Refine the community detection performance by weighted relationship coupling

Dong Min,[1] Kai Yu,[1] and Hui-Jia Li[2,*]

[1]*School of Computer Science and Technology, Xinjiang University of Finance and Economics, Urumqi, Xinjiang 830001, China*

[2]*School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100080, China.*

(Dated: July 1, 2015)

Abstract

The complexity of many community detection algorithms is usually an exponential function with the scale which hard to uncover community structure with high speed. Inspired by ideas of famous Modularity optimization, in this paper, we proposed a proper weighting scheme utilizing a novel k-strength relationship which naturally represents the coupling distance between two nodes. Community structure detection using a generalized weighted Modularity measure is refined based on the weighted k-strength matrix. We apply our algorithm on both famous benchmark network and real networks. Theoretical analysis and experiments show that the weighted algorithm can uncover communities fast and accurately and able to be easily extended to large scale real networks.

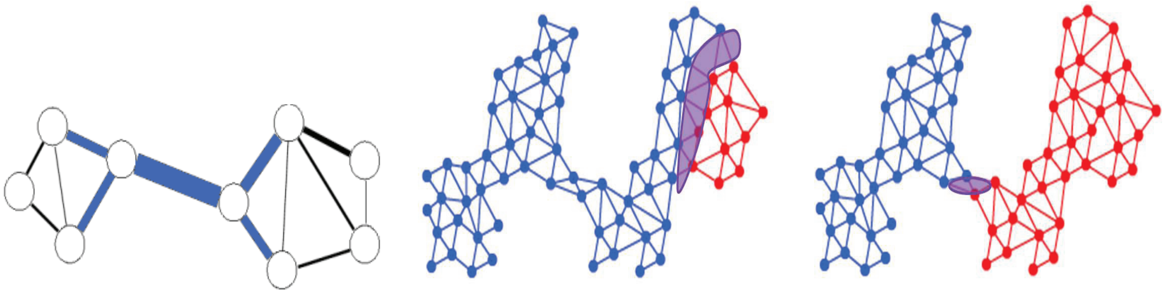
PACS numbers:

*Electronic address: hjli@amss.ac.cn

I. INTRODUCTION

Community structure[1–3] refers to a group of nodes in the network that are more densely connected internally than with the rest of the network. The studies for community detection are potentially useful in real social networks because nodes in a community are more likely to have same properties and all these communities may be functional groups. The methods for detecting community in networks are similar to the graph partitioning in graph theory [4, 5]. For example, in parallel computing, the pattern of required communications can be represented as a graph or network in which the nodes represent processes and edges join process pairs that need to communicate. The problem is to allocate the processes to processors in such a way as roughly to balance the load on each processor, while at the same time minimizing the number of edges that run between processors so that the amount of inter processor communication is maximized. In general, finding an exact solution to a partitioning task of this kind is an NP-complete problem, so it is prohibitively difficult to be solved accurately for large graphs. Inspired of this, a variety of heuristic algorithms have been developed that give acceptably good solutions in many cases, the best known being perhaps the Kernighan-Lin algorithm which runs in time $O(n^3)$ on sparse graphs[11].

Many algorithms on community detection had been proposed recently [6–8] and some of them are designed by the parameters of the networks, for example, eigenvectors of graph matrix, maximal modularity Q, clustering coefficient etc. Some algorithms are designed by dynamical characters of networks, such as random walk and spreading mechanism etc, however, those methods only deal with the binary form, i.e. unweighted network. In real world, the correlation(edge) between two nodes owns different strength, for example, as shown in Fig.1(a), the bottleneck edges between two communities usually own larger influence or betweenness. Furthermore, a network can be cut into several communities by the maximum modularity [9]. Unfortunately, computing the maximum modularity Q is proved to be NP-complete [10]. It means not all the communities are detected by computing the values of Q even though there are many heuristic algorithms. In weighted networks, if the bottleneck edges are not own largest weight, the NP-complete problem appears, just as shown in Fig.1(b). The random walk [12, 13], each node to be a walker and the walker will randomly choose a neighbor and currently stands on to localize in each time, has a probability to reach any other nodes, a dendrogram is got and the communities can be detected with



(a) (b)

FIG. 1: (a) The bottleneck edges between two communities usually own larger weight(influence or betweenness), and the value of weight are illustrated by thickness. (b) The edges with the largest weight are highlighted by enclosed circle, if they are not bottleneck edges(left subgraph), the NP-complete problem appears. Here, different color represents different communities.

the help of modularity Q. But it is difficult to specify the optimum random-walking time. Signal sending [14] is to transfer the topological relationship of nodes into the geometrical structure in n-dimensional Euclidean spaces, how to choose a proper p and partite all nodes into p-cluster is the weakness even it is empirical by the aid of F-statistics. Other methods depend on the probability of the communities in dynamic social networks such as [15, 16], and the values of modularity to find the proper communities such as [7, 8].

Since there are seldom polynomial time algorithms to detect the communities precisely, some valuable researches are focus on how to obtain much lower computing complexity for the detection algorithm compare with much more accuracy. In this paper, in order to design fast and accurate algorithm to detect communities, we proposed a new weighting scheme to enhance the community detection performance based on a novel correlation, i.e. k-strength relationship, which naturally represents the coupling distance between two nodes. Community structure detection algorithm is presented using a generalized Modularity measure based on the k-strength relationship weighted in various types of networks. Finally, we apply our algorithm on both benchmark network and real networks to evaluate its efficiency. Theoretical analysis and experiments show that the algorithm can uncover communities fast and accurately, which able to be easily extended to large scale real networks.

The outline of the paper is as follows. In Section II we introduce the fundamental definitions, such as k-strength relationship and its generalized Modularity measure. In Section III, we present the details of our weighting framework, including the procedures of algorithm and the analysis of computational complexity. Then we give some representative experiments on both benchmark and real networks to validate the effectiveness and efficiency of the algorithm in Section IV . Finally, Section V concludes this paper.

II. DEFINITIONS

In many real world relationships, such as the economic systems, the agents in system influence one another directly: a rush to buy or sell a particular asset can promote the other to do the same. In most cases, the agents are influenced only by their neighbors who joint by direct relationships. All the buyer and seller formed an inseparable structure and have very little interactions outside the structure. Such structure is known as communities in social networks.

The agents by nodes in network and the influence between each other are denoted by a weight between two nodes. In the following, a network is denoted by G with N-node set V, m-link set E and G also is an undirected without loop or multi-edges. The adjacency matrix A of G is a $N \times N$ zero-one matrix denoted by $A = (a_{ij})_{n\times n}$, where a_{ij}=1 if there is a link between i and j, and a_{ij}=0 otherwise. The adjacency matrix in an undirected graph is symmetric. If the network is weighted, we denote the weight of each link by w_{ij} and the weight matrix is $W = (w_{ij})_{N\times N}$. For a given positive integer k, denote a path from node i to j by a k-path if it is a walk with $k+1$ nodes and without cycle on it. $A^k = (a^k_{ij})_{n\times n}, a^k_{ij} = \sum_{l=1}^{N} a^{k-1}_{il} \times a_{lj}$ is the number of k-paths from node i to j $(i \neq j)$, if $i = j$, set a^k_{ij}=0. We denote $S^k = (s^k_{ij})_{N\times N}$ be a matrix of G for a given positive integer k. S^k is defined as a k^{th}**-strength matrix** of G. It was recursively defined as following:

If $k = 0$,

$$S^0 = A, \tag{1}$$

If $k = 1$,

$$S^1 = (w_{ij})_{N\times N}, \tag{2}$$

For all $k \geq 2$, let display

$$S^k = (s^k_{i,j})_{N\times N}, s^k_{i,j} = \sum_{s=1}^{a^k_{i,j}} \frac{1}{k} \sum_{l=1}^{k} w_{i^s_{l-1} i^s_l}, \tag{3}$$

where $i = i^s_0, i^s_1, \cdots, i^s_{k-1}, i^s_k = j$ are k-path for$s = 1, 2, \cdots, a^k_{i,j}$. To compute the values of $s^k_{i,j}$, $S^0 = A$ is fixed as the network determined. All the k-path between each pair of nodes can be obtained by $A^k, S^{i,j}$ is an additive polynomial. Therefore, we can compute the value of $S^k_{i,j}$ precisely.

Each k-strength matrix induces a ***k-strength relationship***: $R_k = \{(i, j, s_{i,j})|s_{i,j} = \sum_{l=1}^{k} s^l_{i,j}\}$. That is, $s_{i,j}$ in R_k are the elements of $S = S^1 + S^2 + \cdots + S^k = (\sum_{l=1}^{k} s^l_{i,j} = s_{i,j})_{N\times N}$. We denote S a ***k-strength matrix of G*** and the networks induced by S is a k-strength relationship networks. It is involved with a global idea of the mean-field theory on the definition of k-strength relationship networks. Each node knows all the others'information (the weights of nodes). It might be quite reasonable in many real systems. For example, the traders on Shanghai Stock Exchange are influenced by others on the same floor, but they can also be reminded by the trading patterns occurring on London or Paris. Therefore, some mature trading behavior patterns will be formed in economic systems. It is also very common in social networks to express the strength of friendship among people. For instance, in acquaintance network, the relationship is the tightness of acquaintance and higher the value is, more often the communication occurs. Another useful definition in our framework is minimal q-cut of a graph, which denotes the cut edges own the smallest sum of weight.

Here, q is a positive integer, $\{C_1, C_2, \cdots, C_q\}$ with $|C_i| = k_i$ and $\cup_{i=1}^{q} C_i \subseteq V(G)$ be a vertex subset such that the remaining of G after deleting all C_i is a disconnected and the sum of link weights among the remaining is the minimum. It was found that a minimum cut is a partition of G when $\cup_{i=1}^{q} C_i \subseteq V(G)$.

Guttmann had designed an algorithm to detect minimum cut in complete graphs[17] and inspired of his idea, we will detect the communities using the strength relationships. Our framework is also based on maximizing Modularity which firstly proposed by Newman, and we generalized it on strength relationship matrix of G. Suppose there are $q(q \leq N/2)$ communities in $G, \mathcal{C} = \{C_1, C_2, \cdots, C_q\}$. The generalized weighted Modularity Q in strength relationship matrix is defined by

$$Q = \max_q \sum_{i=1}^{q} (c_{i,i} - c_i^2) \tag{4}$$

where $c_{i,i} = \sum_{i,j} \frac{s_{i,j}}{\Delta} \delta_{i,j}$,$c_i = \sum_j s_{i,j}$ and $\Delta = \sum_{i,j} s_{i,j}$. $\delta_{i,j} = 1$,if the nodes i and j are in the same community, $\delta_{i,j} = 0$, otherwise. $c_{i,i}$ denotes the fraction of strength with both ends in the same partition C_i, c_i is the proportion of strength with one end in C_i and the other not. If the network is unweighted (binary network), the Q is just Newman's modularity. Based on this form, the new measure can capture the properties of the real social systems. One can find that both direct and undirect information between two nodes can be used within our framework. When two nodes are exchanging their information in social networks, the chains will formed within the same community. Thus, it might be more reasonable to describe relationship between nodes using the strength relationship, such as same ideas in a society are more likely be connected closely and transmitted one by one. It means a tightly connected community implies a faster rate of information transmission or rumor spreading rate than a sparsely connected one, because more paths there are, faster transmissions rate there is.

III. THE FRAMEWORK

A. The weighted k-strength relationship matrix

In this part, we analyze the property of the weighting scheme in detail. As describe above, the k-strength relationship matrix is fundamental to the whole framework. Here, we focus on determining the k-strength relationship matrix. The following theorem not only provides the process of computing all the elements, but also reveal the important time complexity information.

Theorem 1: *The k-strength relationship matrix W^k can be obtained in polynomial time.*

Proof: Suppose adjacent matrix of network G is A, the number of all k-length paths from i to j is $a^k_{i,j}$, and $A^k = A^{k-1} \times A = (a^k_{i,j})$ in [16]. A path is called k-path if its length is k. Denote the k-path by $\{i_0, i_1, \cdots, i_{k-1}, i_k\}$ with $k+1$ nodes and $i_s \neq i_j$ for all s and j, that is, there is no cycle in the path.

In order to get the elements in k-strength relationship matrix, we define an operation $\oplus$ on weight matrix of G. $\oplus$: $W^k = W^{k-1} \oplus W = (w^k_{i,j})_{N\times N}$, where $w^k_{i,j}$ is defined as: If $\sum_{l=1}^{N} (w^{k-1}_{i,l} \times w_{l,j} \neq 0)$, it means there are k links connect node i and j. Equivalent, there are at least one term in $\sum_{l=1}^{N} (w^{k-1}_{i,l} \times w_{l,j} \neq 0)$. Without of generally, we suppose, there are h terms not zeros,$w^{k-1}_{i,l^1} \times w_{l^1,j} \neq 0, w^{k-1}_{i,l^s} \times w_{l^s,j} \neq 0, w^{k-1}_{i,l^h} \times w_{l^h,j} \neq 0$. Then $w^k_{i,j} = \sum_{s=1}^{h} (w^{k-1}_{i,l^s} + w_{l^s,j})$; Otherwise, $w^k_{i,j} = 0$ (that is, there is no link joint i and j).

The value of $w^k_{i,j}$ is the sum of all weights in each k-path from i to j. We can take not so much effort to obtain $s^k_{i,j} = w^k_{i,j}/k$. That is $s^k_{i,j} = \sum_{s=1}^{a^k_{i,j}} \frac{1}{k} \sum_{l=1}^{k} w_{i^s_{l-1},i^s_l} = w^k_{i,j}/k$.

All the k-path can be lay out when $\sum_{l=1}^{N} (w^{k-1}_{i,l} \times w_{l,j})$ is determined. If $w^{k-1}_{i,l^1} \times w_{l^1,j} \neq 0, w^{k-1}_{i,l^s} \times w_{l^s,j} \neq 0, w^{k-1}_{i,l^h} \times w_{l^h,j} \neq 0$ for each positive integer $k \geq 2$; Denote a k-path connect nodes i and j by $P^k_{i,j}$, it is easily to find there are $a^k_{i,j}$ k-paths joint node i and j and hence,$P^k_{i,j} = \{P^{k-1}_{i,l^1} \vee (l^1, j), P^{k-1}_{i,l^2} \vee (l^2, j), \cdots, P^{k-1}_{i,l^h} \vee (l^h, j)\}$, where $P^{k-1}_{i,l} \vee (l, j)$ means the all k-path formed by the $(k-1)$-paths in set $P^{k-1}_{i,l}$ join the link (i, j).

Finally, we can lay out all the k-paths inductively. That is, $\oplus$ is a polynomial time algorithm. Totally, the strength matrix is got by computing the weight matrix with computing complexity time $O(n^2m)$ since the multiplication of each pairs of N-rank matrixes costs at most $N \times N$ and at most m links. Output $S^k = (s^k_{i,j})N \times N$ for a fixed k. Outline all the paths from i to j with length k, $i = i^s_0, i^s_1, \cdots, i^s_{k-1}, i^s_k = j$ for $s = 1, 2, \cdots, a^k_{i,j}$.

The proof is end.

B. Community detection algorithm

A minimal q-cut $\widetilde{E}$ of G is an edge set with minimal sum of weight that the remaining graph of deleting the edges set, $G - \widetilde{E}$, is an isolated graph. A directly method to determine the partitions is investigating all the components of the remaining graph $\mathcal{C} = \{C_1, C_2, \cdots, C_c\}$. We need chose the components such that $\sum_{i<j} w(C_i, C_j)$ is minimum or $\sum_{i=1}^{c} w(C_i, C_i)$ is maximum by maximum flow and minimum cut theorem [18]. However, the minimal q-cut problem is NP-complete and it is difficult to find a polynomial algorithm. Fortunately, Guttmann-Beck and Hassin designed an algorithm in complete graphs and proved the approximate solution is less than three times the optimal [19]. Inspired by this nice idea, we obtain the detailed procedures in Algorithm 1.

C. Computational complexity

For a given positive number q, we can solve the transport problem[21][24] in time $O(N)$ since it is a 0-1 transportation problem. There are $C_N^q O(qN)$ subsets of V. Altogether the time complexity is $O((q+1)N)$. Here, two important claims are proposed which useful to the analysis:

Claim 1: If $\{C_1, C_2, \cdots, C_q\}$ is a partition of $\widetilde{G}$, if and only if it is a partition of G.

Proof: It is easy to verify that the claim holds, since $\widetilde{G}$ and G has the same vertex set.

Claim 2: Suppose $\{C_1, C_2, \cdots, C_q\}$ are q partition of $\widetilde{G}$ such that $\sum_{i<j, C_i, C_j \subset \widetilde{G}}^{q} w(C_i, C_j)$ is a minimum. Then there is a minimum partition of $\widetilde{G}$, say $\{\overline{C_1}, \cdots, \overline{C_q}\}$, is also a minimum partition G such that$\sum_{i<j, \overline{C_i}, \overline{C_j} \subset G}^{q} w(\overline{C_i}, \overline{C_j})$ is a minimum.

Proof: By the definition of $\widetilde{G}, \Delta = \sum_{i<j, C_i, C_j \subset \widetilde{G}}^{q} w(C_i, C_j)$. Since $w(C_i, C_j) = \sum_{i \in C_i, j \in C_j} s_{ij}^1 + \sum_{i \in C_i, j \in C_j} \sum_{k=2} s_{ij}^k$ and the value of Δ is fixed because $\{C_1, C_2, \cdots, C_q\}$ is the minimum partition of $\widetilde{G}$. We know that $\sum_{i \in C_i \subset \widetilde{G}, j \in C_j \subset \widetilde{G}} s_{ij}^1 = \sum_{i \in C_i \subset G, j \in C_j \subset G} s_{ij}^1$ by the definition of k-strength relationship. Therefore, we construct a minimum partition $\{\overline{C_1}, \cdots, \overline{C_q}\}$ in $\widetilde{G}$ by Algorithm 1 such that$\sum_{i \in \overline{C_i} \subset \widetilde{G}, j \in \overline{C_j} \subset \widetilde{G}} s_{i,j}^1$ is a minimum, then,$\{\overline{C_1}, \cdots, \overline{C_q}\}$ is the minimum q partition of G.

The proof is end.

IV. EXPERIMENTS AND RESULTS

In order to verify the effectiveness of our weighting method, we apply it on two famous benchmarks. First, an artificial random network generated by Girvan and Newman, GN network, is used[20][25]. This benchmark was used by many methods for comparing the efficiency of partition result. The establish mechanism is as follows: a 128 nodes network is partition into four communities, and each community owns 32 nodes. Every inner-community edge is linked independently with probability p_{in} and every inter-community edges is linked with probability p_{out}. For each node, the expected inner-community degree is $z_{in} = 31p_{in}$ and the expected inter-community degree is $z_{out} = 31p_{out}$. As Z_{out} increases, the community structures becomes more and more ambiguous, and correspondingly fraction of correctly classified nodes decreases.

To illustrate the efficiency our refinement algorithm, we comparing the percentage of

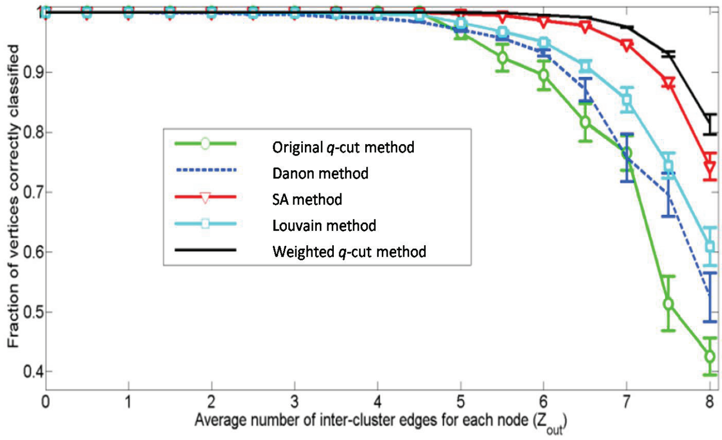


FIG. 2: Computational results by five algorithms including our method as function of Z_{out} in GN network. Each point shows the average and variance over 50 times.

nodes that are classified correctly for different methods including, original minimum q-cut algorithm, refined minimum q-cut algorithm, two famous modularity optimization methods: Louvain method[23] and Danon method[22], and SA heuristic method[27]. As can be observed in Fig.2, the refinement process enhance the performance of community detection a lot. Our algorithm performs the best even Z_{out} increases to 8. Furthermore, we analyze the performance of variance of community partition. As sensitivity to the initial condition, the original minimum q-cut algorithm method shows the largest variation, while the Louvain method and Danon method have less variation comparing to GA heuristic method, and as it was expected, the variance of refined minimum q-cut algorithm is the best among the mentioned approaches.

Next, we test whether the communities are identified completely correct. Different with the GN network, if such a group is not entirely contained in the same community, then all vertices of the group are assumed incorrectly identified. We test our framework in a famous real-world network, i.e. Karate club network [26]. It is a standard network which used to compare the precision between different community detection algorithms. Here, we use Newman Fast method(NF) on weighted network, which need not specify the number of communities. In Fig.3, the dendrogram obtained by using our weighting scheme is represented. We report the modularity measure obtained from original NF method and weighted NF as 0.397 and 0.432. From these results, one can conclude that the weighting scheme improves NF considerably.

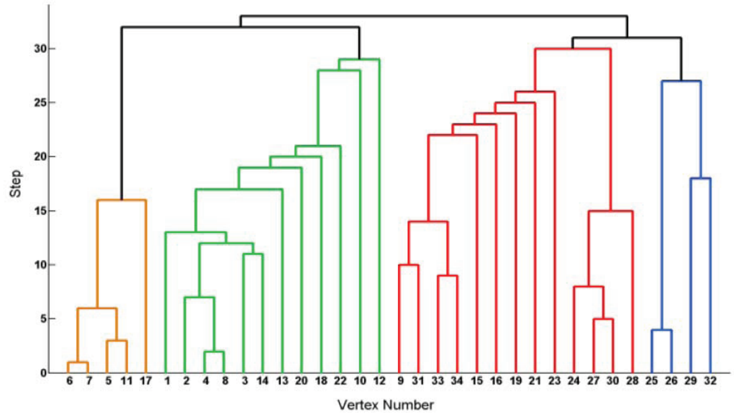


FIG. 3: Using the weighting scheme, we apply our framework on Zachary Karate network identified by Newman Fast Algorithm. Dendrogram of communities are shown and different colors correspond to four community structures

V. DISCUSSION

In this paper, we have designed an efficient algorithm to detect community in social network using a new definition, i.e. k-strength relationship, which naturally represent the coupling degree between two nodes. Theoretical analysis shows this algorithm is polynomial time which much better than most existing ones. Finally, we apply our algorithm on both benchmark network and real networks to evaluate its efficiency. Theoretical analysis and experiments show that the algorithm can uncover the communities fast and accurately, which able to be easily extended to large scale real networks.

Acknowledgments

The authors would like to thank the reviewers for their detailed reviews and constructive comments, which have helped improve the quality of this paper. The research was supported in part by NSFC grants 71401194, 91324203 and "121" Youth Development Fund of CUFE grants QBJ1410.

---

[1] M.E.J.Newman, *Phys. Rev. E.* **69**, (2004): 066133.

[2] M.E.J.Newman, M.Girvan, *Phys. Rev. E.* **69**, (2004): 026113.

[3] M.E.J.Newman, *Proc. Natl. Acad. Sci.* **103**, (2006): 8577-8582.

[4] M. R. Garey and D. S. Johnson, Computers and Intractability: A Guide to the Theory of NP-Completeness, Freeman, San Francisco, (1979).

[5] J. Scott, Social Network Analysis: A Handbook, 2nded, Sage Publications, London, (2000).

[6] M.Girvan and M.E.J.Newman, *Proc. Natl. Acad. Sci.* **99**, (2002): 7821-7826.

[7] X.S.Zhang, R.S.Wang, Y.Wang, J.Wang, Y.Qiu, L.Wang, L.Chen. *Eur.Phys.Lett.* **87** (2009): 38002.

[8] X.S.Zhang, Z.Li, R.S.Wang, Y.Wang. *J.Comb.Optim.* **23(4)** (2012): 425-442.

[9] F. Radicchi, C. Castellano, F. Cecconi. *Proc. Natl. Acad. Sci.*, **101**, (2004): 2658-2663.

[10] M. Latapy and P. Pons. Proceedings of the 20th International Symposium on Computer and Information Sciences, *Lecture Notes in Computer Science.* **3733**, (2005): 284-293.

[11] B.W.Kernighan and S. Lin. *Bell System Technical Journal.* **49**, (1970): 291-307.

[12] H.J.Li and X.S.Zhang. *Eur.Phys.Lett.* **103** (2013): 58002.

[13] H.J.Li, Y.Wang, L.Y.Wu, J.Zhang, X.S.Zhang. *Phys.Rev.E.* **86** (2012): 016109.

[14] U. Brandes, et al. arXiv.org:physics. 0608255, (2006).

[15] L.-C. Huang, T.J. Yen, S. C. T. Chou. 2011 international conference on advances in social networks analysis and mining, IEEE computer society, (2011):110-117.

[16] Peter J. Mucha, et al. *Science*, **328**,(2010): 876-878.

[17] N. Guttmann-Beck and Hassin. *Algorithmica* **27** (2000): 198-207.

[18] M. R. Garey and D. S. Jonson, Computers and intractability: A guide to the theory of NP-completeness, Freeman, San Francisco, CA(1979).

[19] D.P.Maki. Mathematical Models and Applications, with Emphasis on social, life, and Management sciences. Prentice Hall, (1973).

[20] Z.Y.Xia, Z.Bu. *Knowl-Based Syst.* **26** (2012): 30-39.

[21] W.N.E, T.Li, E.Vanden-Eijnden, *Proc. Natl. Acad. Sci.* **105**, (2008):7907-7912.

[22] L.Danon, J,Duch, D.Guilera, A.Arenas, *J. Stat. Mech.* **29**, (2005):P09008.

[23] V.D.Blondel, J.L.Guillaume, R.Lambiotte, E.Lefebvre, *J. Stat. Mech.* **10**, (2005):P10008.

[24] M.Rosvall and C.T.Bergstrom, *Proc. Natl. Acad. Sci.* **105(4)**, (2008):1118-1123.

[25] Z.P.Li, S.H.Zhang, R.S.Wang, X.S.Zhang, L.Chen, *Phys. Rev. E.* **77**, (2008):036109.

[26] A.Lancichinetti and S.Fortunato, *Phys. Rev. E.* **80**, (2009)056117.

[27] R.Guimera and L.A.N. Amaral, *Nature* **433**, (2005):895.

Algorithm 1 Community detection algorithm based on weighting scheme

Input: a network $G = (V, E)$;

Output: a minimum q-partition with maximal modularity value.;

1: **Step 1:** Shrinking each one degree node to its neighbor until there is no one degree node in G. This operation does not affect the community detection, because the one degree node has no other choice but to its unique neighbor, so the one degree node will be in the same community with its neighbor. We still write the network as G.

2: **Step 2:** Set the number of communities q.

3: **Step 3:** For a fixed q , the minimum q-cut problem is polynomial time solvable in $O(|V|^{q^2})$ [17].

4: Therefore, we suppose we had a partition $\mathcal{C} = \{C_1, C_2, \cdots, C_q\}$ in $\widetilde{G} = (V, \widetilde{E})$, and $|C_i| = k_i, \sum_{i=1}^{q} k_i = N, C_i \cap C_j = \emptyset$. We will detect he minimum q-cut in $\widetilde{G}$, and then prove it is also is the minimum q-cut in G. Let $v_i \in C_i$ for $i = 1, 2, \cdots, q$. $x_{i,j} = \begin{cases} 1 \ , \ u_j \in C_i \\ 0 \ , \ otherwise \end{cases}$

Begin

For $\{v_1, v_2, \cdots, v_q\} \subset V, v_i \in C_i$.

For $u_j \in V - \{v_1, v_2, \cdots, v_q\}$,the following transport problem is optimal.

$$min : \sum_{i=1}^{q} \sum_{j=1}^{N-q} w(C_i, u_j)(1 - x_{i,j})$$

$$subject\ to \begin{cases} \sum_{j=1}^{N-q} x_{i,j} = k_i - 1 \ , \ i = 1, 2, \cdots, q \\ \sum_{i=1}^{q} x_{i,j} = 1 \ , \ j = 1, 2, \cdots, N - q \\ x_{i,j} \in \{0, 1\} \ , \ i = 1, 2, \cdots, q \text{ and } j = 1, 2, \cdots, N - q \end{cases}$$

End

$$C_i = C_i \cup \{u_j | x^*_{i,j} = 1, 1 \le j \le N - q\} \text{ for } 1 \le i \le q.$$

End

Back to begin

5: **Step 4:** Output:$\{C_1, C_2, \cdots, C_q\}$ with $v_i \in C_i$ is a minimum q-cut on $\widetilde{G}$.