Research Article

A Community-Based Approach for Link Prediction in Signed Social Networks

Saeed Reza Shahriari,[1] Mohsen Shahriari,[2] and Rafidah MD Noor[1]

[1]*Department of Computer System & Technology, Faculty of Computer Science & Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia*
[2]*Advanced Community and Information System, RWTH Aachen University, Ahornstraße 55, 52056 Aachen, Germany*

Correspondence should be addressed to Saeed Reza Shahriary; shahriarysaeedreza@gmail.com,
Mohsen Shahriari; shahriari@dbis.rwth-aachen.de and Rafidah MD Noor; fidah@siswa.um.edu.my

In signed social networks, relationships among nodes are of the types positive (friendship) and negative (hostility). One absorbing issue in signed social networks is predicting sign of edges among people who are members of these networks. Other than edge sign prediction, one can define importance of people or nodes in networks via ranking algorithms. There exist few ranking algorithms for signed graphs; also few studies have shown role of ranking in link prediction problem. Hence, we were motivated to investigate ranking algorithms availed for signed graphs and their effect on sign prediction problem. This paper makes the contribution of using community detection approach for ranking algorithms in signed graphs. Therefore, community detection which is another active area of research in social networks is also investigated in this paper. Community detection algorithms try to find groups of nodes in which they share common properties like similarity. We were able to devise three community-based ranking algorithms which are suitable for signed graphs, and also we evaluated these ranking algorithms via sign prediction problem. These ranking algorithms were tested on three large-scale datasets: Epinions, Slashdot, and Wikipedia. We indicated that, in some cases, these ranking algorithms outperform previous works because their prediction accuracies are better.

1. Introduction

Recently, social network analysis has attracted great deal of attentions. In social networks, nodes and edges, respectively, indicate people and relationships among them [1]. Social networks are dynamic and evolve over time via registering new members, deleting profiles, and adding/removing some edges or connections among entities [2]. Hence, plenty of studies have investigated this field in order to model these structures. One of the most important problems in social networks is link prediction which can be stated as follows: with how much determinism one can predict forming (lacking) of edge between two people based on available structure of the graph? The importance of this subject is originated from the natural sparsity of social networks [2]. In other words, social networks encompass highly dynamic structure; therefore, available links are just a subset of possible relations among people and some new links will form in future. Link

predication is also widely used in retrieving lost data and it probably helps to construct the graph [3].

One could model social systems by using signed relationships. Inherently in signed graphs, most of relations are positive or negative such as likes and dislikes or trusts and distrusts [4]. Negative edges play an important role in signed networks and these negative links impress greatly on importance of nodes in the system. Studying negative relationships in signed graphs can help in analyzing and better understanding social ecosystems. Link prediction in signed graph appears in the form of predicting sign of edge between two people. Therefore, one important question which comes to mind in signed networks is that how accurately sign of an edge can be predicted according to local and global behavioral patterns in the network. Not only sign prediction enables us to have better understanding of social relations but also it can be utilized in several applications such as recommender systems and online social networks, in which

they offer new friends for users. In these networks, users have capability of expressing their views toward others via binary −1 and +1 values [5, 6].

In this paper, three community-based ranking algorithms for ranking of nodes have been proposed, and we have studied their impacts on the edge sign prediction problem. In order to study the impact of proposed ranking algorithms on signed prediction problem, we extracted the features of the predictor based on reputation and optimism introduced in [7]. Reputation of a node shows how much reputable a node is in the system and optimism denotes voting pattern of the node toward others. We assess our method by utilizing logistic regression classifier and running algorithms on real social network datasets. The structure of this paper is organized as follows.

In Section 2, related works are brought. In Section 3, we mainly introduce proposed algorithms; moreover, the problem of sign prediction is defined and rank-based features are introduced as features for the prediction task. We also separately go through community detection problem, community-based ranking algorithms, and the logistic regression classifier. In Section 4, datasets for experimental purposes are introduced and implementation results are also demonstrated. In Section 5, the discussion is made and finally in Section 6, conclusion and future directions are mentioned.

2. Related Works

There are two major categories of methods used in link prediction: firstly, those approaches that utilize local information of the graph which focus on the local structure of nodes. Among local approaches, [8] has the best performance in link prediction between two specific nodes. Common neighbor index is also known as friend of friend algorithm (FOAF) is used by many online social networks for recommending friends such as Facebook. FOAF determines the similarity of two nodes that tend to communicate with each other on the basis of counting number of joint neighbors [9]. Other metrics for computing similarity are based on preferential attachments, where these measures are calculated based on multiplying or summing of nodes degree. Second category concentrates on global structure of the network and detecting overall features in order to find how strongly two nodes are similar. There are also diverse global approaches which use the whole adjacency matrix in order to predict hidden links, for instance, shortest path algorithm, PWR algorithm, and SimiRank algorithm [1, 10].

In sign prediction, the most notable and remarkable methods are divided into two categories: Belief Matrix Model [5] and machine learning approaches [11]. Belief Matrix Model was introduced by [5] and was proposed for predicting trust or distrust between two particular users in signed networks. It was the fundamental model in sign prediction of edges. Reference [11] employed the idea of signed triads and used logistic regression model and some local feature in order to predict sign of edges in social networks. The features that were introduced by [11] are categorized in two classes: first one is on the base of the positive/negative ingoing/outgoing degree of nodes which basically collect the local information

of nodes. And the second group is based on the extracted principles from social psychology, in which we are able to determine the type of u and v relation by utilizing the information of third party like w.

Ranking of nodes has tight relationship with sign prediction problem so we also investigate ranking in signed networks. Ranking of nodes is the problem of computing how much important or trustable a node is in networks [12]. The centrality measures like betweenness [13], closeness [14], and eigenvector centrality [15] were introduced to compute nodes' importance degree in the network. Other algorithms like HITS [16] and PageRank [17] were added in 1990. All of these ranking algorithms are designed for positive graphs and there are merely several literatures for ranking of nodes in signed networks. The simplest ranking algorithm for signed graphs is prestige, where number of positive and negative incoming links determine ranking of each node [18]. Another ranking algorithm is PageTrust that was introduced by [19]. This method is extension of PageRank, and the main difference is that nodes with negative incoming links will be visited less in random walk process. Exponential ranking is another chief method of ranking for signed graphs [12]. In exponential ranking, the value of ranking vector globally is obtained from local trust values. Another ranking algorithm for signed networks that is greatly similar to HITS was proposed by [20]. This method utilizes the concept of Bias and Deserve which underestimates the vote of optimistic and pessimistic nodes. Reference [3] also proposed new ranking algorithms for signed networks, namely, Modified HITS and Modified PageRank.

Because we propose community-based ranking algorithms, we should go through community detection problem. Community detection algorithms help to prepare more dominant recommendation systems and web page clustering which have great effect on better searches [21]. Community detection algorithms attempt to cluster edges/nodes in order to have minimum number of edges between densely communities [22]. One of the most widely used methods for community detection in unsigned graphs was proposed by [23]. As for signed networks, [24] proposed a two-step spectral approach which was an extension to modularity. The main problem related to modularity is resolution limit in which very small communities might not be detected. In order to address this problem, [22] proposed new method for detecting communities on signed graphs by extending potts model. Reference [25] also introduced useful approach that works on the base of blocking method.

3. Method

In this paper, authors intend to investigate the community-based problem of predicting sign of links in signed social networks. Hence, in this section as well as proposed algorithms and methods, the problems of sign prediction and community detection will be discussed in detail.

3.1. Edge Sign Prediction. In order to define the problem formally, it can be assumed that we have a signed directed graph $G(V, E)$ that V represents set of vertices and E shows

set of edges where customers and users can vote positively and negatively toward each other. So the aforementioned notation V represents users of site and E indicates +1 and −1 relations among them. In all over the paper, the person who gives positive vote and receives it, is named trustor and trustee, respectively [2, 26]. The sign prediction problem can be defined as follows. Suppose that signs of some links in the network are hidden, and the goal is to reliably predict values of these edges by current information in the graph. The sign prediction problem tries to find signs of hidden edges with negligible error [1, 27]. In this work we propose state-of-the-art community-based ranking algorithms and we evaluate their effectiveness via sign prediction problem on three datasets: Epinoins, Slashdot, and Wikipedia.

To this end, [7] already introduced rank-based features named Optimism and Reputation to connect ranking problem with sign prediction. Rank-based reputation of node i indicates patterns of voting toward this node. Meaningfully, rank-based reputation of node i not only considers number of positive/negative incoming links toward node i but also it takes into account ranks of nodes who vote toward node i. In other words, when a person receives several positive incoming links, s/he might not be very reputable because one should consider rank of voters toward node i. If the users who vote toward node i are high rank, then node i can be considered reputable, but if they are not high rank we cannot say that node i is reputable although the number of positive incoming links toward node i is relatively high. The following equation can better describe rank-based reputation [7]:

$$\text{RBR}_i = \frac{\left| R_{\text{in}}^{(+)}(i) \right| - \left| R_{\text{in}}^{(-)}(i) \right|}{\left| R_{\text{in}}^{(+)}(i) \right| + \left| R_{\text{in}}^{(-)}(i) \right|}, \tag{1}$$

where RBR_i is the value of rank-based reputation of node i, $R_{\text{in}}^{(+)}(i)$ indicates sum of rank values of nodes who positively voted toward node i, and, similarly, $R_{\text{in}}^{(-)}(i)$ is sum of rank values of nodes who negatively voted toward node i. In the same vein, one can define rank-based optimism of node i as follows:

$$\text{RBO}_i = \frac{\left| R_{\text{out}}^{(+)}(i) \right| - \left| R_{\text{out}}^{(-)}(i) \right|}{\left| R_{\text{out}}^{(+)}(i) \right| + \left| R_{\text{out}}^{(-)}(i) \right|}, \tag{2}$$

where RBO_i is the value of rank-based optimism of node i, $R_{\text{out}}^{(+)}(i)$ refers to sum of rank values of nodes whom node i positively voted toward them, and similarly, $R_{\text{out}}^{(-)}(i)$ is sum of rank values of nodes in which node i negatively voted toward them. As formula (2) shows, node i which generates several positive outgoing links might not be optimistic because this set might contain nodes in which they are low rank [3]. In order to compute these features, we need algorithms to rank nodes. As for ranking algorithms, we propose three community-based ranking algorithms in the next sections.

3.2. Community-Based Ranking Algorithms to Compute RBR and RBO.

In this section we propose three ranking algorithms in which all of them work based on community detection problem in signed graphs. In other words, firstly, we

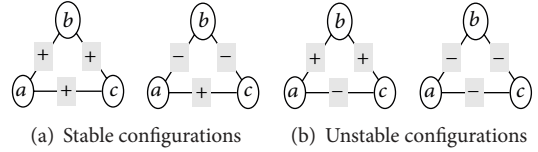(a) Stable configurations   (b) Unstable configurations

FIGURE 1: All possible states of balance theory.

run a community detection algorithm on signed networks. The results will be disjoint communities of nodes. As all community detection algorithms work based on a density based approach in which they try to maximize density of intracluster edges and minimize between cluster edges, so intracluster nodes are more dense and close. From social perspective, intracluster nodes might know each other better (this is the notion behind our community-based ranking algorithm). Meaningfully, nodes in the same community are much more familiar than nodes that are in different communities. Via using this philosophy about intracluster nodes, we change previous ranking algorithm like Prestige, HITS, and PageRank [3] to have influence of intra- and extracluster nodes with parameters α and $(1-\alpha)$, respectively. Then we can use ranking-based features of [7] for the case of sign prediction. Because first phase of the algorithms is community detection, so we investigate community detection problem and a sample community detection algorithm in signed graphs in Section 3.2.1. In this paper a community detection algorithm based on social balance theory is utilized. In Section 3.3, ranking algorithms based on community detection phase are introduced.

3.2.1. Community Detection. The algorithm used in this paper is based on structural balance theory [28]. In balance theory, there are four possible states when nodes are in signed relations in social networks [29]. One can differentiate these states by number of positive and negative edges in each triad [30]. On the base of strong social balance theory, when all of nodes have positive relation or two nodes share the same enemy, these states are called stable. Similarly, cases with all nodes have negative edges or with two positive edges are unstable states [31]. Regarding this definition, a network with more than three nodes is structurally balanced if all the possible triads are stable [32] (Figure 1).

The basic structural theorem states that these triples can be partitioned into two distinct sets in which all the positive relations are inside sets and negative ones are among them [33, 34]. In other words, negative edges connect positive sets. On the basis of this definition, a network is called k-balanced if all positive edges are located in k number of different categories, and these sets are joined with negative relations [35]. In reality, rarely there are structurally balanced networks. There are always some edges that destabilize the graph and transform it into unstable configuration. Therefore, the number of positive edges between clusters and the number of negative edges inside clusters should be minimized [24]. In fact, the problem is like finding the best sets with minimum number of positive relations between partitions and also minimum number of negative edges inside sets [36].

Reference [25] introduced one criterion function which makes decision based on counting number of elements having conflict with k-balanced theory. It can be defined as if one considers N as number of negative edges inside clusters, and P as number of positive edges between clusters, then number of inconsistencies which is denoted by $I(c)$ can be mentioned as

$$I(c) = \beta \times N + (1 - \beta) \times P, \tag{3}$$

where β is the importance factor that is assigned to positive and negative inconsistencies.

If $\beta = 0.5$ then positive and negative relations contribute equally on amount of inconsistency. And for the case that $0.5 < \beta \leq 1$ the negative relations have more impact on result and, when $0 < \beta \leq 0.5$, positive edges have higher influence. The ideal condition is created when P and N have the smallest amount, so better result is achieved. Because $I(c)$ show error, the algorithm tries to find minimum value for N and P via using a hill-climbing optimization technique [25]. Other community detection approaches suitable for signed graphs can also be used in this phase.

3.3. Community-Based Ranking Methods.

In this paper, we propose three new ranking algorithms for signed complex networks that are dependent on community detection. We introduce a method that ties the concept of ranking algorithm and community detection. Suppose that we intend to compute the rank of node i in the network. First of all, we cluster nodes in the network in such a way that each node belongs to one community in the graph (first phase, algorithm referenced in Section 3.2.1), so for calculating rank of node i by taking influence of other nodes in the network, we give priority and high importance to the nodes that belong to the same community, which node i belongs to. These algorithms are described in detail, in the following subsections (second phase of ranking). We will verify rationality of these ranking algorithms in Results section.

3.3.1. PBCD (Prestige Ranking Algorithm Based on Community Detection).

Prestige is the simplest algorithm in signed complex network [18]. In this method, the most important factor for determining ranking of each node in the system is the number of positive and negative incoming nodes that each node receives from others. In other words, if a node has many positive incoming links, therefore, its prestige is high in the network. And it is also true for negative links, if the number of positive incoming links is less than negative ones, the node has low prestige in comparison with the other nodes in the system [3]. The idea of community-based ranking inspires us to incorporate our method with some well-known ranking algorithms like prestige. The proposed prestige can be stated as follows:

$$\Pr(i) = \left[\alpha \times \frac{\left| id_{\text{in}}^{(+)}(i) \right| - \left| id_{\text{in}}^{(-)}(i) \right|}{\left| id_{\text{in}}^{(+)}(i) \right| + \left| id_{\text{in}}^{(-)}(i) \right|} \right] + \left[(1 - \alpha) \times \frac{\left| od_{\text{in}}^{(+)}(i) \right| - \left| od_{\text{in}}^{(-)}(i) \right|}{\left| od_{\text{in}}^{(+)}(i) \right| + \left| od_{\text{in}}^{(-)}(i) \right|} \right], \tag{4}$$

where α is the impact factor to determine degree of importance of nodes that are in the same community that node i belongs to, and $id_{\text{in}}^+(i)$ and $id_{\text{in}}^-(i)$, respectively, indicate set of nodes who positively and negatively voted toward i and these nodes are members of the same community that node i belongs to. Similarly, $od_{\text{in}}^+(i)$ and $od_{\text{in}}^-(i)$ show set of nodes who positively and negatively voted toward node i and these nodes are members of other communities which are different from the community that node i belongs to. Moreover, $| |$ represents magnitude of the set of nodes and the in subscript indicates that we only consider incoming links from $id(i)$ and $od(i)$ sets toward node i. Finally, $i \in c$, $c \in C$: c is the community which node i belongs to and C is the disjoint subgraph of all clusters detected via the algorithm. In this notation, all members of $id(i)$ are in c and none of $od(i)$ members are in c. As intraclusters nodes of communities are more close to each other, ranking algorithm can utilize the influence of intra clusters nodes closeness.

3.3.2. HBCD (HITS Ranking Algorithm Based on Community Detection).

HITS algorithm was introduced by [16], and it was mainly proposed for exploiting helpful information in order to analyze structure of links and has been applied in various applications. This algorithm works based on hub and authority vectors. These vectors are initialized with some predefined (random) values and converge after some recursive iterations [16].

Reference [3] introduced modified version of HITS in which the graph is divided into two positive and negative parts and then run the algorithm on each graph separately. In HITS algorithm, there are two vectors: authority and hub, which finally converge after enough iteration. We propose a new version of HITS in which there is distinction between importance of local neighbor nodes and those members of different communities that node i belongs to. The HITS algorithm based on community detection can be stated as follows:

$$a_i^{(+)}(t+1) = \alpha \times \left(\sum_{j \in id_{\text{out}}^{(+)}(i)} h_j^{(+)}(t) \right) + (1 - \alpha) \times \left(\sum_{j \in od_{\text{out}}^{(+)}(i)} h_j^{(+)}(t) \right),$$

$$a_i^{(-)}(t+1) = \alpha \times \left(\sum_{j \in id_{\text{out}}^{(-)}(i)} h_j^{(-)}(t) \right) + (1 - \alpha) \times \left(\sum_{j \in od_{\text{out}}^{(-)}(i)} h_j^{(-)}(t) \right),$$

$$h_i^{(+)}(t+1) = \alpha \times \left(\sum_{j \in id_{\text{in}}^{(+)}(i)} a_j^{(+)}(t) \right) + (1 - \alpha) \times \left(\sum_{j \in od_{\text{in}}^{(+)}(i)} a_j^{(+)}(t) \right),$$

$$h_i^{(-)}(t+1) = \alpha \times \left(\sum_{j \in id_{\text{in}}^{(-)}(i)} a_j^{(-)}(t) \right)$$
$$+ (1 - \alpha) \times \left(\sum_{j \in od_{\text{in}}^{(-)}(i)} a_j^{(-)}(t) \right), \tag{5}$$

where α indicates importance factor that is given to the local neighbors and $h(t)$ and $a(t)$ show hub and authority vectors at time t. $id_{\text{in/out}}^{(+)}(i)$ and $id_{\text{in/out}}^{(-)}(i)$, respectively, represent set of nodes who have relations (positive/negative or incoming/outgoing) with node i and they are members of the same community that node i belongs to. In a similar manner, $od_{\text{in/out}}^{(+)}(i)$ and $od_{\text{in/out}}^{(-)}(i)$ indicate set of nodes in which they have relations (positive/negative, incoming/outgoing) with node i and they are members of different communities that node i belongs to. Finally, in and out subscripts, respectively, show that $id(i)$ or $od(i)$ represents set of nodes that voted toward node i (incoming links toward node i) or being voted by node i (outgoing links from node i). Hub and authorities are initialized with some random values and they converge after enough iteration.

3.3.3. RBCD (PageRank Algorithm Based on Community Detection).
PageRank is one of the most widely used methods for ranking of nodes [37]. It was extracted from Google Larry page. PageRank uses the concept of random walk that leads to probability distribution which computes the possibility of randomly going from one node to another one, and finally gets to one specific node. This algorithm initially was introduced for graphs with positive and unsigned edges, especially for web pages on the internet. Reference [3] introduced modified version of PageRank in which the graph is divided into two parts and the algorithm is run on each graph separately, and finally for calculating ranking vector, negative ranking vector is subtracted from positive one. Our proposed ranking algorithm states that each node in the network belongs to specific community or cluster, and the general idea of specifying ranking is on the basis of utilizing other nodes' information. In other words, to compute rank of node i, we give priority to the nodes that belong to the same community that node i belongs to. Based on this opinion, PageRank based on community detection can be defined as

$$PR_i^{(+)}(t+1) = \alpha \times \left(\beta \times \sum_{j \in id_{\text{in}}(i)} \frac{PR_j^{(+)}(t)}{\left| d_{\text{out}}^{(+)}(j) \right|} + (1-\beta) \times \frac{1}{N} \right)$$
$$+ (1 - \alpha)$$
$$\times \left(\beta \times \sum_{j \in od_{\text{in}}(i)} \frac{PR_j^{(+)}(t)}{\left| d_{\text{out}}^{(+)}(j) \right|} + (1-\beta) \times \frac{1}{N} \right), \tag{?}$$

$$PR_i^{(-)}(t+1) = \alpha \times \left(\beta \times \sum_{j \in id_{\text{in}}(i)} \frac{PR_j^{(-)}(t)}{\left| d_{\text{out}}^{(-)}(j) \right|} + (1-\beta) \times \frac{1}{N} \right)$$
$$+ (1 - \alpha)$$
$$\times \left(\beta \times \sum_{j \in od_{\text{in}}(i)} \frac{PR_j^{(-)}(t)}{\left| d_{\text{out}}^{(-)}(j) \right|} + (1-\beta) \times \frac{1}{N} \right), \tag{6}$$

where in the above equations α denotes importance factor is assigned to the local neighbor nodes, and t shows the rank at the time of t. β indicates the forgetting factor, and N represents number of nodes in the graph. $id_{\text{in}}(i)$ shows set of incoming links to node i that belong to the same community that node i belongs to. Similarly $od_{\text{in}}(i)$ indicates set of incoming links to node i that belong to different communities that node i belongs to. $d_{\text{out}}^{(+)}(j)$ and $d_{\text{out}}^{(-)}(j)$ involves number of positive and negative outgoing links from node j, respectively.

3.4. Classifier.
Classification is the process of assigning data to one of predetermined classes. Here our classes are the mapped values of positive and negative signs to +1 and −1 values. This process is done via using training set which contains some features to constitute a classifier and then evaluation via using a test set. A brief explanation of the classifier used in our work is brought here.

Logistic regression: logistic regression or sigmoid function is a monotonic, continuous function which lies between −1 and +1 values and is a method of learning function of the form $f : X \rightarrow Y$ or $P(Y \mid X)$. They can be mathematically defined as follows:

$$P(Y = 1 \mid x) = \frac{1}{1 + \exp\left(w_0 + \sum_{i=1}^{n} w_i x_i\right)},$$
$$P(Y = -1 \mid x) = \frac{1}{1 + \exp\left(w_0 + \sum_{i=1}^{n} w_i x_i\right)}. \tag{7}$$

Y is the discrete values of classes which here are +1 and −1, X is the input vector of discrete or continuous values, and w_i are some learning coefficient learnt by the model. Classifiers are evaluated based on accuracy. Accuracy metric is calculated based on TP, FP, TN, and FN as follows [38, 39]:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{8}$$

We used 10-fold validation in order to evaluate generalization of the models. Cross validation is a classical method in which the dataset are partitioned into k folds. The first one is used as test set and the remaining folds are considered as training sets. In the next phase, the second fold is selected as the test set and the remaining are chosen as training sets [40]. We utilized *WEKA* software for computing accuracy that is available through http://www.cs.waikato.ac.nz/ml/weka.

4. Experiments

In order to verify proposed algorithms introduced in Section 3.3 we executed them on three large datasets.

In the following sections these datasets are introduced and achieved results are depicted.

4.1. Datasets. Datasets which are used for experimental purposes are Wikipedia, Epinions, and Slashdot. These online social networks are available through http://snap.stanford edu/data/. In the following, we explain briefly these datasets.

Wikipedia. Volunteers from all around the world collaborate to write this free cyclopedia. A user can take the role of administrator with additional access to some technical features by getting vote of other users. Administrators are responsible for maintenance purposes. A user is nominated as administrator, and then Wikipedia members elect users as administrators via public dialogues and talks. Totally, 7000 users took part in elections, and 100,000 votes were received from 2800 delegated elections. 1200 elections lead to promotion and 1500 elections were not successful and did not produce a winner. Half of voters are current administrators and the remaining are ordinary users [6].

Epinions. Epinions is a review online social network that users can express their views about diverse items like music, TV shows, and hardware. The site members are able to vote positively and negatively toward each other. As a matter of fact, this online social network contains information about who-trusts-whom. The data set is made of 131828 nodes and contains 841372 edges in which 85% of them are positive [41].

Slashdot. Slashdot is a website related to technology and science. It is well known due to its specific users and has introduced itself as user-submitted and science evaluated news. In other words, links of news and summary of different issues are submitted by users, and each story becomes the topic of series of talks. Selected readers that take the role of moderators send ratings to Slashdot. The responsibility of these readers is appointing tags for each comment. Slashdot does not show scores to the users but lets them arrange comments on the base of assigned points. Slashdot also has a service that users have the ability to tag each other as friend or opponent. Therefore, links between users are friend/opponent relationship. The data set is made of 82144 nodes and contains 549202 edges in which 77% of them are positive [42].

4.2. Results. Via using introduced community detection of Section 3.2.1, the communities in Epinions, Slashdot, and Wikipedia are detected. We examine our proposed method on different size of communities, where it can be specified through input parameter of community detection algorithm. Reference [43] introduced a function for determining number of clusters in Epinions, Slashdot, and Wikipedia, and it can be observed that this method has high accuracies when the number of clusters is between four and ten, so we also checked our approach for seven cases starting from four to ten. For ranking nodes, Prestige Based on Community Detection (PBCD), PageRank Based on Community Detection (RBCD), and HITS Based on Community Detection (HBCD) are run on these datasets. In order to differentiate between nodes which belong to various communities, we defined
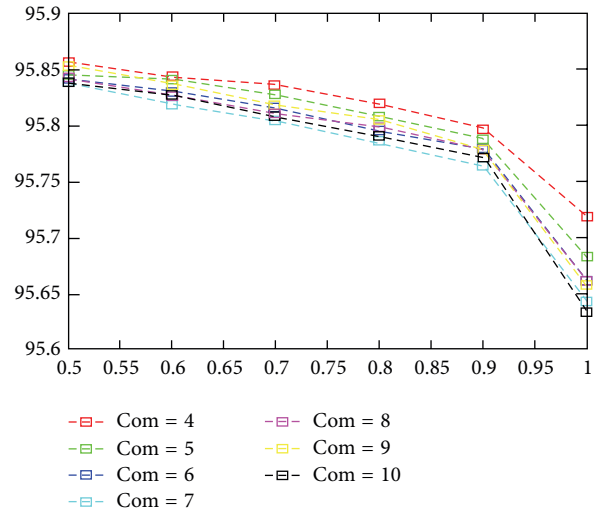

Figure 2: y axis shows prediction accuracy of PBCD ranking algorithm on Epinions dataset and x axis indicates different values of α. Different colors show number of communities detected by the community detection algorithm.

impact factor α for local nodes and $(1 - \alpha)$ for foreign ones where α can contains values of $(0.5, 0.6, \ldots, 0.9, 1)$. When $0.5 < \alpha \leq 1$, local nodes have more privilege than the others, and $\alpha = 0.5$ shows normal ranking algorithms of [3]. In other words, in the case ($\alpha = 0.5$), no community structure is considered. In order to define accuracy of edge sign prediction, rank based features are extracted and in order to evaluate generalization of the models we used 10-fold cross validation. Accuracy of introduced methods is evaluated with various size of communities and different values of α. The results are shown in forms of figures. In the following, significant findings related to each dataset are stated.

4.2.1. PBCD on Datasets. In Epinions and Slashdot, for all size of communities, $\alpha = 0.5$ contains the maximum values. In fact outputs of PBCD on Epinions and Slashdot datasets indicate that using community-based ranking algorithms might slightly degrade prediction accuracy. However, the prediction accuracy is still high for different values of α greater than 0.5 and for all community numbers. This lower accuracy might be because of special property of dataset or community detection algorithm. Results for Epinions and Slashdot are illustrated in Figures 2 and 3, respectively. Results related to Wikipedia are shown in Figure 4. In the case Com (number of communities) equals 4, 6, and 7, prediction accuracy is higher than the case $\alpha = 0.5$. Generally high precisions extract properties of dataset and offers better model for prediction. In Figure 4, where Com = 6, $\alpha = 0.6$ contains maximum value which is equal to 88.7467.

4.2.2. RBCD on Datasets. Analogously to PBCD, we implemented RBCD ranking algorithm on datasets. Outputs for Epinions are illustrated in Figure 5. In Epinions, all the number of communities has acceptable accuracies. Com = 10 with $\alpha = 1$ contains the maximum value of 95.515 among
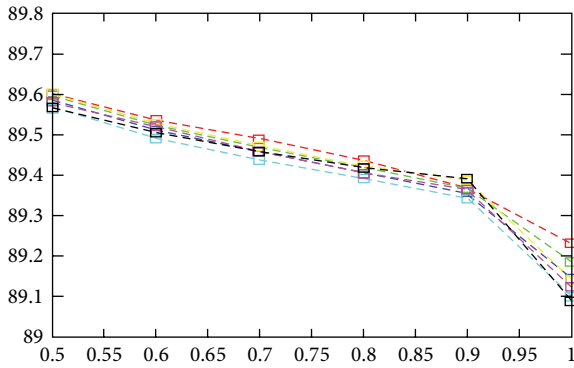
Figure 3: y axis shows prediction accuracy of PBCD ranking algorithm on Slashdot dataset and x axis indicates different values of α. Different colors show number of communities detected by the community detection algorithm.
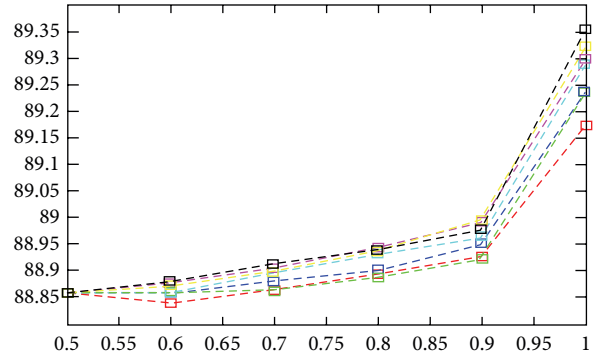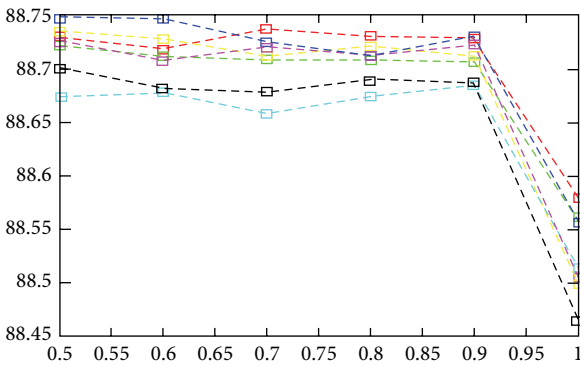

Figure 4: y axis shows prediction accuracy of PBCD ranking algorithm on Wikipedia dataset and x axis indicates different values of α. Different colors show number of communities detected by the community detection algorithm.
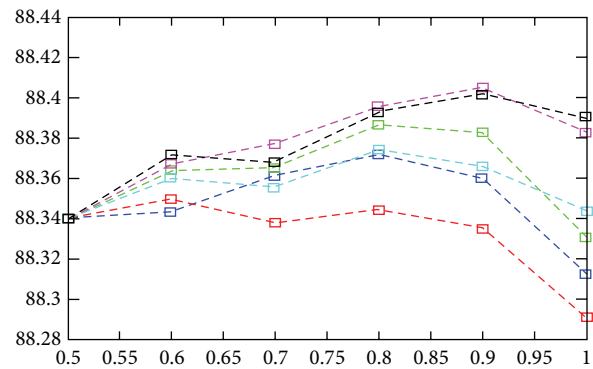

Figure 5: y axis shows prediction accuracy of RBCD ranking algorithm on Epinions dataset and x axis indicates different values of α. Different colors show number of communities detected by the community detection algorithm.
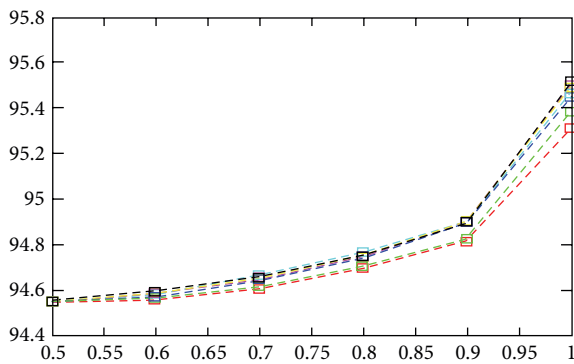
all communities. RBCD have similar outputs in Slashdot; when $\alpha = 1$, it has highest precision with maximum value of 89.335 for 10 communities. Results related to Slashdot are represented in Figure 6. RBCD also produced satisfactory results in Wikipedia. When $\alpha = 0.5$ it contains minimum value of 88.350 among all communities. In case of Com = 8,

Figure 6: y axis shows prediction accuracy of RBCD ranking algorithm on Slashdot dataset and x axis indicates different values of α. Different colors show number of communities detected by the community detection algorithm.

Figure 7: y axis shows prediction accuracy of RBCD ranking algorithm on Wikipedia dataset and x axis indicates different values of α. Different colors show number of communities detected by the community detection algorithm.

maximum accuracy of 88.405 is for $\alpha = 0.9$. When Com = 5, 6, 7, RBCD reach maximum accuracies at $\alpha = 0.8$ in which their values are 88.386, 88.372, and 88.374, respectively.

Similarly, for communities nine and ten maximum values are 88.4021 and 88.402, respectively, with $\alpha = 0.9$. Outputs for Wikipedia are shown in Figure 7.

4.2.3. HBCD on Datasets. We also implemented HBCD on datasets. Achieved results for Epinions are shown in Figure 8. In Epinions, $\alpha = 1$ has best accuracy for all communities. Moreover, community number four with value of 95.620 contains maximum among them. As for Slashdot, $\alpha = 1$ has the best accuracy and Com = 9 generates accuracy of 89.37 which has the highest value among all communities. Outputs for Slashdot are illustrated in Figure 9. There is the same story in Wikipedia. $\alpha = 1$ has the best accuracy and proves our new method. Among all of them community number seven has the best accuracy with value of 88.410. Related results for Wikipedia are presented in Figure 10.

5. Discussion

Random guessing of edge sign prediction on original datasets results in accuracy prediction of approximately 80 percent.
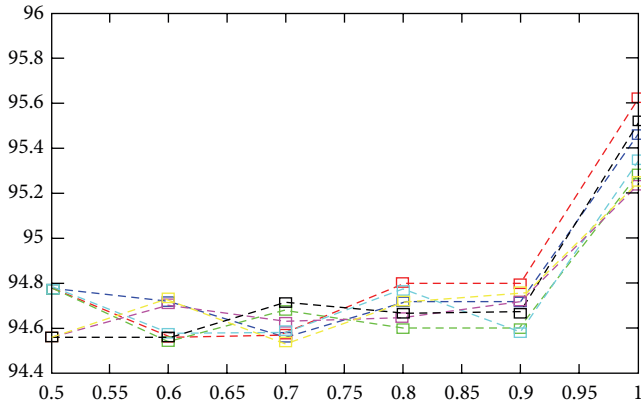
FIGURE 8: y axis shows prediction accuracy of HBCD ranking algorithm on Epinions dataset and x axis indicates different values of α. Different colors show number of communities detected by the community detection algorithm.

FIGURE 9: y axis shows prediction accuracy of HBCD ranking algorithm on Slashdot dataset and x axis indicates different values of α. Different colors show number of communities detected by the community detection algorithm.
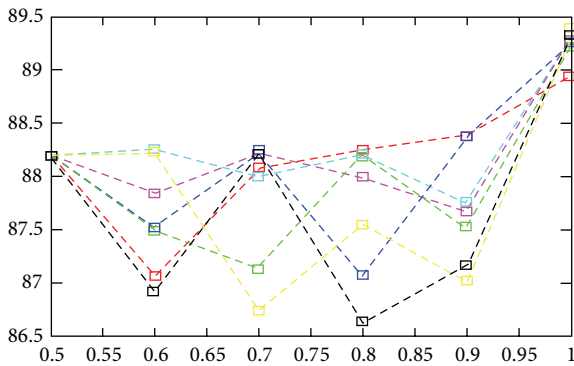

FIGURE 10: y axis shows prediction accuracy of HBCD ranking algorithm on Wikipedia dataset and x axis indicates different values of α. Different colors show number of communities detected by the community detection algorithm.

TABLE 1: Prediction accuracy using various methods on original datasets.

	Epinion	SlashDot	Wikipedia
Prediction accuracy of Leskovec [5]			
Leskovec	90.751	87.117	83.835
Prediction accuracy of Shahriari [3]			
Normal prestige	95.872	89.604	88.747
MPR	94.550	88.859	88.34
MHITS	94.770	88.185	88.359
Prediction accuracy achieved from our proposed algorithms			
PBCD	95.853	89.536	88.747
RBCD	95.515	89.355	88.405
HBCD	95.629	89.380	88.405

As Table 1 indicates accuracy of our work improves the rate of prediction about 10 to 15 percent.

Reference [5] applied degree-based features on Epinions, Slashdot, and Wikipedia and performed sign prediction with precisions of 90.751, 87.117, and 83.835, respectively. It can be perceived in Table 1 that the prediction rate of our work has significant improvement in comparison to [5].

Reference [3] also introduced new ranking algorithms, namely, MPR, MHITS, and improved precision achieved by [5]. In order to compare result of this paper with [3], we can consider $\alpha = 0.5$, as normal Prestige, MPR, and MHITS. In other words, in our proposed formulas, $\alpha = 0.5$ indicates that nodes inside and outside community have the same privilege. All in all, except precision of PBCD on Epinions and Slashdot, in other cases, our method produces better results in comparison with [3]. To put in a nutshell, we find out that our proposed approach outperforms previous works related to sign prediction problem, and it is indicated that local nodes in communities have higher impact on reputation and importance of other nodes in the network. Moreover, number
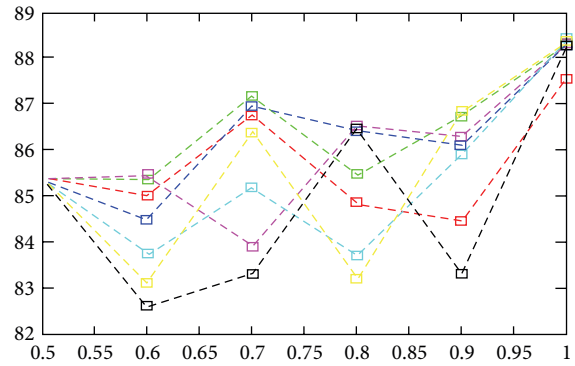
of communities can be estimated in these real-world datasets by analyzing output presented in the previous section. For example, in Wikipedia, PBCD algorithm produces best accuracies for community number six, and community number eight yields the best result for RBCD. Similarly HBCD detect seven communities in Wikipedia. It is very obvious that result of each algorithm is different with another one but it can be easily perceived that all of these numbers are close to each other. Therefore we can deduce that these community-based ranking algorithms are able to approximate the number of communities in these datasets.

6. Conclusion and Future Works

Complex networks have multidisciplinary roles in science comprising artificial intelligence, economics, and chemistry in which their usages have been increasing. Social networks as one branch of complex networks have got a lot of attention recently. One principal topic in social networks is to investigate evolution of graphs; thus researchers are trying to take prediction algorithms in order to find hidden relationships in these networks. A significant factor in predicting relations is ranking of people in societies. The aforementioned concepts

are expressed in social networks as sign prediction and ranking of nodes, respectively.

Nodes ranking algorithms which intend to determine how much a node is reputable in a network are studied in our document. Three community-based ranking algorithms are proposed in this paper. These ranking algorithms have two phases. In the first phase, nodes are assigned to different communities by applying community detection algorithm (in this phase, different community detection algorithms can be applied). In the second phase, rank of each node is computed based on its membership to its neighbors' communities and its incoming/outgoing positive/negative links. So we investigated the effect of community detection on the accuracy of sign prediction problem and compared our work with [3, 5]. Eventually, we deduced that our ranking algorithms outperform both methods and community-based ranking algorithms produce better accuracies in some cases.

Our experiment was performed to check which community number has the best accuracy. In this case, results may be affected by properties of this community detection algorithm. In future research, we intend to compare impact of different community detection methods on accuracy of edge sign prediction problem. The problem of overlapping community detection, especially, has gained much attention. Hence this problem and its effect on signed prediction can be investigated. We are also interested in working on parameter-free community detection methods suitable for signed graphs. Finally, to check the reliability of the method, it is good to test these approaches in person to person recommenders.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] D. Liben-Nowell and J. Kleinberg, " The link prediction problem for social networks," in *Proceedings of the 12th International Conference on Information and Knowledge Management (CIKM '03)*, pp. 556–559, November 2003.

[2] Y. Dong, J. Tang, S. Wu et al., "Link prediction and recommendation across heterogeneous social networks," in *Proceedings of the 12th IEEE International Conference on Data Mining (ICDM '12)*, pp. 181–190, December 2012.

[3] M. Shahriari and M. Jalili, "Ranking nodes in signed social networks," *Social Network Analysis and Mining*, vol. 4, article 172, 2014.

[4] J. Kunegis, A. Lommatzsch, and C. Bauckhage, "The Slashdot Zoo: mining a social network with negative edges," in *Proceedings of the 18th International World Wide Web Conference (WWW '09)*, pp. 741–750, April 2009.

[5] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," in *Proceedings of the 19th International Conference on World Wide Web*, pp. 641–650, New York, NY, USA, April 2010.

[6] K.-Y. Chiang, N. Natarajan, A. Tewari, and I. S. Dhillon, "Exploiting longer cycles for link prediction in signed networks," in *Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM '11)*, pp. 1157–1162, October 2011.

[7] M. Shahriari, O. A. Sichani, J. Gharibshah, and M. Jalili, "Predicting sign of edges in social networks based on users reputation and optimism," *Transactions on Knowledge Discovery from Data*. In press.

[8] L. Adamic and E. Adar, "How to search a social network," *Social Networks*, vol. 27, no. 3, pp. 187–203, 2005.

[9] J. Chen, W. Geyer, C. Dugan, M. Muller, I. Guy, and M. Carmel, "Make new friends, but keep the old: recommending people on social networking sites," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*, pp. 201–210, 2009.

[10] P. Symeonidis, E. Tiakas, and Y. Manolopoulos, "Transitive node similarity for link prediction in social networks with positive and negative links," in *Proceedings of the 4th ACM Recommender Systems Conference (RecSys '10)*, pp. 183–190, September 2010.

[11] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," in *Proceedings of the 19th International Conference on World Wide Web*, pp. 641–650, 2010.

[12] V. A. Traag, Y. Nesterov, and P. Van Dooren, "Exponential ranking: taking into account negative links," in *Social Informatics*, vol. 6430 of *Lecture Notes in Computer Science*, pp. 192–202, Springer, Berlin, Germany, 2010.

[13] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[14] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215–239, 1978.

[15] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *The Journal of Mathematical Sociology*, vol. 2, no. 1, pp. 113–120, 1972.

[16] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[17] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer Networks and ISDN Systems*, vol. 56, no. 18, pp. 3825–3833, 2012.

[18] K. Zolfaghar and A. Aghaie, "Mining trust and distrust relationships in social Web applications," in *Proceedings of the 6th IEEE International Conference on Intelligent Computer Communication and Processing*, pp. 73–80, 2010.

[19] C. de Kerchove and P. van Dooren, "The PageTrust algorithm: how to rank web pages when negative links are allowed?" in *Proceedings of the 8th SIAM International Conference on Data Mining*, pp. 346–352, April 2008.

[20] A. Mishra and A. Bhattacharya, "Finding the bias and prestige of nodes in networks based on trust scores," in *Proceedings of the 20th International Conference on World Wide Web (WWW '11)*, pp. 567–576, April 2011.

[21] P. Anchuri and M. M. Ismail, "CommunityDetection in Signed Networks".

[22] V. A. Traag and J. Bruggeman, "Community detection in networks with positive and negative links," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 80, no. 1, Article ID 036115, 7 pages, 2009.

[23] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 23, pp. 8577–8582, 2006.

[24] P. Anchuri and M. Magdon-Ismail, "Communities and balance in signed networks: a spectral approach," in *Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 235–242, August 2012.

[25] P. Doreian and A. Mrvar, "Partitioning signed social networks," *Social Networks*, vol. 31, no. 1, pp. 1–11, 2009.

[26] T. Zhang, H. Jiang, Z. Bao, and Y. Zhang, "Characterization and edge sign prediction in signed networks," *Journal of Industrial and Intelligent Information*, vol. 1, no. 1, pp. 19–24, 2013.

[27] S. H. Yang, A. J. Smola, B. Long, H. Zha, and Y. Chang, "Friend or frenemy? Predicting signed ties in social networks," in *Proceedings of the 35th Annual ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '12)*, pp. 555–564, August 2012.

[28] D. Cartwright and F. Harary, "Structural balance: a generalization of Heider's theory," *Psychological Review*, vol. 63, no. 5, pp. 277–293, 1956.

[29] P. Doreian, "Evolution of human signed networks," *metodološki Zvezki*, vol. 1, no. 2, pp. 277–293, 2004.

[30] M. Szell, R. Lambiotte, and S. Thurner, "Multirelational organization of large-scale social networks in an online world," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 31, pp. 13636–13641, 2010.

[31] S. Maniu, B. Cautis, and T. Abdessalem, "Building a signed network from interactions in Wikipedia," *Proceedings of the 1st ACM SIGMOD Workshop on Databases and Social Networks (DBSocial '11)*, pp. 19–24, 2011.

[32] S. R. T. Antal, P. L. Krapivsky, and S. Redner, "Social balance on networks: the dynamics of friendship and enmity," *Physica D. Nonlinear Phenomena*, vol. 224, no. 1-2, pp. 130–136, 2006.

[33] P. Doreian, "A multiple indicator approach to blockmodeling signed networks," *Social Networks*, vol. 30, no. 3, pp. 247–258, 2008.

[34] S. A. Marvel, J. Kleinberg, R. D. Kleinberg, and S. H. Strogatz, "Continuous-time model of structural balance," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 5, pp. 1771–1776, 2011.

[35] N. P. Hummon and P. Doreian, "Some dynamics of social balance processes: bringing Heider back into balance theory," *Social Networks*, vol. 25, no. 1, pp. 17–49, 2003.

[36] G. Adejumo, P. R. Duimering, and Z. Zhong, "A balance theory approach to group problem solving," *Social Networks*, vol. 30, no. 1, pp. 83–99, 2008.

[37] L. Page, S. Brin, R. Motawani, and T. Winograd, "The page rank citation ranking:bringing order to the web," 1999.

[38] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2006.

[39] T. M. Mitchell, *Machine Learning*, McGraw Hill, 1st edition, 1997.

[40] J. Ye, H. Cheng, Z. Zhu, and M. Chen, "Predicting positive and negative links in signed social networks by transfer learning," in *Proceedings of the 22nd International Conference on World Wide Web (WWW '13)*, pp. 1477–1488, 2013.

[41] J. Kunegis, "What is the added value of negative links in online social networks?" in *Proceedings of the 22nd International Conference on World Wide Web*, pp. 727–736, 2013.

[42] K.-Y. Chiang, C.-J. Hsieh, N. Natarajan, I. S. Dhillon, and A. Tewari, "Prediction and clustering in signed networks: a local to global perspective," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1177–1213, 2014.

[43] A. Javari and M. Jalili, "Cluster-based collaborative filtering for sign prediction in social networks with positive and negative links," *ACM Transactions on Intelligent Systems and Technology*, vol. 5, no. 2, article 24, 2014.