A Novel Local Community Detection Algorithm Based on Contribution of Common Neighbor Nodes

Tianhong Wang[1], Xing Wu[2] and Wangsen Lan[3]

[1] Department of Math , Xinzhou Teachers University, Xinzhou 034000, China
[2] School of Computer Engineering and Science, Shanghai University, Shanghai 200072, China
[3] Department of Math , Xinzhou Teachers University, Xinzhou 034000, China
E-mail: [1]thwang2008@163.com, [2]xingwuvip@gmail.com, [3]ws_lan@163.com

Abstract

Most of the local community detection algorithms based on node similarity often simply count the number of common neighbors as the basis of selecting members of the community that cannot accurately measure the value of a common neighbor node in the information transmission of nodes. For this, we use the new concept of a common neighbor contribution; borrowed from the idea about the local modularity, put forward a new fast community detection algorithm. The algorithm accurately selects candidate nodes to join the community, according to the contribution of the common neighbor node, also without calculating local modularity for each common neighbor node, and greatly improved the accuracy and efficiency in merging Members. the experimental results of the computer-generated network and the real networks verified reliability and efficiency of the algorithm.

Keywords: Community detect;complex network;local community; Common neighbor contribution; Local modularity

1. Introduction

In recent years, complex network has become one of the hottest research fields, which has wide application for a great number of subjects [1-3]. With the in-depth research, researchers found that a lot of the common nature of the actual network, such as community structure. it becomes a hot spot how to detect communities in a complex network.

In detecting the community structure of complex networks, researchers have proposed many community detection algorithms, such as spectral bisection algorithm [4-5], GN algorithm [6-7] and FN algorithm which is based on the global information network to detect communities of the network. However, when the network size is too big, the access to global information is very difficult, especially in dynamic networks, such as the Internet. In addition, in many cases, researchers just focus on the local community structure of the network. For example, in social network, people usually only care about someone's community, and it is not necessary to understand the community structure of the entire social networks. Or in the academic network, researchers need to understand the subject of a journal. In these cases, we do not need to look for the global community structure of the network, and just search local communities. Recently, researchers have proposed some local community detection algorithms. For example, Clauset proposed the concept about local modularity as a measure of a local community [8]. Chen *et al*. [9] proposed a new method based on the local degree central node to detect the local community. Qi *et al*. [10] proposed a novel algorithm for community delection based on density drop, optimal clusters are automatically output by applying Max-flow Min-cut theorem. Zhu *et al*. [11] proposed a local algorithm for finding well-connected clusters

based on random-walk theory. Wang *et al*. [12] proposed a novel fuzzy overlapping community detection algorithm based on the new distance measurement of the dissimilarity index. Inspired by Claus*et al*gorithm, this paper uses the new concept of a common neighbor contribution, denoted as CNC, and proposes a new fast local community detection algorithm. The algorithm overcomes the fault of the Claus*et al*gorithm that the new node is merged by computing local modularity of each neighbor node of the community, and directly chooses the largest CNC node as a candidate node, which greatly improves accuracy and efficiency in community detection algorithm. This paper takes a common neighbor nodes contribution to information transmission as a measure of relationship between nodes to determine a new node of community.

2. Related Work

Since community structure is a qualitative definition, so many researchers try different methods to quantify it. Since the high time complexity of global modularity, Clauset proposed a local modularity [8]. The modularity is defined as follows:

$$Q_g = \frac{L_{in}}{L_{in} + L_{out}} \tag{1}$$

L_{in} means the number of interior edges in the community, L_{out} means number of edges with only one node in the community. According to the definition of community, The larger Q_g indicates more closely community structure. Since the number of edges connected to the node is required by local modularity only, thus the time of calculating the local modularity will be greatly reduced.

In the community detection, we often do not know the full connection of the network, just the local networks known. Some nodes that connect the community are called neighbor node of the commnity , defined as follows:

$$S(C) = \bigcup_{i=1}^{T} S(i) - \bigcup_{i=1}^{T} i \tag{2}$$

Where S(i) denotes the neighbor node set of the node i, S(C) denotes the neighbor node set of the community C, T denote the scale of C. In social network analysis, if the two bodies share some friends, they have in common. Then, some scholars put forward the concept of similarity between nodes [16, 18-20].

$$s_{jaccard}(i, j) = \frac{|s(i) \cap s(j)|}{|s(i) \cup s(j)|} \tag{3}$$

$$s_{cosine}(i, j) = \frac{|s(i) \cap s(j)|}{\sqrt{|s(i)\| s(j)|}} \tag{4}$$

$$s_{min}(i, j) = \frac{|s(i) \cap s(j)|}{\min(|s(i)|, |s(j)|)} \tag{5}$$

All of the above papers take the common neighbor number of nodes as a measure. It means that the more number of common neighbor, the greater similarity between nodes. It's actually not the case. It is necessary to measure the contribution of all common neighbor nodes for the transmission of information between nodes. For example, there is a common neighbor h between nodes A and B, whose degree is k. It means h is connected k nodes, but only two sides connected with A, B, respectively, so the contribution of the common neighbor node h to the transmission of information between nodes A and B is 2/k, so we come to the following definition of common neighbor contribution,denoted as CNC:

$$\delta_{ij} = \sum_{h \in S(i) \cap S(j)} \left[\frac{2}{k(h)} \right]^2 \tag{6}$$

where S(i) denotes the set of neighbors of i, and h denotes the common neighbor of i and j, and k denotes the degree of h. Therefore, according to equation (6), we can calculate CNC of any pair of nodes. However, it could result in inaccurate community detection to distinguish similarity between nodes by this measure regardless of their direct connection or indirect connection. Thus, in the case of direct or indirect connection, the similarity between a pair of nodes is different in the community detection. Thus, we calculate CNC of i and j:

$$
\delta_{ij} = \begin{cases} \sum_{h \in S(i) \cap S(j)} \left[\dfrac{2}{k(h)} \right] & \text{if } i, j \text{ connected} \\ 0 & \text{otherwise} \end{cases}
$$

3. Algorithm

Input: an undirected network, G = <V, E>, i = 1.

Output: communities.

Step 1 initializing, The CNC of all common neighbors between pairs of nodes are calculated to form a CNC matrixδ.

Step 2 If V is empty, go to step 10; otherwise, the initial local Modularity is defined, $Q_g = 0$, $Q_{max} = 0$, $C_i = \emptyset$.

Step 3 Find the largest degree node VD,$C_i = C_i + VD$,$V = V - VD$

Step 4 updating the neighbor node set S of community C_i by equation (2).

Step 5 If all neighbor nodes of the node V_j are in the community C_i, V_j will be added into directly C_i , $C_i = C_i + V_j$, go to Step 9.

Step 6 extracting the CNC crosstabs R of the community C_i and the neighborhood set S from δ.

Step 7 extracting the multiple pairs nodes Set Φ with the largest CNC from R ,$\Phi = \{\delta_1(RC_1, RS_1), \delta_2(RC_2, RS_2)....\}$, If each$\delta(RC, RS) = 0$ then go to step 10. Otherwise, the values of Φ all are set to 0, that is, $\delta(RC, RS) = 0$.

Step 8 taking a $\delta_t(RC_t, RS_t)$ from Φ, $\Phi = \Phi - \Phi_t$, $C_i = C_i + RS_t$。

Step 9 calculating Q_g according to the formula (1).

If $Q_g > Q_{max}$ then $V = V - RS_t$, update Q_{max}, return to Step 4;

If $Q_g <= Q_{max}$ and Φ is not empty then $C_i = C_i - RS_t$, return to Step 8

If $Q_g <= Q_{max}$ and Φ is empty then $C_i = C_i - RS_t$, return to Step 7

Step 10 $i = i + 1$, the detection of communities C_i complete,return to step 2

Step 11 the algorithm ends

This algorithm takes the 19 node network as the example to detect communities for illustrating the algorithm process. The network is shown in Fig 1.
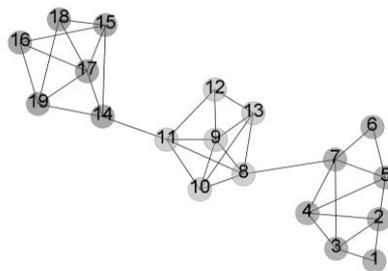

Figure 1. The 19 Node Network

According to the formula (6), a CNC matrix δ can be obtained,as show in talbe 1. Then the algorithm select the node set VD with the largest degree from V, VD={7,8,9, 11,17}, arbitrarily taking a node 7 as the initial community C1. The community detection begins. First,the algorithm updated the C1 neighbor set {3,4,5,6,8} , extracted CNC crosstabs R

from δ, selected the maximum value set Φ from R,where just a maximum value $\delta1$ (7,5) = 1.25 is in Φ, added 5 into C1, and then calculated Qg ,Qg =0.125000> Qmax =0, C1={7,5}. Repeating the process, the nodes 6 4,2,3,1 are successively added with Qmax = 0.923. The CNC crosstabs R of the community C1{7,5,6} and the neighborhood set S is shown in Table 2.

Table 1. Part of the CNC Table of the 19 Node Network

node	1	2	3	4	5	6	7
1	0	0.25	0.25	0	0	0	0
2	0.25	0	1.25	0.5	0.25	0	0
3	0.25	1.25	0	0.41	0	0	0.25
4	0	0.5	0.41	0	0.41	0	0.5
5	0	0.25	0	0.41	0	0.25	1.25
6	0	0	0	0	0.25	0	0.25
7	0	0	0.25	0.5	1.25	0.25	0

Table 2. The CNC Crosstabs R of the Community C1 and its Neighborhood Set S

S	C1			
	3	4	2	8
7	0.25	0.5	0	0
5	0	0.41	0.25	0
6	0	0	0	0

At last , the neighbor node set of C1 is S = {8}, with six pairs of nodes in the set Φ while$\delta1$(7,8) =$\delta2$(5, 8) =$\delta3$(6,8)=$\delta4$(4,8) =$\delta5$(2,8)=$\delta6$(3,8) =0. This meet the community termination condition , go to step 10. The community C1 ends ,C1 = {7,5,6,4,2,3,1}, i = i + 1 = 2, go to step 2. The community C2 began.
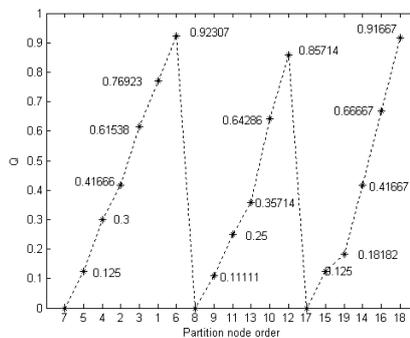

Figure 2. The Changes of the Modularity in the 19 Node Network

the above process is repeated until the end of the algorithm, with two communities, C2 = {8,9,11,13,10,12} , C3 = {17,15,19,14, 16, 18}. The changes of the modularity are shown in Fig 2. The found Order of each Node with the corresponding CNC is shown in Table.3.

Table 3. The Found Order of each Node with the corresponding CNC

node order	first	second	third	four	five	six	seven
$\Phi c1$	7	(7,5)	(7,4)	(4,2)	(2,3)	(2,1)	(7,6)
CNC of C1		1.25	0.5	0.5	1.25	0.25	0.25
$\Phi c2$	8	(8,9)	(9,11)	(9,13)	(8,10)	(9,12)	_____
CNC of C2		0.66	0.85	0.85	0.57	0.41	_____
$\Phi c3$	17	(17,15)	(17,19)	(17,14)	(17,16)	(17,18)	_____
CNC of C3		1.14	1.14	0.5	0.5	0.5	_____

Since there is few common neighbor between some pairs of nodes, the CNC of those nodes are set to 0. This problem can be solved by two methods, either performs the step 5 to add the node early into the community C1, or instead of step 5, perform step 9. The question is worth to be discussing which a gooder way to reduce the complexity is. Such as node 6, when the community C1 = {7,5}, the two neighbors of the node 6 belong to C1, which meet the requirements of step 5. So the node 5 is added into C1. Of course, for the latter, the node 5 can also be added into C1. By comparison, because the home of each neighbor node is computed by the former and this will greatly increase the time complexity of the algorithm, and therefore in this case, the algorithm gives the process as follows:

(1) When the degree of the node is one in the network, and its number of common neighbor is 0, the node can be added into a community whose neighbor is in.

(2) Those nodes satisfy the conditions of Step 5, but whose common neighbors are more than 0, go to step 9.

In our algorithm, the time complexity to compute the CNC of all the nodes is O (MK), where M is the number of nodes in the network, K is the average degree of the nodes. Next, there are the following key steps: The time complexity of those steps is O (1) , such as Looking for the initial community core, finding community neighbors, creating the crosstabs R between the community and its neighbors,　selecting the largest CNC from R. Then, the time complexity of selecting the largest CNC is O (M) in the worst case. The time complexity of Calculating the local modularity of each node to join the community is O(K). The time consumed for detecting the community is proportional to the size L of the community, L<N. So the total time complexity is O ((1 + 1 + 1 + 1 + M + K) L + MK) < O (MN). Real networks are usually sparse networks, so O (M) is equivalent to O (N). The complexity of the algorithm is O (N2).

4. Experimental Analysis

4.1. Zachary Karate Club Network

From 1970 to 1972, sociologists Zachary observed an unofficial karate club from an American university, and build the corresponding social relation network according to the mutual exchanges among members [13]. Zachary Karate Club network has become a benchmark network of testing community mining algorithm performance. It can only be used to test whether or not the algorithm accurately predicts the final divide situation based on the observed network structure. The nodes in the network denote the club members, and the connection between two nodes indicates that It has a close relationship between the corresponding members. The network has 34 nodes and 78 sides.

The director and the president of the club have had a dispute over whether to raise the club charges, and leading to the club divide into two smaller groups, With the director and president as the core, respectively. The node 34 and 1 are the club president and director respectively.
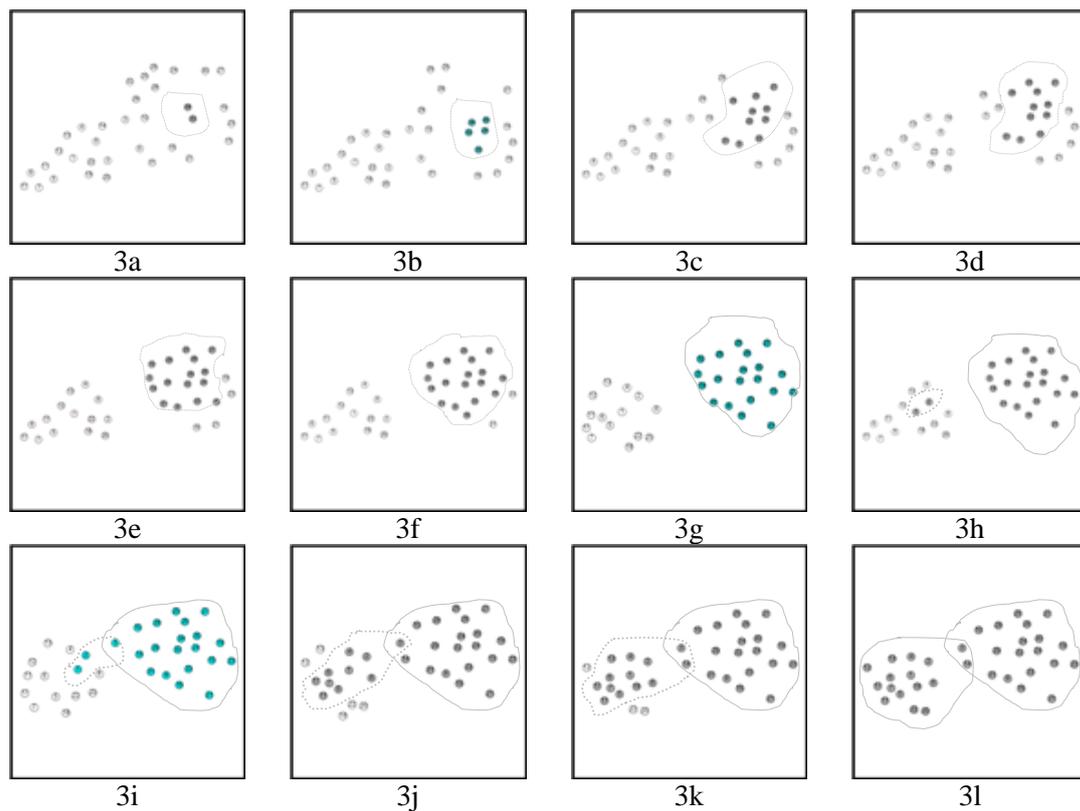
Figure 3. 3a - 3g Shows Iteration Results of the Community C1 for the 1,4,9,11,28,46,64 times, Multiple pairs of nodes were selected, for the nodes: (34, 33), (34, 32), (34,31), (9, 28), (33 , 15), (33,19), (33,23). Figure 3h ~ 3l shows iteration results of the community C2 for the 1,3,9,14,18 times, where in the 3rd and 14th iteration, the node 3 and 14 of C1 were added to C2. So C1 and C2 become overlapping communities

Firstly, the CNC matrix is calculated according to the equation (6), and then the maximum degree node 34 and its neighbor of the network are found. By calculating the node 33 are identified which has the largest CNC with node 34 from its neighbor nodes, so the initial community is set up ,C1 = {34,33}. A new community modularity Q_g is Calculated, $Q_g = 0.035714 > Q_{max} = 0$, return to Step 4.Then , the algorithm updated the community neighbor node set, extracted the CNC crosstabs, and got the new CNC maximum node 30 at last. The node 30 is added, determining its Q_g, $Q_g = 0.100000 > Q_{max}$. The above process is repeated, and finally we got the local community C1= {3,9,10,14,15,16,19,21,23,24,25,26,27,28,29,30,31,32,33,34}. Fig 3a ~ 3g is the formation of the community C1. Executing the algorithm in the remaining network nodes,the community C2 is obtained, C2={1,2,3,4,5,6,7,8,11,12,13,14,17,18, 20,22}. The node 3 has the largest CNC among the C2 neighbors in the second iteration,its Q_g = 0.041667> Q_{max}, so the node 3 becomes the Share node of C2 and C1. Similarly, node 14 also has become a shared node. Fig 3h ~ 3k shows the division process of community C2. The ownership of the node 3 has been very controversial, because it is samely close contact with the two communities from the topological relation. So it is reasonable to make the node 3 a shared node.

4.2. American Political Book Network

The American political books networks [14] is constructed by V. Krebs network, according to the sales pages associated with political books in the Amazon online bookstore, as shown in Fig 4a. Nodes in the network on behalf of the book, the network edge means that a certain number of readers both buy the two books at the same time. The network nodes are divided into three categories: "liberal," "conservative" and "centrist", as shown in Fig 4a, yellow nodes, red nodes, blue nodes, respectively. These factions are divided by Mark Newman's analysis, According to the book views as well as evaluation on the Amazon online bookstore. The network consists of 105 nodes and 441 edges. Firstly , this algorithm calculated all CNC values to obtain the CNC matrix, then set the value Qmax = 0, found the node 9 with the largest degree from all nodes as the initial node of the community, Looked for initial community neighbor node sets S, S= {4, 10,11,12,13,14,15,21,22,23,24,25,27,28,33,34,36,38,41,42,43,44,45,46,47},and extracted the CNC crosstabs R, as shown in Table 4.

Table 4. The CNC Crosstabs R of the Initial Community

Neighbor	4	10	11	12	13	14	15	21	22
CNC	0.6044	0.3881	0.3573	0.6119	0.9194	0.6298	0.0913	0.0973	0.0871
Neighbor	23	24	25	27	28	33	34	36	38
CNC	0.0323	0.4327	0.1313	0.2421	0.1598	0.1288	0.3152	0.4318	0.1072
Neighbor	41	42	43	44	45	46	47		
CNC	0.7846	0.0838	0.1960	0.1748	0.0824	0.0897	0.0064		

As shown in Table 4, the Node 13 is found from the R with a maximum value 0.9194, then, calculating Qg, Qg = 0.0204> Qmax = 0, so adding 13 into the community C1, C1 = {9,13}. So C1 constantly is updated. the program is terminated when the Qg stopped growing. the "conservative" community is finally obtained with gray background nodes, as shown in Figure 4b.
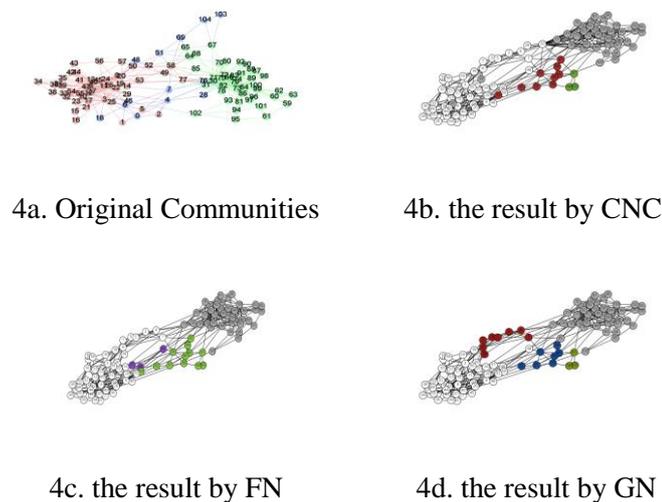

4a. Original Communities   4b. the result by CNC

4c. the result by FN   4d. the result by GN

Figure 4. Original Communities and Communities divided by Algorithms about the American Political Book Network

Compared with the GN algorithm and FN algorithms, the proposed algorithm divides the network into four communities, while five communities for GN algorithm. Next, we take the liberals community as an example to introduce the accuracy of CNC. The

proposed algorithm divides the centrist nodes 4, 6, 7, 18, 28, 46, 48, 51 and the liberal node 30 into conservatives. Another five conservative nodes 29, 52, 53, 57, 77 failed to be found. The analysis result of American political books networks by GN algorithm is shown in Fig 4d. the GN algorithm divided the network into five communities. There are 10 nodes failed to find, and five nodes divided mistakenly. The analysis result of the networks by FN algorithm is shown in Fig 4c. There are 8 nodes failed to find, and 7 nodes divided mistakenly. Specifically, as shown in Table 5.

Table 5. The Contradiction Node set of CNC and other Algorithms

algorithms	Misclassification	Not classified
CNC	4 6 7 18 28 30 46 48	29 32 52 53 57 77
FN	4 6 7 18 28 30 48	29 32 49 50 52 53 58 77
GN	4 18 28 46 48 51	1 2 3 5 8 29 32 52 53 77

From the above results, the error divided nodes of this paper are slightly more, but undivided nodes are far less than the above two methods, so the correct rate of the proposed algorithm is better than the above two algorithms.

4.3. American University Football Network

The American college football network shows American University football game between a season. The network includes 115 nodes and 613 edges, whose node represents the team, whose edge mean a game between two teams. These teams are divided into 12 Union, The games within the Union are more than the games among Unions. The proposed algorithm divided 115 nodes into 13 communities, where the Sun Belt is divided into two communities. The divided communities are shown in Fig 5. In addition, seven communities were entirely correct classification by the proposed algorithm, as follows Mountain West, Atlantic Coast, Conference USA, Pacific Ten, Big Ten, Southeastern, Big Twelve, with the time 0.172s Only,as show Table 6. In contrast, the GN algorithm divides the network into 10 communities, and only two communities are divided correctly, Atlantic Coast and Big Ten, whose modularity is 0.599629, and whose running time is 0.7s. Although the speed of FN slightly faster than the proposed algorithm, it only detected six right communities. In addition, the CS algorithm [15] divided the network into 13 communities, where the node LouisianaTech and MiddleTennesseeState were divided into two communities respectively, whose running time is up to 1.31s. For comparison, not only our algorithm has the highest modularity, but also In terms of accuracy and speed the proposed algorithm has a greater advantage than other algorithms.

Table 6. Comparison of CNC and other Algorithms

algorithms	The total number of communities found	The number of correct communities	Run time	Modularity
CNC	13	7	0.172	0.6010089
GN	10	2	0.7	0.599629
FN	6	2	0.11	0.5497407
CS	13	5	1.31	0.5959712

5a. the result by CNC   5b. the result by FN

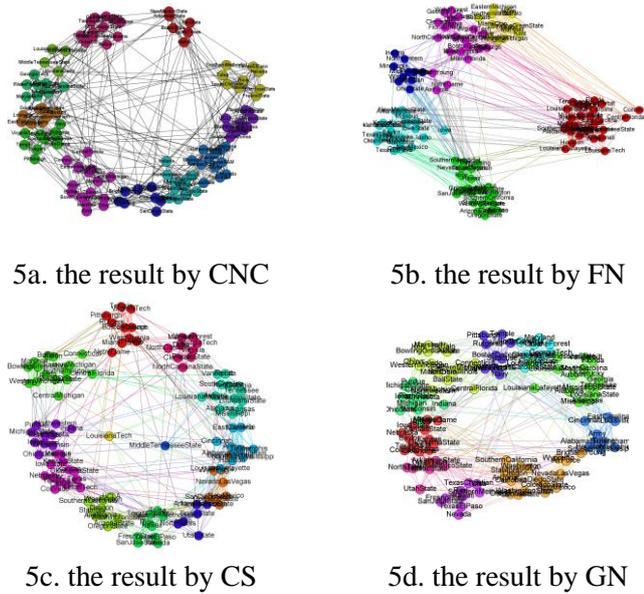5c. the result by CS   5d. the result by GN

Figure 5. Communities Divided by Algorithms about American University Football Network

4.4. Computer Generated Network

In order to test the performance of our algorithm, we apply our algorithm in an LFR reference dataset by comparing with other algorithms. Modularity Q is a parameter that can measure community characteristics, defined as [16]

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \qquad (7)$$

let k denotes the degree, Ci denotes the community of i, and m denotes the total number of edges in the network. If ci = cj, = 1 , otherwise 0. Q is between 0 and 1, In general, when Q is greater than 0.3, the network has an obvious community structure. Modularity is the most widely indicator used to judge community features. The original modularity is defined only for non-overlapping communities, but our algorithm can detect overlapping community, Chen *et al*. [17] proposed an improved modularity function on overlapping communities.

$$Q_{ov}^{O} = \frac{1}{2|E|} \sum_{c \in C} \sum_{i,j \in c} \left[A_{ij} - \frac{k_i k_j}{2|E|} \right] a_{i,c} a_{j,c} \qquad (8)$$

the belonging coefficient defined as:

$$a_{i,c} = \frac{\sum_{k \in c} A_{ik}}{\sum_{c' \in C_i} \sum_{k \in c'} A_{ik}} \qquad (9)$$

where Ci is the set of communities to which the node i belongs. It measures how tightly i is connected to C.

$$Q_{ov}^{O} = \frac{1}{2|E|} \sum_{c \in C} \sum_{i,j \in c} \left[A_{ij} - \frac{k_i k_j}{2|E|} \right] \frac{\sum_{k \in c} A_{ik}}{\sum_{c' \in C_i} \sum_{k \in c'} A_{ik}} \frac{\sum_{k \in c} A_{jk}}{\sum_{c' \in C_j} \sum_{k \in c'} A_{jk}} \qquad (10)$$

The network size can be changed by modifying the input parameters of the LFR reference network. the size of the network were 1000,10000,50000,100000. the community size is from 10 to 100.the average degree of nodes is 15. The maximum degree nodes are 30. Table 7 shows the number of communities detected in our algorithm

and other algorithms; the number of communities detected by our algorithm is closer to the real number of the LFR reference dataset.

Table 7. The Number of Communities Detected in LFR Data Sets

Algorithm	Scal=1000	Scal=10000	Scal=50000	Scal=100000
Number of Communities	17	184	1020	1962
GN	13	-	-	-
FN	30	284	1108	3091
CS	20	216	1083	2234
CNC	16	181	992	1987

The results are shown in Table 8. our algorithm runs as good as other algorithms. When the size of the community is large, Newman's algorithm can not run on our computer.

Table 8. The Modularity in LFR Data Sets

Algorithm	Scal=1000	Scal=10000	Scal=50000	Scal=100000
GN	0.9068	-	-	-
FN	0.849	0.8518	0.8416	0.8492
CS	0.8772	0.8681	0.8749	0.8679
OUR	0.8881	0.8763	0.8804	0.8673

5. Conclusion

This paper proposes a local community detection algorithm that started at degree maximum vertices and can discover the community structures of networks based on CNC. The algorithm can accurately select the candidate neighbor node to join the community, without visiting each neighbor node, and calculating the local modularity, so it can improve efficiency. In our algorithm, only partial information of networks is required to discover and to expand the local communities, the global information of entire network need not be known in advance. The overlapping communities can also be identified by using our method. This paper gives the experimental results on typical datasets and benchmark data sets. The experiment results show that the identification of community structure of our algorithm is as good as the other algorithms, while the running time of our algorithm is less than that of the other algorithms. Therefore, our algorithm is efficient and suitable for finding community structures in large real networks.

Acknowledgments

References

[1] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks", nature, vol. 393, no. 6684, **(1998)**, pp. 440-442.
[2] L. A. Adamic and B. A. Huberman, "Power-law distribution of the world wide web", Science, vol. 287, no. 5461, **(2000)**, pp. 2115-2115.
[3] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks", Proceedings of the national academy of sciences, vol. 99, no. 12, **(2002)**, pp. 7821-7826.
[4] M. Fiedler, "Algebraic connectivity of graphs", Czechoslovak mathematical journal, vol. 23, no. 2, **(1973)**, pp. 298-305.

[5] A. Pothen, H. D. Simon and K. P. Liou, "Partitioning sparse matrices with eigenvectors of graphs", SIAM journal on matrix analysis and applications, vol. 11, no. 3, **(1990)**, pp. 430-452.

[6] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks", Proceedings of the national academy of sciences, vol. 99, no. 12, **(2002)**, pp. 7821-7826.

[7] M. E. J. Newman, "Fast algorithm for detecting community structure in networks", Physical review E, vol. 69, no.6, **(2004)**, pp. 066133.

[8] A. Clauset, "Finding local community structure in networks", Physical review E, vol. 72, no. 2 **, (2005)**, pp. 026132.

[9] Q. Chen, T. T. Wu and M. Fang, "Detecting local community structures in complex networks based on local degree central nodes", Physica A: Statistical Mechanics and its Applications, vol. 392, no. 3 , **(2013)**, pp. 529-537.

[10] X. Qi, "Optimal local community detection in social networks based on density drop of subgraphs", Pattern Recognition Letters, vol. 36, **(2014)**, pp. 46-53.

[11] Z. A. Zhu, S. Lattanzi and V. Mirrokni, "A local algorithm for finding well-connected clusters", Proceedings of the 30th International Conference on Machine Learning (ICML-13), **(2013)**, pp. 396-404.

[12] W. Wang, "Fuzzy overlapping community detection based on local random walk and multidimensional scaling", Physica A: Statistical Mechanics and its Applications, vol. 392, no. 24, **(2013)**, pp. 6578-6586.

[13] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks", Proceedings of the national academy of sciences, vol. 99, no. 12**, (2002)**, pp. 7821-7826.

[14] M. E. J. Newman, "Modularity and community structure in networks", Proceedings of the national academy of sciences, vol. 103, no. 23, **(2006)**, pp. 8577-8582.

[15] J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection", Physical Review E, vol. 74, no. 1, **(2006)**, pp. 016110.

[16] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks", Physical review E, vol. 69, no. 2, **(2004)**, pp. 026113.

[17] D. Chen, "Detecting overlapping communities of weighted networks via a local algorithm", Physica A: Statistical Mechanics and its Applications, vol. 389, no. 19, **(2010)**, pp. 4177-4187.

[18] P. Jaccard, "Distribution de la Flore Alpine: dans le Bassin des dranses et dans quelques régions voisines", Rouge, **(1901)**.

[19] G. Salton, "Automatic text processing: The transformation, analysis, and retrieval of Reading: Addison-Wesley", **(1989)**.

[20] E. Ravasz, "Hierarchical organization of modularity in metabolic networks", science, vol. 297, no. 5586, **(2002)**, pp. 1551-1555.

Authors
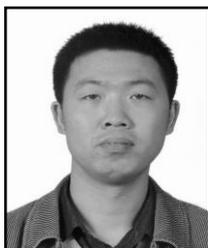
Wang Tianhong, he received his M.S. degree in software from Taiyuan University of Technology in Taiyuan, China. He is currently a lecturer in Xinzhou Teachers University. His research interest is mainly in the area of Complex Network. He has published several research papers in scholarly journals in the above research area and has participated in several books.