

Research on Complex Network Algorithm Optimization Community Detection Problem

ZHENG Zhiqing^{1, a}

¹Information Network Center, Shandong Institute of Business and Technology, Yantai, China

^aemail:Yantaizzqing@163.com

Key words: community structure; supernetwork; GN algorithm; standardization centrality

Abstract. The community structure is one of the most universal and important attributes of complex network, and the community division research of complex network is to reasonably divide the community structure practically existed in the complex network. Apply the thought and theoretical method of supernetwork into the community division research of complex network. Aiming at the deficiencies of present GN algorithm, combining theoretical method of standardization centrality from the perspective of supernetwork, a kind of new community division algorithm of complex network is constructed, and the new algorithm is verified and analyzed through the example. The experimental result reveals that the new algorithm is improved and perfected to some extent in the division result compared with GN algorithm.

Introduction

The complex network is one of the present most important cross-discipline research fields [1-3]. Paralleling with worldlet [1], scale free [2] and other features, the community structure is one of the most important attributes of complex network [3]. The community division research of complex network is to reasonably divide the community structure practically existed in the complex network. It is of great importance and value to analyze the topological structure of complex network, understand the functions of complex network, discover the concealed rules of complex network and predict the complex network behaviors [4].

Though many community division methods of complex network are proposed at present, it is still a problem that has not been solved well for how to further improving the community division precision, especially how to divide the structure of network communities more reasonably under the premise that the quantity of network communities is unknown[4]. For this point, it proposes to adopt the thought and method of supernetwork for the research in this paper. When Nagurney was managing the blending network, the network superior to the existing network was called as the “supernetwork” and the meaning of supernetwork became clear. The academician Wang Zhongtuo and professor Wang Zhiping further proposed the features of supernetwork. At present, the researches of supernetwork mainly include the research based on variational inequality, which transfers the multi-hierarchy and multi-criteria balance model of supernetwork to optimization problem, and then is solved by evolution of variational inequality. The research based on system science mainly comprehensively and partially research on the supernetwork, including research of relation among networks in supernetwork, research of inter-network relationship with the network and the research of overall performance. The work in this paper is mainly inspired and helped in this aspect. The research based on supergraph is considered to be another relatively important research direction.

Algorithm Research

GN algorithm is a kind of splitting algorithm. Compared with other methods, GN algorithm can be called as a kind of classical community division algorithm, it often can get comparatively good result after matching with the modularity of Q , and the limiting conditions of GN algorithm are few with high adaptability and wide application. However, after research, recently it reveals that

there are some deficiencies for such algorithm in the aspect of precision. The division classification of GN algorithm for connected even points, errors often occur and the division result is not ideal, including the wrong division of node 3 when dividing the Zachary karate club network. Zachary karate club network is a common used classical test network in the fields of complex network and social network analyses. Wayne Zachary spent several a few years observing the social relations among members of a karate club in a university and constructed the social network of the members of the club. The network includes 34 nodes, each node represents a member of the club and the lines connecting the nodes represent the friend relationship between two members. During the investigation, due to dispute between the supervisor and the coach, the club broke up into two small communities centered on these two people. As a true social network, the network is commonly used for testing the precision and effectiveness of community division method.

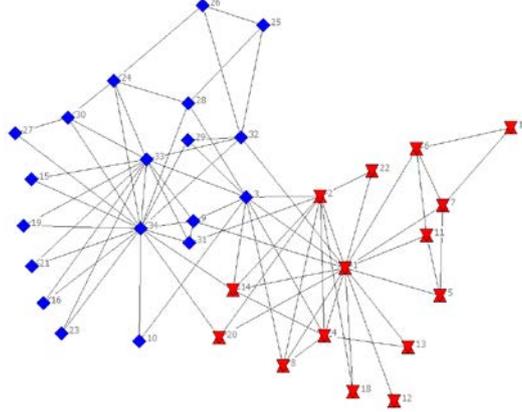


Fig.1 Division of GN Algorithm on Zachary Network

Fig.1 is the division result of GN algorithm on the network of Zachary karate club, and thereinto the node 3 is wrongly divided in the left community, which is inconsistent with the actual situation. The degree centrality algorithm is currently a kind of key method to test the importance of the network node. Individuals with higher degree centrality also play an important role in this group or network. The formula is as shown below, thereinto the formula (1) is the absolute figure, i.e. totaling the correlation coefficient of one individual. Formula (2) is the normalized value, i.e. divide it by the maximum possible correlation coefficient of the individual in the network for the convenience of comparison among different networks.

$$C_D(n_i) = d(n_i) = \sum_j X_{ij} = \sum_j X_{ji} \quad (1)$$

$$C'_D(n_i) = \frac{d(n_i)}{g-1} \quad (2)$$

X_{ji} is the numerical value of 0 or 1, and it represents whether the individual j is related with the individual i . g is the individual number in one network [23,13]. The basic process of the new algorithm is below:

Given the network $G=(V, E)$, thereinto the node v is any node in the network, i.e. $v \in V$; the connecting line e is any connecting line in the network, i.e. $e \in E$. Define the object network that is to managed initially from the perspective of supernetwork as full network $G_S, G_S=(V, E)$, and define the community after division as sub-network $G_Z, G_Z=(V_Z, E_Z)$, and $G_Z \subseteq G_S$.

(1) Step 1: calculate the quantity of the shortest path $B(e)$ passing each line in the network, $B(e) = \sum_{v_i \neq v_j \in V} n_{ij}(e)$, and $n_{ij}(e)$ represents the quantity including the line e in the shortest path among nodes $v_i, v_j \in V$.

(2) Step 2: find the line with the most quantity of passing the shortest path and remove it from the network, and respectively record the vertex of the removed line as per the different Q value of the corresponding modularity, i.e. generate the set V_Q of vertexes of the lines among the

corresponding communities as per different Q value of the modularity.

(3) Step 3: repeat step two till each node is a degraded community. Then select the community division result with local peak of modularity Q for analysis. Analyze whether the vertex v_Q of connecting line among communities in the selected result is in conformity with the connecting line equalization of each connected community, and define the node with equal connecting line of each connected community as equal node v_{jd} , i.e. judge whether $v_Q \in V_{jd}$ is supported at this time.

(4) Step 4: in case that $v_Q \in V_{jd}$, compare the standardization centrality $SC'_D(v_Q) = \frac{d_s(v_Q)}{g_s - 1}$ of the full network with the standardization centrality $Z_a C'_D(v_Q) = \frac{d_{G_{Za}}(v_Q)}{g_{G_{Za}} - 1}$ of the sub-network within the community ($v_Q \in G_{Za}$ at the moment). In case that standardization centrality of the full network is bigger than the standardization centrality of the sub-network within the community, i.e. $SC'_D(v_Q) > Z_a C'_D(v_Q)$. Turn to step 5, or end it.

(5) Step 5: duplicate the node v_Q into other communities with the selected division results, calculate the standardization centrality $Z'_x C'_D(v_Q) = \frac{d_{G_{Zx+v_Q}}(v_Q)}{g_{G_{Zx+v_Q}} - 1}$ ($x \neq a$) of the sub-network within in the new community, compare $Z'_x C'_D(v_Q)$ with the standardization centrality $SC'_D(v_Q)$ of the full network, and find the result that is bigger than the standardization centrality of the full network and with the maximum value. The corresponding division result at the moment is the desired result.

Verification And Analysis

Firstly take the Zachary karate club network as one example for verification and analysis, and it can get the result that the node 3 meets the condition $v_Q \in V_{jd}$ through the first three steps in this paper, i.e. the quantity of the connecting lines of node 3 with the connecting two communities in the left and right and both are 5. In step 4, the calculation and analysis of the standardization centrality of the node 3 are made from the perspective of supernetwork, the result of full network standardization centrality is as shown in Fig. 2, and the result of sub-network standardization centrality is as shown in Fig. 3.

		1	2	3
		Degree	NrnDegree	Share
34	34	17.000	51.515	0.109
1	1	16.000	48.485	0.103
33	33	12.000	36.364	0.077
3	3	10.000	30.303	0.064
2	2	9.000	27.273	0.058
4	4	6.000	18.182	0.038
32	32	6.000	18.182	0.038
14	14	5.000	15.152	0.032
9	9	5.000	15.152	0.032
24	24	5.000	15.152	0.032
6	6	4.000	12.121	0.026
8	8	4.000	12.121	0.026
30	30	4.000	12.121	0.026
7	7	4.000	12.121	0.026
28	28	4.000	12.121	0.026
31	31	4.000	12.121	0.026
26	26	3.000	9.091	0.019
29	29	3.000	9.091	0.019
5	5	3.000	9.091	0.019
20	20	3.000	9.091	0.019
25	25	3.000	9.091	0.019
11	11	3.000	9.091	0.019
19	19	2.000	6.061	0.013
18	18	2.000	6.061	0.013
21	21	2.000	6.061	0.013
22	22	2.000	6.061	0.013
27	27	2.000	6.061	0.013
18	18	2.000	6.061	0.013
16	16	2.000	6.061	0.013
13	13	2.000	6.061	0.013
23	23	2.000	6.061	0.013
15	15	2.000	6.061	0.013
17	17	2.000	6.061	0.013
12	12	1.000	3.030	0.006

Fig. 2 Degree Centrality of Nodes in Full Network

		1	2	3
		Degree	NrnDegree	Share
19	34	15.000	83.333	0.188
18	33	12.000	66.667	0.150
1	3	5.000	27.778	0.063
17	32	5.000	27.778	0.063
9	24	5.000	27.778	0.063
13	28	4.000	22.222	0.050
2	9	4.000	22.222	0.050
15	30	4.000	22.222	0.050
16	31	3.000	16.667	0.038
11	26	3.000	16.667	0.038
14	29	3.000	16.667	0.038
10	25	3.000	16.667	0.038
4	15	2.000	11.111	0.025
12	27	2.000	11.111	0.025
3	10	2.000	11.111	0.025
7	21	2.000	11.111	0.025
8	23	2.000	11.111	0.025
5	16	2.000	11.111	0.025
6	19	2.000	11.111	0.025

Fig. 3 Node Degree Centrality of the Left Community

We can see from Fig. 2 and Fig. 3 that $SC_D(3)=10$, $SC'_D(3)=\frac{d_s(3)}{g_s-1}=0.30303$, $Z_0C_D(3)=5$,

$$Z_0C'_D(3)=\frac{d_{G_{Z_0}}(3)}{g_{G_{Z_0}}-1}=0.27778, \text{ i.e. it is in conformity with the conditions of } SC'_D(3) > Z_0C'_D(3). \text{ In}$$

step 5, duplicate the node 3 into other sub-networks other than the belonged sub-network, i.e. duplicate the node 3 into the right community in Fig. 1, calculate the sub-network standardization centrality in the new community, and the result is as shown in Fig. 4.

		1	2	3
		Degree	NrnDegree	Share
1	1	14.000	93.333	0.212
2	2	8.000	53.333	0.121
4	4	6.000	40.000	0.091
3	3	5.000	33.333	0.076
7	7	4.000	26.667	0.061
6	6	4.000	26.667	0.061
8	8	4.000	26.667	0.061
12	14	4.000	26.667	0.061
5	5	3.000	20.000	0.045
9	11	3.000	20.000	0.045
11	13	2.000	13.333	0.030
14	18	2.000	13.333	0.030
13	17	2.000	13.333	0.030
16	22	2.000	13.333	0.030
15	20	2.000	13.333	0.030
10	12	1.000	6.667	0.015

Fig. 4 Degree Centrality of the Right Community after Introducing Node 3

We can see from Fig. 2 and Fig. 4 that $SC_D(3)=10$, $SC'_D(3)=\frac{d_s(3)}{g_s-1}=0.30303$, $Z_1C_D(3)=5$,

$$Z_1C'_D(3) = \frac{d_{G_{Z_1+3}}(3)}{g_{G_{Z_1+3}}-1} = 0.33333, \text{ and it is in conformity with the conditions that}$$

$Z'_x C'_D(v_Q) > SC'_D(v_Q)$ and the value is the maximum. In other words, according to the algorithm herein, for the current network, it will be more suitable if incorporating the node 3 into the right sub-network. The division result of the karate club network by the algorithm in this paper is as shown below and it is in conformity with the actual situation of the Zachary network.

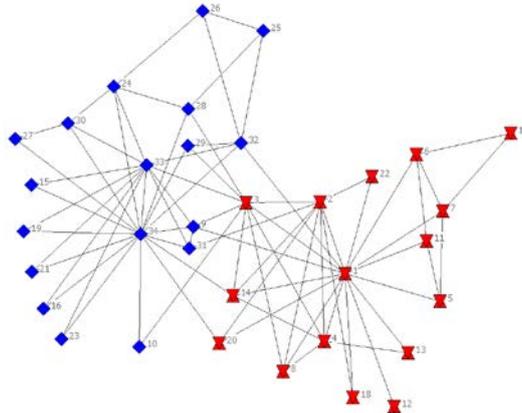


Fig. 5 Division of Zachary Network by the Algorithm in This Paper

Conclusions

It applies the supernetwork thought, theory and method of supernetwork into the community division research of complex network, Aiming at some deficiencies of the present algorithm, a kind

of community division algorithm of complex network based on supernetwork perspective is constructed, and the verification and analysis are made. The experimental result reveals that compared with GN algorithm, the new algorithm has been improved and perfected in the division result. Our current work is relatively rough, and we hope the experts and peers give more guidance and help.

Reference

- [1] Y. Geng, J. He, H. Deng and K. Pahlavan, Modeling the Effect of Human Body on TOA Ranging for Indoor Human Tracking with Wrist Mounted Sensor, 16th International Symposium on Wireless Personal Multimedia Communications (WPMC), Atlantic City, NJ, Jun. 2013.
- [2] Y. Geng, J. He, K. Pahlavan, Modeling the Effect of Human Body on TOA Based Indoor Human Tracking[J], International Journal of Wireless Information Networks 20(4), 306-317
- [3] Su, Tianyun, Zhihan Lv, Shan Gao, Xiaolong Li, and Haibin Lv. "3D seabed: 3D modeling and visualization platform for the seabed." In Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on, pp. 1-6. IEEE, 2014.
- [4] Li, Xiaoming, Zhihan Lv, Baoyun Zhang, Weixi Wang, Shengzhong Feng, Jinxing Hu. "WebVRGIS Based City Bigdata 3D Visualization and Analysis." In Pacific Visualization Symposium (PacificVis), 2015 IEEE. IEEE, 2015.
- [5] Li, Xiaoming, Zhihan Lv, Baoyun Zhang, Ling Yin, Weixi Wang, Shengzhong Feng, Jinxing Hu. "Traffic Management and Forecasting System Based on 3D GIS" Cluster, Cloud and Grid Computing (CCGrid), 2015 15th IEEE/ACM International Symposium on. IEEE, 2015.