Journal of Engineering Science and Technology Review 7 (3) (2014) 68 – 73

JOURNAL OF
Engineering Science
and Technology Review

www.jestr.org

Research Article

Community Mining Method of Label Propagation Based on Dense Pairs

WENG Wei [1,*], ZHANG Nian[1] and XU Huarong[2]

[1]College of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China
[2] Dept. of Manage. Inf. Syst., Univ. of Arizona, Ttucson, Arizona ,CO.85721, United States

Received 15 June 2013; Accepted 23 March 2014

Abstract

In recent years, with the popularity of handheld Internet equipments like mobile phones, increasing numbers of people are becoming involved in the virtual social network. Because of its large amount of data and complex structure, the network faces new challenges of community mining. A label propagation algorithm with low time complexity and without prior parameters deals easily with a large networks. This study explored a new method of community mining, based on label propagation with two stages. The first stage involved identifying closely linked nodes according to their local adjacency relations that gave rise to a micro-community. The second stage involved expanding and adjusting this community through a label propagation algorithm (LPA) to finally obtain the community structure of the entire social network. This algorithm reduced the number of initial labels and avoided the merging of small communities in general LPAs. Thus, the quality of community discovery was improved, and the linear time complexity of the LPA was maintained.

Keywords: Community Mining; Label Propagation; Social Network

1. Introduction

A social network is a collection of individuals and their reciprocal relationships. Along with the rapid development of the Internet and the emergence of social networking systems in the 1990s, users became drawn to the virtual Internet, and studies on social networks started to progress steadily. Numerous social networking platforms, such as Internet forums, social networking sites, and instant communication networks, have emerged, accumulating large amounts of social networking data. These social networks reflect real lives directly or indirectly and affect the social behavior of people. For instance, visitors who review and mark your blog and then connect it to their personal sites are likely to be your good friends in real life. Wikipedia is considered the biggest and most popular reference on the Internet, and movies reviews on the IMDB forum provide guidance when booking tickets.

Community [1] is one of the basic features of social networks. Let the social network be represented by a graph, the objects in it as a node, and the mutual relationship between objects as the edge. The community can be seen as a sub-graph, within which a dense connection exists, but its connection with the external sub-graph is sparse. Similar to “birds of the same feather flock together,” we require prior knowledge of network structures to understand the organizational chart of a complex network structure and thus understand the interaction, evolution, and organization function of relationships among actors. Extensive research has been conducted on community discovery in various social and ecological fields, such as the World Wide Web [2], scientific cooperation network [3], and molecular biological analysis [4]. The findings have been success fully applied in link prediction [5] and in influence maximization in network advertising [6].

However, community mining suffers from the lack of an accurate definition for community [7]. Thus, community mining is a challenging task [8-9]. There are many community definitions [10-11] and community mining algorithms, but according to any kind of definitions (such as a certain quantitative function for community quality), community mining remains an NP-complete problem [12]. Two methods may be used to classify community quality using a community mining algorithm. One method involves direct evaluation using known baseline data for community structure. The other method involves evaluation by a certain quantitative function to measure community quality. The former method obtains a moderate amount of data, because numerous human observations and statistics are involved in the collection of baseline data; thus, it is hard to represent the virtual communities that are identified in large numbers. The quantitative function also suffers from resolution limits [13]. Thus, the computation of the results obtained from the Internet must be analyzed comprehensively. Time performance problems for large-scale data should also be considered in data structures [14] or algorithms [15].

Social networks illustrate the relationship among nodes, but the degree of the connection among nodes within their own communities varies. Fig.1 presents the relationship between nodes a and b and that between a and c from the local domain, where set a and its adjacent nodes are denoted as $\{a, b, c, d\}$, set b and its adjacent nodes are denoted as $\{a, b, c, d, e\}$, and set c and its adjacent nodes are denoted as $\{a, b, c, d\}$. Set a is almost the same as set b but differs from set c. In other words, a and b reach basic consensus on the range of corresponding relationships, and their relationship is

[*] E-mail address: xmutwei@163.com

superior to that of *a* and *c*. In the present work, community mining is conducted base on the quantitative description of the aforementioned relationship.

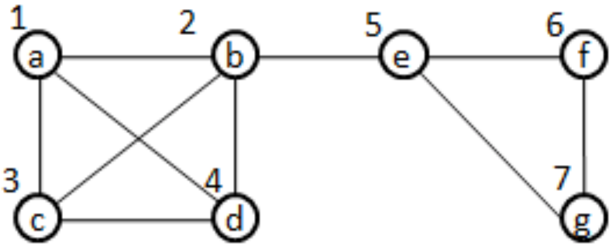

Fig. 1 Example of a network

2. Related Work

The modularity proposed by Newman and Girvan to measure community quality marked a great step forward in community mining algorithms [16]. Subsequently, a series of algorithms for modularity optimization were generated; examples include greedy algorithm [17] and extreme optimization [18]. Reichardt et al [19] introduced the simulated annealing algorithm to community mining to prevent the greedy algorithm from falling into the local optimum, but the time cost obviously increases. In [20], the original community is first selected through the random walk strategy, and the final community is divided by modularity optimization. Experiments have proven that this method achieves a fair compromise in time efficiency and the quality of community division. However, Fortunato et al [13] point out that the modularity optimization algorithm has resolution limits, that is, it fails to produce small communities but tends to merge them. Take Fig. 1 as an example. In this figure, two communities can be observed, but algorithms can generally identify only one community, that is, {*a*, *b*, *c*, *d*, *e*, *f*, *g*}. Other methods are available in the literature [21-23].

Most community mining algorithms have high time complexity, but the label propagation algorithm (LPA) [24] is a simple and effective method whose near-linear time complexity makes it suitable for handling large-scale network data. The basic method of the LPA is as follows: each vertex is given a unique original label. Based on the voting principle, each vertex then adopts the labels to which the maximum number of its neighbors belongs. This process continues iteratively until the labels of all vertices are stable. At this time, the vertices assigned by the label form a community. However, resolution limits occur easily in this basic LPA. Zhao Zhuoxiang et al [25] contend that numerous scattered small communities are formed during the iterative process because each vertex in the network is assigned a unique label, and the most meaningful labels cannot be discovered during the initial stage; these conditions slow down the pace of label convergence. According to the definition of the influence of nodes in [26], the authors in [24] assign the initial tags to the first *k* nodes with the greatest effect, but this method fails to directly determine the value of *k* and requires predetermining other parameters for calculating the influence degree. In extreme cases, if the value of *k* is less than the actual number of communities, the label convergence algorithm fails to identify all the communities.

To overcome the defects of the current label algorithm, we develop a label propagation community mining method based on local similarity. This method identifies closely related local domains by determining if the connection of the network structure is intense. We consider these local domains as the rudiments of a community and thus assign them the original label. On the basis of the voting principle, we expand and adjust the region through label propagation before forming a community. This method prevents numerous tags that are common in the traditional LPA and eliminates the accuracy limitation without using artificial parameters.

3. Algorithm Description

We formally define the basic concept as follows. Let $G=(V, E)$ be a network, where $V=\{v_1, v_2, \ldots, v_n\}$ is the node set, and $E=\{e_1, e_2, \ldots, e_m\}$ is the edge set with $e_i \in V \times V$. If there exists a edge between v_i and v_j, then $w_{ij}=1$, otherwise $w_{ij}=0$. We denote the neighbor of v_i as $N(v_i)$, that is, $N(v_i)=\{v_j|w_{ij}=1\}$, and the star neighbourhood of v_i as $St(v_i)=\{v_i\}$□$N(v_i)$. $K(v_i)=w_{i1}+w_{i2}+\ldots+w_{in}$ is the sum of the edges relative to node v_i and is defined as the degree of v_i.

3.1 Algorithm Steps

The proposed community mining method based on local similarity includes the following steps (Fig.2).

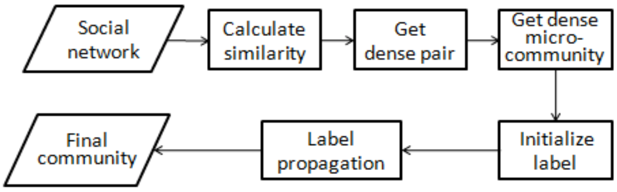
Social network
Calculate similarity
Get dense pair
Get dense micro-community
Initialize label
Label propagation
Final community

Fig. 2 Algorithm process

(1) Calculate the similarity of each pair of nodes according to the connection of network nodes.

In this case, the calculation formula [27] for the similarity of nodes (v_i, v_j) is as follows:

$$S(v_i, v_j) = \frac{\sum_{v_l \in St(v_i) \cap St(v_j)} \frac{1}{K(v_l)}}{\sqrt{\sum_{v_m \in St(v_i)} \frac{1}{K(v_m)}} \sqrt{\sum_{v_n \in St(v_j)} \frac{1}{K(v_n)}}} \tag{1}$$

Based on the above formula, the similarity of node v_i and v_j depends on the degree of the nodes in their star neighbourhood intersection set. A great degree indicates that the node makes little contribution to the similarity of v_i and v_j because the role of betweenness in connection is dispersed. The denominator plays a role in uniformization. Thus, if $S(v_i, v_j)=0$, then $S(v_i, v_j) \in [0, 1]$ demonstrates that nodes v_i and v_j neither connect with each other nor share the same neighbors. If $S(v_i, v_j)=1$, then $St(v_i)$ is equal to $St(v_j)$. Thus, we can conclude that if nodes v_i and v_j share the same neighbor (whether the two nodes are connected or not), then the relationship between these two nodes is always greater than zero. This condition avoids the underestimation common in other measuring methods; for example, in the case of two nodes connected with each other but with no neighbors in common, their similarity is computed by some indicators as 0, which obviously underestimates their relationship. In this relationship, we only calculate the similarity between nodes and edges.

As a typical model of a small-sized network [28], examining the connection between these two nodes and their neighbours is reasonable when calculating the similarity of the two nodes. Assuming we expand the scope of consideration, which is time consuming, the calculated nodes probably contain most of the nodes in the entire network.

(2) Identify all the dense pairs of nodes.

If a node pair (u, v) has the largest similarity among the adjacent nodes, then it is called a dense pair [29], that is, $\sigma(u, v)=max\{S(x, y)|(x=u, y\in N(u)) \vee (x=v, y\in N(v))\}$, which is denoted by $u\leftrightarrow_{\varepsilon} v$, where $\varepsilon=\sigma(u, v)$. Identify the nodes pairs that constitute the set of dense pairs.

(3) Identify all the dense micro-communities.

If nodes within the node set constantly have other nodes to constitute respective dense pairs and if their corresponding nodes are always within the set, these nodes can be aggregated and built into a micro-community [29]. In fact, the dense micro-community is a sub-graph and can be marked as $C(w)=(V, E, \varepsilon)$, which meets the following three conditions: (a) $w\in V$; (b) for all $u\in V$, $\exists v\in V(u\leftrightarrow_{\varepsilon} v)$; and (c) $u\in V((u\leftrightarrow_{\varepsilon} v \wedge u\in V \wedge v\notin V)$ is non-existent.

(4) Assign all dense micro-communities with the original label. Nodes in the same dense micro-community share the same initial label, but the initial label varies between dense micro-communities.

(5) Place the nodes of the network into set X according to their degree in descending order.

(6) Let t=1.

(7) Obtain each node $x\in X$ as the order in X, and divide the adjacent nodes of x into groups according to the type of label. The group with the maximum number of members is assigned x. If several identical maximum values exist, choose the label of the group with the largest sum of node degrees. We can adopt the asynchronous update label strategy to avoid the label concussion generated in bigraphs [15].

(8) If the labeling values remain unchanged in the next step, then end the algorithm. Otherwise, given t=t+1, repeat repeat Step (7).

3.2 Time Complexity Analysis

A complex network comprises thousands of nodes, but a node only has connections with some small nodes. Thus, matrix storage wastes space. In the present work, we adopt a linked storage for the adjacency relationship between nodes. Given the maximum degree of network nodes d_{max}, the number of dense pairs is p, the number of nodes in the largest dense micro-community is t, the time complexity of the similarity among nodes is $O(nd_{max}^{2})$, the time complexity for dense pairs is $O(nd_{max})$, the time complexity of the dense micro-community is $O(pt)$, and the complexity in label propagation is $O(n)$. For a large and complex network, d_{max} is smaller than n. In later experiments, p is found to be smaller than n, and the average size of the dense micro-community is smaller than n. Therefore, our algorithm is close to the algorithm of linear time complexity.

3.3 Example

We elaborate this algorithm with the combination of networks shown in Fig. 3.

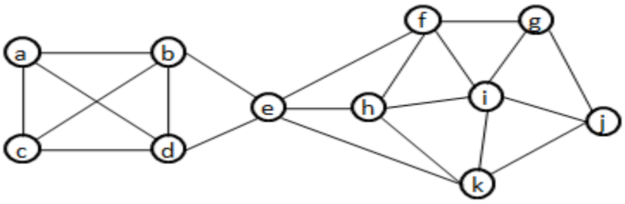

Fig. 3 Sample network

(1) Compute the similarity of each node pair based on the adjacency matrix, as shown in Fig. 4.

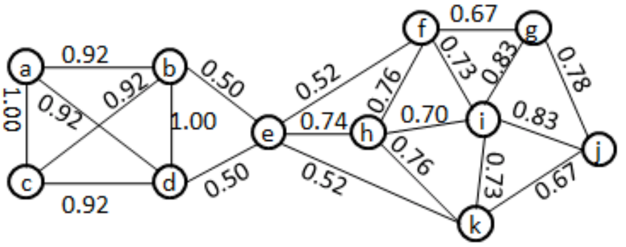

Fig. 4 Similarity assigned on the edges

(2) Six dense pairs can be found in Fig.4: (a, c), (b, d), (f, h), (g, i), (h, k) , and (i, j).

(3) Four dense micro-communities can also be found in Fig.4: (a, c), (b, d), (f, h, k), and (g, i, j).

(4) In Fig. 4, four kinds of symbols are used to represent four different labels. The original labeling is shown in Fig. 5.

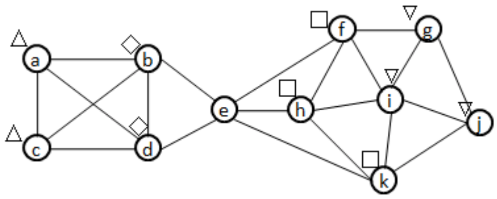

Fig. 5 Original labeling

(5) Propagate the label under the voting principle until the label of any node stops updating.

A new round of label propagation based on Fig.5 is necessary. The first updated node is e, as shown in Fig.6. The Second updated node is i, as shown in Fig.7. The other updates are omitted.

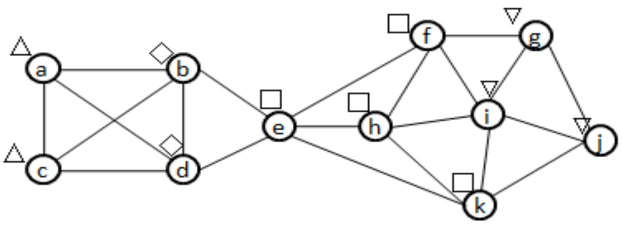

Fig. 6 First update of a node

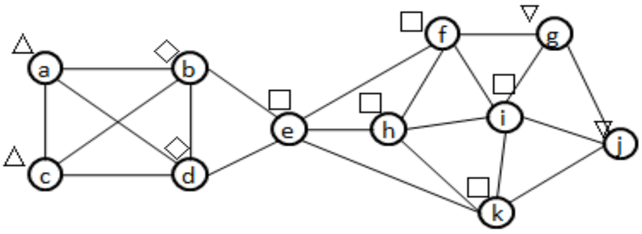

Fig. 7 Second update of a node

(6) Form the final community structure after the first update of each node (shown in Fig.8).

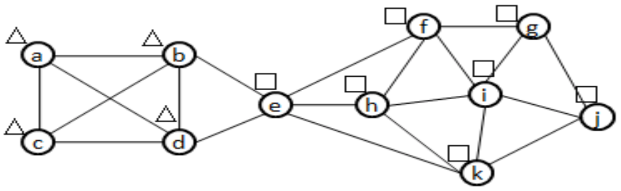

Fig. 8 Final community structure

The results show that the final community structure is consistent with the observed structure and that the society is divided reasonably.

4. Experimental Analyses

On the basis of data on social networks, such as the Karate Club Network (hereinafter abbreviated as Karate) [30] and the U.S. Undergraduate Football Competition 2000 (hereinafter abbreviated as Football) [31], which are widely used to test the effectiveness of community mining, we compared the proposed algorithm with the traditional LPA [15] and the LPA based on the junction influence (hereinafter abbreviated as LIB) [24]. The comparison data are obtained directly from previous work [24]. The proposed algorithm is denoted as improved LPA (ILPA).

Tab. 1 Modularity comparison

Dataset	LPA	LIB	ILPA
Karate	0.3653	0.3715	0.3715
Football	0.4665	0.5945	0.6000

As shown in Tab.1, the modularity of the ILPA is consistent with that of the LIB for the Karate dataset because both algorithms can correctly divide all nodes into two communities. However, the modularity of the ILPA is 1.7% higher than that of the LPA. For the Football dataset, the modularity of the ILPA is higher than that of the LPA by 28.62% and slightly higher than that of the LIB. Although the outcome of the proposed algorithm is similar to that of the LIB, the latter requires a large number of parameters that are difficult to set.

We chose three large datasets from different fields to further study the nature of this algorithm. The first dataset is a high energy physics theory collaboration network (hereafter referred to as High Energy Physics) [32]. This dataset includes 9,987 nodes and 25,998 edges. The network describes the cooperation among authors in the domain of high energy physics theory on Arvix net between January 1993 and April 2003. If author *i* and *j* write an article together, then the network will have an undirected edge from node *i* to *j*. Likewise, if *k* number of authors work together to create a thesis, then the network will cover a complete sub-graph that consists of the corresponding nodes of these *k* authors. The second dataset is the peer-to-peer network Gnutella [33], which represents the connection among hosts and contains 26,518 nodes and 65,369 edges. If user *i* casts a vote for user *j*, then the network will include a direct edge from node *i* to *j*. This algorithm is seen as an undirected graph. The third dataset is the shopping information on Amazon (com-amazon) [34], which includes 334,863 nodes and 926,872 edges. If node *i* is linked to node *j*, then commodity *i* and *j* are sold out together.

Tab.2 summarizes the operation of large datasets in the proposed algorithm. For these datasets, the sum of the number of nodes in the initial community accounts for 56% to 73% of the total nodes, and the number of initial communities is about two to four times of the final number of communities. At the same time, the number of iterations is so low that time complexity, unlike the number of nodes and edges, can be ignored. This algorithm can also discover numerous small communities (with less than 10 nodes). Such cases are common in real life. For example, in a paper on cooperative networks, these small communities indicate that only a few experts can cooperate with a large number of people and make achievements in many fields. However, most experts cooperate with only a few researchers.

Tab. 2 Result of the ILPA for the High Energy physics, Gnutella, and com-amazon datasets

Dataset	Nodes	Edges	Initial communities	Nodes in initial communities	Iteration	Final communities	Communities with nodes>=10
High Energy Physics	9877	25998	3271	7208	12	1043	214
Gnutella	26518	65369	5777	16603	6	2441	438
com-amazon	334863	926872	90163	188958	10	21716	10902

As shown in Figs. 9, 10, and 11, the respective community size and quantity generated by the ILPA and the original algorithm LPA are compared further. In these figures, the horizontal ordinate stands for the interval of community size, 1 to 9 represents the range of node numbers within the community (i.e., 1 to 9, 10 to 19, 20 to 29, 30 to 39, 40 to 49, 50 to 99, 100 to 149, 150 to 299, and > 300). The vertical ordinate represents the corresponding numbers of communities (logarithm value). The picture shows that a large number of communities can be discovered in different large datasets through the use of the ILPA. By contrast, the ILPA can discover many communities in most of the range intervals (especially in parts close to the small interval) in the same dataset, and the number of communities even exceeds 10 in some intervals. Therefore, the LPA has the accuracy limitation, whereas the ILPA eliminates the limits effectively. Thus, the label algorithm proposed in this study is another improvement of the traditional LPA and advances the quality of community division effectively.

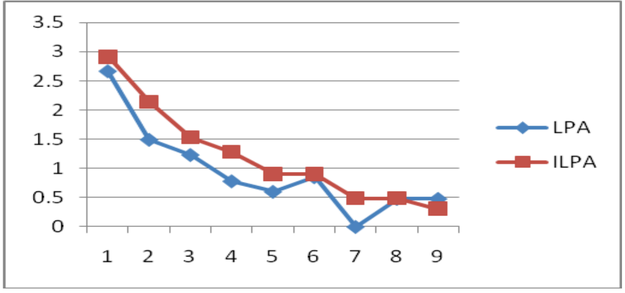

Fig. 9 Comparison of two algorithms for the High Energy Physics dataset

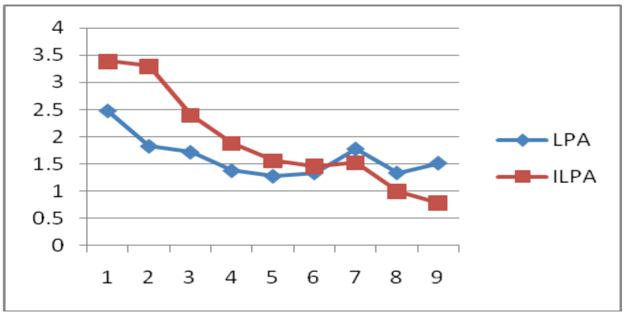

Fig. 10 Result of the comparison of two algorithms for the Gnutella dataset

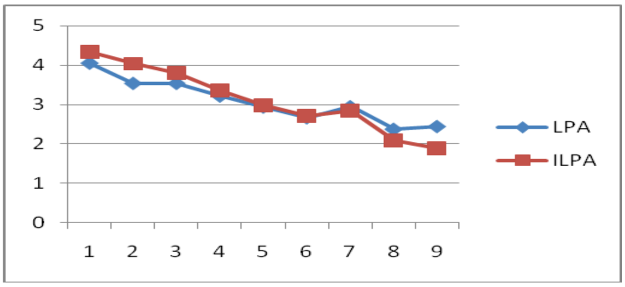

Fig.11 Result of the comparison of two algorithms for the com-amazon dataset

5. Conclusion

The proposed algorithm relies solely on the social network structure and requires neither function value optimization nor any artificial parameter. Our algorithm is divided into two phases. The first phase aims to identify dense micro-communities, that is, the prototype of a community. The second phase involves some adjustments in the subsequent label propagation to obtain the final communities. However, labels update quickly because only adjacent nodes affect the selection of node labeling. The proposed algorithm maintains the linear time complexity of the LPA. The main contribution of the proposed method is that a simple assignment of initial labeling in the LPA is discovered and that the method reduces the number of original labels and eliminates the accuracy limitation in traditional LPA. As shown in Fig. 2, this algorithm can mine two communities, namely, {*a*, *b*, *c*, *d*} and {*e*, *f*, *g*}, and thus provide a good way to solve the problem of accuracy limitation.

In some social networks, an individual may belong to multiple communities. For instance, in the paper cooperative network, someone may have more than one research field and may be connected with several research communities; this structure is referred to as an overlapping community [35]. In a social network such as a blog, we cannot simply think that the relationship among participants is identical. If different weights are adopted to represent the affinities of nodes, then a weighted network is formed. Thus, the next step is to determine how the ILAP can be applied in overlapping and weighted community mining.

Acknowledgment

This work is partially supported by the Xiamen University of Technology's International Cooperation and Exchange Project under Grant Nos. E201301300 and E201300200, the Fujian Province Department of Education Category A Projects under Grant Nos. JA13238 and JA12468, and the Project of Xiamen Science and Technology Program under Grant Nos. 3502Z20133041 and 3502Z201330447.

References

1. Girvan M., Newman M. E. J., "Community structure in social and biological networks", *Proceedings of the National Academy of Sciences*, 99(12), 2002, pp. 7821-7826.
2. Albert R., Jeong H. and Barabási A. L., "Error and attack tolerance of complex networks", *nature* 406(6794), 2000, pp. 378-382.
3. Hopcroft J., Khan O., Kulis B. and Selman B., "Tracking evolving communities in large linked networks", *Proceedings of the national academy of sciences of the United States of America,* 101(Suppl 1), 2004, pp. 5249-5253.
4. Spirin V., Mirny L. A., "Protein complexes and functional modules in molecular networks", *Proceedings of the National Academy of Sciences*, 100(21), 2003, pp. 12123-12128.
5. Liben□Nowell D., Kleinberg J., "The link□prediction problem for social networks", *Journal of the American society for information science and technology*, 58(7), 2007, pp. 1019-1031.
6. Chen W., Yuan Y. and Zhang L., "Scalable influence maximization in social networks under the linear threshold model", *IEEE 10th International Conference on Data Mining (ICDM)*, 2010, PP. 88-97.
7. Fortunato S., "Community detection in graphs", *Physics Reports*, 486(3), 2010, pp.75-174.
8. Leskovec J., Lang K. J. and Mahoney M., "Empirical comparison of algorithms for network community detection", *Proceedings of the 19th international conference on World Wide Web*, ACM, 2010, pp. 631-640.
9. Leskovec J., Lang K. J. and Dasgupta A, Mahoney M W, "Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters", *Internet Mathematics*, 6(1), 2009, pp. 29-123.
10. Danon L., Diaz-Guilera A., Duch J. and Arenas A., "Comparing community structure identification", *Journal of Statistical Mechanics: Theory and Experiment*, 09, 2005:P09008.
11. Radicchi F., Castellano C., Cecconi F., Loreto V. and Parisi D., "Defining and identifying communities in networks", *Proceedings of the National Academy of Sciences of the United States of America*, 101(9) , 2004, pp. 2658-2663.
12. Schaeffer S. E., "Graph clustering", *Computer Science Review*, 1(1), 2007, pp. 27-64.
13. Fortunato S., Barthelemy M., "Resolution limit in community detection", *Proceedings of the National Academy of Sciences*, 104(1), 2007, pp. 36-41.
14. Clauset A., Newman M. E. J. and Moore C., "Finding community structure in very large networks", *Physical review E*, 70(6), 2004:066111.
15. Brandes U, "A faster algorithm for betweenness centrality", *Journal of Mathematical Sociology*, 25(2), 2001, pp.163-177.
16. Newman M.E.J., Girvan M., "Finding and evaluating community structure in networks", *Physical review E* 69(2), 2004:026113.
17. Newman M.E.J., "Fast algorithm for detecting community structure in networks", *Physical review E* 69(6), 2004:066133.
18. Duch J., Arenas A., "Community detection in complex networks using extremal optimization", *Physical review E,* 72(2), 2005:027104.
19. Reichardt J., Bornholdt S., "Statistical mechanics of community detection", *Physical Review*, E74 (1), 2006:016110.
20. Pujol J.M., Béjar J. and Delgado J., "Clustering algorithm for determining community structure in large networks", *Physical Review E*, 74(1) , 2006:016107.
21. Sarkar S., Dong A., "Community detection in graphs using singular value decomposition", *Physical Review E*, 83(4), 2011:046114.

22. J. Leskovec, K.j. Lang, A. Dasgupta, and M.W. Mahoney, "Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters", *Internet Mathematics*, 6(1), 2009:29-123.
23. Jin D., He D., Hu Q., Baquero C. and Yang B., "Extending a configuration model to find communities in complex networks", Journal of Statistical Mechanics: Theory and Experiment, 09, 2013: P09013.
24. Raghavan U.N., Albert R. and Kumara S., "Near linear time algorithm to detect community structures in large-scale networks", *Physical Review E*, 76(3) , 2007:036106.
25. Zhao Z.-X., Wang Y.-T., Tian J.-T, and Zhou Z.-X., "A Novel Algorithm for Community Discovery in Social Networks Based on Label Propagation", *Journal of Computer Research and Development*, 48(Suppl.), 2011, pp.8-16. (in Chinese)
26. Wang Y., Feng X., "A potential-based node selection strategy for influence maximization in a social network", *Proceedings of Advanced Data Mining and Applications*, Chengdu, China, 2009, pp. 350-361.
27. Liu X.,Yi D.-Y., "Complex network community detection by local similarity", *ACTA AUTOMATICA SINICA*, 37(12) , 2011, pp.1520-1529. (in Chinese)
28. Watts D.J., Strogatz S.H., "Collective dynamics of small world networks", *Nature*, 393(6684), 1998, pp.440-442.
29. Huang J., Sun H., Han J., Deng H., Sun Y. and Liu Y., "SHRINK: a structural clustering algorithm for detecting hierarchical communities in networks", *Proceedings of the 19th ACM international conference on Information and knowledge management*, ACM, 2010, pp. 219-228.
30. Zachary W., "An Information Flow Model for Conflict and Fission in Small Groups1", *Journal of anthropological research*, 33(4), 1977, pp. 452-473.
31. Girvan M., Newman M.E.J, "Community structure in social and biological networks", *Proceedings of the National Academy of Sciences,* 99(12), 2002, pp. 7821-7826.
32. Leskovec J., Kleinberg J. and Faloutsos C., "Graph evolution: Densification and shrinking diameters", *ACM Trans. KDD*, 1(1), 2007, pp.1-41.
33. Ripeanu, M., Foster, I, Iamnitchi, A, "Mapping the Gnutella network: properties of large-scale peer-to-peer systems and implications for system design", *IEEE Internet Computing Journal*, 6(1), 2002, pp. 50-57.
34. Yang J., Leskovec J., "Defining and evaluating network communities based on ground-truth", *Proc. ACM SIGKDD Work shop on Mining Data Semantics*, 2012, pp.745-754.
35. Lancichinetti A., Fortunato S. and Kertész J., "Detecting the overlapping and hierarchical community structure in complex networks", *New Journal of Physics,* 11(3), 2009:033015.