

## **A HIERARCHICAL CLUSTERING COMMUNITY ALGORITHM WHICH MISSED THE SIGNAL IN THE PROCESS OF TRANSMISSION**

Ran Jin, Lvshi Hong, Cui Wang, Lifei Wu and Wangchun Si

School of Computer Science and Information Technology, Zhejiang Wanli University, Zhejiang, China.

Email: ran.jin@163.com

### **ABSTRACT**

Community identification is a basic task of social network analysis, meanwhile the community structure detection is a key problem of community identification. Each node in the community structure is regarded as the signal source. A hierarchical clustering community algorithm is proposed in order to settle the problem of signal missing in the process of signal transmission. The algorithm measures the probability of receiving signals of nodes by degree centrality to quantify the signal missing values. After the signal transmission, the topology of the network is transformed into geometric relationships among the vectors. On the basis, the hierarchical clustering algorithm is used to find the community structure. In order to validate the proposed method, we compared it with SHC algorithm, CNM algorithm, GN algorithm and similar algorithm. Under three real networks, the Zachary Club, American Football and Netscience, the experimental results indicate that SMHC algorithm can effectively improve precision.

**Keywords:** Community identification, Signaling process, Signal missing, Degree centrality, Hierarchical clustering.

### **1. INTRODUCTION**

Many systems can be indicated by network in the real world, for example, the link relations of WWW, the friend relations in Facebook, food web and social relationship networks, etc. In a general way, a network is composed of a vertex set and an edge set. For social networks, every individual in real life is a vertex which constitutes the vertex set with other individuals in social networks, and the relationship between individuals forms a network edge set. Individuals can be composed of different circles, groups, and populations based on interests. For example, the moments in Wechat, the mutual concern friends in Microblog. In the study of the network system, some people regarded the individual as the node, the relationship as the edge, the constituted groups are called the Community.

Community identification is a basic task of social network analysis which helps to realize other social computing tasks and can be used to solve many practical problems. Currently, the community identification has been widely used in social networks, the protein function relation networks, Web networks, metabolism networks. It can achieve social network client division by similar interests of users, then recommend products to the relevant customers, transform the information into vector space information, which can improve the success rate of trade. Through clustering the users in similar Internet location, it can provide personalized services to different groups; also, for large network clustering, it can improve the data storage structure and make it easy to query.

At present, the community detection algorithm is roughly divided into those based on optimization, genealogy theory and hierarchical clustering, etc. The classic algorithm is K-L

algorithm and Extremal optimization algorithm; spectral bisection method and spectrum method based on standard matrix; the most classic one of hierarchical clustering algorithm is the GN algorithm based on the node split mind of edge betweenness, the Newman greedy algorithm based on condensed ideas (CNM algorithm), the similarity measure of coherence algorithm (Similar algorithm). This paper presents an algorithm based on ideological cohesion, initially regards each node in the network as a separate community, then sets certain conditions in order to merge the nodes which meet the conditions, this cycle until all the nodes are in a community.

Hu [10] proposed a method which detects community structures by gaining signal propagation through the network (Signal Algorithm). First of all, the nodes in the network are processed by signal transmission, it means that one network which has  $n$  nodes are represented by an  $n$ -dimensional vector space, and then cluster the  $n$ -dimensional vector by using the K-means clustering method, finally the network partitions can be obtained. The literature [11] puts forward that it can detect communities by the concept of signal transmission and hierarchical clustering (SHC "algorithm"). This paper has reference of the process of signal transmission by Hu [10], and considers the signal missing during the process of signal transmission at the same time, then divides the networks by the signaling process of degree centrality and the hierarchical clustering algorithm, finally verifies this method through the network data set.

### **2. THE CLUSTERING ALGORITHM BASED ON NODES SIMILARITY**

The hierarchical clustering algorithm mainly merges or splits nodes by similarity measurement between nodes to nodes, and one of the key issues that must be resolved in the community identification is how to construct the similarity measurement between nodes.

### 2.1 Node vectorization and similarity measurement

A signal transmission method proposed in the literature [10]: regard a network which has  $m$  nodes as a system, the individuals in the vertex set ( $V = \{V_1, V_2, \dots, V_m\} (i = 1, 2, \dots, m)$ ) of the system all have the functions of sending, receiving and recording the signals, transfer the signal itself and adjacent node, the signal transmitting process are presented by formula as following:

$$V = (I + A)^T \quad (1)$$

Wherein,  $A$  is the adjacency matrix and  $T$  is the number of iterations.

After signaling, the space topology information of the nodes in the network transfers into vector space information, so the similarity can be calculated by the space distance. The similarity measurement gets to Angle cosine formula.

As formula [2] shows,  $i=1, 2, \dots, m; j=1, 2, \dots, m$

$$r(V_i, V_j) = \frac{\left| \sum_{k=1}^m v_{ik} v_{jk} \right|}{\sqrt{\sum_{k=1}^m v_{ik}^2 v_{jk}^2}} \quad (2)$$

From formula [2], the cosine of the Angle between two nodes: the greater the value  $r(V_i, V_j)$ , that is the greater the vector product of the two nodes, it indicates the more similar that nodes and it tends to be in the same community.

### 2.2 The measurements of signal missing

In transfer process, the node signal may cause signal missing because of external interference. When the node  $i$  passes signals to  $j$ , node  $j$  will received a signal from the node  $i$  according to certain probability value. The importance of a node is relevant to the numbers of the adjacent node—a

greater degree of Node, the node is more and more important. The degree centrality of the node  $C(v_i)$  can be defined by formula [3]:

$$C(v_i) = d_i = \sum_j A_{ij} \quad (3)$$

The definition of that Node  $j$  receives a transmission signal from Node  $i$  is as following:

$$p(j, i) = \frac{d_i}{(m-1)} \quad (4)$$

Wherein,  $d_i$  represents the degree of node  $i$ ,  $m$  represents the total number of the vertex set  $V$ 's nodes.

Suppose the Signal loss matrix  $P = \{P_1, P_2, \dots, P_m\}$ , ( $i=1, \dots, m$ ), according to formula [4], each element in lack of definition in the matrix is  $P_i = p(j, i)A_i$ , and  $P_i = \{P_{i1}, P_{i2}, \dots, P_{im}\}$ .

Adjacency matrix of the network is  $A = \{A_1, A_2, \dots, A_m\}$ ,  $A_i = \{a_{i1}, a_{i2}, \dots, a_{im}\} (i = 1, 2, \dots, m)$ . If there is an edge between the node  $i$  and  $j$ , it means,  $a_{ij} = 1$ , when  $i = j$ , then  $a_{ij} = 0$ .

The definition of the matrix after the signal transmission is as following:

$$V = p^T A + p^{T-1} I + p^{T-2} I + \dots + I \quad (5)$$

In the formula [5], after  $T$  time transmission, the vertex set is got:  $V' = \{V'_1, V'_2, \dots, V'_m\}$ , wherein  $V'_i = \{v'_{i1}, v'_{i2}, \dots, v'_{im}\} (i = 1, 2, \dots, m)$ . In order to get its relative influence quantity, it is standardized to  $U_i = \{u_{i1}, u_{i2}, \dots, u_{im}\} (i = 1, 2, \dots, m)$ . Wherein,

$$u_{ij} = v'_{ij} / \sqrt{\sum_{j=1}^m v'_{ij}{}^2}. U_i \text{ represents } i \text{ node degree of influence on}$$

the network, which will convert the network topology to the relationship of  $M$  dimensional vector geometry. The transmission process of signal loss is as chart [1]:

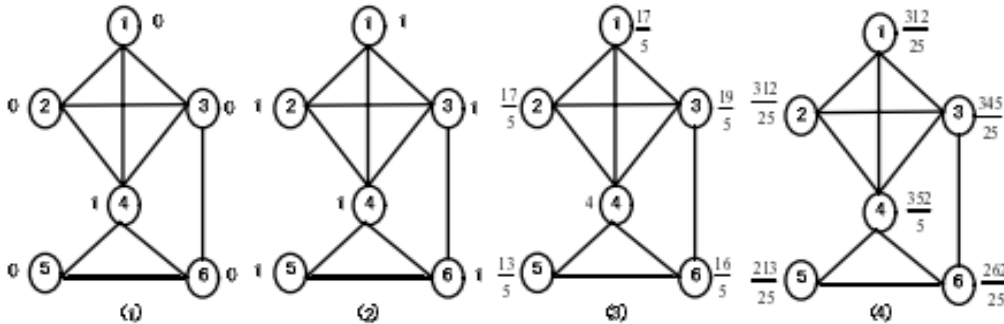


Figure 1. Signal transmission process in the presence of signal missing

The semaphore owned by  $s_i$  signal node gives the node 4 an initial signal value in (1) of figure (1). In (2) part, node 4 sends signal to its adjacent node to make each node 1, 2, 3, 5,

6 has a signal value. In (3), taking Node 1 for example: when  $T = 2$ , the semaphore of it is

$$s_1 = s_1 + p(2,1)s_2 + p(3,1)s_3 + p(4,1)s_4 \dots \text{Despite signal missing, the signal value of node 1 is ... Due to the different}$$

probability of the node receiving signals,  $\Delta s (\Delta s = s'_i - s)$  means the difference of the signal amount owned by the nodes under two cases, namely the signal which node did not receive, and that is also the signal loss in the process of signal transmission.

### 3. THE HIERARCHICAL CLUSTERING COMMUNITY BASED ON MODULARITY

For hierarchical clustering method, how to choose properly is a key issue. Currently, about community discovery algorithm, the standards of the community quality division is based on modularity.

#### 3.1 Modularity

Modularity is a measurement of the network division quality. In this paper, in this algorithm, we use the following formula [1].

$$Q = \frac{1}{2m} \sum_{i,j} (A_{ij} - \frac{d_i d_j}{2m}) \quad (6)$$

In formula (6),  $\frac{d_i d_j}{2m}$  means the expectations of edge between the node  $i$  and the node  $j$ , and  $d_i$ ,  $d_j$  represent the degree of nodes  $i$  and  $j$  respectively. From  $(A_{ij} - \frac{d_i d_j}{2m})$ , there is a difference between the actual number of connected side edge and the desired extent of node  $i$  and  $j$ . Modularity can be quantitatively represented the strength of the community structure.

#### 3.2 The hierarchical clustering community based on modularity

We combine signal transmission and hierarchical clustering then community found SMHC algorithm are designed as follows.

**Input:** network  $G(V, E)$

**Output:**  $G(V, E)$  Hierarchical clustering tree diagram

**Algorithmic procedure:**

**Step1**  $m = |V|$ , The  $G(V, E)$  is represented as an adjacency matrix, then the formula (5) is used to give vertex set after  $T$  time passes, the vertex set is  $V = \{V_1, V_2, \dots, V_m\}$  ( $V_i = \{v_{i1}, v_{i2}, \dots, v_{im}\}, i=1, 2, \dots, m$ ), then transfer into the standard form  $U_i = \{u_{i1}, u_{i2}, \dots, u_{im}\}$  ( $i=1, 2, \dots, m$ ).

**Step2** Construction process of hierarchical clustering tree;

**Step2.1** If  $m$  is the number of vertices in  $G$  and  $m \leq 0$ , perform Step3;

**Step2.2** Using Equation (2) to calculate the similarity  $r(V_i, V_j)$  between vertice  $i$  and  $j$ ;

**Step2.3** Select biggest similarity  $r(V_s, V_t)$  then merge two vertices  $s$  and  $t$ ;

**Step2.4**  $m = m - 1$ ;

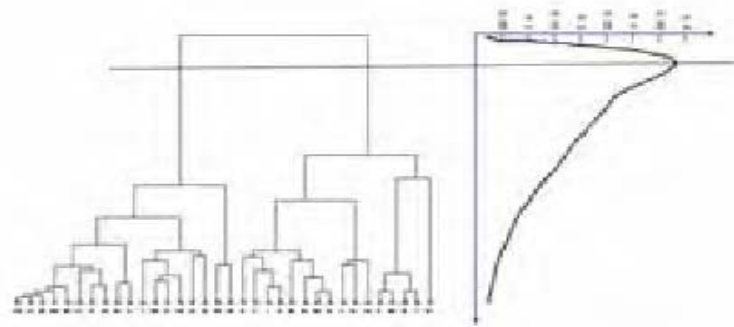
**Step2.5** Calculate the module of network  $G$  (value  $Q$ ) after the combined;

**Step2.6** Save the result of the merger of the hierarchical clustering tree;

**Step3** In step2, according to merge all modules in the process of the value of  $Q$ , the corresponding graph can be gotten. Each level in Hierarchical clustering diagram obtained by the merge order of the vertices corresponds to a module degree. We select the maximum value in all the modules, cut at the corresponding point of the tree graph to obtain classified network.

**Step4** Algorithm end

We use SMHC algorithm and take Karate Club data sets for example, then the resulting graph module degrees and hierarchical clustering diagram is shown in figure 2.



**Figure 2.** Hierarchical clustering tree diagram and module degree curve

**Description:** In figure 2, the height of the vertical axis in the left side represents the distance between two child nodes in the cluster, and the horizontal axis shows the order of the node merge each other. Module are represented on the right side of the graph.

By figure 2, the module value reaches the maximum at a certain time point, then the time point of the network structure is divided into one of the best community structure. Cutting the hierarchy tree in the position of the module maximum degree can obtain the community classification results.

### 4. DATA SET AND CLASSIC ALGORITHMS

#### 4.1 Artificial data set

By literature [7], the artificial data set rules is as following:

Each network consists of 128 nodes and divided into four communities. Each community includes 32 nodes. And each node randomly generated 16 connection, including  $Z_{out}$  nodes to connect communities, and  $16 - Z_{out}$  node connected inside the community.

Reference [14] in this paper makes this condition: when  $Z_{out} \in \{1, 2, \dots, 8\}$ , and  $Z_{out} = 1$ , the connection in a community is the closest, the community structure is the most obvious; When  $Z_{out} = 8$ , each point has the same number of connections both within and outside the community, and the community structure is not obvious. The greater  $Z_{out}$  is, the fuzzier the community structure is. The experiment regards the final results and the real dividing accuracy as the evaluation index. Because the connection between the node is randomly generated, we choose the average result of 10 times as the experimental data in the end.

#### 4.2 Network dataset

##### (1) Standard data set

The following two data sets have clear results and those are what we called the standard data sets.

- 1) Karate Club network is the classic data sets of social network analysis, whose data set contains 34 nodes and 78 edge [15].
- 2) American Football Network is a network formed between the United States in 2000 to attend university football team. If there is an edge between two nodes, the corresponding at least once between the two teams had a game [14].

(2) Netscience network is scientific research and the author network constructed by Newman from et al., Cornell

University, “high-energy” physics of electronic literature research, including 1,589 authors [7].

#### 4.3 Classical algorithms

GN algorithm [7]: Using the method successively to removes the largest side of betweenness centrality;

CNM algorithm [8]: Using the rapid greedy rule to merge division method;

Similar algorithm [9]: Using a new local similarity measurement, the application of a similarity index decreasing function and Ward clustering method.

Signal algorithm [10]: the use of signaling and K-means clustering algorithm.

SHC algorithm [11]: the use of signal transmission and hierarchical clustering algorithm (without considering the signal missing)

### 5. EXPERIMENTAL RESULTS AND ANALYSIS

To test the effectiveness of the algorithm and the correct rate of communities division in this paper, we have the experiments 1-6. Refer to reference [10], this paper suppose  $T = 3$  in SMHC.

#### 5.1 Experiment 1-based on the artificial dataset

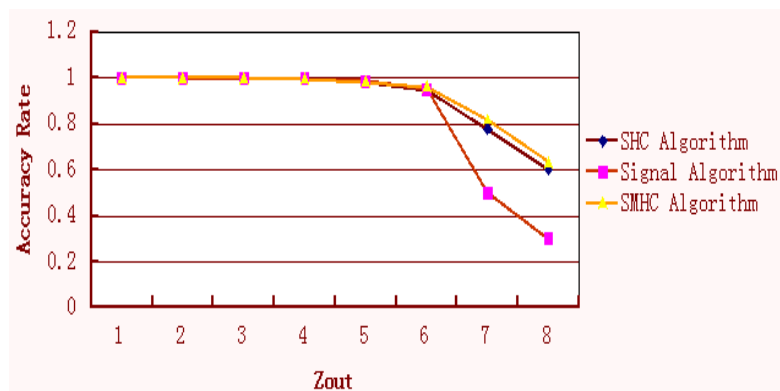


Figure 3. Comparison of accuracy rate

In order to verify the validity of the algorithm SMHC artificial datasets, it was compared with the Signal method and SHC algorithms. Finally the results are shown in Figure 3.

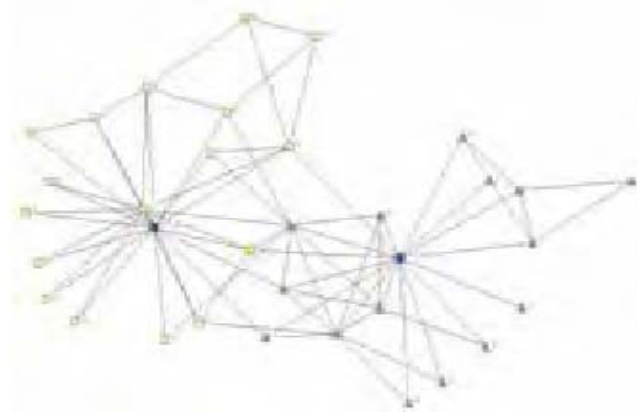
According to Figure 3:

- 1) When  $Z_{out} \leq 4$ , correct division of the network can be obtained entirely through the three algorithms. Then the accuracy rate is 100%.
- 2) When  $4 < Z_{out} \leq 6$ , the performance trend of three kinds of algorithms gradually decreased, but the accuracy of their divided communities are similar.
- 3) When  $Z_{out} \geq 6$  time, Signal Algorithm curve decline rapidly with  $Z_{out}$  increases. This indicates that the community structures at this time quickly become blurred. The SHC algorithm and SMHC algorithm are on a similar curve downward trend, and also are placid compared with the algorithm of Signal. At the same  $Z_{out}$  value, SMHC algorithm accuracy is higher than

SHC algorithm and Signal algorithm. The main reason is that SHC “algorithm and Signal algorithm only considers the adjacency link between nodes but neglects the relationship of nodes in the network. In addition, SMHC algorithm uses local information of the network and also considered the global information of the network on the basis of SHC “algorithm. Then signal missing value is put forward in order to make its better performance than other two methods. This shows that the new measure value can improve the accuracy of the community division, and make the community structure is more obvious.

#### 5.2 Experiment 2 - based on benchmark data sets

- (1) SMHC algorithm in this article was applied to Karate Club data sets, the division of results are obtained as shown in figure 4.



**Figure 4.** Partition result graph of Karate Club dataset

**Table 1.** Partition result of American Football dataset

Community	Nodes in community
1	1,2,10,17,24,42,94,105
2	2,26,34,38,46,90,104,106,110
3	3,7,14,16,33,40,48,61,65,101,107
4	4,6,11,41,53,73,75,82,85,99,103,108,98
5	8,9,22,23,52,69,78,79,109,112
6	12,25,51,70,91,29
7	13,15,19,27,32,35,39,44,55,62,72,86,100,37,43
8	18,21,28,57,63,66,71,77,88,96,97,114,60,64
9	20,30,31,36,56,80,95,102,81,83
10	47,50,54,59,68,74,84,89,115
11	45,49,58,67,76,87,92,93,113,111

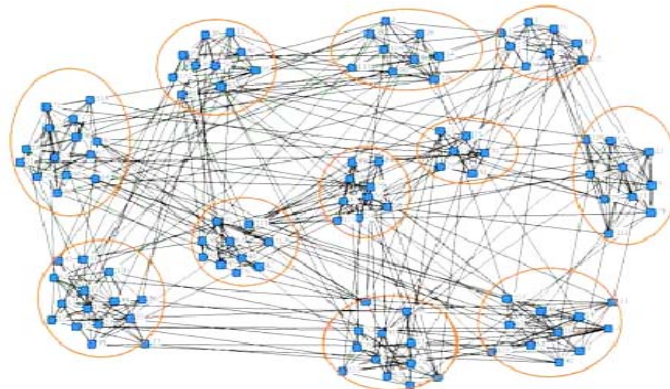
In figure 4, the two different community are represented with triangle and square respectively, and the dot is the center of the two communities respectively. The sizes of two communities are 16 and 18 respectively, and are also the same as the actual community structure. Among them, the nodes 1 and 34 said the club manager and coaches respectively which are the center of the two communities, and the results of a SMHC algorithm and actual situation are completely consistent.

(2) The SMHC algorithm in this paper was applied to American Football data sets, and the grouping situation is as

shown in table as well as the results of the division are shown in figure 5.

From table 1 and figure 5:

In actual division football team network structure, the node {37, 43, 81, 83, 91} is classified into the same community. The five nodes in table 1 were divided into three different communities. Their number of edges is sparse in the dividing diagram of figure 5. From the original data analysis, there is no mutual matches between teams which represented the five nodes. But the algorithm is used to divide it to its most times group, then the grouping different actual situation is appeared.



**Figure 5.** Partition result of American Football dataset

(3) In American Football data set, the SMHC algorithm and five kinds of algorithm are compared about the accuracy of each grouping situation. It is shown in table 2.

From table 2 know:

1) In 12 communities, SMHC algorithm and SHC algorithm accuracy has four higher than that of Signal algorithm and ten higher than those of the other three algorithms. It illustrates that the thought of signal

transmission is more effective than the other algorithms.

- 2) Numbers for 1, 2, 3, 5, 6, 7, 9 community in SMHC algorithm has the same result with that divided by SHC algorithm.
- 3) The division number of community according to SMHC algorithm is 11, and those which were divided

into the right community has a number of 106. According to table 1: on the condition of considering the signal missing in the signaling process, SMHC algorithm divided 59 and 111node into number 10 and 11 community respectively, which has the same result as the actual one.

**Table 2.** Comparison of the accuracy of SMHC algorithm and the classical algorithm in the American Football dataset

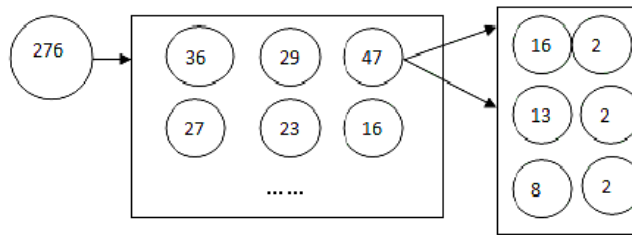
Community number	SMHC algorithm	SHC algorithm	Signal algorithm	GN algorithm	CNM algorithm	Similar Algorithm
1	1.000	1.000	1.000	0.000	0.514	0.375
2	1.000	1.000	1.000	0.900	1.000	1.000
3	1.000	1.000	1.000	1.000	0.983	1.000
4	0.923	1.000	1.000	0.923	1.000	0.914
5	1.000	1.000	1.000	0.555	0.937	0.712
6	0.666	0.666	0.443	0.443	0.043	0.123
7	0.866	0.867	0.536	0.861	0.885	0.832
8	0.857	0.800	0.800	1.000	0.834	0.560
9	0.800	0.800	0.801	1.000	0.835	0.561
10	0.900	0.801	0.723	0.723	0.723	0.509
11	0.000	0.000	0.181	0.000	0.000	0.181
12	1.000	0.900	0.900	0.900	0.842	0.706

### 5.3 Experiment 3-the network based on Netscience

(1) Data preprocessing:

Netscience network has 379 nodes and 914 edges. In order to find the community better, the data needs to be preprocessed before the experiment. In scientific

collaboration network, common reference papers can be viewed as the signal transmitted in the process of signal transmission. As shown in figure 6.

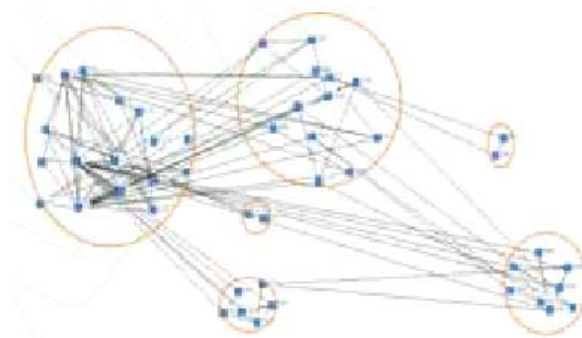


**Figure 6.** Maximal connected branch in Netscience network

There are 276 nodes which were divided into 21 communities in figure 6. Then the communities including 47 nodes are randomly selected. Next, 6 communities are gotten by SMHC algorithm. The dividing diagram is shown in figure 7.

Module degrees is regarded as a measurement of community classification quality. In this paper, the module values of signal missing and no missing in the process of Netscience network signal transmission are compared (the division of figure 7). The results are shown in table 3.

(2) Module value of signal missing and no missing



**Figure 7.** Partition graph of maximal connected branch in Netscience network

**Table 3.** Modular degree Q of Netscience dataset

Dataset	SMHC algorithm	SHC algorithm
Netscience	0.4835	0.4712

From table 3, the value of the module degrees that belongs to the community found by SMHC algorithm increased by 1.77%. This illustrates that it can improve the community quality when signal missing is considered.

#### 5.4 Experiment 4- the number of module and community

In order to show the effectiveness of the algorithm, this experiment compares the numbers of module and community

**Table 4.** Comparison of modularity and number of community division for five algorithms on three datasets

Network	Nodes	Edges	SMHC algorithm	SHC algorithm	CNM algorithm	GN algorithm	Similar algorithm
Karate	34	78	0.389/2	0.38/5	0.38/3	0.40/5	0.39/4
Football	115	613	0.546/11	0.45/12	0.55/6	0.60/10	0.60/12
Netscience	379	914	0.82/34	0.81/38	0.84/19	0.84/18	0.82/17

As can be seen in Table 4:

(1) In the three kinds of networks, the value of the module degrees obtained by the GN algorithm is very high, but the algorithm also very complex, and spends more time.

(2) In the Karate Club network, the value of the module degrees based on Similar algorithms and CNM algorithm was very close to the value obtained by SMHC algorithm, and the former two algorithms has more divided communities. From the actual characteristics of the network, it is reasonable to divide into two communities.

(3) In American Football in the network, the value of the module degrees based on Similar algorithm is higher than that of SMHC algorithm, and the number of divided communities is consistent with the actual division. But the table 2 shows that the accuracy of classified nodes was 75.7%, which was below the algorithm in this paper. The module degree of CNM algorithm is slightly higher than that of SMHC, but its number of divided communities is only half of the correct division, meanwhile the community structure is not obvious.

(4) In Netscience networks, the value of the module degrees and the number of communities obtained by SMHC algorithm are reasonable.

## 6. CONCLUSION

In this paper, the theory of signaling process are used to regard each node in the network as signals which can be received and sent, and the semaphore in the process of signal transmission are recorded, then the semaphore distribution of nodes is seemed to be the influence of the entire network after T time passes. In the network, the node always has impact on its communities, and then the entire network, therefore it can be known that the impact of the similar network nodes should belong to the same community. In the process of signal transmission, it is necessary to consider not only the node adjacency relations, but also the positions of the nodes in the network and its influence. We start from the local information then couple with the overall network topology, finally put forward the probability of the node signal received in the network which is described by degree centrality. It can improve the accuracy to add signal missing values on the basis of signal transmission after the experimental comparison.

by SMHC algorithm, SHC algorithm, CNM algorithm, GN algorithm and Similar algorithms in three data sets with the results shown in table 4.

**Description:** Nodes in Table 4 represent the network nodes, and Edges represent to the number of network connections. For each algorithm, we compare the value of its modularity and the number of divided communities. For example, 0.38 / 3 means that the value of the module degrees by CNM algorithm is 0.38 degrees and the number of divided communities is 3 in the Karate Club network.

The algorithm in this paper is not only for small social networks, but also for weighting network. For the division of large networks, there are questions of that uneven community division and high computational complexity. In the experiment, the node numbers in every community divided by Netscience network has a certain gap with each other, which may be the cause of instability in the community. As a result, next we need further research in this area. The literature [16] can also obtain the community structure of the network based on the analysis of the edge. This paper puts forward the new measures from the angle of the nodes of the similarity measure, and how to merge the node and edge together is also worth to pay close attention in our future direction.

## ACKNOWLEDGMENT

This work was supported by the Ningbo Natural Science Foundation under grant No.2015A610141, and by the National Undergraduate Training Programs for Innovation and Entrepreneurship under grant No. 201410876011.

## REFERENCES

1. Cheng Xueqi, Shen Hua-Wei, Community Analysis in Social Signal Network, *Communication of Computer Science in China*, vol. 12(7), pp. 12-20, 2011.
2. Strogatz S. H., Exploring Complex Networks, *Nature*, vol. 410(6825), pp. 268-276, 2001.
3. Cheng Xue-qi, Shen Hua-wei, Community Structure of Complex Networks, *Complex Systems and Complexity Science*, vol. 8(1), pp. 57-70, 2011.
4. Kernighan B. W., Lin S., An Efficient Heuristic Procedure for Partitioning Graphs, *Bell System Technical Journal*, vol. 49(2), pp. 291-307, 1970.
5. Yuan Chao, Chai Yi. Method for Local Community Mining in the Complex Networks [J], *Acta Automatica Sinica*, vol. 40(5), pp. 921-934, 2014.
6. Lin Youfang, Wang Tianyu, Tang Rui. An Effective Model and Algorithm for Community Detection in Social Network, *Journal of Computer Research and Development*, vol. 49(2), pp. 337-345, 2012.

7. Newman M. E. J., Girvan M., Finding and Evaluating Community Structure in Networks, *Physical Review E*, vol. 69(2), pp. 026113, 2004.
8. Girvan M., Newman M. E. J., Community Structure in Social and Biological Networks, *Proceedings of the National Academy of Sciences*, vol. 99(12), pp.7821-7826, 2002.
9. Liu Xu, Yi Dongyun, Complex Network Community, Detection by Local Similarity [J], *Acta Automatica Sinica*, vol. 37(12), pp. 1520-1529, 2011.
10. Hu Y., Li M., Zhang P., et al., Community Detection by Signaling on Complex Networks, *Physical Review E*, Vol. 78(1), pp. 016-115, 2008.
11. Zheng Fengni, Research on Network Community-finding Method on Complex Network Based on Similar Network Node Clustering [J], *Computer and Modernization*, vol.46(5), pp. 231- 234, 2013.
12. Gregory S., A Fast Algorithm to Find Overlapping Communities in Networks [M], *Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg*, 2008: 408-423.
13. Newman M. E. J., Girvan M., Finding and Evaluating Community Structure in Networks [J], *Physical Review E*, 2004, 69(2):26-113.
14. Zachary W. W., An Information Flow Model for Conflict and Fission in Small Groups, *Journal of Anthropological Research*, vol. 33, pp. 452-473, 1977.
15. Ahn Y. Y., Bagrow J. P., Lehmann S., Link Communities Reveal Multiscale Complexity in Networks [J], *Nature*, vol. 466(7307), pp. 761-64, 2011.
16. Zhou Lin, Yan Li, Shen Xiangjun, Partition Method for Community Structure in Complex Networks based on Edge Density [J], *Computer Applications and Software*, vol. 30(12), pp. 8-11, 2013.