Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 31 (2014) 85 – 94

Procedia Computer Science

2nd International Conference on Information Technology and Quantitative Management, ITQM 2014

Weighted graph clustering for community detection of large social networks

Ruifang Liu[a], Shan Feng[a], Ruisheng Shi[b,c,*], Wenbin Guo[a]

[a]*School of Information and Communication Engineering, Beijing University of Posts and Telecommunications,Beijing 100876, China*
[b]*Education Ministry Key Laboratory of Trustworthy Distributed Computing and Service,, Beijing University of Posts and Telecommunications,Beijing 100876, China*
[c]*School of Humanities, Beijing University of Posts and Telecommunications,Beijing 100876, China*

Abstract

This study mainly focuses on the methodology of weighted graph clustering with the purpose of community detection for large scale networks such as the users' relationship on Internet social networks. Most of the networks in the real world are weighted networks, so we proposed a graph clustering algorithm based on the concept of density and attractiveness for weighted networks, including node weight and edge weight. With deep analysis on the Sina micro-blog user network and Renren social network, we defined the user's core degree as node weight and users' attractiveness as edge weight, experiments of community detection were done with the algorithm, the results verify the effectiveness and reliability of the algorithm. The algorithm is designed to make some breakthrough on the time complexity of Internet community detection algorithm, because the research is for large social networks. And the another advantage is that the method does not require to specify the number of clusters.

Selection and peer-review under responsibility of the Organizing Committee of ITQM 2014.

Keywords: Micro-blog; User Attractiveness; Community Detection; Weighted Graph Clustering

1. Introduction

Community detection has received significant attention in all kinds of networks,such as the World Wide Web[1], collaboration networks[2,3], biological networks[4], and social networks[5].

With the rapid development of the Internet, it recently have attracted the attention of researches with different algorithms to discover and analyze the potential communities in the Internet. It was known that, the variety of physical social circles, in some levels, could reflect the relationship among people. People in a physical social circle usually also have some contacts in the Internet. Through analyzing community structure of the online social networks, such as FaceBook, Twitter, Sina micro-blog, Renren, we could probably find the potential relationship exists among people. Many applications on Internet, such as recommendation systems, can also benefit from such social network analysis.

* Corresponding author.
E-mail address: shiruisheng@bupt.edu.cn

Selection and peer-review under responsibility of the Organizing Committee of ITQM 2014.
doi:10.1016/j.procs.2014.05.248

Community detection problem has been studied as the graph partitioning problem in computer science for decades and is known to be a NP-hard problem. Many algorithms have been proposed, including hierarchical clustering[5,6,7], random walk based methods[8,9], spectral clustering[10,11], modularity based methods[4,7,12], user profile based methods[13,14]. These methods are all popular methods, but in the real world, most of the networks contain weighted information, however, there are only a few algorithms designed for weighted networks, and most algorithms are difficult to extend to weighted networks unfortunately. Another problem is that many algorithms are not fit for the large-scale networks community detection because of the high computational complexity.

In the paper, we proposed the concept of community attractiveness, with this definition, a clustering algorithm is constructed, named attractiveness-based community detection(ABCD) algorithm, which are introduced in section 3. In section 4, with the analysis on the micro-blog user network, we define the concepts of node weight and edge weight for the network, and present the experimental results, and the performance and execution time compare were done between ABCD algorithm and CNM(Clauset-Newman-Moore) algorithm[12]. Section 5 shows the experimental results on the College Football Team dataset, and section 6 shows the experimental results of a social network called Renren. Conclusions appear in Section 7. The research is for large social networks, and the another advantage is that the method does not require to specify the number of clusters, this number is usually not known in advance and is difficult to estimate.

2. Related Works

GN algorithm[6] is historically important, because it marked the beginning of a new era in the field of community detection, but it requires a time $O(n^3)$ on a sparse graph. CNM algorithm[12] is an improved algorithm, it has essentially linear running time $O(n \log^2 n)$.

Some works were done for Internet social networks. ISCoDe[13] is a framework based on methods for detecting communities over weighted graphs, where graph edge weights are defined based on measures of similarity between individuals interests tag. Slah Alsaleh et al.[14] provide a system using a clustering technique to create sets of communities based on users information, and then similar communities are matched based on users activities. For twitter dataset in a paper[15], twitters are nodes, the count of retweets between A and B is the weight of edge, and then to accentuate clusters with variable density, the experimental result is not good, only a small number of communities were detected in the Twitter dataset.

3. Weighted Graph Clustering

3.1. Problem Statement

Community structure finding can be considered as a graph clustering problem. And this problem can be considered as an optimization problem[16].

We suppose each person or a community has a density value, and each pair of persons or communities has an attractiveness value. The social network is a graph, each person is a node, edges are the relationship between people. Given a sparse graph $G(V, E, W_V, S_E)$ which consists of the node set V, the edge set E, the weight of node set W_V, and the weight of edge set S_E, we are interested in finding the clusters of G as communities.

Undirected graphs are the most common models of networks, where the directions of the connections are unimportant and can be safely ignored. Here we considered only undirected graph. Figure 1 shows such a graph, the weight of node implies the core degree of the person in the network, and the weight of edge means the attractiveness between the two nodes.

The result of graph clustering should partition a graph into several sub-graph(clusters), each part has a weight value, what's more, there are attractiveness values between clusters which similar with the edge weights. The candidate communities should have weights higher than the attractiveness with other clusters.

The optimization objective function is equation (1), $\mathcal{P}$ is the partitions of a graph.

$$\text{argmax}_{\mathcal{P}}\{\sum_{k\in\mathcal{P}} W(k) - \sum_{i,j\in\mathcal{P}} S(i,j)\} \tag{1}$$

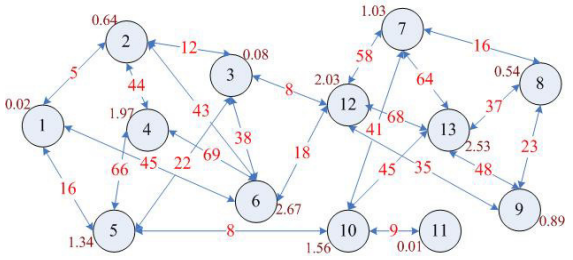

Fig. 1. An example of weighted graph

3.2. Preliminaries of Community in Weighted Graph

1) Definition one: density of cluster

Cluster density is the average of all the weights of nodes in the cluster. That is to say, if cluster i has Q_i nodes, each node's weight is $W_a, a \in 1, 2, ..., Q_i$, then the cluster density of i is:

$$W_i = \frac{\sum_{a=1}^{Q_i} W_a}{Q_i} \tag{2}$$

2) Definition two: attractiveness between clusters

Attractiveness between clusters is the ration of the sum of all the edges' weights between the two clusters and the product of the node number of the two clusters. The number of edges between cluster i and j is q, the edge weight is $S_e, e \in 1, 2, ..., q$, community i has Q_i nodes, and community j has Q_j nodes, then the attractiveness between cluster i and j is:

$$S_{ij} = \frac{\sum_{e=1}^{q} S_e}{Q_i \times Q_j} \tag{3}$$

3) Definition three: inter-interested clusters

If cluster i and cluster j are inter-interested clusters, then they must satisfy the following conditions:

$$q \geq Q_i, q \geq Q_j \tag{4}$$

4) Definition four: community

A cluster i can be a community, it must satisfy that:

$$S_{ij} < W_i + W_j, \forall j \tag{5}$$

Cluster j is the inter-interested cluster of cluster i.

3.3. Clustering Algorithm

In initial, each node is looked as a single cluster. The algorithm is an agglomerative algorithm.

If we want to do the merger for cluster i, firstly, we need to find which cluster among all its inter-interested clusters would get the highest attractiveness with it, which will be denoted by j.

But after finding the two clusters, we can not do the merger directly, because the attractiveness between them may be very small, meaning that they may not be of the same community, so we have to make some other judgment. Only S_{ij} meets the condition:

$$S_{ij} \geq W_i + W_j \tag{6}$$

The cluster i and j will be merged.

In addition, there may be two special cases during the merger. The first is that cluster i may not have inter-interested clusters, then cluster i will not merge with any other clusters, it will be a community; the second is that there are more

than one clusters have the highest attractiveness with cluster i, and satisfies the expression (6) at the same time, then we merge cluster i with any one of them.

Cluster attractiveness matrix S is a k-order matrix, where $S_{ij} = S_{ji}$ denotes the attractiveness between the cluster i and j, k is changing in every iteration. Assuming that the total number of node is n, then the attractiveness matrix S is a n-order matrix at the begin. The matrix S below shows the attractiveness of the graph shown in Fig. 1. Since the matrix S is sparse, so we can use the triplet to store the elements of the matrix.

$$S = \begin{pmatrix}
0 & 5 & 0 & 0 & 16 & 45 & 0 & 0 & 0 & 0 & 0 & 0 \\
5 & 0 & 12 & 44 & 0 & 43 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 12 & 0 & 0 & 22 & 38 & 0 & 0 & 0 & 0 & 8 & 0 \\
0 & 44 & 0 & 0 & 66 & 69 & 0 & 0 & 0 & 0 & 0 & 0 \\
16 & 0 & 22 & 66 & 0 & 0 & 0 & 0 & 0 & 8 & 0 & 0 \\
45 & 43 & 38 & 69 & 0 & 0 & 0 & 0 & 0 & 0 & 18 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 16 & 0 & 41 & 0 & 58 \\
0 & 0 & 0 & 0 & 0 & 0 & 16 & 0 & 23 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 23 & 0 & 0 & 0 & 35 \\
0 & 0 & 0 & 0 & 8 & 0 & 41 & 0 & 0 & 0 & 9 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 9 & 0 & 0 \\
0 & 0 & 8 & 0 & 0 & 18 & 58 & 0 & 35 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 64 & 37 & 48 & 45 & 0 & 68 \\
\end{pmatrix}$$

$$\begin{matrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 64 \\ 37 \\ 48 \\ 45 \\ 0 \\ 68 \\ 0 \end{matrix}$$

The number of clusters would be reduced after merging, the density of clusters will change, and the attractiveness between clusters will also change accordingly, so it is necessary to update the matrix S. When we update the new attractiveness matrix, we can make use of the old one to reduce the amount of calculation. Before the merge, the collection of clusters is CM_{pre}, the corresponding attractiveness matrix is S' , the number of clusters is k' ; after the merger, the collection of clusters is CM_{cur}, and the corresponding attractiveness matrix is S, the number of clusters is k. We use the following mathematical expression to denote CM_{pre}.

$$CM_{pre} = \{cm_l | l = 1, 2, ..., k'\} \tag{7}$$

Where cm_l denotes the l-th cluster in CM_{pre} .

If the cluster p in CM_{cur} contains a number of clusters in CM_{pre} , and the number is m, that is cluster p in CM_{cur} is formed by the merger of m clusters in CM_{pre}, the mathematical expression is:

$$CM_{cur}^{p} = \{cm_t | cm_t \in CM_{pre}, t = 1, 2, ..., m\} \tag{8}$$

Where CM_{cur}^{p} denotes the cluster p in CM_{cur} . Then, the attractiveness between community i and community j, that is the element S_{ij} in matrix S, can be updated by formula (9):

$$S_{ij} = \frac{\sum_{cm_r \in CM_{cur}^{i}, cm_t \in CM_{cur}^{j}} S'_{cm_r, cm_t} \times Q_{cm_r} \times Q_{cm_t}}{Q_i \times Q_j} \tag{9}$$

Updating the elements in S one by one with formula (9), then we get the new attractiveness matrix.

ABCD algorithm can be divided into two main steps, iterating between the two steps to get clusters:

1. Merge the pair of clusters which has the largest attractiveness.
2. Calculate or update the cluster density and cluster attractiveness matrix;

Executing the update of cluster density and attractiveness matrix, and the cluster merger process iteratively, until the structure of clusters does not change, or there is only one cluster left.

In initial, the time required to calculate attractiveness matrix is $O(n\overline{k})$, where n is the number of nodes, and $\overline{k}$ denotes the average number of inter-interested nodes for all nodes. The time consuming of merger of each iteration is $O(m_i)$, the time for updating attractiveness matrix is $O(m_i^2)$, so the time complexity of each iteration is $O(m_i^2)$, where m_i denotes the number of clusters at the beginning of i-th iteration. The maximum number of iterations is t , so the total time complexity of ABCD algorithm is $O(n\overline{k} + t\overline{m}^2)$. Based on the experimental results, the number of iterations is much smaller than the number of nodes, especially for large-scale network, the number of merger is of several orders of magnitude smaller than the number of nodes.

4. Community Detection of Micro-blog

4.1. User Characteristics of Micro-blog

There are several micro-blog service systems in China, Sina Weibo is one of the biggest, registered users are more than 300 millions. Weibo user publishes any topics and follows other users to receive their tweets, just like Twitter. In

The ABCD Algorithm

1) Initialize community attractiveness matrix S, S is a n-order matrix, the matrix elements S_{ij} denotes the attractiveness between the node i and j;
2) According to the newest attractiveness matrix S, for each cluster i, find cluster j which would get the highest attractiveness with i, and record their attractiveness with S_{ij} , and calculate the density of the cluster i and cluster j, respectively denoted by W_i, W_j;
3) Merge the communities satisfy the expression(6);
4) If any of the following happens, skip to step 7):
Case one: the structure of clusters does not change
Case two: only one cluster left
5) Update the cluster attractiveness matrix S;
6) Repeat from step 2) to 5);
7) Stop the iterative process, save the result of clusters as communities and return.

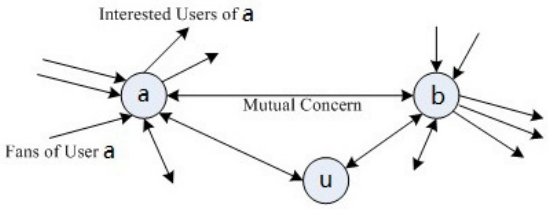

Fig. 2. The user's following relationship of Micro-blog

order to verify the validity of our algorithms, we tested them on the data set got from Sina micro-blog, which contains 70 thousand users and 0.6 million bi-connect links(following each other) among these users.

Each user of micro-blog has four specific attributes: number of interested users, number of fans, number of micro-blog and verified or not. What we are mainly concerned is to discover the potential relationship among the micro-blog users, thus we only consider two attributes, number of interested users and number of fans, shown in Fig. 2. The number of fans, as in-degree, shows the user's popularity. The number of interested users, as out-degree, shows the user's activity.

Micro-blog is a special social network, in this network, if a is interested in b, then the relationship between a and b is established, but b may not be interested in a, so b has the absolute right upon the relationship between a and b. In addition, another point needs to be highlighted, due to the trend of "mutual concern" in micro-blog, some users pay attention to a lot of people in order to increase the number of their fans, the characteristics of these users is that the ratio of the number of fans and number of interested users is less than or equal to one. While the ratio of the number of fans and number of interested users of the truly core users is far greater than one, so we can use this feature to check a user is core or not, and then decrease the influence of such users. In Fig. 2, user a and user u are mutual-concern, the same with user b and u. The number of users like user u is a very important parameter for the connection of node a and node b.

4.2. Definitions of Weights

1) Definition one: node weight

Each Weibo user is a node in the weighted graph of the users relationship network. The node weight is the user's core degree in the network.

If F_a is the number of fans of user a, P_a is the number of interested users of user a, then the user's core degree of user a is:

$$W_a = \frac{F_a}{P_a^2} \tag{10}$$

The higher the user core degree, the more important the user, then the greater the probability that he is the core member of a community.

2) Definition two: edge weight

Attractiveness between users is looked as the edge weight.

If q is the number of users that both a and b are mutual-concern with.

a) When a and b are interested in each other, then the attractiveness between them is:

$$S_{ab} = q \times (W_a + W_b) \tag{11}$$

b) When a is interested in b, but b is not interested in a, then the attractiveness between them is:

$$S_{ab} = q \times (W_a - W_b) \tag{12}$$

c) When b is interested in a, but a is not interested in b, then the attractiveness between them is:

$$S_{ab} = q \times (W_b - W_a) \tag{13}$$

4.3. The Experimental Results

We tested both ABCD algorithm and CNM (Clauset-Newman-Moore) algorithm[12] on this data set. The CNM algorithm we used is from SNAP (Stanford Network Analysis Platform)[17], SNAP is a social network analysis toolkit which developed by Stanford University, this algorithm is an implementation of CNM algorithm proposed by Clauset, Newman and Moore et al. The results of these two algorithms are shown in Table 1.

Table 1. The clustering results of ABCD alg. and CNM alg.

the number of members	detected by ABCD Algorithm	detected by CNM Algorithm
> 50	166 communities	32 communities
> 100	34 communities	28 communities
> 400	0 communities	13 communities
> 1000	0 communities	6 communities

From the result, we find that the communities our algorithm identifies is much smaller than CNM algorithm, the communities CNM identifies is too big, some of them have members more than 1000, even close to 10000, so the communities CNM algorithm identifies are usually composed of several real-world communities, but the communities our ABCD algorithm identifies are basically consistent with the real world.

Table 2. The comparing of ABCD alg. and CNM alg.

CNM alg.	one community with 1018 members	considered only the relationship between users
ABCD alg	13 communities with more than 20 members(total 743 users), 27 communities with members more than 10 and less than 20, 9 communities with members less than 10	considered more information of users

Fig. 3 (a) shows one community with 1018 members which CNM algorithm identified. We think it's too big. With these users, ABCD algorithm could get better result, shown in Fig. 3 (b) and table 2, the circles have the same color indicate the members in the same community. The bigger the circle in one clique, the more members link to it, which means the member it denotes is more important in the community. The wider the edge between two circles, the closer the relationship between the two members. The figures are drawing by Protovis[18], which draws users and the number of inter-interested users as weight of connection between users.

Comparing the two figures, we can clearly see that the validity of our algorithm is much better than CNM algorithm. The reason is that CNM algorithm considers only the bi-connections between users, and ABCD algorithm considers more information, such as the number of interested users, the number of fans, the number of inter-interested users.

Fig. 4 shows the community size distribution diagram of the clustering results of ABCD algorithm.

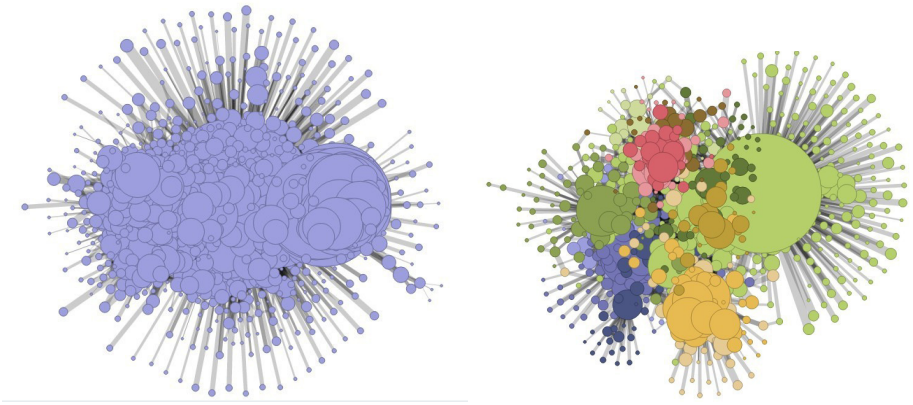

Fig. 3. (a) A community that CNM algorithm identified; (b) communities that ABCD algorithm identified.

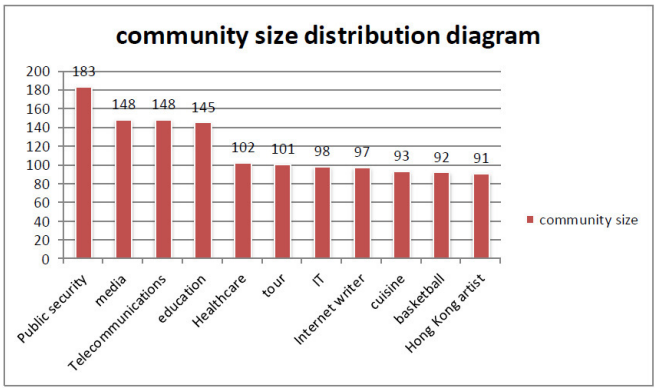

Fig. 4. community size distribution diagram

From the point of time efficiency, the two algorithms almost the same, but the memory usage of CNM algorithm is higher than ABCD algorithm. With the same computer, which has 4G RAM, when process the data set, ABCD algorithm can be successfully finished, however, CNM algorithm would run out of memory.

5. Community Detection of College Football teams

College Football data set[6] represents a regular season of the U.S. college football game in 2000. The node in the network represents a team, and the link represents the game between two teams. The community structure of the network is known: all the teams were divided into 12 conferences, and each conference owns different number of teams. Games are more frequent between the teams in the same conference.

The data set is widely used to test the effectiveness by many unweighted community detection algorithms, here we should define the node weight and edge weight for the network at first.

1) Definition one: node weight

Each team is a node in the weighted graph of the teams relationship network. The node weight is the team's core degree in the network, but the teams are thought have the same importance with each other, so the weights are always assigned with a same value.

2) Definition two: edge weight

Attractiveness between teams is looked as the edge weight.

If q is the number of teams who competed with of both a and b, F_a is the competed teams number of a, and F_b is the competed teams number of b, then the attractiveness between a and b is:

$$S_{ab} = q \times (\frac{1}{F_a} + \frac{1}{F_b}) \tag{14}$$

With these definitions of weights, the relationships of football teams become a weighted network, we would like to detect communities with ABCD algorithm. The results are shown in Fig. 5 and different colors represent the community structure that our method detects.

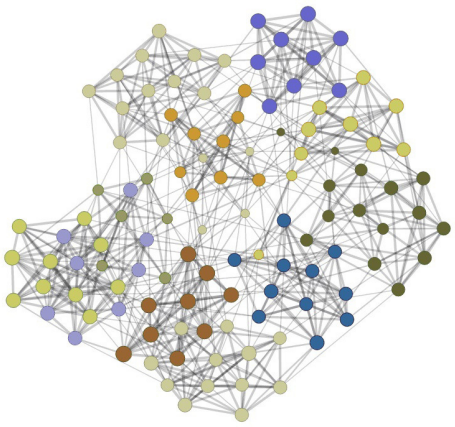

Fig. 5. Communities of football teams identified by ABCD algorithm

There are 11 communities identified by ABCD algorithm, because ABCD algorithm did not specify the number of communities before detecting and we suppose we did not know the number of communities before detecting. We can find that the communities are almost consistent with the original conference from Table 3.

Table 3. The clustering results of ABCD alg. on the teams

community number	teams detected by ABCD Algorithm(the original community number)
1	0(8),4(8),9(8),16(8),23(8),41(8),93(8),104(8)
2	1(1),25(1),33(1),37(1),45(1),89(1),103(1),105(1),109(1)
3	2(3),6(3),13(3),15(3),32(3),39(3),47(3),60(3),64(3), 100(3),106(3)
4	3(4),5(4),10(4),40(4),52(4),72(4),74(4),81(4),84(4), 97(11),98(4),102(4),107(4)
5	7(9),8(9),21(9),22(9),51(9),68(9),77(9),78(9),108(9),111(9)
6	11(11),24(11),28(12),50(11),69(11),90(6)
7	12(7),14(7),18(7),26(7),31(7),34(7),36(6),38(7),42(6), 43(7),54(7),61(7),71(7),85(7),99(7)
8	17(10),20(10),27(10),56(10),58(12),59(11),62(10),63(11), 65(10),70(10),76(10),87(10), 95(10),96(10),113(10)
9	19(2),29(2),30(2),35(2),55(2),79(2),80(6),82(6),94(2),101(2)
10	44(5),48(5),57(5),66(5),75(5),86(5),91(5),92(5),112(5)
11	46(12),49(12),53(12),67(12),73(12),83(12),88(12), 110(5),114(12)

6. Community Detection of Renren

6.1. The Characteristics of Renren Users

Renren is a social network service web site, which can provide the functions like FaceBook.Renren.com is founded in 2005, with a claimed 170 million registered users [19], Renren is the largest online social network in China. Different

from micro-blog systems, a mutual friendship between two users on Renren is built if and only if one sends a request and the other approves the request.

In order to identify the communities of Renren with the ABCD algorithm, we should define the node weight and edge weight for the network at first.

1) Definition one: node weight

Each Renren user is a node in the weighted graph of the users relationship network. The node weight is the user's core degree in the network, but the registered users on Renren are all real name users, they are thought have the same importance with each other, so the weights are always assigned with a same value.

2) Definition two: edge weight

Attractiveness between users is looked as the edge weight.

If q is the number of users who are friends of both a and b, F_a is the friend number of a, and F_b is the friend number of b, then the attractiveness between a and b is:

$$S_{ab} = q \times (\frac{1}{F_a} + \frac{1}{F_b}) \tag{15}$$

With these definitions of weights, the relationships of Renren users become a weighted network. The definitions are according as the weight definitions of Collage Football Teams network, it's reasonable. What's more, we checked a part of our experiment results, the identified communities are match with the ground trues.

In order to verify the validity of our algorithms, we tested them on the data set got from Renren about BUPT. We collected Renren users information started with several BUPT public users, from the friendship we can get more users related with BUPT, we crawled with breadth first search and considered only three levels. The data set contains 86 thousand users and 4.8 million connections(friend relationship) among these users.

Table 4. The clustering results of ABCD alg. on Renren

the number of members	detected by ABCD Algorithm
> 50	172 communities
> 100	75 communities
> 200	23 communities
> 400	1 communities
> 1000	1 communities

The clustering results of ABCD algorithm on Renren data set are shown in Table 4. From the result, we find that the big communities with more than 400 members are only two. They are not actual communities, the reason is just the missing of friendship according the privacy protection of Renren users. There are many small actual communities identified. Fig. 6 shows some of the communities identified by ABCD algorithm. The figures are drawing by Protovis[18].

7. Conclusion

For weighted graph clustering, we propose an attractiveness-based community detection algorithm. It is an amalgamation algorithm, the merge between clusters could be considered while the attractiveness of clusters (as the edge weight) is bigger than the densities of clusters (as the node weight). ABCD algorithm is designed to make some breakthrough on the time complexity of community detection for large social networks. The algorithm does not require to specify the number of clusters, because the number is usually not known in advance and is difficult to estimate in actual applications. Three datasets are used to test the effectiveness and reliability of the algorithm. For large social network, how to combine with user profiles improving the algorithm and enhance the performance is one of our works in the future.

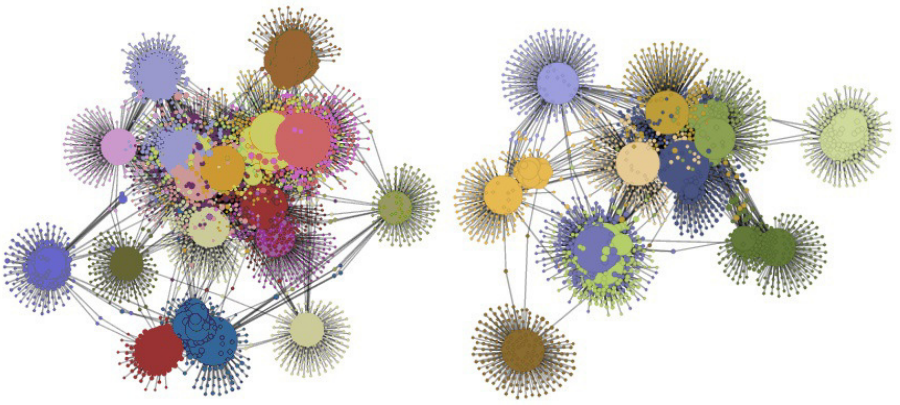

Fig. 6. (a) Renren Communities with members from 150 to 200; (b) Renren Communities with members from 200 to 250.

Acknowledgements

This work was supported by National Grand Fundamental Research 973 Program of China under Grant No.2013CB329606; National Natural Science Foundation of China under Grant No.91124002; National Science and Technology Support Program of China under Grant No.2013BAH43F00-01; Chinese Universities Scientific Fund(BUPT2014RC0701).

References

1. A.Broder, R.Kumar, F.Maghoul, P.Raghavan, S.Rajagopalan, R.Stata, A.Tomkins, and J.Wiener. ”Graph structure in the web,” In Proc. of the Ninth International Conference on the World Wide Web, 2003:15-19.
2. M.E.J.Newman. The structure of scientific collaboration networks. *PNAS*,2001,98:404-409.
3. Malik Magdon-Ismail, Jonathan Purnell. ”SSDE-cluster: fast overlapping clustering of networks using sampled spectral distance embedding and GMMs,” 2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing.
4. J. Ruan and W. Zhang. ”An efficient spectral algorithm for network community discovery and its applications to biological and social networks,” Proceedings of the 2007 Seventh IEEE International Conference on Data ining,ICDM-07, Omaha, NE, USA, Oct. 28-31, 2007.
5. Partha Basuchowdhuri and Jianhua Cheny. ”Detecting communities using social ties,” 2010 IEEE International Conference on Granular Computing.
6. M. Girvan and M.E.J. Newman. ”Community structure in social and biological networks,” Proceedings of the National Academy of Sciences of the United States of America, 2002.99(12), P.7821.
7. M.E.J. Newman. Fast algorithm for detecting community structure in networks. *Physical Review E*, 2004.69(6), P.066133.
8. B.Cai, H.Wang, H.Zheng, H.Wang. ”Evaluation repeated random walks in community detection of social networks,” Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao, 11-14 July 2010.
9. P. Pons and M. Latapy. Computing communities in large networks using random walks. *J. Graph Algorithms Appl.*, vol. 10, no. 2, pp. 191C218, 2006.
10. Jie Chen and Yousef Saad. Dense subgraph extraction with application to community detection. *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, NO. 7, JULY 2012. pp1216-1230.
11. S. White and P. Smyth. ”A spectral clustering approach to finding communities in graphs,” Proceedings of the 5th SIAM International Conference on Data Mining, Philadelphia, USA, 2005:76-84.
12. A. Clauset, M.E.J. Newman and C. Moore. Finding community structure in very large networks. *Physical Review E*, 2004.70(6), P.066111.
13. Eva Jaho, Merkouris Karaliopoulos and Ioannis Stavrakakis. ”ISCoDe: a framework for interest similarity-based community detection in social networks,” 2011 IEEE conference on Computer communications workshops(INFOCOM WKSHPS).
14. Slah Alsaleh, Richi Nayak and Yue Xu, ”Finding and matching communities in social networks using data mining,” 2011 International Conference on Advances in Social Networks Analysis and Mining.
15. Kumar Subramani, Alexander Velkov, Irene Ntoutsi, Peer Kroger, Hans-Peter Kriegel. ”Density-based community detection in social networks,” 2011 IEEE 5th International Conference on Internet Multimedia Systems Architecture and Application, IMSAA2011.
16. Yang Yang, Yizhou Suny, Saurav Pandit, Nitesh V. Chawla and Jiawei Han. ”Is objective function the silver bullet?” 2011 International Conference on Advances in Social Networks Analysis and Mining.
17. Stanford Network Analysis Project. http://snap.stanford.edu/, 2012.
18. Protovis: A graphical approach to visualization. http://mbostock.github.com/protovis/, 2012.
19. Jiali Lin, Zhenyu Li, Dong Wang, Kave Salamatian and Gaogang Xie, ”Analysis and comparison of interaction patterns in online social network and social media,” IEEE International Conference on Computer Communications and Networks, 2012