Role Assorted Community Discovery for Weighted Networks

Ruixin Ma
School of software, Dalian University of Technology, Dalian, China
Email: teacher_mrx@126.com

Guishi Deng and Xiao Wang
School of Management Science and Engineering, Dalian University of Technology, Dalian, China
Email: Denggs@dlut.edu.cn, kara0807@126.com

Abstract—**This paper considers the difficulties in community discovery, and comes up with a community discovery algorithm on the basis of role assorted thoughts. Previous work indicates that a robust approach to community detection is the maximization of inner communication and the minimization of the in-out interaction. Here we show that this problem can be solved accords to the role assorted method which give distinguish labels to vertices in the same community. This method leads us to a number of possible algorithms for detecting community structures in both unweighted and weighted networks. The applicability and expandability of algorithms proposed are illustrated with application to a variety of computer-generated networks and real-world complex networks.**

Index Terms—**community discovery, role assorted thoughts, robust approach, distinguish label**

I. INTRODUCTION

Networks have attracted considerable recent attention in biology and other fields as foundation for the mathematical presentation of a variety of complex systems. [1-4]. It seems that all complex systems can be abstracted as networks which are combined by all kinds of interacting units. Therefore, network provides a totally novel, intuitional method for complex system research. It is of great value to find community structures in complex network. E.g. communities in social network are used to reveal the groups of users have similar interests, habits and background[5-7]. Community discovery in citation system helps readers to quickly find the papers they want. Communities in biochemical networks help researchers to find the functional related units[8-11], to name but a few.

The research of community discovery is first proposed by Girvan and Newman[12,13]. They put forward some classic CDAs such as G-N and fast G-N, besides, Newman also proposed a method to measure the results of CDAs which was called modularity. As we all know, modularity is indeed a good way to test whether the discovery results is good or not, however, it is limited in unweighted networks while not suit for weighted networks. Especially now that the development of SNS[14] has become an inevitable tendency all over the world, CDA should try its best to work for good social network services. In that case, some scholars are not content with singly find community structures in the entire network, but prefer to explore the structure inside a community and refine the community division results.

Advertising practitioner Fang Shouxing said that there is a "law of special" in the procedure of website promotion. That is to say, there are three kinds of people play very important roles during the course of information dissemination. First, authorities with great stores of knowledge in one or more frontiers. For websites and application platforms, lots of successful founders on the internet are either authorities or the contact of authorities, this kind of people have special perspectives to convene experts. In other words, they are the leaders in their communities. Second, liaisons, this kind of people have talents in social interactions. Generally speaking, liaisons keep in touch with different communities at the same time. Six degrees of separation tells us that every two people in the world can reach each other within six degrees. However, it is not means any one in the world is able to reach the others but only some special ones who have access to different communities can. Most people keep in touch with world by these few special ones. Third, recommender, they take charge of "the last mile" which means they persuade people to buy the things they recommend. Most recommenders are just ordinary users, they may far away from the authoritative experts and the smooth liaisons, but they can persuade our target user to accept their idea. Whether information can be propagated like virus or not depends on how many powerful recommenders are working for you. Make every user be your recommender maybe every website's dream. Recommenders on SNS may be the target user's most trusted people, or their nearest neighbors.

The above mentioned three kinds of people are often called the opinion leaders on the website. Most users take leaders' opinions as their references to browse pages or to

doi:10.4304/jsw.6.12.2441-2448

buy goods while they do not sure what they want from the Internet.

As far as our goals in this paper are concerned, a very useful method is that taken by social network analysis with the set of knowledge known as collaborative filtering[15-19] and the theory of space vector model[20]. This knowledge is aimed at mining natural structures in social networks, based on both the adjacent relationship and the strength of connection between vertices which is called role assorted community discovery algorithm. RACDA is a kind of agglomerative algorithm who focuses on the addition of vertices to different communities. In this agglomerative method, similarities are calculated by one method or another between vertices and communities.

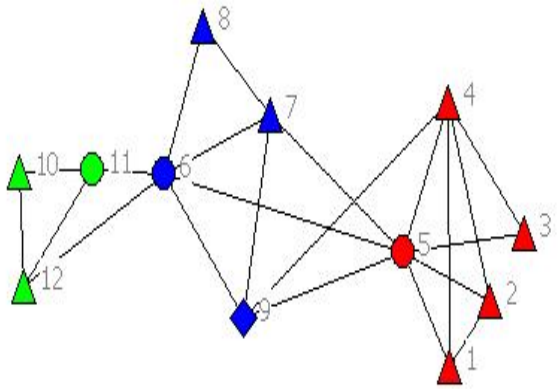

Figure 1. A small network with community structure of the type considered in this paper.

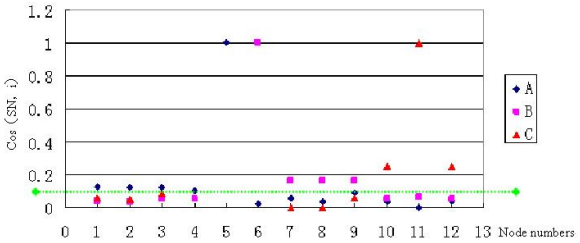

Figure 2. The diagram illustrates the type of output generated by the our algorithms.

Frankly speaking, agglomerative methods have some problems on the procedure of finding the right community structure. One concern is that they fail with some frequency to find the correct communities in networks where the community structure is known, which makes it difficult to place much trust in them in other cases. Another is their tendency to find only the cores of communities and leave out the periphery[21]. Besides, in cases those vertices only have a single link to a specific community, agglomerative methods often fail to place such vertices correctly.

However, in this paper, we come up with a novel agglomerative CDA which based on the theory of space vector model. This method has been relatively little studied in the previous literature, either in social network theory or elsewhere. Here we introduce the concept of community seed, community eigenvector and liaisons. In this agglomerative method, we start with finding the community seed and use community eigenvectors to represent the adjacent circumstances in each community, take figure 1 as example; community seeds 5, 6 and 11 are primarily found. By calculating the similarity between free vertices and communities repeatedly, we get the entire network's natural structures and the illustration diagram for the communities we get as figure 2 shows.

The approach we take follows roughly these lines, but adopts a some what different heuristic viewpoint. Rather than starting with randomly choosing vertices in the network, we make a list of ranking nodes in decreasing order of fitness, which makes sure the priority of vertices with better position. We pay much more attention to the similarity between vertices and communities instead of similarities in node pairs. Besides, by doing some marks to the vertices which are divided into the same communities, we can also find the liaisons among different communities. How this idea works out in practice will become clear in the course of the presentation.

Briefly then, the outline of this paper is as follows. In section II we describe the crucial concepts behind our algorithm and show how these concepts works in the implementation of our method. In section III, we describe the detailed implementation of our algorithm. In section IV, we give a number of applications of our algorithms to particular networks. At the end of this paper, we give a conclusion.

II. COMMUNITY DETECTING FOR SOCIAL NETWORK SERVICE

In this paper, we present a new algorithm for network clustering. Our discussion focuses primarily on networks with a type of weighted edges while generalization to less complicated network types is certainly possible.

There are three central features that distinguish our algorithms from those that have preceded before. First, our algorithms are aimed at finding three kinds of special people on the network and use the influence of the special to find community structures behind the network. Second, we introduce the application of space vector model in community discovery. Third, characteristics of community are represented by community eigenvectors which is totally different from the metric's eigenvector.

To make things more concrete, we use example to illustrate how the space vector model works in this algorithm.

Both the adjacent relation and the intimacy relation are c represented by vectors. Before further explanation, we firstly define that the adjacent matrix of network is *Matrix* and the intimacy relation matrix of network is *R-Matrix*. Each row in adjacent matrix shows a vertex's adjacent relation with others while each row in *R-Matrix* shows the intimacy degree of each pair of vertices, which is showed as $V_i =< e_{i,1}, e_{i,2} \ldots e_{i,n} >$. The intimacy relation vector of communities in weighted networks can be

calculated as formula (1). Values on the edge are added to the community eigenvector to show the intimacy between vertex and the existing community. Both the adjacent topology and the intimacy between vertices become the factors to judge the adscription of vertices.

foreach vertex i in SN

$$V_{SN} = V_{SN} + V_i < e_{i,1}, e_{i,2}, \ldots e_{i,n} > \quad (1)$$

The totting-up of relation during the procedure of structure division makes the close-knit vertices come closer while expand the gaps between different communities.

This measure is only a suggestion; many others are possible and may well be appropriate for specific applications.

Another way in which our method differ from previous ones is in the procedure of doing distinguish marks for vertices with different characteristics, which means network managers are able to provide different levels of protection for vertices in the same community. We put vertices into three classes in the procedure of community detection: community seed, liaison and direct recommender.

Definition 1: Community seed. seeds are the initial users in a community; it must be trust worthy and also have a large number of neighbors.

Definition 2: Liaison. They take charge of information dissemination among different communities. Liaisons belong to the overlapped area of different communities and they help people from different groups to communicate and share messages.

Thus the general form of our community structure discovery algorithm is as follows:

(1) Construct the adjacent map for network;

(2) Add intimacy value to each edge accords to the intimacy between vertices pairs and save this intimacy relation matrix;

(3) Calculate the fitness value for each vertex and build the fitness list;

(4) Find the community seeds, liaisons according to the chosen principle bellowed and divide the whole network.

In fact, it appears that the last two steps are the most important features of our algorithm as far as getting satisfactory results are concerned. As our studies mentioned above, once the community seed and the liaisons are found, the exact community structures will be very clear.

Seeds in a weighted network do not only need a large number of adjacent neighbors but also need to be trusted. We suppose that there is a network N, A_N and T_N are separately N's adjacent matrix and intimacy matrix, in that way, we use formula (2) to measure vertex i's average value of being trusted.

$$C_i = \frac{\sum_{j=1}^{n} T_{i,j}}{\sum_{j=1}^{n} A_{i,j}} \quad (2)$$

To better weigh the vertices' position in the network, in other words, to set a criterion to rank the vertices, we set a fitness function.

$$F_i = \alpha \times \sum_{j=1}^{n} A_{i,j} + \beta \times C_i \quad (3)$$

As we can see, fitness value is affected both by the number of adjacent neighbors and the average intimacy value. α and β decide the weight proportion of adjacent neighbors and the intimacy value. To distinguish the community seed and the liaison, we also set a minimum intimacy threshold ε which will be showed how it works in the following part.

Community seeds and liaisons are chosen as below shows.

All vertices are sorted in decreasing order of fitness which constitutes a list L_{fit}. Community seeds set *S* and liaison set L are both initialized to empty. Vertices in L_{fit} are checked in turn from the beginning to the end of the list. Calculate the similarity between free vertex *i* and the existing community *SN* , if similarity*(i, SN)*< δ and C_i > ε , it becomes a new seed and is added to *S*; else if C_i < ε , i becomes a liaison and is put into L; if there are many existing communities, and only similarity(SN,i)> δ, *i* becomes a member of *SN*, else if both similarity(SN1,i) > δ and similarity(SN2,i) > δ, we mark i as liaison and put i into the community with which it has larger similarity; Iteratively calculate the similarities until the end of the list. At the end, take out of the vertices in L and put them into suited communities.

The similarity between community *SN* and vertex *i* is calculated as formula (4) which is enlightened by CF.

$$Sim(SN,i)=cos(SN,i)=\frac{V_S \cdot V_i}{\sqrt{V_S^2}\sqrt{V_i^2}} \quad (4)$$

We can also use the Pearson coefficient to measure the relation similarity between free vertices and existing communities. For the simple networks we studied in this paper, we set the minimum similarity threshold as 1/n(n is the size of network). The flow chart of our algorithm is as figure 3 shows.

In the next part of our paper, we demonstrate the efficiency and expandability of our algorithm with a number of examples and show that our algorithm can be reliably and sensitively extract community structures from both artificially generated and real-world networks with known community structures. Besides, we also prove how our algorithm can be used to analyze networks whose structure is otherwise difficult to comprehend. The networks studied include a collaboration network of scientists, in which our method allows us to generate similarity comparison diagrams of the whole network.

III. IMPLEMENTATION

In theory, the description of the preceding section completely define the methods we consider in this paper, but in practice there are a number of subtleties to their

implementation that are important for turning the description into a workable computer algorithm.

```
Begin RACDA
    Initialize;
    Take out the first member in Lfit and put it into S;
    for i=1:n
        for j=1:k
            Calculate the Similarity value between vertex i and community SNj;
            If similarity(i,SNj) is higher than the minimum similarity threshold
                Calculate the trust value of vertex i;
                If Ci is higher than the minimum intimacy value
                    i becomes a seed;
                else i becomes a liaison
            else i becomes a member of community SNj
        end
    end
End RACDA
```

Figure 3. The detailed procedure of RACDA.

Essentially most of the work in this algorithm is in the calculation of the similarity between free vertices and the existing communities; the job of choosing appropriate α and β for fitness function becomes the most pressing matter of the moment for RACDA. Figure 3 uses pseudocode to show the general flow of our algorithm.

A. Undirected, weighted networks

This algorithm dynamically set the number of communities in a complex network accords to the community seeds. To illustrate the fundamental principles and running results of our algorithm, we constructed a small weighted network *N* like Figure 4 shows. The adjacent matrix and trust matrix of this network are showed below. From Figure 4 we can see that each edge has a weight to label the intimacy between vertices.

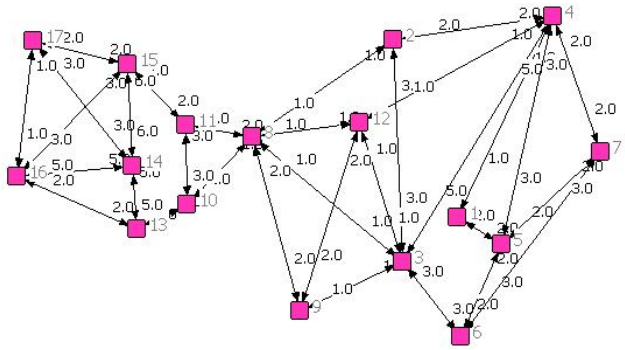

Figure 4. A weighted network

During the course of study we find that most scholars pay much attention to the degree of nodes which results in the misleading of communities. For example, vertex 8 will becomes a member of the left side community and plays a core role in it if we don't think about the weight on edges. However, we find that the values on edges that contact vertex 8 to others is very small which means vertex 8 is not very trust worthy. In other words, vertex 8 is a liaison instead of a seed. Here we show the adjacent matrix and the intimacy matrix of N, and present the results of RACDA. We use A_N to represent the adjacent matrix of Figure 4, T_N represent the trust matrix.

$A_N =$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0
3	0	1	1	1	0	1	0	1	1	0	0	1	0	0	0	0	0
4	1	1	1	1	1	0	1	0	0	0	0	1	0	0	0	0	0
5	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
6	0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0
7	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0
8	0	1	1	0	0	0	0	1	1	1	1	1	0	0	0	0	0
9	0	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0
10	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0
11	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0
12	0	0	1	1	0	0	0	1	1	0	0	1	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	1	0
14	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1
15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
16	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
17	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1

$T_N =$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	3	2	0	0	0	1	0	0	0	0	0	0	0	0	0
3	0	3	0	5	0	3	0	1	1	0	0	1	0	0	0	0	0
4	1	2	5	0	3	0	2	0	0	0	0	1	0	0	0	0	0
5	2	0	0	3	0	2	2	0	0	0	0	0	0	0	0	0	0
6	0	0	3	0	2	0	3	0	0	0	0	0	0	0	0	0	0
7	0	0	0	2	2	3	0	0	0	0	0	0	0	0	0	0	0
8	0	1	1	0	0	0	0	0	2	1	2	1	0	0	0	0	0
9	0	0	1	0	0	0	0	2	0	0	0	2	0	0	0	0	0
10	0	0	0	0	0	0	0	1	0	0	3	0	5	0	0	0	0
11	0	0	0	0	0	0	0	2	0	3	0	0	0	0	0	0	0
12	0	0	1	1	0	0	0	1	2	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	5	0	0	0	5	0	2	0
14	0	0	0	0	0	0	0	0	0	0	0	0	5	0	6	5	3
15	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	3	2
16	0	0	0	0	0	0	0	0	0	0	0	0	2	5	3	0	1
17	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	1	0

We set α =0.6 and β =0.4, the fitness table of Figure 4 is as Table I shows.

TABLE I.
THE FITNESS OF VERTICES IN NETWORK *N*

ID	F	ID	F	ID	F
1	4.4668	7	5.2	13	5.32
2	4.7	8	5.6284	14	5.88
3	5.9716	9	4.6	15	5.4
4	5.9716	10	5	16	5.24
5	5.08	11	4.8	17	4.7
6	4.9	12	4.76		

Formula (5) is used to set the minimum trust threshold.

$$\varepsilon = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} T_{i,j}}{2\times|E|} \qquad (5)$$

Formula (5) tells us that threshold ε is the average value of weight on each edge, for network *N*, ε = 2.45 and the minimum similarity threshold is 1/17. We use ○, △ and ◇ to separately represent the community seed, ordinary user and liaisons in a community. Figure 5

shows the dividing results of Figure 4 and Figure 6 is the output of RACDA.

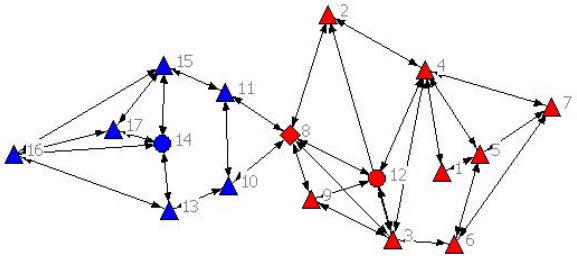

Figure 5. Dividing results of figure 4.

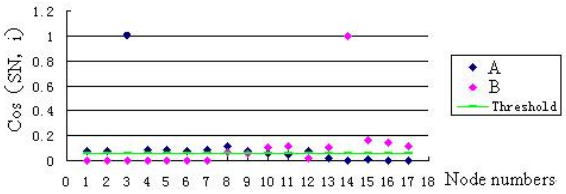

Figure 6 . The output of RACDA

The topology structure of Figure 5 is too simple to demonstrate the effectively of RACDA. It has no single vertices on the margin of different communities. Here we use another real-world network who has a little more complex relationship, the *Les Miserables* dataset. Using the list of character appearances by scene complied by Knuth[22], the network was constructed in the situation that the vertices represent characters and the edge between two vertices represent coappearance of the corresponding characters in the same one or more scenes, the value on edges represent the times of their coappearance in the same scene. Generally, there are six most important people on the network: Jean Valjean(vertex 11), detective Javert(vertex 27), father Bishop Myriel(vertex 0), grisette Fantine(vertex 23) and her daughter Cosette(vertex 26). 77 vertices and 508 edges are showed in the network altogether like Figure 7 shows.

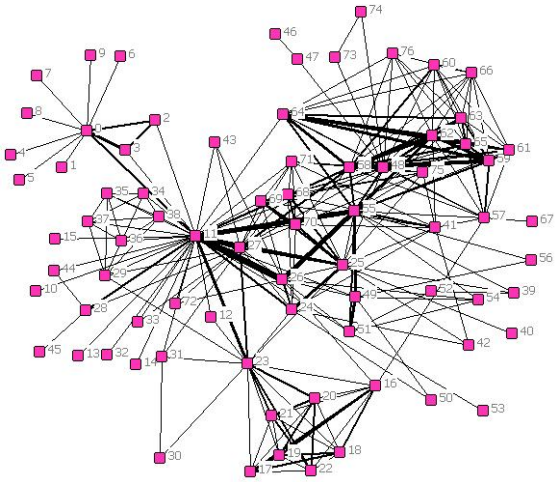

Figure 7 . The topology of Les Miserables.

From Figure 7 we can easily get the central vertices in *Les Miserables*. Compared vertex 43 and 72 we can see that their roles are different, and the only difference between 43 and 72 is that the intimacy between 43 and 11 are much stronger than the intimacy between 72 and 11 while the intimacy between 72 and 26 are stronger than 43 and 26. Figure 8 shows the dividing result of G-N to *Les Miserables*. Figure 9 illustrates the differences between community seeds and liaisons, although vertex 27 has a lot of neighbors, the intimacy between 27 and its neighbors is rather distant which limits 27 to a liaison rather a seed. Another difference between figure 9 and 10 is the situation of vertex 16 and 57, in figure 10 we can see that the relationship in communities is much closer than the relationship between different communities.

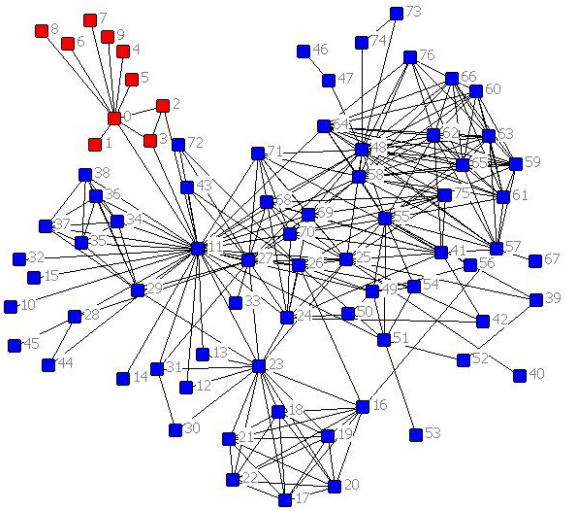

Figure 8 . The results of G-N to Les Miserables.

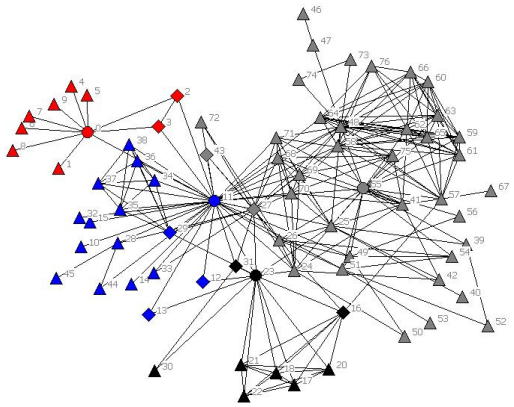

Figure 9 . Dividing results of RACDA to unweighted Les Miserables

Here we set α =0.6 and β =0.4 to calculate the fitness for vertices in *Les Miserables*. Figure 9 is the dividing result of unweighted *Les* while Figure 10 represents the dividing result to weighted *Les Miserables*.

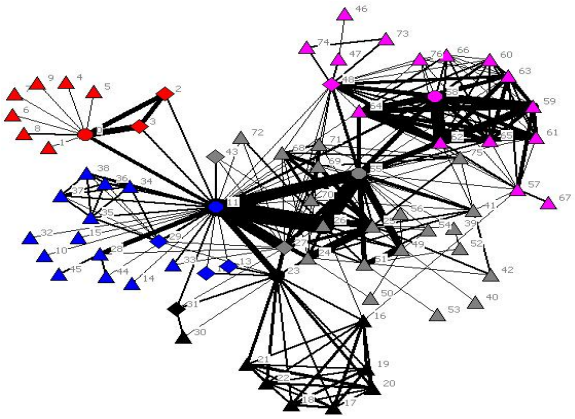

Figure 10 . The dividing result of RACDA to weighted Les Miserables.

On the procedure of clustering, we use intimacy relation-weighted community eigenvectors and get exciting results. If we do not add intimacy values to edge, some vertices who located at the margin area of different communities will be divided into wrong groups. Such as vertex 2 and 3, because vertex 11 is much more central than vertex 0 and it also has much more sophisticated relationships, which result in V_3 and V_2 fall into the community that vertex 11 in. In that way, vertices 2 and 3 were divided into the wrong community. However, if we use intimacy weighted eigenvectors, the denominator of similarity between vertex 3 and community A(vertex 11 lead) becomes very small while the denominator of similarity between 3 and community B(vertex 0 lead) stays almost the same. Taking both the adjacent relationship and the intimacy between pairs into consideration, we get exactly what we want.

B. Undirected,unweighted networks

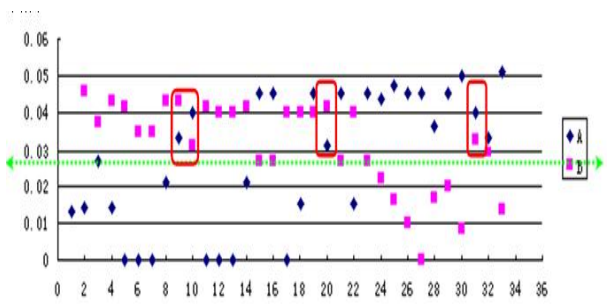

Figure 11 . Output of RACDA to Zachary club[23].

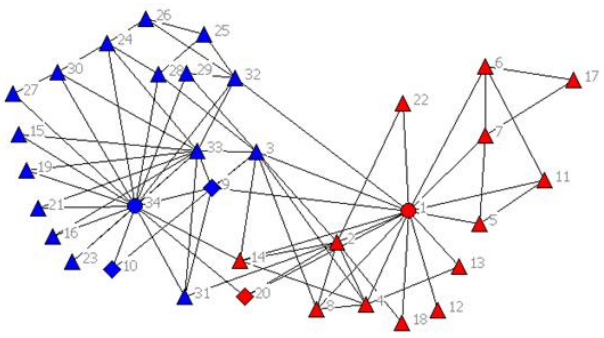

Figure 12 . Dividing result of RACDA to Zachary Club

Figure 11 is the output of RACDA to Zachary Club and Figure 12 is the dividing result. There is no weight of intimacy in unweighted networks, so we set $\alpha=1$ and $\beta=0$. However, we can see from Figure 11 that vertex 10 actually community A's inner member. The reason for this mistake is that we calculate the similarity only on the basis of whether the edge exist or not while lacking weight for edges to measure which edge plays more important roles on the network and which edge should be given a prior consideration. During some following experiments we prove that this mistake can be avoid if we adjust the community eigenvector.

IV. COMPARISON

Here we show the comparison results of modularity on different real-networks of various kinds of community discovery algorithms.

TABLE II.
THE COMPARISON OF DIVISION RESULTS OF DIFFERENT CDAS

Name	Football	Karate Club	Dolphins	Les Miserables
GN	0.377	0.395	0.381	0.082
Fast G-N	0.376	0.360	0.379	0.074
Polish	0.327	0.358	0.226	0.107
RACDA	0.484	0.401	0.379	0.486

From Table II we know that RACDA gets the highest modularity on all kinds of social networks, especially on complex networks. Fast G-N has smaller complexity while sacrifices the division accuracy.

The results of RACDA showed above use unweighted edges in the network which means the fitness function is as formula (6) shows.

$$F_i = \sum_{j=1}^{n} A_{i,j} \tag{6}$$

Under this situation, the degree of vertices becomes the only element to measure the centrality of vertices.

The primary remaining difficulty with our algorithm is how to use modularity to test the quality of our algorithm. For weighted networks, link-in and link-out numbers are not the only standard to judge the cohesion level of a community, but also the weight of edges are also needed to be taken into consideration, especially edges between different communities. Besides, the methods to define the minimum similarity threshold and the minimum trust threshold need to be optimized as well. Most networks with complicate topologies and indistinct levels have their own characteristics and we should set special values for them. However, a better approach would be to find some improvement in the algorithm itself to optimize the results of clustering.

V. CONCLUSION

In this paper, we describe a new algorithm for social community discovery, the task of extracting the natural

community structures from networks of vertices and edges. This is a problem long studied in computer science, applied mathematics, physics, and the social science. Especially with the development of SNS, community discovery is playing a more and more important role in E-commence and SNS websites. Whether from the viewpoint of social network service or network security, community discovery would become a hot-topic for the next development of complex adaptive system. We believe the methods described here give a good solution to community discovery. Our algorithm is defined by two crucial features. First, we use three labels to distinguish the users in the same community which provides a promise for personalized recommendation. Besides, the differentiation also makes it possible to give special protections for different users which enhance the robustness of the network. Second, our methods include a recalculation step until the vertices are all put into the right communities. This step, which determines the ascription of vertices, turns out to be of primary importance to the success of our algorithm. Without the similarity recalculation, this algorithm fails miserably at even the simplest clustering tasks.

REFERENCES

[1] M. E. J. Newman. "Finding Community Structure in Networks Using the Eigenvectors of Matrices." Physics. 2006, 1–22.

[2] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Complex networks: "Structure and Dynamics." Physics Reports 424, 2006, 175–308.

[3] Leicht E A, Newman M E J. "Community Structure in Directed Networks." Physical Review Letters, 2008, 100:118703.

[4] S.N.Dorogovtsev and J.F.F.Mends, "Evolution of Networks: From biological Nets to the Internet and WWW". Oxford University Press, Oxford, 2003.

[5] G.W. Flake, S.R.Lawrence, C.L.Giles, and F.M.Coetzec, IEEE computer 35, 2002, 66.

[6] Rosvall M, Bergstrom C T. "An Information-theoretic Framework for Resolving Community Structure in Complex Networks." PNAS, 2007, 104(18): 7327-7331.

[7] A.Broder, R.Kumar, F.Maghoul, P.Raghavan, S.Rajagopalan, R.stata, A.Tomkins and J.Wiener, Computer. Netw. 2000, 33,309.

[8] B. Hu, X.Y. Jiang, J.F. Ding, Y.B. Xie, B.H. Wang. "A Weighted Network Model for Interpersonal Relationship Evolution." Physica A. 2005, 353:576-594.

[9] C.Moore and M.E.J.Newman, "Epidemics and Percolation in Small-world Networks." Phys. 2000, Rev. E61, 5678-5682

[10] LIU Ji, DENG Gui- shi. "A Collaborative Recommendation Method Based on User Network Community with Weighted Spectral Analysis". Journal of Dalian University of Technology. 2010, 50(3): 438-442.

[11] Cun-rui， Xiao-dong，LIU Xiang-dong，LI Zhi-jie. "A Physical Community Discovery Algorithm." MICROELECTRONICS&COMPUTER. 2010. 27(9): 33-36.

[12] M.E.J. Newman. "Scientific Collaboration Networks: II. Shortest paths, Weighted Networks, and Centrality." Phys. 2004, Rev. E64, 016132.

[13] M.E.J.Newman. "Mixing Patterns in Networks." Phys. 2003, Rev. E67, 026126.

[14] "Research of the Present Situation of Operation and the Future Tendency of SNS Website." http://media.people.com.cn/GB/22114/119489/140165/8454258.html（2009）

[15] LUO Ze-bi, XIE Qink-sheng. "Collaborative Filtering Recommendation Algorithm Based on Web Data Mining." Journal of Guizhou University(Natural Science). 2009, 26:40-43

[16] Guo Yan-hong, DENG Gui-shi, LUO Chun-yu. "Collaborative Filtering Recommendation Algorithm Based on Factor of Trust." Computer Engineering, 2008, 34(20): 1-3

[17] DAI Ya-e, GONG Song-jie. "Collaborative Filtering Recommendation Based on Fuzzy Clustering in Personalization Services." Computer Engineering & Science, 2009, 31(4): 110-116

[18] Ashish Sureka and Pranav Prabhakar Mirajkar. "An Empirical Study on the Effect of Different Similarity Measures on User-Based Collaborative Filtering Algorithms." Springer-Verlag Berlin. 2008, LNAI 5351: 1065–1070.

[19] Ahn, H.J. "A New Similarity Measure for Collaborative Filtering to Alleviate the New User Cold-starting Problem." Information Sciences: an International Journal. 2008, January.

[20] FAN Cong-xian XU Ting-rong FAN Qiang-xian. "Research and Improved Algorithm of HITS Based on Web Structure Mining." Computer Information. 2010, 26:160-162.

[21] M.E.J. Newman, M. Girvan. "Finding and Evaluating structure in networks." Physics Review[C]. 2004, E69, 026113

[22] D.E.Knuth, The Stanford Graphbase: "A Platform for Combinatorial Computing". Addison –Wesley, 1993, Reading, MA.

[23] Zachary W W. "An Information Flow Model for Conflict and Fission in Small Groups." Journal of Anthropological Research, 1970, 33:452–473.

Ruixin Ma, (1975--). Lecturer of Dalian University of Technology. Research area: E-commence, community discovery and swarm intelligence.

Guishi Deng, (1945--). Professor of Dalian University of Technology. Research area: E-commence, decision analysis and analysis of complex system.

Wang Xiao, (1988--). Ph.D. candidate at Institute of Automation, Chinese Academy of Science. Research area: E-commence, community discovery and swarm intelligence.