

Signed Network Label Propagation Algorithm with Structural Balance Degree for Community Detection

Lei Fang^(✉), Qun Yang, Jiawen Wang, and Weihua Lei

College of Computer Science and Technology,
Nanjing University of Aeronautics and Astronautics, Nanjing, China
fanglei_nuaa@139.com, Qun.Yang@nuaa.edu.cn,
{287568706, 623659607}@qq.com

Abstract. Social networks are usually modeled as signed networks. The community detection is an important problem for the research of signed networks. The time complexity of signed label propagation algorithm is lower than most existing algorithms for community detection in the signed networks. However, bad performance on robustness and accuracy in the algorithm should not be ignored. Thus, we propose a structural balance degree to measure the balance of an edge in the local network and the local network density. Then a novel signed network label propagation algorithm with structural balance degree is proposed for community detection in signed networks. Besides, the algorithm is tested on several real-world social networks. Experimental results prove that the optimized algorithm can enhance both the robustness and the effectiveness. Its convergence rate is also faster than current algorithms.

Keywords: Community detection · Signed networks · Structural balance degree · Label propagation algorithm

1 Introduction

There are various interpersonal relationships in real world, such as friends, relatives and even enemies. The network with binary relationship cannot simply be modeled as an unsigned network, which can cause the core information loss of the edge. For this reason, the network is usually modeled as a signed network with both positive and negative edges. The community detection problem [1, 2] is a key research point in social network analysis. Nodes located in the same community are usually closely connected, but they may be sparsely connected to other nodes in different communities. Many approaches, such as user interaction based algorithm [3], clustering algorithm [4, 5], information cascade-based model [6], and semantic network-based algorithm [7] have been proposed to tackle the community detection problem. Raghavan et al. proposed the label propagation algorithm (LPA) for community detection in unsigned networks [8]. It is easy to implement and it has linear time complexity [9]. Šubelj et al. [10] found that random node update orders within the algorithm severely hamper its robustness, and consequently also the stability of the identified community structure.

So they proposed a balanced propagation that counteracts for the introduced randomness by utilizing node balancers. It is for the unsigned network which is not our research point in this paper.

Triangle structural balance theory [11], which is considered as the basis of the social psychology analysis of the signed network theory, has a great impact on the research of the signed network. Nakanishi et al. [12] research whether software agents can influence human relations by using balance theory in agent-mediated communities. It shows that the balance theory can be used in social network analysis. But the problem which they research on is not a community detection problem.

Compared to unsigned networks, it is necessary to consider about the sign information of edge when detecting communities in the signed network. As for the problem, we presented the signed network label propagation algorithm based on LPA first. Then, we presented the structural balance degree which is used to measure the balance of each edge in the signed network. Last, we presented the signed network label propagation algorithm with structural balance degree. The algorithm can detect communities in signed networks effectively and its results are stable.

2 Related Works

Researchers tend to adopt two kinds of methods to solve the community detection problem in signed network now.

2.1 Whole-Network-Based Algorithm

It sets an objective function of the signed network as the measurement of the network partitioning result and it is based on the whole network information. It finds optimal partition ways by optimizing the objection function solution of the case. The modularity is put forward and used to measure different network partitioning results in unsigned complex networks. Gomez et al. [13] developed the modularity function to discover communities in signed networks. It splits network into a positive part and a negative part and then the corresponding subgraph modularity is calculated respectively. The weighted sum of plus and minus modularity is the overall modularity of the signed network. Modularity is a measurement function originated from the concept of network connection density. Thus, this classification method is rather closer to the definition of the community.

Traag et al. [14] extend an existing Potts model to incorporate negative links as well, resulting in a method similar to the clustering of signed graphs. They opted for simulated annealing to minimize an objective function. Its time complexity is much higher than the time complexity of LPA and the improved algorithms.

Li et al. [15] proposed a two-stage signed network community discovery algorithm. In the first phase, the GN algorithm is used to get a preliminary division in subnet composed of positive edges. In the second phase, a hierarchical clustering method is used to combine the negative edges between different positive subnets.

Bansal et al. [16] introduced the correlation clustering algorithm into the document network and proposed two clustering goals. One is to maximize the sum of the number of internal plus edges in clusters and the number of negative edges between different clusters. The other is to minimize the sum of the number of internal negative edges in cluster and the number of positive edges between different clusters. They also proved that the two kinds of clustering goals in the signed network are NP- hard problems.

2.2 Local-Information-Based Algorithm

It is a local community discovery algorithm. Yang et al. [17] presented finding extracting community (FEC) algorithm of local random walks on the signed network to find communities. The time complexity of FEC algorithm is linear and it performs well in terms of time efficiency. But the weakness is that the experimental results are not stable because of the random selection of initial node in the process of random walks and the uncertainty of random walks steps.

Shama et al. [18] proposed a two-step cluster CRA (clustering re-clustering algorithm) algorithm. Compared to FEC, CRA is more stable because the two step clustering could be affirmatory without the artificial set of parameters.

3 Signed Network Label Propagation Algorithm

In some particular cases, if there are two strong communities in the signed network, which are only connected by negative edges, labels cannot spread across them. The problem is also related to label propagation in two unsigned subnets and it suggests that the label propagation algorithm is feasible on the signed network. For this reason, we develop a sign network label propagation algorithm (SLPA). It does not require the division of the positive and negative subsets in the clustering process. Combining label propagation algorithm with edges' sign, SLPA detects communities directly instead of dividing network into positive and negative subnets.

$G(V, E_+, E_-)$ is an unweighted and undirected graph which can represent the signed social network. V stands for all the nodes in graph G , and E_+ is the set of all positive edges; E_- is the set of all negative edges. e_{ij} is the edge that connects the node v_i with the node v_j .

l_i is the label of the node $v_i (v_i \in V)$ and the set of the node v_i 's neighbors is represented by N_i . N_i^+ is the set of nodes connected with the node v_i by positive edges, and N_i^- is the set of nodes connected with the node v_i by negative edges. $N(v) = N_i^+ \cup N_i^-$, and $N_i^+ \cap N_i^- = \emptyset$.

Each node is only affected by its neighbors. If there is a positive edge between the node and its neighbor, it would get the label from the neighbor node. But if there is a negative edge between the node and its neighbor, it would reject the label from the neighbor node. After several iterations, the nodes, which are closely-connected to positive edges, receive the same label. Finally, the nodes with the same label are clustered into the same community. The formula of l_i is shown below:

$$l_i = \arg \max_l (|N_i^+(l)| - |N_i^-(l)|) \quad (1)$$

$N_i^+(l)$ is the set of nodes with the label l in N_i^+ . $N_i^-(l)$ is the set of nodes with the label l in N_i^- . The formula shows that if there are nodes with the label l in neighbor nodes of v_i and the difference value of nodes with the label l in N_i^+ and N_i^- is the maximum, the node v_i selects the label l as its new label.

SLPA can detect communities in signed networks, but it does not perform well in terms of robustness due to its strong randomness. Besides, neither network density relations nor the structural balance of the signed network are taken into consideration. Thus, the label propagation algorithm remains to be improved to solve the problems.

4 The Structural Balance Degree of Signed Network

In an unsigned network, if two adjacent nodes have a common adjacent node, the two nodes and their common adjacent node would constitute a triangle. When calculating the similarity of the two nodes, the number of their common adjacent nodes is an essential part of the operator. According to the definition of the community, the joint situation between adjacent nodes of the two nodes reflects the connection density of the local network which the two nodes belong to. Since each edge in the signed network can be positive or negative, there exist two kinds of relationships between two connected nodes: attraction and mutex. And their common adjacent node can have various relations with the two nodes.

There may be a positive edge or a negative edge between two nodes in the signed network. If the edge e_{12} is a positive edge, there are four types of triangles, which the edge maybe belong to, as shown in Fig. 1:

According to the triangle structural balance theory of the signed network, Fig. 1(a) and (d) are balanced, but Fig. 1(b) and (c) are unbalanced. In these unbalanced triangles, the relation among the three nodes still changes dynamically and their final status remains unknown. For example, in Fig. 1(b), if the positive relationship of edge e_{13} and e_{12} is very strong, but the negative relationship of node 2 and node 3 is not strong, the negative relationship of edge e_{23} is likely to be converted to the positive relationship. Thus, in unbalanced triangle, the positive or negative nature of the edges may gain more possibility to change. The unbalanced triangle in the signed network does not take part in the calculation of the structural balance degree of edges.

The structural balance degree is used to measure the balance degree of each edge in the signed network. Its value can use the number of balanced triangles which are constituted of the edge and its adjacent edges. For the edge e_{ij} , its structural balance degree is K_{ij} .

If the structural balance degree of the edge e_{ij} is higher, its sign nature would be more stable, and its local network density would be bigger. For a positive edge, if the structural balance degree is higher, the two nodes connected by the edge would be closer and they gain more probability to be in the same community. For a negative

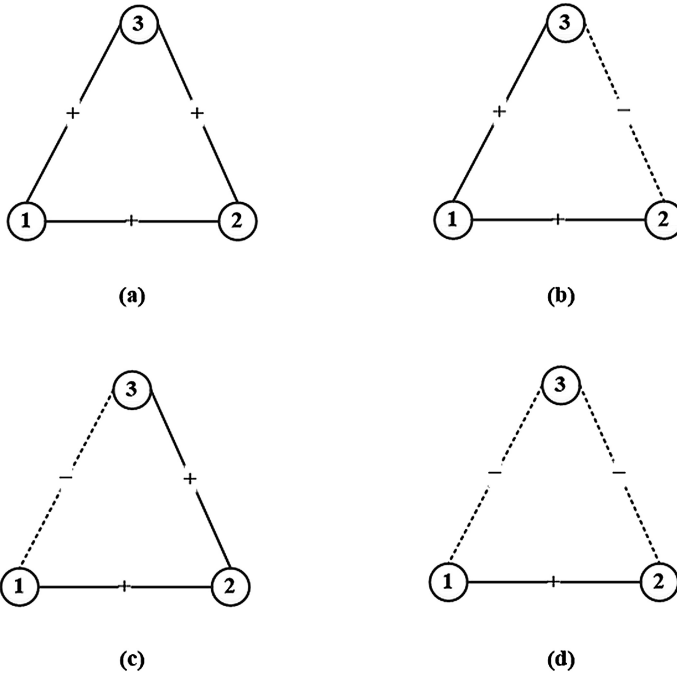


Fig. 1. Four types of triangle with a positive edge e_{12} .

edge, if the structural balance degree is higher, the two nodes connected by the edge would be more distant and they gain less probability to be in the same community.

As edges have the nature of plus or minus sign, the structural balance degree of the positive edge should be positive and that of negative edge should be negative. The product decision method helps to determine whether a triangle is balanced. For a triangle, if the product of the three edges' signs is $+1$, the triangle would be balanced else the triangle would be unbalanced. Figure 2 shows the calculation examples of the structural balance degree of the four basic triangles.

Concrete steps of the structural balance degree are as follows:

1. In a signed network, $e_{ij} \in E$ and its structural balance degree is set by its sign. If it's a positive edge, its default structural balance degree K_{ij} would be $+1$. If it's a negative edge, its default structural balance degree K_{ij} would be -1 .
2. For the edge $e_{ij} \in E$, the nodes connected to v_i and v_j make up a set N_{ij} .
3. For each $v_k \in N_{ij}$, v_i , v_j and v_k constitute a signed triangle. If the product of e_{ij} , e_{ik} and e_{jk} is $+1$, the structure of the triangle would be balanced. If the product of e_{ij} , e_{ik} and e_{jk} is -1 , the structure of the triangle would be unbalanced. When the triangle is balanced, if e_{ij} is positive, K_{ij} add 1; if e_{ij} is negative, K_{ij} minus 1.
4. The resulting K_{ij} is the structural balance degree of e_{ij} .

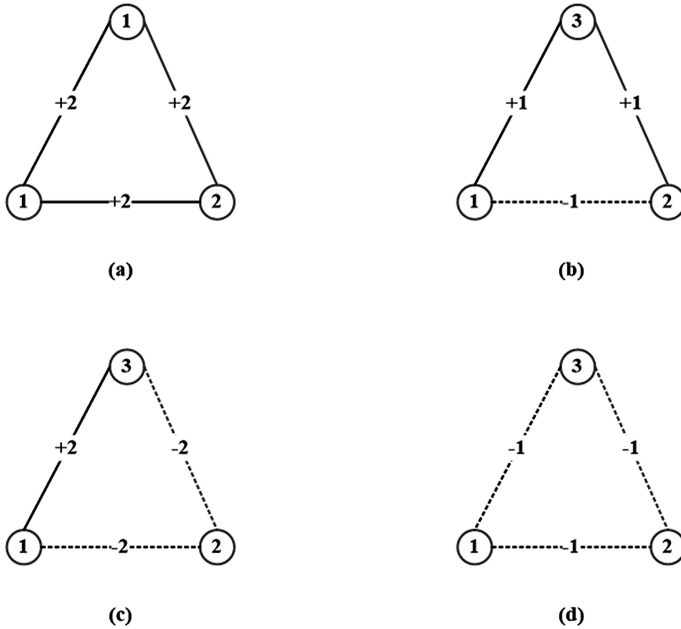


Fig. 2. The structural balance degree of four signed triangles

5 Signed Network Label Propagation Algorithm with Structural Balance Degree

Neither the network density nor the structural balance of the signed network is taken into consideration in SLPA. Thus, the structural balance degree is proposed to measure the balance of each edge in the signed network. It not only reflects the stability of the sign, but also measures local network density. This is primarily because the structural balance degree also has sign. It can cause the label to spread easily in dense positive subnets but difficultly in dense negative subnets. For this reason, we propose the sign network label propagation algorithm with structural balance degree (SBDSLPA). The label updating strategy is shown as the following formula.

$$l_i = \arg \max_l \sum_{l_j=l, v_j \in N_i} K_{ij} \tag{2}$$

l_i is the label of the node $v_i (v_i \in V)$. If v_i is connected with $v_j, e_{ij} \in E$, and its structural balance degree is K_{ij} .

The steps of SBDSLPA are as follows:

1. Each node in the signed network is labeled by a unique integer value. This label represents the community that the node belongs to.
2. Compute the structural balance degree of each edge in the signed network.

3. All nodes in V are arranged in a random order, and node's labels are updated in turn. The new label of the node v_i is calculated according to the formula (2).
4. Repeat step 3 until labels of all nodes no longer change and the algorithm convergence. At last nodes with the same label belongs to the same community.

If there are both positive and negative edges between v_i and the nodes which belong to the Community C , it would be easy to judge whether the positive relationship or the negative relationship is stronger between Community C and v_i , according to the sign of the sum value of the edges' structural balance degree. This determines whether the node v_i belongs to the community C . Label propagation algorithm is an intelligent clustering algorithm based on the label preference of nodes. The structural balance degree can give the spontaneous clustering a control guide.

6 Experiments and Discussion

In order to compare SLPA with SBDSLPA, some experiments are conducted on three different benchmark signed network data sets. The algorithms are implemented by the standard C language.

The experimental environment is shown in Table 1:

Table 1. The experimental environment

Operating system	Ubuntu 12.10
Compiler	gcc 4.8
Memory	4 GB
CPU	3.20 GHz

The experiment network data sets are three true signed networks. Their specific data attributes are as shown in Table 2.

Table 2. The signed social network datasets

Network	Nodes	Edges	Edges(+)	Edges(-)
Gahuku-Gama Subtribes Network (GGSN)	16	116	58	58
Sampson Monastery Network (SMN)	18	158	80	78
Slovene Parliamentary Network (SPN)	10	45	18	27

We run SLPA and SBDSLPA for 50 times on the three datasets. For the generating communities, the average signed network modularity (SQ) and average normalized mutual information degree are computed and the statistical results are shown in Tables 3 and 4 respectively. The average running time is displayed in Table 5.

The average signed network modularity of SLPA and SBDSLPA on the three different network data sets is recorded in the Table 3. The average signed network

Table 3. The average signed network modularity

Network	SLPA	SBDSLPA
GGSN	0.35	0.43
SMN	0.44	0.57
SPN	0.41	0.45

Table 4. The average normalized mutual information

Network	SLPA	SBDSLPA
GGSN	0.83	0.97
SMN	0.85	0.92
SPN	0.93	1.00

Table 5. The average execution time (ms)

Network	SLPA	SBDSLPA
GGSN	45	37
SMN	41	29
SPN	43	35

modularity of SBDSLPA is larger than SLPA by 20.7 percent. It shows that the result of SBDSLPA is closer to the real community classification.

Table 4 provides the average normalized mutual information of the experimental results on the three signed network data sets. The larger the value of NMI is, the more stable the results from the community detection algorithms are. From the table, the average value of NMI of the SBDSLPA is larger than SLPA on all the three test network datasets. On average, SBDSLPA's NMI is larger than SLPA's NMI by 10.9 percent. So SBDSLPA is more stable than SLPA.

The average execution time of SLPA and SBDSLPA on the three datasets is given in Table 5. From the table we can see that the average execution time of SBDSLPA is shorter than the SLPA. SBDSLPA's average execution time is shorter than SLPA by 21.9 percent. This shows that the convergence rate of SBDSLPA is much faster than SLPA on the same datasets.

7 Conclusions

Label propagation algorithm is a linear time algorithm for community detection without using any prior parameters. It is widely used for solving community detection problems in unsigned networks. In this paper, we analyze the triangle structural balance theory of signed networks and propose a community detection algorithm in the signed network called Signed Network Label Propagation Algorithm with Structural Balance Degree (SBDSLPA). First, we present a novel measurement to the structural balance degree of each edge in the signed network. Then, this measure is used to control the label propagation process. It makes labels propagate easily in the subset where the positive

edges are intensive, but difficultly in the subnet where the negative edges are intensive. The experimental results show that SBDSLPA is more effective and stable than SLPA in the signed network.

References

1. Girvan, M., Newman, M.E.: Community structure in social and biological networks. *Proc. Natl. Acad. Sci.* **99**, 7821–7826 (2002)
2. Newman, M.E.: The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003)
3. Dev, H.: A user interaction based community detection algorithm for online social networks. In: *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, pp. 1607–1608 (2014)
4. Lin, W.Q., Kong, X.N., Yu, P.S., Wu, Q.Y., Jia, Y., Li, C.: Community detection in incomplete information networks. In: *Proceedings of the 21st International Conference on World Wide Web*, pp. 341–350 (2012)
5. Jin, H., Wang, S.L., Li, C.Y.: Community detection in complex networks by density-based clustering. *Phys. A Stat. Mech. Appl.* **392**, 4606–4618 (2013)
6. Barbieri, N., Bonchi, F., Manco, G.: Cascade-based community detection. In: *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, pp. 33–42 (2013)
7. Xia, Z.Y., Bu, Z.: Community detection based on a semantic network. *Knowl. Inf. Syst.* **26**, 30–39 (2012)
8. Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. *Phys. Rev. E* **76**, 036106 (2007)
9. Šubelj, L., Bajec, M.: Unfolding communities in large complex networks: combining defensive and offensive label propagation for core extraction. *Phys. Rev. E* **83**, 885–896 (2011)
10. Šubelj, L., Bajec, M.: Robust network community detection using balanced propagation. *Phys. Condens. Matter* **81**, 353–362 (2011)
11. Heider, F.: Attitudes and cognitive organization. *J. Psychol.* **21**, 107–112 (1946)
12. Nakanishi, H., Nakazawa, S., Ishida, T., et al.: Can software agents influence human relations?: balance theory in agent-mediated communities. In: *International Joint Conference on Autonomous Agents & Multiagent Systems, AAMAS 2003*, 14–18 July 2003, Melbourne, Victoria, Australia, *Proceedings*, pp. 717–724 (2003)
13. Sergio, G., Pablo, J., Alex, A.: Analysis of community structure in networks of correlated data. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **80**, 016114 (2009)
14. Traag, V.A., Bruggeman, J.: Community detection in networks with positive and negative links. *Phys. Rev. E* **80**, 036115 (2009)
15. Li, X., Chen, H., Li, S.: Exploiting emotions in social interactions to detect online social communities. In: *PACIS* (2010)
16. Bansal, N., Blum, A., Chawla, S.: Correlation clustering. *Mach. Learn.* **56**, 89–113 (2004)
17. Yang, B., Cheung, W.K., Liu, J.: Community mining from signed social networks. *IEEE Trans. Knowl. Data Eng.* **19**, 1333–1348 (2007)
18. Sharma, T., Charls, A., Singh, P.K.: Community mining in signed social networks—an automated approach. In: *Proceedings of the International Conference on Computer Engineering and Applications*, pp. 152–157. IACSIT Press, Singapore (2009)