ARTICLE IN PRESS
JID: NEUCOM [m5G;February 13, 2017;10:23]

Neurocomputing 000 (2017) 1–10

Contents lists available at ScienceDirect

Neurocomputing

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

Sign prediction in social networks based on tendency rate of equivalent micro-structures☆

Abtin Khodadadi[a], Mahdi Jalili[b,*]

[a] Department of Computer Science and Information Technology, Institute for Advanced Studies in Basic Sciences, Zanjan, Iran
[b] School of Engineering, RMIT University, Melbourne, Australia

ARTICLE INFO

Article history:
Received 16 June 2016
Revised 26 November 2016
Accepted 3 December 2016
Available online xxx

Keywords:
Distrust
Equivalence relation
Probabilistic method
Sign prediction
Signed social networks
Trust

ABSTRACT

Online social networks have significantly changed the way people shape their everyday communications. Signed networks are a class of social networks in which relations can be positive or negative. These networks emerge in areas where there is interplay between opposite attitudes such as trust and distrust. Recent studies have shown that sign of relationships is predictable using data already present in the network. In this work, we study the sign prediction problem in networks with both positive and negative links and investigate the application of network tendency in the prediction task. Accordingly, we develop a simple algorithm that can infer unknown relation types with high performance. We conduct experiments on three real-world signed networks: Epinions, Slashdot and Wikipedia. Experimental results indicate that the proposed approach outperforms the state of the art methods in terms of both overall accuracy and true negative rate. Furthermore, significantly low computational complexity of the proposed algorithm allows applying it to large-scale datasets.

1. Introduction

Communication has experienced an advanced stage by online social networks and billions of people are in contact around the world by means of them. They tend to exchange information according to some similarities such as common interests, cognation, friendship and acquaintanceship. However, like other societies, there may exist conflicts among individuals. Signed networks are a class of social networks which aim to model this opposition. In such networks, existing relations are explicitly categorized as positive or negative. Positive interactions may represent friendship, interest or trust while negative interactions specify enmity, disinterest or distrust [7,18,20]. Some popular signed networks are Epinions, Slashdot, Wikipedia, Amazon, Ebay and Advogato.

One of the challenges in signed networks is to infer the type of unknown relations that is often referred to as sign prediction [18,19]. Sign prediction is similar to the link prediction which is a well-studied problem in social network analysis [2,21,22]. However, it has not been investigated until recent years. This problem was first introduced by Guha et al. [11] and later developed by Leskovec et al. [19,20]. Generally, approaches in this area can be categorized into some groups. Some algorithms are based on social theories and apply their predefined constraints to decide on unknown relations [20,24]. Some others are based on matrix-factorization to extract the kernel of adjacency matrix and to perform algebraic operators to get the prediction results [1,5,11,14,18]. Other algorithms are based on machine learning techniques. They build different sets of features using implicit network characteristics and use them in a supervised classification task [6,19,25]. Finally, There are some studies which follow recommender system's methodologies [15,26]. The main challenge addressed in these studies is to find a solution which can infer the signs with less complexity and better accuracy that scales well with the properties of real networks. In addition, as these networks mostly comprise positive relations, the solution should distinguish negative relations as well.

To tackle these challenges, in this work we developed an algorithm based on tendency rate of triple micro-structures in a signed network. We studied different possible configurations that can be made using signed relations between three individuals. Then, we devised a probabilistic framework to compute network tendency towards each configuration. Having this global trend, we measured implicit forces that direct the sign of each relation in its neighborhood. Our work shows the important role of reciprocated relations and inefficiency of social theories to accurately model them. It defines closed triple micro-structures and discriminates them in equivalence classes. The probabilistic approach is introduced to measure the interplay between various configurations. Finally the

☆ This research was supported by Australian Research Council through grant No. DE140100620 to Mahdi Jalili.
* Corresponding author.
E-mail address: mahdi.jalili@rmit.edu.au (M. Jalili).

http://dx.doi.org/10.1016/j.neucom.2016.12.069

performance of this approach is analysed and compared with a number of naive and state of the art approaches.

1.1. Related works

Considering trust and distrust explicitly to predict the type of relations between users in a real large network, was first implemented by Guha et al. [6,9,11]. They used trust and distrust between users as a real number between 0 and 1 in two distinct matrices and built *belief matrix* based on them. Then, they defined four types of *atomic propagation* as basic operations to model trust propagation between nodes. They encoded these atomic propagations as a matrix operator and applied them to the belief matrix in a sequence to get the final results [11]. Inspired by spreading activation models, Ziegler et al. proposed *Appleseed* propagation model as a classification scheme for trust metrics. Later, they extended this metric for distrust and compared it with *Advogato*, another trust metric, to compute trust neighborhoods [28]. Shahriari and Jalili investigated node ranking algorithms in signed networks. They defined *optimism* and *reputation* based on node ranks. Using these metrics for each side of a relation, they built a four dimension feature vector and used it in logistic classification approach to predict signs [25].

Leskovec et al. studied the structure of signed networks and investigated compatibility of two social theories with the structure of signed networks [20]. Based on these theories they built some feature vectors and used it in a machine learning approach to solve the sign prediction problem. Dubois et al. considered trust values present in the network as the probability of having an edge between the nodes in a random graph. They performed a reverse mapping where trust values of each pair corresponds to the probability of having a path between them. Path probability and spring embedding values were used in a two dimensional vector to position trust value of each edge [9]. Chiang et al. defined some measures of social imbalance based on *higher order cycles* (i.e. cycles with length more than three) and used it to predict relation types. They showed that their imbalance measure plays the role of *Katz* measure in signed networks. They also performed a supervised learning classifier (like the one proposed in [19]) based on these measures and reported the results [6]. Yang et al. devised a latent factor model (called *Behavior Relation Interplay*) which exploits users' activity (e.g. movie ratings) to determine social signed ties in unsigned networks. They used *Pearson correlation* score to assess similarity between ratings [26].

Some other recent studies are connected to Social theories. Patidar et al. initially generated a decision tree using C4.5 and user attributes (e.g., gender, career) to induce relation categories. Then, they employed balance index (proportion of balanced triads to total triads) to identify ties which increase the balance [23]. Qian and Adali extended balance theory by considering weak/strong ties. They discussed different configurations and participants behavior with regards to their tolerance and network's stress. Accordingly, they suggested a convergence model in the form of Metric Multidimensional Scaling optimization problem and predicted edge signs based on this model [24]. In another work, Javari and Jalili followed a collaborative filtering approach by first discovering the community structure of the signed networks, and then, applying the collaborative filtering framework. Their approach, being much less complex than machine learning methods, gives comparable results with them [15].

There are some studies which adopt a matrix kernel approach. Kunegis et al. investigated some network characteristics on Slashdot, such as clustering coefficient, centrality and popularity of nodes. They used these measures to identify unpopular users and predicted the link signs. They used exponentials of the adjacency matrix to predict the sign of edges. They also used dimensionality reduction and Laplacian matrix as alternative ways to do the task [18]. Ye et al. employed transfer learning to use information available in a source network to predict link signs of a target network. They built explicit (e.g., node degree, edge embeddedness) and latent topological features and employed an AdaBoost-like algorithm for the prediction [27]. In other studies, sign prediction is investigated as a matrix completion problem and a matrix factorization technique is employed to solve it [1,5,13,14]. For example, Chiang et al. verified the low rank property of signed networks studying the rank of adjacency matrix of a complete k-weakly balanced network. Then, they employed a gradient-based matrix factorization method to provide an approximate completion of the network. [5].

2. Preliminaries

We follow the formulation introduced in [6] to formally define the sign prediction problem. We model signed networks by a graph $G = (V, E, \Sigma)$ where $V = \{1, 2, \ldots, |V|\}$ is a set of *nodes* or *vertices* and $E = \{e_1, e_2, \ldots |E|\}$ is a set of *directed edges* in the form (i, j) for $i, j \in V$ and Σ is a mapping $\Sigma : E \to \{+1, -1\}$ giving a *sign* to each edge. The sign prediction problem can be formally defined as follows. Given a graph $G = (V, E, \Sigma)$ and a test edge $e_{test} \in E$, we want to predict $\Sigma(e_{test})$ based on the edges in $E - \{e_{test}\}$.

2.1. Related social theories

Two social theories have had dominant roles in studies about signed networks: *balance theory* and *status theory*. Balance theory was first introduced by Heider [12] and then formulated by Cartwright-Harary [3] in the language of graphs. Having a relation between two known individuals as friendship or enmity, every three individuals can connect to each other in four distinct configurations. Balance Theory states that configurations with odd number of friendship (i.e, three friends or two friends with a mutual enemy) are more probable based on psychological stress. However, Davis refined this theory to *Weak balance* by adding three-enemy relationship as a probable configuration [8]. Balance theory poses certain conditions on the network. One can say if a network is fully balanced, it can be partitioned into two clusters such that nodes within a cluster are friends and nodes in different clusters are enemies [7,19].

Status theory, however, has different interpretation of signed links. It states that a positive link from a to b means that in a's viewpoint, b has higher status while a negative link has the opposite meaning. This theory first appeared implicitly in the work of Guha et al. [11] and then accentuated in Leskovec et al. [19,20]. Status theory poses certain conditions on the network. First, it entails all the edges to be directed and implies that converting the direction of an edge reverses its sign. Second, if a network is completely consistent with this theory, there must be a global ordering of nodes, so we can rank them [19,20]. In directed networks, a node a can connect to node b with respect to a node c in 16 distinct ways. In some of the configurations status can predict the sign of relations while in others additional information is needed.

Leskovec et al. performed the sign prediction based on these theories in three different networks [19]. They conducted experiments based on local and global presentation of both theories. In a different setup, they also used social theory related features beside 16-triads, in a machine learning approach and obtained the results. They finally conjectured that status has better explanation of the network global properties, while in local prediction tasks both theories performed well. In general, machine learning classifiers performed much better [19]; however they are computationally much more expensive than simple social theory based methods.

Please cite this article as: A. Khodadadi, M. Jalili, Sign prediction in social networks based on tendency rate of equivalent micro-structures, Neurocomputing (2017), http://dx.doi.org/10.1016/j.neucom.2016.12.069