

Supervised User Ranking in Signed Social Networks

Xiaoming Li, Hui Fang,* Jie Zhang

School of Computer Science and Engineering, Nanyang Technological University, Singapore

*Research Institute for Interdisciplinary Sciences and School of Information Management and Engineering

*Shanghai University of Finance and Economics, China

lixiaoming@ntu.edu.sg, fang.hui@mail.shufe.edu.cn, zhangj@ntu.edu.sg

Abstract

The task of user ranking in signed networks, aiming to predict potential friends and enemies for each user, has attracted increasing attention in numerous applications. Existing approaches are mainly extended from heuristics of the traditional models in unsigned networks. They suffer from two limitations: (1) mainly focus on global rankings thus cannot provide effective personalized ranking results, and (2) have a relatively unrealistic assumption that each user treats her neighbors' social strengths indifferently. To address these two issues, we propose a supervised method based on random walk to learn social strengths between each user and her neighbors, in which the random walk more likely visits "potential friends" and less likely visits "potential enemies". We learn the personalized social strengths by optimizing on a particularly designed loss function oriented on ranking. We further present a fast ranking method based on the local structure among each seed node and a certain set of candidates. It much simplifies the proposed ranking model meanwhile maintains the performance. Experimental results demonstrate the superiority of our approach over the state-of-the-art approaches.

Introduction

Signed social networks have become increasingly popular (Tang et al. 2016), in which relationship between online users is not limited to be positive (e.g. friend and trust) anymore, but also includes the negative one which is much more consistent with the real social life. More and more online platforms are built based on signed structures, such as Slashdot (friend or foe), Epinions (trust or distrust), and Wikipedia RFA (agree or disagree). The ever-growing interest in signed networks has heightened the need to rethink the user ranking problem, which becomes non-trivial because of the existence of the negative links.

Traditional ranking methods in unsigned networks only consider positive links, which rank user pairs by the probability of forming positive links (i.e. link prediction). In signed networks, however, the task of ranking users transforms to that of ranking potential 'friends' on the top of the list whereas ranking 'enemies' on the bottom. Therefore, traditional methods in unsigned networks cannot be directly applied to the signed scenarios. For signed networks, several

approaches have been proposed by revising the traditional models, which can be summarized into two main types. One representative type of approaches ranks user pairs based on heuristic similarity scores (Symeonidis and Tiakas 2014), which merely consider local node/edge attributes but fail to capture the global network structure. Thus, they cannot guarantee a satisfying performance. The other type is derived from random walk, which is a dominant technique for user ranking in networks. For example, Shams et al. (2016) revise random walk to be computable and applicable in signed networks by firstly converting signed networks into unsigned ones and then obtaining ranking scores accordingly.

However, these approaches aim to generate a global ranking list for the whole network, which could easily lead to a relatively unfair scenario where some users might have a large number of potential links in the ranking list while most users have very few or even no links. In this case, they cannot be easily adapted for many real-world applications such as social recommendation or social-aware product recommendation. In contrast, personalized user ranking, which generates a ranking list for each individual, is more practical and realistic (Jung et al. 2016). Besides, the ranking list for a user provided by existing random walk methods is fixed given a certain network snapshot (i.e. the network structure). They inappropriately assume all the links have the same weights (i.e. social strengths, a.k.a. link strengths). In other words, they cannot learn each individual's own opinions towards her neighbors, such as what kind of user link (i.e. neighbors) is more important.

To fill the research gap, we propose Signed Supervised Random Walk (SSRW), through which we learn social strengths that capture a user's different preferences towards different neighbors, and thus to better facilitate the task of personalized user ranking. More specifically, instead of considering the random walk in a given network snapshot (i.e. training data), we split the training data into two parts in terms of the timestamp (denoted as A and B), and learn social strengths (i.e. transition probabilities) so that random walk more likely visits those newly positively connected nodes (i.e. in B compared to A) whereas more reluctantly visits the newly negatively connected nodes. We conduct experiments on four real-world datasets and the results show that SSRW's performance has an improvement of 6.05% compared to the state-of-the-art approaches. To im-

prove SSRW’s efficiency but simultaneously maintain its effectiveness, we also design a fast ranking method (F-SSRW) based on the local structure among each seed node and a certain set of candidates of the seed node. It has been demonstrated that F-SSRW can maintain the performance in contrast with the original SSRW when the ranking candidates of a user satisfy the requirement of having substantial common neighbors with the user.

Related Work

In this section, we briefly review related works on user ranking in signed networks. We summarize the literature into two parts: the traditional link prediction and the task of personalized user ranking.

For traditional link prediction in signed networks, feature-based approaches are dominant, which design topological features with a regression model (Leskovec et al. 2010; Chiang et al. 2011). Regression results are then deployed to distinguish positive and negative links. Another representative type is low-rank models. For example, Hsieh et al. (2012) propose a matrix factorization model to infer link signs. Li et al. (2018) design the FILE model to rank all user pairs by the order of positive, no-relation and negative links. However, the aforementioned approaches aim to distinguish positive and negative links, or rank user pairs globally. In this case, the users who have limited social connections in the past are put in an unfavorable position by these approaches and will receive few potential links.

On the contrary, the personalized user ranking in signed networks tries to provide a ranking list for each individual user. It is worth noting that there are lots of existing works for personalized user ranking in unsigned networks, and the representative approaches include similarity-based ones (Sarkar et al. 2011; Brzozowski and Romero 2011), random walk based models (Yin et al. 2010; Zhao et al. 2018) and low-rank models (Man et al. 2016; Wang, Shi, and Yeung 2017; Nelakurthi and He 2017). However, these existing methods cannot be directly applied in signed networks because of the existence of negative links. Therefore, a few works have strived to extend the traditional methods into the signed scenarios, which can also be summarized into two categories: similarity-based approaches and the random walk based ones.

For similarity-based approaches, Symeonidis et al. (2014) propose a similarity metric based on users’ out/in degree of positive and negative links. A higher similarity score between two users indicates a higher chance to establish a positive link, while a lower score indicates a possible negative link. Zhu et al. (2017) use the number of common friends minus the number of common enemies as the similarity metric. However, these studies adopt heuristic similarity settings, and cannot gain good performance. Song et al. (2015) aim to rank user pairs as the order of positive, no-relation and negative. They learn users’ latent vectors by adopting matrix factorization technique, and model the ranking score as the inner product between the corresponding user vectors.

For random walk based approaches, Shahriari et al. (2014) firstly split the signed graph into two graphs: a positive and a negative one, and then apply random walk with restart on

each graph. They finally combine the results from two random walks to generate one ranking list for each user. In (Wu et al. 2016), a signed network is converted into a positively weighted graph, and then obtain the ranking list using the random walk technique. Jung et al. (2016) propose a model named SRWR, which introduces a sign into a random surfer so that negative links can be also considered by changing the sign of walking.

In summary, current approaches mainly focus on global ranking rather than the personalized perspective. Besides, they assume all the links in the network have the same weights. In other words, they ignore the difference in social strengths, which actually play a key role in personalized user ranking. In this paper, we propose a supervised method to learn social strengths, which can be used to obtain a better personalized ranking performance.

Problem Formulation and Transformation

We first define the user ranking problem as: given a seed node i in a signed social network $S \in \mathbb{R}^{n \times n}$ (n is the number of users) with $S_{ij} \in \{1, 0, -1\}$, we aim to rank all the users $m \in \{m | S_{im} = 0\}$ in the present, by the probability of transforming (i, m) to a positive link, maintaining no-relation, or transforming to a negative link in the future. We strive to answer that: “Of user pairs (i, m_1) and (i, m_2) , which pair is more likely to become friends (or enemies)?”

As aforementioned, social strengths have been ignored by existing approaches in the literature. In fact, the intuition that a user’s preferences towards other users (even towards the set of already formed friends) are different, has been widely explored and leveraged in the unsigned networks (Xiang et al. 2010; Katsimpras et al. 2015). We thus adopt the idea and consider that social strengths can also impact link formation in signed networks. Therefore, we transform the user ranking problem in signed networks into a supervised learning problem, by which we learn social strengths to better facilitate user ranking task.

Formally, for any link (i, j) (i.e. user pair), we learn its link strength $f_{w_i}(x_{ij})$, in which $f(\cdot)$ is a differentiable function parameterized by w_i and x_{ij} is the observable feature vector of the link. By doing this, we obtain a weighted network with different edge strength $f_{w_i}(x_{ij})$. We use r_{im} to represent m ’s ranking score given seed node i , which is the probability obtained from random walk based on the weighted network. Thus, the problem is reduced to:

“Given a seed node i and any nodes $m, n \in C_i$, we aim to find the optimal w_i , which satisfies: if there is a new positive link generated from i to m in future meanwhile there is no positive link (i.e. no-relation or a negative link) between i and n , the ranking score should follow $r_{im} \geq r_{in}$. Similarly, $r_{im} \leq r_{in}$ if a negative link is formed between i and m .”

The main notations are summarized in Table 1. For the seed user i , we aim to optimize the following function:

$$\begin{aligned} \underset{w}{\text{Minimize}} \quad & F(w_i) = \|w_i\|^2 + \frac{1}{\theta} \sum e_{mn} \\ \text{s.t.} \quad & r_{in} - r_{im} + e_{mn} \geq 1, \forall m, n \in C_i \end{aligned} \quad (1)$$

where $\theta = |N_i| \cdot |P_i| + |U_i| \cdot |P_i| + |N_i| \cdot |U_i|$, in which $|P_i|$, $|U_i|$ and $|N_i|$ are the number of nodes in the corresponding

Table 1: Notations

S	adjacency matrix, $S \in \mathbb{R}^{n \times n}$, $S_{ij} \in \{1, 0, -1\}$
i	seed node
(i, j)	link $i \rightarrow j$
C_i	node i 's candidate set, $C_i = P_i \cup U_i \cup N_i$
m, n	node $m, n \in C_i$
x_{ij}	feature vector of (i, j)
$f_{w_i}(x_{ij})$	strength function of (i, j) parameterized by w_i
P_i	$P_i = \{m S_{im}^t = 0 \ \& \ S_{im}^{t+1} = 1, m \in C_i\}$
N_i	$N_i = \{m S_{im}^t = 0 \ \& \ S_{im}^{t+1} = -1, m \in C_i\}$
U_i	$U_i = \{m S_{im}^t = 0 \ \& \ S_{im}^{t+1} = 0, m \in C_i\}$
δ_i	node i 's bias on distrust
r_{im}^+	seed node's positive ranking score towards m
r_{im}^-	seed node's negative ranking score towards m
r_{im}	the final ranking score
q	unit vector with $q_i = 1$

set respectively, and $\sum e_{mn}$ is equivalent to:

$$\alpha \sum_{m \in N_i, n \in P_i} e_{mn}^1 + \beta \sum_{m \in U_i, n \in P_i} e_{mn}^2 + \gamma \sum_{m \in N_i, n \in U_i} e_{mn}^3 \quad (2)$$

The corresponding weights α, β, γ are user-specified and application-dependent, denoting the penalties of different types of errors, where e_{mn}^1 is type 1 error that $m \in N_i, n \in P_i$, e_{mn}^2 is type 2 error that $m \in U_i, n \in P_i$, and e_{mn}^3 is type 3 error that $m \in N_i, n \in U_i$. The objective of the optimization equation 1 is to find the optimal parameter set w and can be proceeded once the ranking score r and $\frac{\partial r}{\partial w}$ are obtained. Therefore, our main research question is reduced to ‘‘how to design the function r and then calculate its derivation accordingly’’. In this paper, we extend the supervised random walk technique (Backstrom and Leskovec 2011) to signed networks, i.e. *signed supervised random walk*, for obtaining the ranking score r .

SSRW: Signed Supervised Random Walk

In SSRW, we first follow the idea of the sign surfer (Jung et al. 2016) to make random walk workable in signed networks. A surfer begins with a positive sign (since it will always trust itself) and then visits other nodes, and the sign flips if it meets a negative link, otherwise the sign remains unchanged.

The intuition of the sign flips is adopted from balance theory (Antal et al. 2006), which can be explained as ‘‘my friend’s friend is my friend’’ or ‘‘my enemy’s friend is my enemy’’, and considered solidly effective in signed networks (Leskovec et al. 2010). As a personalized ranking approach, the surfer will restart with a probability c , and the sign will be reset to positive. When the surfer visits a certain node, the sign can be either positive or negative since it can reach the node via different routes. Therefore, for the seed node i , each node m in C_i will eventually get two ranking scores: a positive one (i.e. r_{im}^+) and a negative one (i.e. r_{im}^-), from which we can obtain the final ranking score r as:

$$r_{im} = r_{im}^+ - \delta_i r_{im}^- \quad (3)$$

where δ_i is user i 's bias on distrust (i.e. negative relationship), as some users will be more likely to distrust others, while some might be more reluctantly to distrust others.

Next, we investigate the connection between link strength $f_{w_i}(\cdot)$ and r . Let a_{ij} be the normalized link strength of (i, j) , which equals to 0 if there is no direct link from node i to j :

$$a_{ij} = \frac{f_{w_i}(x_{ij})}{\sum_z f_{w_i}(x_{iz})}, \exists(i, j) \quad (4)$$

In this case, the matrix of transition probability Q is:

$$Q_{ij} = \begin{cases} a_{ij} & \exists(i, j) \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

and we split Q to two matrices Q^+ and Q^- according to the link sign between users. Specifically, $Q_{ij} \in Q^+$ if $S_{ij} = 1$, and $Q_{ij} \in Q^-$ if $S_{ij} = -1$. We thus have:

$$Q = Q^+ + Q^- \quad (6)$$

Based on the setting and sign surfer, seed node i 's ranking score matrices r_i^+ and r_i^- towards other users are recursively entangled and derived as follows:

$$r_i^+ = (1 - c)(Q^+ r_i^+ + Q^- r_i^-) + cq \quad (7)$$

$$r_i^- = (1 - c)(Q^+ r_i^- + Q^- r_i^+)$$

where q is the unit vector with $q_i = 1$. Thus, for a node $m \in C_i$, its ranking score can be written as:

$$r_{im}^+ = (1 - c) \sum_j (r_{ij}^+ Q_{jm}^+ + r_{ij}^- Q_{jm}^-) \quad (8)$$

$$r_{im}^- = (1 - c) \sum_j (r_{ij}^- Q_{jm}^+ + r_{ij}^+ Q_{jm}^-)$$

We take the corresponding derivations to compute $\frac{\partial r_{im}}{\partial w_i}$:

$$\begin{aligned} \frac{\partial r_{im}^+}{\partial w_i} &= (1 - c) \sum_j Q_{jm}^+ \frac{\partial r_{ij}^+}{\partial w_i} + r_{ij}^+ \frac{\partial Q_{jm}^+}{\partial w_i} + Q_{jm}^- \frac{\partial r_{ij}^-}{\partial w_i} + r_{ij}^- \frac{\partial Q_{jm}^-}{\partial w_i} \\ \frac{\partial r_{im}^-}{\partial w_i} &= (1 - c) \sum_j Q_{jm}^+ \frac{\partial r_{ij}^-}{\partial w_i} + r_{ij}^- \frac{\partial Q_{jm}^-}{\partial w_i} + Q_{jm}^- \frac{\partial r_{ij}^+}{\partial w_i} + r_{ij}^+ \frac{\partial Q_{jm}^+}{\partial w_i} \end{aligned} \quad (9)$$

where $\frac{\partial Q_{jm}}{\partial w_i}$ can be derived as:

$$\frac{\partial Q_{jm}}{\partial w_i} = \frac{\frac{\partial f_{w_i}(x_{jm})}{\partial w_i} (\sum_z f_{w_i}(x_{jz})) - f_{w_i}(x_{jm}) (\sum_z \frac{\partial f_{w_i}(x_{jz})}{\partial w_i})}{(\sum_z f_{w_i}(x_{jz}))^2} \quad (10)$$

Similarly, $r_{im}^+, r_{im}^-, \frac{\partial r_{im}^+}{\partial w_i}$, and $\frac{\partial r_{im}^-}{\partial w_i}$ are recursively entangled, and can be computed iteratively as in Algorithm 1.

Based on the computed $r_{im}^+, r_{im}^-, \frac{\partial r_{im}^+}{\partial w_i}$, and $\frac{\partial r_{im}^-}{\partial w_i}$, we then apply the gradient descent method (Chapelle and Keerthi 2010) to find a local minimum for Equation 1.

The advantage of SSRW is to well capture the global structure of a network, and thus to obtain the global optimum, but it is also computationally expensive. Through SSRW, every single user $m \in \{m | S_{im} = 0\}$ will eventually get a ranking score with respect to the seed node i . However, it is intuitive that most users will not connect to the seed node ultimately. In other words, we do not need to consider all the nodes in the network, since the link formation probabilities between the seed node and them are negligible. Inspired by this, we thus propose a fast ranking method, denoted as F-SSRW and detailed in the next section.

Algorithm 1 Computation of r_{im}^+ , r_{im}^- , $\frac{\partial r_{im}^+}{\partial w_i}$, and $\frac{\partial r_{im}^-}{\partial w_i}$

Initialize: $r_{im}^{+(0)}$, $r_{im}^{-(0)}$, $\frac{\partial r_{im}^{+(0)}}{\partial w_i} = 0$, $\frac{\partial r_{im}^{-(0)}}{\partial w_i} = 0$,

$t = 1$

repeat

 Calculate $r_{im}^{+(t)}$, $r_{im}^{-(t)}$ based on Equation 8;

$t = t + 1$;

until Converge

$r_{im}^+ = r_{im}^{+(t-1)}$, $r_{im}^- = r_{im}^{-(t-1)}$

$t = 1$

repeat

 Calculate $\frac{\partial r_{im}^{+(t)}}{\partial w_i}$, $\frac{\partial r_{im}^{-(t)}}{\partial w_i}$ based on Equation 9;

$t = t + 1$;

until Converge

$\frac{\partial r_{im}^+}{\partial w_i} = \frac{\partial r_{im}^{+(t-1)}}{\partial w_i}$, $\frac{\partial r_{im}^-}{\partial w_i} = \frac{\partial r_{im}^{-(t-1)}}{\partial w_i}$

Output: r_{im}^+ , r_{im}^- , $\frac{\partial r_{im}^+}{\partial w_i}$, $\frac{\partial r_{im}^-}{\partial w_i}$

F-SSRW

As aforementioned, given a seed node, we aim to speed up the ranking procedure by only considering a certain set of candidates who have much larger probability to form a link with the seed node. Intuitively, we select candidates by two criteria: (1) hop distance between the seed node and the candidates; (2) the number of their mutual neighbors.

In this section, we first empirically investigate these two factors which may influence link formation. We mainly conduct the analysis in Epinions dataset¹, which contains the timestamp of every link formation over 30 months. We also perform analysis in other three real-world signed networks: Slashdot², Wikipedia RFA³ and Bitcoins⁴.

Hop distance. We check the hop distance between two users when they are linked. As can be seen in Table 2, two-hop distance is dominant with which a much larger number of links are configured compared to all other ones. Besides, as shown in Figure 1(a), the majority of connected links have direct common neighbors (i.e. the corresponding two users are within two-hop distance). In other words, users are more likely to get linked if they are within two-hop distance. This intuition has also been well validated in unsigned networks (Sun et al. 2005).

Table 2: The hop distance between users when they form new links in Epinions dataset.

Hop distance	Link counts	Ratio
2	187,990	71.65%
3	32,014	12.20%
4+	42,371	16.15%

¹<http://www.trustlet.org/epinions.html>

²snap.stanford.edu/data/soc-Slashdot0902.html

³snap.stanford.edu/data/wiki-RFA.html

⁴cs.umd.edu/~srijan/wsn

Number of mutual neighbors. We further check the relation between link formation and the number of mutual neighbors. Figure 1(b) depicts the cumulative distribution of the number of mutual neighbors between linked users. Specifically, more than 90% of links have at least 3 mutual neighbors, and more than 82% of them have at least 5 mutual neighbors. Therefore, in signed networks, users with more common neighbors are more likely to form links (either positive or negative).

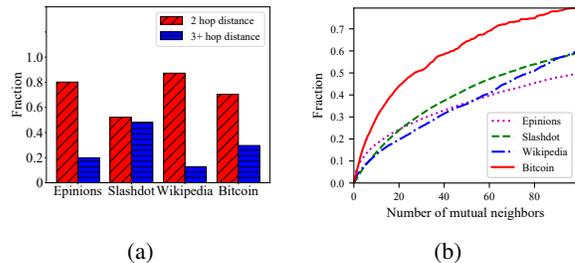


Figure 1: (a) Distribution of hop distance; (b) Distribution of the number of mutual neighbors.

The results of data analysis further inspire our design of the fast ranking model which mainly focuses on local structure among the seed node and certain set of candidates. Specifically, for a seed node i , we prune the graph to include only $\{i, \{j\}, m\}$, in which j is a common neighbor of i and m , and $j \in \{j | \exists(i, j) \& (j, m)\}$. In other words, we only consider the candidates who are two-hop distance with the seed node i . Besides, we further reduce the number of candidates by constraining each to have at least T mutual neighbors with i . In this case, we keep the candidates which are more likely to obtain a higher r_{im}^+ or r_{im}^- . In the pruned graph, based on random walk, a candidate node m 's ranking score r_{im} can be estimated as:

$$r_{im} \propto \sum_j r_{ij} a_{jm} \quad (11)$$

Because the pruned graph only contains i 's two-hop neighbors, we can further approximate r_{im} as

$$r_{im} \propto \sum_j r_{ij} a_{jm} \propto \sum_j a_{ij} a_{jm} \quad (12)$$

where we strategically ignore the routes with more than 2 hops. We argue that the gap between the approximation and global optimum (obtained from random walk) is actually small since the contribution of 3-hop routes is limited in term of hitting probabilities compared to 2-hop routes. The intuition has been shown in data analysis meanwhile being well validated in the literature (Sun et al. 2005).

Similarly, grounded on balance theory, we will also obtain two ranking scores, r_{im}^+ and r_{im}^- in the signed network:

$$\begin{cases} r_{im}^+ \propto \sum a_{ij} a_{jm} & \text{if } S_{ij} S_{jm} = 1 \\ r_{im}^- \propto \sum a_{ij} a_{jm} & \text{if } S_{ij} S_{jm} = -1 \end{cases} \quad (13)$$

Finally, we obtain r_{im} following Equation 3, denoted as:

$$r_{im} = \sum_j (r_{im}^{j+} - \delta_i r_{im}^{j-}) \quad (14)$$

where r_{im}^{j+} and r_{im}^{j-} denote the respective ranking score contributed by the route $i \rightarrow j \rightarrow m$, and following Equation 4:

$$r_{im}^j = \frac{f_{w_i}(x_{ij}) \cdot f_{w_i}(x_{jm})}{\sum_k f_{w_i}(x_{ik}) \sum_k f_{w_i}(x_{jk})} \quad (15)$$

By taking derivation on equation 14, we thus obtain:

$$\frac{\partial r_{im}}{\partial w_i} = \sum_j \left(\frac{\partial r_{im}^{j+}}{\partial w_i} - \delta_i \frac{\partial r_{im}^{j-}}{\partial w_i} \right) \quad (16)$$

Based on the computed r_{im} and $\frac{\partial r_{im}}{\partial w_i}$, we then apply the gradient descent method to find a local minimum for Equation 1.

Discussion

F-SSRW is a simplified version of SSRW, which works on the pruned graph and only focuses on selected candidates. Therefore, in F-SSRW, we can only obtain approximate ranking scores for these candidates. In contrast, SSRW works on the global graph and every single user $m \in \{m | S_{im} = 0\}$ will eventually get a ranking score. For both of them, the time complexity of the optimization method is $O(|C_i|^2)$. For each iteration to get the derivation, SSRW takes $O(|E| + |V|)$, in which $|E|$ is the number of links and $|V|$ is the number of nodes in the graph, whereas in each iteration F-SSRW takes $O(|C_i|)$. As the candidate set and the graph in F-SSRW are much smaller than those in SSRW, accordingly the efficiency of the F-SSRW is much more largely improved than that in SSRW.

Experiments

We conduct experiments on four real-world datasets and compare our approaches with the state-of-the-art methods.

Experimental Settings

Data. We employ the four datasets (i.e. Epinions, Slashdot, Wikipedia RFA and Bitcoin), which are the only public available datasets with signed structure. In this study, we focus on the users (i.e. seed nodes) who are active in the social networks, where the activeness is measured by user’s degree. Specifically, to conform with the previous studies (Backstrom and Leskovec 2011), the selected users’ degree is larger than 20. We randomly select 200 of them as seed nodes. Besides, to make a more comprehensive evaluation, we adopt different criteria for candidate node selection. Based on the data analysis, we only select two-hop neighbors as candidates, meanwhile further filter them by the number of mutual neighbors with the seed node. Specifically, we use $d \geq 3$, $d \geq 5$ and $d \geq 10$. The major statistics of the datasets are listed in Table 3.

Table 3: Dataset statistics.

		Epinions	Slashdot	Wikipedia	Bitcoin
	Users	131, 828	82, 140	9, 654	3, 783
	Positive links	717, 667	425, 072	87, 766	22, 650
	Negative links	123, 705	124, 130	16, 788	1, 536
$d \geq 3$	Candidate users	800.4	137.4	243.9	201.1
(Per seed)	Linked users	54.4	9.8	34.1	29.9
$d \geq 5$	Candidate users	382.3	41.6	106.5	85.6
	Linked users	44	5.7	23.8	20.6
$d \geq 10$	Candidate users	168.4	14.9	22.9	26.7
	Linked users	31.9	3.5	9	11.4

Evaluation Metrics. We use 2-fold cross-validation for training and testing, and utilize GAUC (Generalized AUC) (Song and Meyer 2015) to measure the ranking performance, which is formulated as:

$$\frac{1}{|P| + |N|} \left(\frac{1}{|U| + |N|} \sum_{m \in P} \sum_{n \in U \cup N} I(r(m) > r(n)) + \frac{1}{|U| + |P|} \sum_{m \in N} \sum_{n \in U \cup P} I(r(m) < r(n)) \right)$$

Another metric is precision@top k , by which we evaluate the performance of the link recommendation. Specifically, we use PPrec@ k (NPrec@ k) to denote the ratio of positive (or negative) links in the top (or bottom) k prediction.

Benchmarking approaches. We compare with state-of-the-art approaches, including similarity-based models: Similarity with Positive and Negative Relations (SPNR) (Zhu et al. 2017), friend Transitive Node Similarity (TNS) (Symeonidis and Tiakas 2014), Social Feature Model (SFM) (Li et al. 2017), and random walk based model: Signed Random Walk with Restart (RWR) (Jung et al. 2016). We also compare our personalized models with the global version of our model (G-SSRW), which is based on SSRW but strives to minimize the sum of losses over all seed nodes in the network, instead of loss minimization over every seed node.

Parameter settings. In this experiment, we consider 5 features for the vector x to describe a user pair. Two of the features are two users’ degrees respectively, which imply their activeness; and the rest three are the number of their common friends, enemies, and frenemies (one’s friend but the other one’s enemy) respectively, which describe the social patterns within their joint relationship. we utilize the linear model to represent the link strength, i.e., $f_w(x) = w^T x$, where w can be seen as the weight vector of the features, and denote importance degrees of the corresponding features.

For the benchmark approaches, we set the parameters recommended in the literature. In SSRW, there are five hyperparameters: δ , α , β , γ and the restart probability c . The first four are application-dependent, and in view of simplicity and fair comparisons, we make them equal to 1 respectively. Besides, we set $c = 0.2$ for SSRW in the comparative experiments considering favourable performance of our model in this setting.

Table 4: Performance of different methods. The best performance is highlighted in bold, and the second-best one (except SSRWs) is marked by *. ‘Improvement’ indicates the improvement of SSRW over the model having the highest performance among existing models.

Datasets	SPNR	TNS	SFM	RWR	G-SSRW	F-SSRW	SSRW	Improvement
Epinions@3	0.619	0.462	0.609	0.628*	0.644	0.639	0.678	7.96 %
Epinions@5	0.621	0.475	0.633	0.671*	0.660	0.676	0.702	4.62 %
Epinions@10	0.598	0.499	0.677*	0.647	0.673	0.729	0.743	9.75 %
Slashdot@3	0.619	0.567	0.541	0.633*	0.601	0.626	0.645	1.90 %
Slashdot@5	0.605	0.563	0.537	0.627*	0.612	0.634	0.659	5.10 %
Slashdot@10	0.554	0.628	0.569	0.645*	0.658	0.718	0.715	11.32 %
Wikipedia@3	0.549	0.545	0.487	0.568*	0.583	0.587	0.633	11.44 %
Wikipedia@5	0.544	0.582	0.574	0.579*	0.612	0.619	0.638	10.38 %
Wikipedia@10	0.558	0.598	0.679*	0.596	0.647	0.658	0.681	0.29 %
Bitcoin@3	0.589	0.472	0.554	0.613	0.574	0.588	0.601*	-0.33 %
Bitcoin@5	0.596*	0.490	0.585	0.583	0.599	0.601	0.614	6.67 %
Bitcoin@10	0.573	0.557	0.640*	0.615	0.621	0.663	0.657	3.59 %

Experimental Results

Here, we show the comparison results under different scenarios and the impact of different parameters on the approaches.

Overall Performance. Table 4 depicts the experimental results under different scenarios in terms of GAUC. Overall, SSRW achieves the best performance when compared with other approaches across all the datasets, and the improvement is 6.05% on average. The results of t-test demonstrate that the improvement of our approach is statistically significant ($p\text{-value} < 0.01$).

Particularly, among all these approaches, similarity-based models (SPNR and TNS) perform the worst under almost all scenarios, indicating that traditional similarity metrics cannot be easily extended into signed networks. The global ranking approaches, including SPNR, TNS and SFM, perform worse than the personalized approaches (e.g. RWR). On the contrary, SSRW and F-SSRW perform much better than RWR, validating the effectiveness of the supervised approach and personalized link strengths. Besides, SSRW performs much better than G-SSRW, implying the reasonability of personalized user ranking compared to the global one.

With regard to the SSRW and F-SSRW, we can see that SSRW is better than F-SSRW, which performs almost better than all the rest approaches except SSRW. In addition, as the increase of d , the performance gap between F-SSRW and SSRW becomes smaller, further demonstrating the effectiveness of our heuristic intuitions.

Impact of candidate selection by d . To demonstrate the robustness of the proposed approaches, we check the performance of different approaches in terms of GAUC as the change of d in the range of $[1, 10]$. We compare SSRW and F-SSRW with RWR as it performs the best among all the benchmarks. As shown in Figure 4, we can find that SSRW consistently performs better than RWR and F-SSRW. As d increases, the performance gap between F-SSRW and SSRW becomes smaller, validating the soundness of our argument that a greater d can assure a better approximation

of F-SSRW compared to SSRW. In other words, considering the efficiency, we can adopt the F-SSRW model in those applications where the candidate nodes have substantial common neighbors with the seed node.

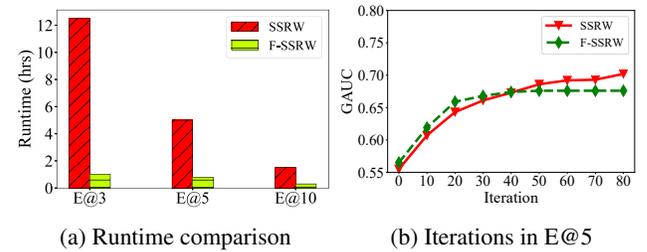


Figure 2: Comparison between SSRW and F-SSRW.

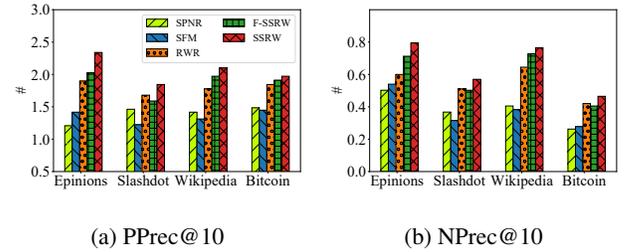


Figure 3: Comparative performance in terms of ranking top 10 positive links and negative links.

Runtime. We then further empirically check the actual runtime of our approaches conducted on a four CPU 3.7GHz machine with 16GB memory. Figure 2(a) shows the runtime comparison between SSRW and F-SSRW on the Epinions dataset, and we can see that F-SSRW is significantly efficient than SSRW. Figure 2(b) demonstrates the convergence

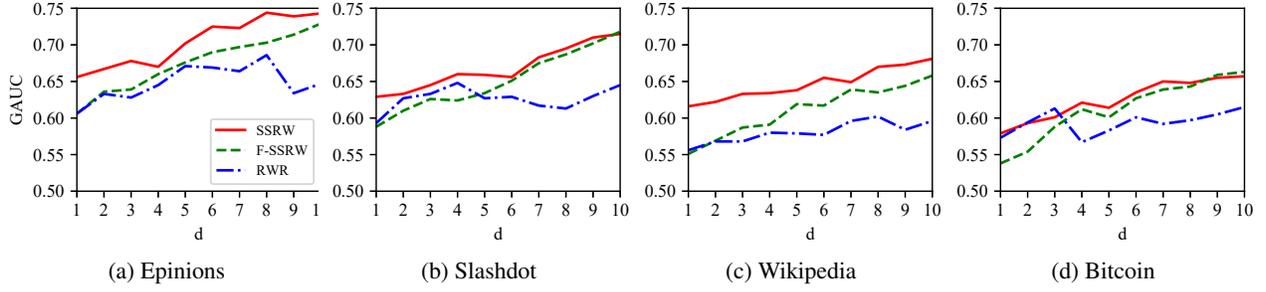


Figure 4: The comparative performance with the change of d .

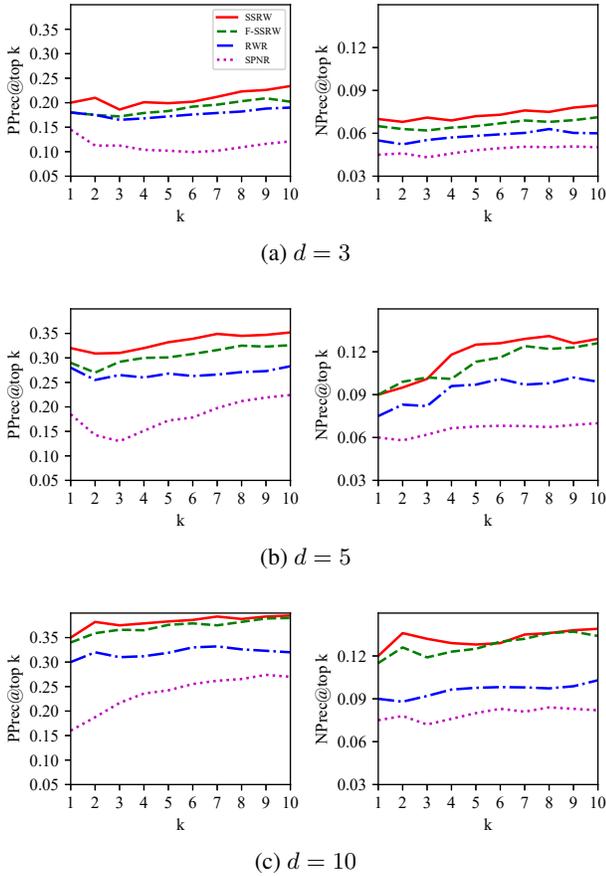


Figure 5: PRec@top k (left) and NRec@top k (right) in the Epinions dataset with different d .

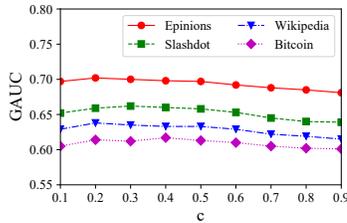


Figure 6: Impact of the parameter c on SSRW ($d = 5$).

of our approaches in dataset E@5, where an iteration represents an update of the parameter w .

Precision@Top- k . We investigate the ranking performance of different approaches in terms of PRec@ k and NRec@ k . Figure 3 shows the comparison results in Epinions on top 10 precision when d equals to 3. We can see that SSRW consistently achieves the best results across all datasets. Random-walk based approaches (i.e. SSRW, F-SSRW and RWR) obtain better performance than other benchmarks, implying that simply taking local attributes into consideration for similarity-based metrics cannot assure satisfying performance in personalized user ranking task. The better performance of SSRW and F-SSRW compared with others also indicates the reasonability of taking personalized social strengths into account.

We also examine the performance of top k by varying k in the range $[1, 10]$, along with different d . We show the experimental results in Epinions in Figure 1, which demonstrate the consistent superior of SSRW over benchmarks. Besides, the performance of F-SSRW becomes better as d increases. Overall, the results imply the effectiveness of our model on top@ k ranking, where positive top @ k can be used for link recommendation, whereas the negative top k can be used in security-related domains.

Impact of the parameter c . The restart probability c is an important parameter for random walk. A smaller c will allow the model ‘walk’ far away from the seed node while a larger c will force the model to walk within the local structure. We thus check the impact of c on SSRW in terms of GAUC by varying c in the range of $[0.1, 0.9]$. As shown in Figure 6, c indeed affects SSRW’s performance, and we can obtain a relatively better performance when $c \in [0.2, 0.4]$. When $c \geq 0.4$, SSRW performs slightly worse with the increase of c . However, the performance variance is insignificant, indicating that SSRW is relatively insensitive and robust in terms of the restart probability c .

Conclusions and Future Work

User ranking is a fundamental and key research problem in signed networks, which has wide applications in real-world scenarios such as recommendation systems and security-related platforms. In this paper, we propose the SSRW model which learns social strengths to optimize the user

ranking list for each individual user. Specifically, we apply supervised random walk in signed scenarios and learn link strengths to guide more effective random walk. Based on the heuristics from data analysis, we further design a simplified and efficient ranking method (F-SSRW), which only focuses on certain candidate nodes and runs the learning algorithm within the local graph of the seed node. A comprehensive evaluation demonstrates the superiority of the proposed models over state-of-the-art approaches, and the robustness in terms of parameters and experimental settings.

In the future, we will try more complex functions to represent the ranking score function by simultaneously incorporating more explicit features. Besides, we strive to apply SSRW in real-world scenarios such as social recommendation to further validate their effectiveness.

Acknowledgments

This work is supported by the MOE AcRF Tier 1 funding (M4011894.020) and the Telenor-NTU Joint R&D funding awarded to Dr. Jie Zhang, and by NSFC projects (No. 71601104 and 71601116) and the Basic Academic Discipline Program for Shanghai University of Finance and Economics.

References

- Antal, T.; Krapivsky, P. L.; and Redner, S. 2006. Social balance on networks: The dynamics of friendship and enmity. *Physica D: Nonlinear Phenomena* 224(1):130–136.
- Backstrom, L., and Leskovec, J. 2011. Supervised random walks: predicting and recommending links in social networks. In *Proceedings of the fourth ACM international conference on Web Search and Data Mining (WSDM)*, 635–644. ACM.
- Brzozowski, M. J., and Romero, D. M. 2011. Who should i follow? recommending people in directed social networks. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM)*.
- Chapelle, O., and Keerthi, S. S. 2010. Efficient algorithms for ranking with svms. *Information retrieval* 13(3):201–215.
- Chiang, K.-Y.; Natarajan, N.; Tewari, A.; and Dhillon, I. S. 2011. Exploiting longer cycles for link prediction in signed networks. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM)*, 1157–1162. ACM.
- Hsieh, C.-J.; Chiang, K.-Y.; and Dhillon, I. S. 2012. Low rank modeling of signed networks. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 507–515. ACM.
- Jung, J.; Jin, W.; Sael, L.; and Kang, U. 2016. Personalized ranking in signed networks using signed random walk with restart. In *Proceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM)*, 973–978. IEEE.
- Katsimpras, G.; Vogiatzis, D.; and Paliouras, G. 2015. Determining influential users with supervised random walks. In *Proceedings of the 24th International Conference on World Wide Web (WWW)*, 787–792. ACM.
- Leskovec, J.; Huttenlocher, D.; and Kleinberg, J. 2010. Predicting positive and negative links in online social networks. In *Proceedings of the 19th International Conference on World Wide Web (WWW)*, 641–650. ACM.
- Li, X.; Fang, H.; and Zhang, J. 2017. Rethinking the link prediction problem in signed social networks. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI)*, 4955–4956.
- Li, X.; Fang, H.; and Zhang, J. 2018. File: A novel framework for predicting social status in signed networks. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI)*, 330–337.
- Man, T.; Shen, H.; Liu, S.; Jin, X.; and Cheng, X. 2016. Predict anchor links across social networks via an embedding approach. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI)*, volume 16, 1823–1829.
- Nelakurthi, A. R., and He, J. 2017. Finding cut from the same cloth: Cross network link recommendation via joint matrix factorization. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI)*, 1467–1473.
- Sarkar, P.; Chakrabarti, D.; and Moore, A. W. 2011. Theoretical justification of popular link prediction heuristics. In *Proceedings of the Twenty-second International Joint Conference on Artificial Intelligence (IJCAI)*, volume 22, 2722.
- Shahriari, M., and Jalili, M. 2014. Ranking nodes in signed social networks. *Social Network Analysis and Mining* 4(1):172.
- Shams, B., and Haratizadeh, S. 2016. Sibrank: Signed bipartite network analysis for neighbor-based collaborative ranking. *Physica A: Statistical Mechanics and its Applications* 458:364–377.
- Song, D., and Meyer, D. A. 2015. Recommending positive links in signed social networks by optimizing a generalized auc. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI)*, 290–296.
- Sun, J.; Qu, H.; Chakrabarti, D.; and Faloutsos, C. 2005. Neighborhood formation and anomaly detection in bipartite graphs. In *Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM)*, 8–pp. IEEE.
- Symeonidis, P., and Tiakas, E. 2014. Transitive node similarity: predicting and recommending links in signed social networks. *World Wide Web* 17(4):743–776.
- Tang, J.; Chang, Y.; Aggarwal, C.; and Liu, H. 2016. A survey of signed network mining in social media. *ACM Computing Surveys (CSUR)* 49(3):42.
- Wang, H.; Shi, X.; and Yeung, D.-Y. 2017. Relational deep learning: A deep latent variable model for link prediction. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI)*, 2688–2694.
- Wu, Z.; Aggarwal, C. C.; and Sun, J. 2016. The troll-trust model for ranking in signed networks. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM)*, 447–456. ACM.
- Xiang, R.; Neville, J.; and Rogati, M. 2010. Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World Wide Web (WWW)*, 981–990. ACM.
- Yin, Z.; Gupta, M.; Weninger, T.; and Han, J. 2010. A unified framework for link recommendation using random walks. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 152–159. IEEE.
- Zhao, H.; Xu, X.; Song, Y.; Lee, D. L.; Chen, Z.; and Gao, H. 2018. Ranking users in social networks with higher-order structures. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI)*, 330–337.
- Zhu, T.; Peng, Z.; Wang, X.; and Hong, X. 2017. Measuring the similarity of nodes in signed social networks with positive and neg-

ative links. In *Asia-Pacific Web and Web-Age Information Management Joint Conference on Web and Big Data (APWeb-WIM)*, 399–407. Springer.