Information Fusion 46 (2019) 1–10

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full Length Article

Graph kernel based link prediction for signed social networks

Weiwei Yuan, Kangya He, Donghai Guan*, Li Zhou, Chenliang Li

[a] College of Computer Science & Technology, Nanjing University of Aeronautics and Astronautics, China
[b] Collaborative Innovation Center of Novel Software Technology and Industrialization, China

ARTICLE INFO

Keywords:
Link prediction
Graph kernel
Sign prediction
Signed social network

ABSTRACT

By revealing potential relationships between users, link prediction has long been considered as a fundamental research issue in singed social networks. The key of link prediction is to measure the similarity between users. Existing works use connections between target users or their common neighbors to measure user similarity. Rich information available for link prediction is missing since use similarity is widely influenced by many users via social connections. We therefore propose a novel graph kernel based link prediction method, which predicts links by comparing user similarity via signed social network's structural information: we first generate a set of subgraphs with different strength of social relations for each user, then calculate the graph kernel similarities between subgraphs, in which Bhattacharyya kernel is used to measure the similarity of the k-dimensional Gaussian distributions related to each k-order Krylov subspace generated for each subgraph, and finally train SVM classifier with user similarity information to predict links. Experiments held on real application datasets show that our proposed method has good link prediction performances on both positive and negative link prediction. Our method has significantly higher link prediction accuracy and F1-score than existing works.

1. Introduction

Link prediction predicts the positive sign or negative sign for links of signed social networks. It has long been considered as a fundamental research issue in the singed social network research [1,2]. By revealing the potential relationships between users, link prediction can be used in many applications. For example, it can be used in the recommender system to find users that may have opposite preferences with the active user. Using recommendations given by these users, the recommender system can avoid recommending items that the active user may not interest in.

The key of link prediction is to measure the similarity between users of signed social networks [3]. The more similar two users are, the more likely there exists a positive link between them [1], in which positive link means the source node of the link trusts, believes or likes the destination node of the link. The less similar two users are, the more likely there exists a negative link between them [1], in which negative link means the source node of the link distrusts, disbelieves or dislikes the destination node of the link. The researches of link prediction regard the signed social network as a graph, where users are regarded as nodes and their social relations are regarded as edges. To predict links between users is to predict signs of edges between nodes of the graph.

Existing works use the connection information between users or their common neighbors to measure user similarity. Some works use the number of users' common neighbors to evaluate user similarity. Two users are regarded as more similar if they have more common neighbors in signed social networks. Other works use the triangles including the target link and the connections between users and their common neighbors to measure user similarity. Based on Structural Balance Theory [4,5] or/and Status Theory [6], these works assign signs to the target link to keep the balance of the triangles. However, the similarity between two end nodes of the target link is not only influenced by their common neighbors, but also influenced by other users in the signed social network via social connections. By only considering the connections between the target nodes and their common neighbors, existing works lose rich information available for link prediction.

To solve the problems of existing works, we propose a novel graph kernel based link prediction method, which predicts the sign of links by comparing user similarity via signed social network's structural information. Our proposed method first generates a set of subgraphs for each user of target links. Since the distance of the shortest path between users reflects the strength of their social relations, our method uses this distance to control the generation of subgraphs, letting subgraphs have different strength of social relations. We then calculate the graph kernel similarities between subgraphs of the target link's two end users. Our method creates the k-order Krylov subspace, which is a set of vectors generated from the power iteration of a subgraph's corresponding adjacency matrix, for each subgraph. Bhattacharyya kernel, which is a

* Corresponding author.
E-mail address: dhguan@nuaa.edu.cn (D. Guan).

https://doi.org/10.1016/j.inffus.2018.04.004
Received 9 May 2017; Received in revised form 20 April 2018; Accepted 22 April 2018
1566-2535/

measure to evaluate the similarity of two normal distributions, is used to calculate the similarity of the k-dimensional Gaussian distributions related to each k-order Krylov subspace. SVM classifier is then used to learn relations between signs of links and the graph kernel based user similarities. This trained SVM classifier is used to predict links for signed social networks. Experiments held on real application datasets show that our proposed method has good link prediction performances on both positive and negative link prediction. Our method has significantly higher link prediction accuracy and F1-score than existing works.

The contributions of this work mainly lie in two aspects:

1. Our method proposes to use the structural information of signed social networks to predict links. Connections between users of the target link and their common neighbors, which are considered in existing link prediction works only reflect partial signed social network information. The target link is comprehensively influenced by many nodes and links in the signed social network. This is more properly to be represented in structural information.
2. Our method uses the graph kernel to calculate and compare the structural information of signed social networks. Since the structural information of signed social networks cannot be directly compared by machine learning methods, we map the structural information of signed social networks into mathematical spaces. This is achieved by generating the k-order Krylov subspace and calculating Bhattacharyya kernel of subgraphs.

The rest of this paper is organized as follows: Section 2 introduces the related works; Section 3 presents our proposed graph kernel based link prediction method in details; Section 4 shows the experimental results of our proposed method, which are compared with the performances of existing works; Section 5 concludes this paper and points out the future works.

2. Related works

A common method to predict the link between two nodes is to measure the similarity between these two nodes. The more similar these two users are, the more likely there exists a positive link between these two users. The less similar two users are, the more likely there exists a negative link between these two users. For example, in the recommender system, if two users have similar age, same sex, similar education background, similar job and similar living area, they will probably have the same preference. i.e., for the network measuring the preferences of users, there will probably exist a positive link between these two users.

Related works calculate the similarity of nodes or links to predict the sign of the target link. We summarize several most popular methods as follows:

(1) Node similarity based methods

A. Common Neighbors (CN)

CN [7,8] measures the similarity of users by their common neighbors. The more common neighbors two nodes have, the more similar they are.

Suppose node v_i and node v_j are two nodes of graph G, the similarity of v_i and v_j is:

$$S^{CN}(v_i, v_j) = |N_1^G(v_i) \cap N_1^G(v_j)| \tag{1}$$

where $N_1^G(v_i)$ is the neighbors of v_i in G, $N_1^G(v_j)$ is the neighbors of v_j in G, and $|\bullet|$ means the number of •.

B. Resource Allocation (RA)

RA [9] is based on the idea of resource allocation. As mentioned in [13], the resource of each node is regarded as a unit; each node allocates its resource evenly to its neighbors, and the resource between each pair of nodes are transferred via their common neighbors. The similarity of two nodes are defined as the resource one node can get from the other node.

For node v_i and node v_j of graph G, their similarity is calculated as:

$$S^{RA}(v_i, v_j) = \sum\nolimits_{z \in N_1^G(v_i) \cap N_1^G(v_j)} \frac{1}{d(z)} \tag{2}$$

where $N_1^G(v_i)$ and $N_1^G(v_j)$ are the neighbors of v_i and v_j in G respectively, $d(z)$ is degree of the selected common neighbor.

The difference of CN and RA is that: CA does not differentiate the common neighbors, i.e., each common neighbor is supposed to have the same contribution to the similarity calculation; while RA differentiates common neighbors by their degrees. i.e., the higher degree a common neighbor has, the less important of this selected common neighbor is. This is because the higher degree a common neighbor has, the less resource it can allocate to the target node. RA sets the importance of the common neighbor linearly relate to the reciprocal of the common neighbor's degree.

C. Adamic-Adar index (AA)

AA [10] is similar as RA: they both differentiates common neighbors by their degrees. The difference is that AA uses the logarithm of degrees to differentiate the contribution of common neighbors to user similarity, while RA directly uses the degrees to differentiate the contribution of common neighbors to user similarity. In some networks, the degrees of nodes tend to be very high, if the user similarity calculation uses the reciprocal of degrees directly, some similarity tends to be very small. AA therefore improves RA by enlarging the value of similarity. For node v_i and node v_j of G, their similarity is calculated as:

$$S^{AA}(v_i, v_j) = \sum\nolimits_{z \in N_1^G(v_i) \cap N_1^G(v_j)} \frac{1}{\log(d(z))} \tag{3}$$

(2) Link similarity based methods

Besides the degree information, it is very important to involve network information in user similarity calculation. Existing works [11–13] focus on using link information to evaluate user similarity. These works are based on based on Structural Balance Theory [4,5], Status Theory [6] or both.

Structural Balance Theory [4,5] considers four distinct triad relations between node v_i, node v_j and their common neighbor v_c, as shown in Fig. 1. Structural Balance Theory regards the triad relations shown in Fig. 1(a) and Fig. 1(b) as balanced. Fig. 1(a) means three nodes are mutual friends. Fig. 1(b) means two nodes are friends, and they are mutual enemy of the third node. Structural Balance Theory regards the

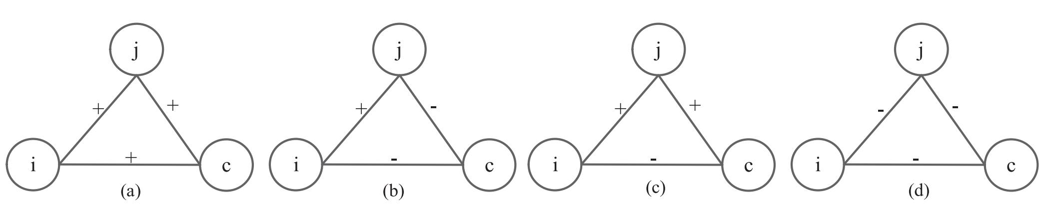

Fig. 1. Triad relationships considered in Structural Balance Theory.