76 nodes and 426 edges. These nodes and edges constitute 8 of the 11 divisions. The community structure obtained through renormalization technology in NACC rugby network is shown in Fig. 4.

The community structure shown in Fig.4 is acted as the input of the new proposed algorithm of this paper. The NACC community discovered by this new algorithm is shown in Fig.5, and the detailed statistics of the results are shown in Table 1.
Table 1 Algorithm accuracy on NACC rugby network

| Community number | Node number (algorithm) | Precision rate | Recall rate | F-measure value |
|------------------|------------------------|----------------|-------------|-----------------|
| 1                | 9                      | 0.8889         | 1.0000      | 0.9412          |
| 2                | 9                      | 0.8889         | 0.8889      | 0.8889          |
| 3                | 11                     | 1.0000         | 1.0000      | 1.0000          |
| 4                | 11                     | 0.9091         | 1.0000      | 0.9524          |
| 5                | 14                     | 0.8571         | 1.0000      | 0.9231          |
| 6                | 9                      | 1.0000         | 0.9000      | 0.9474          |
| 7                | 9                      | 1.0000         | 0.9000      | 0.9474          |
| 8                | 13                     | 0.9231         | 1.0000      | 0.9600          |
| 9                | 8                      | 1.0000         | 1.0000      | 1.0000          |
| 10               | 14                     | 0.9286         | 1.0000      | 0.9630          |
| 11               | 8                      | 0.5000         | 0.6667      | 0.5714          |
| Total            |                        | **0.8996**     | **0.9505**  | **0.9225**      |

It can be seen in table 1 that although some of the teams in this algorithm are not very accurate, the team can be well identified in most of the games. In general, the effect of this algorithm on this actual data set is satisfactory, the average precision rate is 89%, and the recall rate is 95%.

This algorithm estimates the correlation dimension of the NACC Rugby network by 3.0225, and the effect is shown in Fig.6. It can be seen in Fig.6 that the relationship between distance and correlation function is approximately a straight line, indicating that the NACC rugby network also has the approximate self similarity, that is, it has fractal characteristics.
Fig.6 Correlation dimension of NACC rugby network

5 Conclusions
A multi-scale complex network community discovery algorithm based on fractal feature evolution is proposed in this paper. The characteristic of the algorithm is to make full use of the fractal features and the evolution rules of complex networks, and take the renormalization process as a bridge linking different evolution scales, and realize a complex network dynamic community discovery algorithm. The effectiveness of the algorithm is verified by experiments.

The main contributions of this paper are not only exploring the topological structure of complex networks, but also providing useful information retrieval methods in complex networks. It also helps people understand the functions of complex networks, discover hidden rules in complex networks, and provide a new way for people to predict the behavior of complex networks.

References
[1] Chunying, TANG Zhikang, TANG Yong. Community Detection Algorithm with Local-First Approach in Social Networks[J].Journal of Frontiers of Computer Science & Technology.2018,(12)8: 1263-1277(in Chinese)
[2] GUO Kun, LI Guo-hui, CHEN Yu-zhong, WU Ling. Incremental Parallel Community Detection Algorithm Considering the Stability of Community Structure[J]. Journal of Chinese Computer Systems,2018,39(7): 1548-1553 (in Chinese)
[3] Pawan Kumar, Ravins Dohare A. neighborhood proximity based algorithm for overlapping community structure detection in weighted networks[J].Frontiers of Computer Science, 2019, Vol.13 (6), pp.1353-1355
[4] CHENG Xueqi, JIN Xiaolong. Technological progress and trends of big data [J]. Science & Technology Review, 2016 , 34 (14) :49-59(in Chinese)
[5] Wilson James D,Palowitch John,Bhamidi Shankar,Nobel Andrew B. Community Extraction in Multilayer Networks with Heterogeneous Community Structure.[J]. Journal of machine learning research : JMLR,2017,18 :5458-5506
[6] Alexander Goryashko, Leonid Samokhine, Pavel Bocharov. About complexity of complex networks[J]. Applied Network Science, 2019 Vol.4 (1) pp.1-21
[7] Mehdi Azaouzi,Delel Rhouma,Lotfi Ben Romdhane. Community detection in large-scale social networks[J] state-of-the-art and future directions Social Network Analysis and Mining, 2019, Vol.9 (1), pp.1-32
[8] Song, C., Havlin, S., Makse, H. A. Self-similarity of complex networks[J]. Nature, 2005, 433:392-395