Measuring the Similarity of Nodes in Signed Social Networks with Positive and Negative Links

Tianchen Zhu[1], Zhaohui Peng[1(✉)], Xinghua Wang[1] and Xiaoguang Hong[1]

[1]School of Computer Science and Technology, Shandong University, Jinan, China
ztc@mail.sdu.edu.cn, pzh@sdu.edu.cn,
wang.xingh@foxmail.com, hxg@sdu.edu.cn

Abstract. Similarity measure in non-signed social networks has been extensively studied for decades. However, how to measure the similarity of two nodes in signed social networks remains an open problem. It is challenging to incorporate both positive and negative relationships simultaneously in signed social networks due to the opposite opinions implied by them. In this paper, we study the similarity measure problem in signed social networks. We propose a basic node similarity measure that can utilize both positive and negative relations in signed social networks by comparing the immediate neighbors of two objects. Moreover, we exploit the propagation of similarity in networks. Finally, we perform extensive experimental comparison of the proposed method against existing algorithms on real data set. Our experimental results show that our method outperforms other approaches.

Keywords: similarity measure, signed networks, positive and negative links

1 Introduction

Measuring the similarity of nodes in signed social networks[1, 2, 3] is an important but still not fully explored problem due to the following challenges. First, the existence of negative links in signed networks challenges many concepts for unsigned networks. In addition, it is challenging to incorporate both positive and negative relationships simultaneously in signed networks. As shown in Fig. 1, users can express trust or distrust on others in Epinions.com, in which “+” indicates trust and “-” indicates distrust.

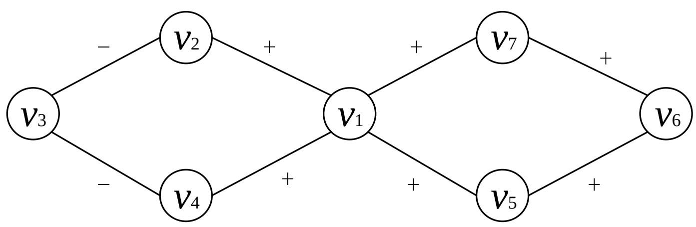


Fig. 1. Example of Signed Social Network.

adfa, p. 1, 2011.
© Springer-Verlag Berlin Heidelberg 2011

We can see from Fig.1, user v_1 and user v_3 have two common neighbors, user v_1 and user v_6 have two common neighbors as well. If we ignore the polarity of this social network, we can conclude that user v_3 and user v_6 are both similar to user v_1. However, when we take into account the polarities of the links, things go differently. From the Fig.1, we can see that user v_1 and user v_3 have different evaluations for the same persons, while user v_1 and user v_6 have identical evaluations for the same persons. So we can deduce that user v_1 is more similar to user v_6, and user v_1 is more dissimilar to user v_3.

In this paper, we study the problem of measuring the similarity of nodes in signed social networks. Compared to existing approaches, our method considers the influence of negative relationships in addition to positive relationships. In particular, we define a basic similarity measure that captures effectively the proximity of two nodes in signed social networks. Besides, we also exploit the propagation of similarity in social networks. For example, considering the user v_3 and user v_6 in Fig. 1, these two users have no immediate neighbors in the signed social network, however, we can infer that user v_3 is dissimilar to user v_6 because user v_3 is dissimilar to user v_1 and user v_1 is similar to user v_6.

The major contributions of this paper are summarized as below.

- It investigates the similarity measure problem in signed social networks, a new but increasingly important issue due to the continuous development of social networks.
- It proposes a basic similarity measure method in signed social networks, which can utilize both positive and negative relationships.

The rest of this paper is organized as follows. Section 2 summarizes the related work, whereas section 3 denotes a node similarity measure in signed social networks. Experimental results are given in section 4. Finally, section 5 concludes this paper.

2 Related Work

Unsigned social networks have been studied for decades [4, 5, 6]. There are many similarity measure method based on unsigned social networks. Traditional approaches to measure the similarity of nodes are mainly based on their feature values, such as Jaccard coefficient, Cosine similarity and Euclidean distance. However, these similarity measures are mainly focus on the nodes features, they do not consider link structures among objects.

The link based approaches measure the similarity of nodes defined on networks, such as shortest path algorithm, RWR (Random Walk with Restart) [7], SimRank [8], Personalized PageRank [9] and etc. Liben and Kleinberg [10] claimed that the identification of the shortest path between any pair of nodes in a graph can be used for friend recommendation. RWR iteratively explores the global structure of the network to estimate the proximity between two nodes. Starting at a node, the walker faces two choices at each step: either moving to a randomly chosen neighbor, or jumping back to the starting node. SimRank is a symmetric similarity measure that says “two objects are considered to be similar if they are referenced by similar objects”. Personalized PageRank is an asymmetrical similarity measure that evaluates the probability starting from source nodes to target nodes by RWR. However, all above these methods do not consider the negative links, i.e., these studies are built on a fundamental assumption that all links in networks are positive.

FriendTNS [11, 12] discussed the similarity measure problem in signed social networks based on status theory. FriendTNS evaluates the similarity of two nodes in terms of their positive in-degree, negative out-degree, positive out-degree and negative in-degree. However, FriendTNS considers the degree of nodes only, it does not consider the relationship between two nodes. Besides, FriendTNS cannot apply in undirected signed social networks.

3 Node Similarity Measure

In this section, we define a basic node similarity measure to determine the proximity between a pair of nodes in signed social networks. For node v_i and node v_j, we define a specific function $sim(v_i, v_j)$ to express their corresponding similarity.

3.1 Basic Idea

To capture proximity between two nodes, we consider the example shown in Fig. 2. Fig. 2(a) means user u_1 and user u_2 have many common friends, and Fig. 2(b) means that user u_1 and user u_2 have many common enemies. We can infer that u_1 and u_2 are likely to be similar, because they have the same evaluations on the same people. In contrast, Fig. 2(c) and Fig. 2(d) show that the evaluations of user u_1 and user u_2 on the same people are diametrically opposite. So, we can infer that u_1 and u_2 are most likely not similar.

This inspires us that if a pair of users will have high similarity in a signed social network, they should be satisfied the following condition:

- They have many common friends.
- They have many common enemies.

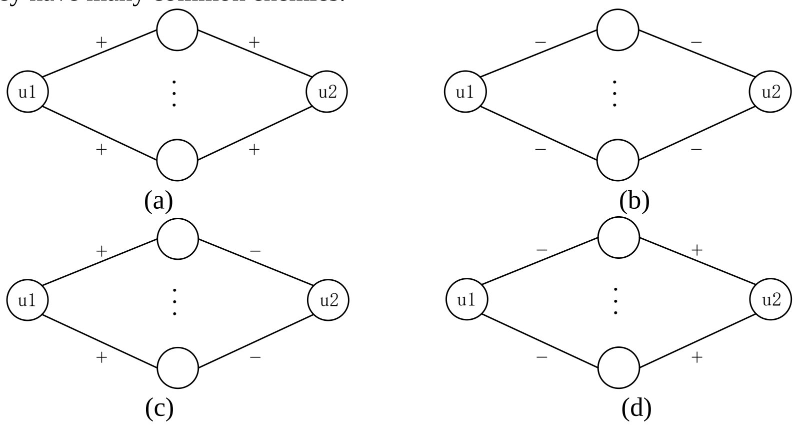


Fig. 2. Example of two neighbor connected users (nodes) in a signed social network.

3.2 Similarity Measure in Undirected Signed Social Networks

Based on the above intuition, in an undirected signed social network, given a pair of nodes v_i and v_j, we can determine their similarity by considering their immediate neighbor sets. The followings are the immediate neighbor nodes, positive immediate neighbor nodes and negative immediate neighbor nodes of node v_i.

$$N_i = \{v_k \mid (v_i, v_k) \in \mathcal{E}\} + \{v_i\} \tag{1}$$

$$N_i^+ = \{v_k \mid (v_i, v_k) \in \mathcal{E}^+\} + \{v_i\} \tag{2}$$

$$N_i^- = \{v_k \mid (v_i, v_k) \in \mathcal{E}^-\} \tag{3}$$

We have discussed above that the more same evaluations on common neighbors two nodes have, the more similar they are. In contrast, the more contrary evaluations on common neighbors two nodes have, the more dissimilar they are.

Given a pair of nodes v_i and v_j, we can get the sets of nodes that they hold same evaluations and different evaluations.

$$C_s(v_i, v_j) = \{v_k \mid \left(v_k \in N_i^+ \ \wedge \ v_k \in N_j^+\right) \vee (v_k \in N_i^- \ \wedge \ v_k \in N_j^-)\} \tag{4}$$

$$C_d(v_i, v_j) = \{v_k \mid \left(v_k \in N_i^+ \ \wedge \ v_k \in N_j^-\right) \vee (v_k \in N_i^- \ \wedge \ v_k \in N_j^+)\} \tag{5}$$

The similarity of these two nodes can be determined as follows.

$$sim\left(v_i, v_j\right) = \frac{|C_s(v_i,v_j)|-|C_d(v_i,v_j)|}{|N_i \cup N_j|} \tag{6}$$

The node similarity measure in undirected signed social networks returns values into the interval [-1, 1]. Note that the maximum value of similarity (equal 1) and minimum value of similarity (equal -1) can be reached. When the two nodes have totally the same neighbors and the evaluation on each neighbor is the same, then the similarity of these two nodes is equal 1. However, when the two nodes have totally the same neighbors, but the evaluation on each neighbor is completely different, then the similarity of these two nodes is equal to -1. The larger the value, the more similar these two nodes are. The similarity value 1 indicates that the two nodes are extremely similar, and the value -1 indicates that the two nodes are extremely dissimilar.

Now, let us calculate some similarity values on the graph of Fig. 1 using Eq. 6. Considering the node v_1 and node v_3, N_1 = { v_1, v_2, v_4, v_5, v_7}, N_3 = {v_2, v_3, v_4}, $N_1^+ = N_1$, $N_1^- = \emptyset$, N_3^+={ v_3}, N_3^-={ v_2, v_4}, $C_s(v_1, v_3) = \emptyset$, $C_d(v_1, v_3)$ = {v_2, v_4}, so $sim(v_1, v_3) = \frac{0-2}{6} = -\frac{1}{3}$. Similarly, the similarity between node v_1 and v_6 is: $sim(v_1, v_6) = \frac{1}{3}$. Thus, the similarity score between nodes v_1, v_3 is less than that of v_1, v_6, and the result is also consistent with our intuitive experience.

3.3 Similarity Measure in Directed Signed Social Networks

In this section, we present how to measure similarity of a pair of nodes in a directed signed social network. In a directed social network, a user can express his attitudes to others, but can also receive the evaluations by other people. So, we separate neighbor set into four parts: (1) positive input neighbor set; (2) negative input neighbor set; (3) positive output neighbor set; (4) negative output neighbor set. Taking Fig. 3 as an example, we consider the similarity between node u_1 and node u_2.

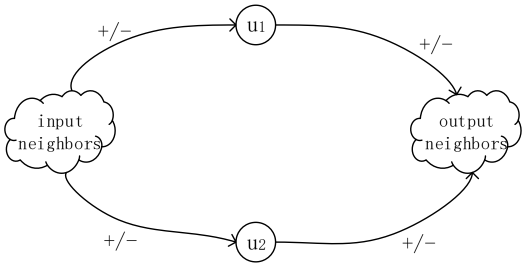


Fig. 3. Illustration of the intuition that measure similarity in a directed signed social network.

Based on the above intuition, u_1 and u_2 may be similar if:

- The evaluations on the two users that given by a lot of people are the same.
- The evaluations on a group of persons that given by the two users are the same.

Besides, if the two users hold different evaluations on many same persons and many people have different opinions about the two users, they may be dissimilar. It is an extension of the intuition we described in section 3.1.

The followings are the positive input neighbor set, negative input neighbor set, positive output neighbor set and negative output neighbor set of node v_i in a directed signed social network.

$$I_i^+ = \{v_k \mid (v_k, v_i) \in \mathcal{E}^+ \} + \{v_i\} \tag{7}$$

$$I_i^- = \{v_k \mid (v_k, v_i) \in \mathcal{E}^- \} \tag{8}$$

$$O_i^+ = \{v_k \mid (v_i, v_k) \in \mathcal{E}^+ \} \tag{9}$$

$$O_i^- = \{v_k \mid (v_i, v_k) \in \mathcal{E}^- \} \tag{10}$$

So, given a pair of nodes v_i and v_j in a directed signed social network,

$$C_s^I(v_i, v_j) = \{v_k | (v_k \in I_i^+ \wedge v_k \in I_j^+) \vee (v_k \in I_i^- \wedge v_k \in I_j^-)\} \tag{11}$$

$$C_d^I(v_i, v_j) = \{v_k | (v_k \in I_i^+ \wedge v_k \in I_j^-) \vee (v_k \in I_i^- \wedge v_k \in I_j^+)\} \tag{12}$$

$$C_s^O(v_i, v_j) = \{v_k | (v_k \in O_i^+ \wedge v_k \in O_j^+) \vee (v_k \in O_i^- \wedge v_k \in O_j^-)\} \tag{13}$$

$$C_d^O(v_i, v_j) = \{v_k | (v_k \in O_i^+ \wedge v_k \in O_j^-) \vee (v_k \in O_i^- \wedge v_k \in O_j^+)\} \tag{14}$$

where $C_s^I(v_i, v_j)$ means the common neighbors that hold the same evaluations on v_i and v_j, $C_d^I(v_i, v_j)$ means the common neighbors that hold different evaluations on v_i and v_j, $C_s^O(v_i, v_j)$ means the common neighbors that v_i and v_j hold the same evaluations on, and $C_d^O(v_i, v_j)$ means the common neighbors that v_i and v_j hold different evaluations on. So, the similarity between v_i and v_j is:

$$sim(v_i, v_j) = \frac{|C_s^I(v_i,v_j)| - |C_d^I(v_i,v_j)| + |C_s^O(v_i,v_j)| - |C_d^O(v_i,v_j)|}{|N_i \cup N_j|} \tag{15}$$

where $N_i = I_i \cup O_i$, $N_j = I_j \cup O_j$.

The basic similarity measure defined in directed signed social networks returns values into the interval [-1, 1] as well. Interval (0, 1] indicates two nodes are similar and interval [-1, 0) indicates two nodes are dissimilar. Specifically, 0 is a special value that denotes neither similar nor dissimilar. The larger the similarity value, the more similar they are.

3.4 Propagation of Node Similarity

Based on the above discussion, the similarity values between all non-neighbor nodes in a graph $\mathcal{G}$ are zero. It is unreasonable. By propagating the similarity in the network, we can solve this problem. Notice that, we only propagate the similarity in the network in two hops according to the principle of balance theory [13, 14] that says "*the friend of my friend is my friend*", "*the enemy of my enemy is my friend*". That is to say, if two users have neither immediate common neighbors nor common related users (i.e. the similarity value determined by the basic similarity measure is not zero), the similarity between them is zero eventually. We define a extended similarity between two nodes v_i and v_j, denoted as *exsim*(v_i, v_j).

$$exsim(v_i, v_j) = \begin{cases} sim(v_i, v_j), \text{ if } \mathrm{v_i}, \mathrm{v_j} \text{ have immediate common neighbors} \\ \dfrac{1}{|\mathcal{R}(v_i, v_j)|} \sum_{v_k \in \mathcal{R}} sim(v_i, v_k) \cdot sim(v_k, v_j), otherwise \end{cases} \tag{16}$$

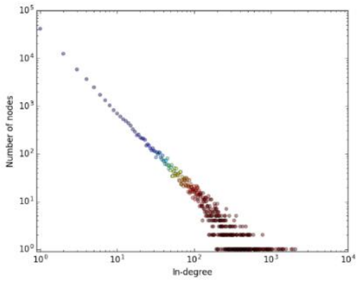
where $\mathcal{R}(v_i, v_j)$ is the set of common related users between user v_i and user v_j, $\mathcal{R}(v_i, v_j) = \{v_k \mid sim(v_i, v_k) \neq 0 \wedge sim(v_k, v_j) \neq 0\}$.

4 Experiments

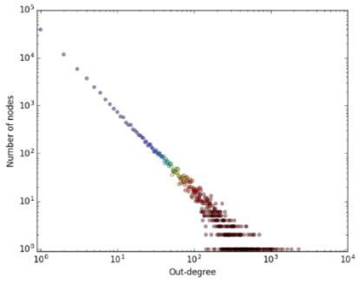
In this section, we compare experimentally our approach with existing similarity measure algorithms.

4.1 Data Sets

We use the Epinions [15] data set, which is a signed social network.We compute the in-degree and out-degree distributions of Epinions graph, treating both the positive and negative edges alike, as shown in Fig. 4.



(a) In-degree



(b) Out-degree

Fig. 4. Degree distributions of Epinions data set.

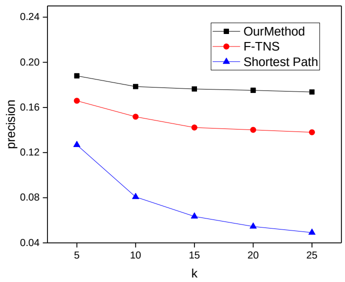
4.2 Experimental Setup and Evaluation

To evaluate the proposed approach, we recommend a list of k friends (top-k list) to a target user. Our evaluation considers the division of neighbors of each target user into two sets: (1) the training set $\mathcal{E}_T$ is treated as known information and, (2) the verification set $\mathcal{E}_V$ is used for verification. We use *precision* and *accuracy* metric as performance measures for recommendations. In order to validate the effectiveness of our approach in signed social networks, we compare our model with FriendTNS algorithm and the Shortest Path algorithm, denoted as F-TNS and Shortest Path, respectively.

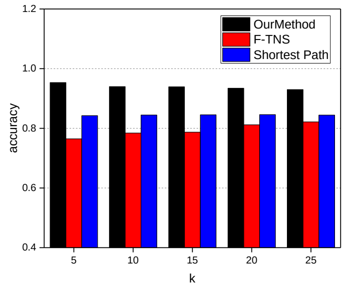
4.3 Experimental results

In this section, we quantitatively compare our proposed method with baselines. Fig. 5 shows the results conducted in Epinions data set.

The horizontal axis of Fig.5 represents the number of friends that we recommend to a target user. From the Fig.5, we can see that our method can achieve better performance than baselines. The main reason is that our proposed method takes into account both the positive relationships and negative relationships. It shows from a side view that the negative links are important and valuable.



(a) Precision



(b) Accuracy

Fig. 5. Comparison of proposed method with baselines in Epinions.

5 Conclusion

In this paper, we study the similarity measure problem in signed social networks. We proposed a novel method to measure the similarity of nodes in signed social networks. Our method incorporates both positive and negative relationships simultaneously in signed networks. Extensive evaluations demonstrate the effectiveness of our approach.

Acknowledgement. This work is supported by NSF of China (No. 61602237), 973 Program (No. 2015CB352501), NSF of Shandong, China (No. ZR2013FQ009), the Science and Technology Development Plan of Shandong, China (No. 2014GGX101047, No. 2014GGX101019).

References

1. J. Kunegis, A. Lommatzsch, C. Bauckhage: The slashdot zoo: Mining a social network with negative edges. WWW 2009: 741-750
2. J. Leskovec, D. P. Huttenlocher, J. M. Kleinberg: Predicting positive and negative links in online social networks. WWW 2010: 641-650
3. J. Tang, Y. Chang, C. Aggarwal, H. Liu: A Survey of Signed Network Mining in Social Media. ACM Comput. Surv. 49(3): 42:1-42:37 (2016)
4. S. Wang, X. Hu, P. S. Yu, Z. Li: MMRate: inferring multi-aspect diffusion networks with multi-pattern cascades. KDD 2014: 1246-1255
5. S. Wang, Z. Yan, X. Hu, P. S. Yu, Z. Li: Burst Time Prediction in Cascades. AAAI 2015: 325-331
6. C. Shi, Y. Li, J. Zhang, Y. Sun, P. S. Yu: A Survey of Heterogeneous Information Network Analysis. IEEE Trans. Knowl. Data Eng. 29(1): 17-37 (2017).
7. J. Pan, H. Yang, C. Faloutsos, P. Duygulu: Automatic multimedia cross-modal correlation discovery. KDD 2004: 653-658
8. G. Jeh, J. Widom: SimRank: a measure of structural-context similarity. KDD 2002: 538-543
9. G. Jeh, J. Widom: Scaling personalized web search. WWW 2003: 271-279
10. D. Liben-Nowell, J. M. Kleinberg: The link prediction problem for social networks. CIKM 2003: 556-559
11. P. Symeonidis, E. Tiakas, Y. Manolopoulos: Transitive node similarity for link prediction in social networks with positive and negative links. RecSys 2010: 183-190
12. P. Symeonidis and E. Tiakas: Transitive node similarity: predicting and recommending links in signed social networks. World Wide Web 17.4 (2014): 743-776
13. F. Heider. Attitudes and Cognitive Organization. The Journal of Psychology, 21: 107-112, 1946
14. D. Cartwright and F. Harary. Structure Balance: A Generalization of Heider's Theory. The Psychological Review, 63(5): 277-293, 1956
15. J. Leskovec, D. P. Huttenlocher, J. M. Kleinberg: Signed networks in social media. CHI 2010: 1361-1370