Personalized Ranking in Signed Networks using Signed Random Walk with Restart

Jinhong Jung*, Woojeong Jin†, Lee Sael‡ and U Kang§
*§ Department of Computer Science and Engineering, Seoul National University
† Department of Electrical and Computer Engineering, Seoul National University
‡ Department of Computer Science, The State University of New York (SUNY) Korea
*jinhongjung@snu.ac.kr, †woojung211@snu.ac.kr, ‡sael@sunykorea.ac.kr, §ukang@snu.ac.kr

***Abstract*—How can we rank users in signed social networks? Relationships between nodes in a signed network are represented as positive (trust) or negative (distrust) edges. Many social networks have adopted signed networks to express trust between users. Consequently, ranking friends or enemies in signed networks has received much attention from the data mining community. The ranking problem, however, is challenging because it is difficult to interpret negative edges. Traditional random walk based methods such as PageRank and Random Walk with Restart cannot provide effective rankings in signed networks since they assume only positive edges. Although several methods have been proposed by modifying traditional ranking models, they also fail to account for proper rankings due to the lack of ability to consider complex edge relations. In this paper, we propose SIGNED RANDOM WALK WITH RESTART, a novel model for personalized ranking in signed networks. We introduce a signed random surfer so that she considers negative edges by changing her sign for walking. Our model provides proper rankings reflecting signed edges based on the signed surfer. Through extensive experiments, we demonstrate that SRWR achieves the best accuracy (up to 87%) for sign prediction, and predicts trolls 4× more accurately than other ranking models.**

***Keywords*—Personalized ranking in signed networks; Signed random walk with restart**

I. INTRODUCTION

How can we measure trust or distrust between users in signed networks? In many social networks, users are allowed to make opinions to indicate their trust or distrust to other's opinions. Those expressions are represented as positive and negative edges in graphs, and such graphs are called *signed networks*. Ranking nodes in signed networks has received much interest from data mining community to reveal trust and distrust between users [7] inducing many useful applications such as link prediction [9] and community detection [15] in signed networks.

Traditional ranking models, however, do not provide satisfactory global or personalized rankings in signed networks. Existing random walk based methods such as PageRank [11] and Random Walk with Restart [13] assume only positive edges; thus, they are inappropriate in the networks containing negative edges. Many researchers have proposed heuristics on the classical methods to make them computable in signed networks [7], [12]. However, the approaches are also unsatisfactory because they cannot consider complex relationships of consecutive edges such as friend-of-enemy or enemy-of-enemy. This problem also results from the non-interpretable negative edges in traditional random walks. In addition, most existing ranking models in signed networks focus only on global rankings, although personalized rankings are more useful for individuals in many contexts such as recommendation.

TABLE I: Table of symbols.

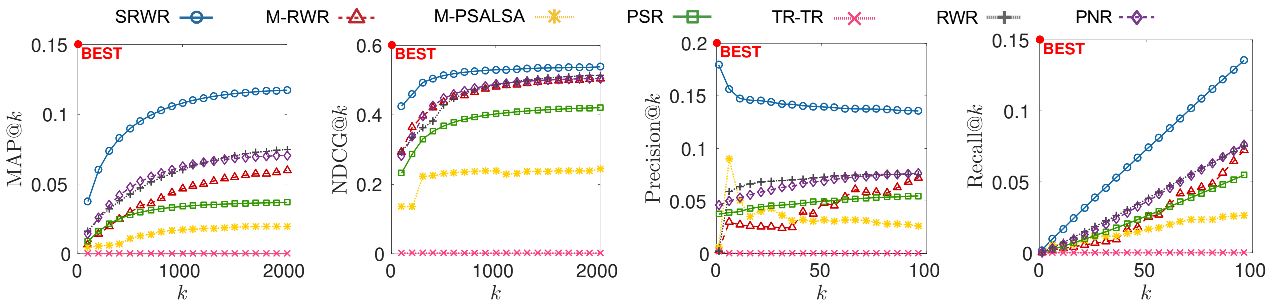
Symbol	Definition
G	signed input graph
n	number of nodes in G
m	number of edges in G
s	seed node (= query node, source node)
c	restart probability
$\overleftarrow{\mathbf{N}}_u$	set of in-neighbors to nodes u
$\overrightarrow{\mathbf{N}}_u$	set of out-neighbors from nodes u
$\mathbf{A}$	$(n \times n)$ signed adjacency matrix of G
$\lvert\mathbf{A}\rvert$	$(n \times n)$ absolute adjacency matrix of G
$\mathbf{D}$	$(n \times n)$ out-degree matrix of $\lvert\mathbf{A}\rvert$, $\mathbf{D}_{ii} = \sum_j \lvert\mathbf{A}\rvert_{ij}$
$\tilde{\mathbf{A}}$	$(n \times n)$ semi-row normalized matrix of $\mathbf{A}$
$\tilde{\mathbf{A}}_+$	$(n \times n)$ positive semi-row normalized matrix of $\mathbf{A}$
$\tilde{\mathbf{A}}_-$	$(n \times n)$ negative semi-row normalized matrix of $\mathbf{A}$
$\mathbf{q}$	$(n \times 1)$ starting vector (= sth unit vector)
$\mathbf{r}^+$, $\mathbf{r}^-$	$(n \times 1)$ trust and distrust SRWR score vector, resp.
$\mathbf{r}^d$	$(n \times 1)$ relative trustworthy vectors, $\mathbf{r}^d = \mathbf{r}^+ - \mathbf{r}^-$

In this paper, we propose SIGNED RANDOM WALK WITH RESTART (SRWR), a novel model for effective personalized rankings in signed networks, and an iterative algorithm for computing personalized rankings efficiently. The main idea of SRWR is to introduce a sign into a random surfer in order to let the surfer consider negative edges. Consequently, our model considers complex edge relationships, and makes random walks interpretable in signed networks. Through extensive experiments, we demonstrate the effectiveness of SRWR as shown in Figure 1. Our main contributions are as follows:

- **Novel ranking model.** We propose SIGNED RANDOM WALK WITH RESTART (SRWR), a novel model for personalized rankings in signed networks (Definition 1).
- **Algorithm.** We propose an iterative algorithm for computing SRWR scores efficiently in signed networks (Algorithm 2).
- **Experiment.** We show that SRWR achieves the best accuracy (up to 87%) for sign prediction, and predicts trolls 4× more accurately than other ranking models (Figure 1).

The code of our method and datasets used in this paper are available at http://datalab.snu.ac.kr/srwr. The rest of this paper is organized as follows. We describe our proposed model and an iterative algorithm for computing SRWR scores in Section II. After presenting our experimental results in Section III, we provide a review on related works in Section IV. Lastly, we conclude in Section V.

2374-8486/16 $31.00 © 2016 IEEE
DOI 10.1109/ICDM.2016.140



(a) MAP@k (b) NDCG@k (c) Precision@k (d) Recall@k

Fig. 1: We evaluate the performance of models for the troll identification task through various measurements: MAP@k (1(a)), NDCG@k (1(b)), Precision@k (1(c)), and Recall@k (1(d)). SRWR shows the best performance for all the measurements compared to other competitors.

II. Proposed Method

We propose Signed Random Walk with Restart (SRWR), a novel model for personalized ranking in signed networks, and an iterative algorithm for SRWR scores w.r.t. a query node. Table I lists the symbols used in this paper.

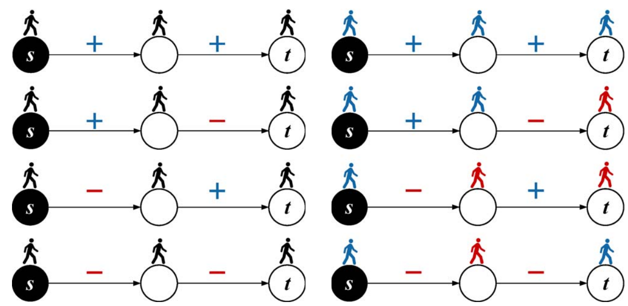
A. Signed Random Walk with Restart Model

As discussed in Section I, complicated relationships of signed edges are the main obstacles for providing effective rankings in signed networks. Most existing works on signed networks have not focused on personalized rankings. In this work, our goal is to design a novel ranking model which resolves those problems in signed networks. The main ideas of our model are as follows:

- We introduce a signed random surfer. The sign of the surfer is either positive or negative, which means favorable or adversarial to a node, respectively.
- When the random surfer encounters a negative edge, she changes her sign from positive to negative, or vice versa. Otherwise, she keeps her sign.
- We introduce balance attenuation factors into the surfer to consider the uncertainty for friendship of enemies.

There are four cases according to the signs of edges as shown in Figure 2: 1) friend's friend, 2) friend's enemy, 3) enemy's friend, and 4) enemy's enemy. Suppose a random surfer starts at node s toward node t. A traditional surfer just moves along the edges without considering signs as seen in Figure 2(a) since there is no way to consider the signs on the edges. Hence, classical models cannot distinguish those edge relationships during her walks. For instance, the model considers that node s and node t are friends for the second case (friend's enemy), even though node t are more likely to be an enemy w.r.t. node s.

On the contrary, our model in Figure 2(b) has a signed random surfer who considers those complex edge relationships. If the random surfer starting at node s with a positive sign encounters a negative edge, she flips her sign from positive to negative, or vice versa. Our model distinguishes whether node t is the friend of node s or not according to her sign at node t. As shown in Figure 2(b), the results for all cases from our model are consistent with structural balance theory [2]. Thus, introducing a signed random surfer enables our model to discriminate those edge relationships.



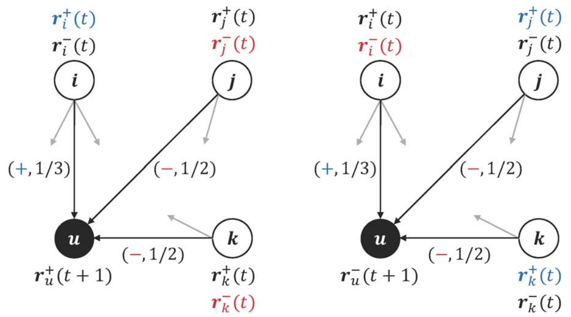
(a) Traditional random walks (b) Signed random walks

Fig. 2: Examples of traditional random walks and signed random walks. Each case represents 1) friend's friend, 2) friend's enemy, 3) enemy's friend, or 4) enemy's enemy from the top. A random surfer has either a positive (blue) or a negative (red) sign on each node in Figure 2(b). When the signed random surfer traverses a negative edge, she changes her sign from positive to negative or vice versa.

Trust or distrust relationships between a specific node s and other nodes are revealed as the surfer is allowed to move around a signed network starting from node s. If the positive surfer visits a certain node u many times, then node u is trustable for node s. On the other hand, if the negative surfer visits node u many times, then node s is not likely to trust node u. Thus, rankings are obtained by revealing a degree of trust or distrust between people based on the signed random walks. Here, we formally define our model on signed networks in Definition 1. Note that Definition 1 involves the concept of restart which provides personalized rankings w.r.t. a user.

Definition 1 (Signed Random Walk with Restart): A *signed random surfer* has a sign, which is either positive or negative. At the beginning, the surfer starts with $+$ sign from a seed node s because she trusts s. Suppose the surfer is currently at node u, and c is the restart probability of the surfer. Then, she takes one of the following actions:

- **Action 1: Signed Random Walk.** The surfer randomly moves to one of the neighbors from node u with probability $1-c$. The surfer flips her sign if she encounters a negative edge. Otherwise, she keeps her sign.
- **Action 2: Restart.** The surfer goes back to the seed node s with probability c. Her sign should become $+$ at the seed node s because she trusts s. ■



(a) An example of a positive probability, $\mathbf{r}_u^+(t+1)$

(b) An example of a negative probability, $\mathbf{r}_u^-(t+1)$

Fig. 3: Examples of how $\mathbf{r}_u^+$ and $\mathbf{r}_u^-$ are defined in SRWR.

We measure two probabilities on each node through SIGNED RANDOM WALK WITH RESTART (SRWR) starting from the seed node s. The two probabilities are represented as follows:

- $\mathbf{r}_u^+ = \Pr_s(u, +)$: the probability that the positive surfer is at node u after SRWR from the seed node s.
- $\mathbf{r}_u^- = \Pr_s(u, -)$: the probability that the negative surfer is at node u after SRWR from the seed node s.

If $\mathbf{r}_u^+$ is high, then node u is trustable for node s. On the other hand, if $\mathbf{r}_u^-$ is high, node u is not reliable for node s. $\mathbf{r}^+$ is a trust SRWR score vector and $\mathbf{r}^-$ is a distrust SRWR score vector for all nodes. Both are used for personalized rankings w.r.t. the seed node s. If we regard $\mathbf{r}^+$ and $\mathbf{r}^-$ as score vectors, $\mathbf{r}^d = \mathbf{r}^+ - \mathbf{r}^-$ is considered as a relative trustworthiness vector of nodes w.r.t. s, which is also used as a personalized ranking. For instance, if $\mathbf{r}_u^d$ is positive, then node u is trustable for node s. Otherwise, node u is not trustable for node s.

B. Formulation for Signed Random Walk with Restart

We formulate the probability vectors, $\mathbf{r}^+$ and $\mathbf{r}^-$, following SIGNED RANDOM WALK WITH RESTART. First, we explain how to define $\mathbf{r}^+$ and $\mathbf{r}^-$ using the example shown in Figure 3. In the example, we label a (sign, transition probability) pair on each edge. For instance, the transition probability for the positive edge from node i to node u is $1/3$ because node i has 3 outgoing edges. This edge is denoted by $(+, 1/3)$. Other pairs of signs and transition probabilities are also similarly defined. In order that the random surfer has a positive sign on node u at time $t+1$, a positive surfer on one of u's neighbor at time t must move to node u through a positive edge, or a negative surfer must move through a negative edge according to the signed random walk action in Definition 1. Considering the restart action of the surfer with the probability c, $\mathbf{r}_u^+(t+1)$ in Figure 3(a) is represented as follows:

$$\mathbf{r}_u^+(t+1) = (1-c)\left(\frac{\mathbf{r}_i^+(t)}{3} + \frac{\mathbf{r}_j^-(t)}{2} + \frac{\mathbf{r}_k^-(t)}{2}\right) + c\mathbf{1}(u = s)$$

where $\mathbf{1}(u = s)$ is 1 if u is the seed node s and 0 otherwise. In Figure 3(b), $\mathbf{r}_u^-(t+1)$ is defined similarly as follows:

$$\mathbf{r}_u^-(t+1) = (1-c)\left(\frac{\mathbf{r}_i^-(t)}{3} + \frac{\mathbf{r}_j^+(t)}{2} + \frac{\mathbf{r}_k^+(t)}{2}\right)$$

Note that we do not add the restarting score $c\mathbf{1}(u = s)$ to $\mathbf{r}_u^-(t+1)$ in this case because the surfer's sign must become positive when she goes back to the seed node s. The recursive equations of our model are defined as follows:

$$\begin{aligned}
\mathbf{r}_u^+ &= (1-c)\left(\sum_{v \in \overleftarrow{\mathbf{N}}_u^+} \frac{\mathbf{r}_v^+}{|\overrightarrow{\mathbf{N}}_v|} + \sum_{v \in \overleftarrow{\mathbf{N}}_u^-} \frac{\mathbf{r}_v^-}{|\overrightarrow{\mathbf{N}}_v|}\right) + c\mathbf{1}(u = s) \\
\mathbf{r}_u^- &= (1-c)\left(\sum_{v \in \overleftarrow{\mathbf{N}}_u^-} \frac{\mathbf{r}_v^+}{|\overrightarrow{\mathbf{N}}_v|} + \sum_{v \in \overleftarrow{\mathbf{N}}_u^+} \frac{\mathbf{r}_v^-}{|\overrightarrow{\mathbf{N}}_v|}\right)
\end{aligned} \quad (1)$$

where $\overleftarrow{\mathbf{N}}_i$ is the set of in-neighbors of node i, and $\overrightarrow{\mathbf{N}}_i$ is the set of out-neighbors of node i. Superscripts of $\overleftarrow{\mathbf{N}}_i$ or $\overrightarrow{\mathbf{N}}_i$ indicate signs of edges between node i and its neighbors (e.g., $\overleftarrow{\mathbf{N}}_i^+$ indicates the set of positively connected in-neighbors of node i). We need to introduce several symbols related to an adjacency matrix $\mathbf{A}$ to vectorize Equation (1).

Definition 2 (Signed adjacency matrix): The signed adjacency matrix $\mathbf{A}$ of G is a matrix such that $\mathbf{A}_{uv}$ is positive or negative when there is a positive or a negative edge from node u to node v respectively, and zero otherwise. ■

Definition 3 (Semi-row normalized matrix): Let $|\mathbf{A}|$ be the absolute adjacency matrix of $\mathbf{A}$, and $\mathbf{D}$ be the out-degree diagonal matrix of $|\mathbf{A}|$ (i.e., $\mathbf{D}_{ii} = \sum_j |\mathbf{A}|_{ij}$). Then semi-row normalized matrix of $\mathbf{A}$ is $\tilde{\mathbf{A}} = \mathbf{D}^{-1}\mathbf{A}$. ■

Definition 4 (Positive or negative semi-row normalized matrix): The positive semi-row normalized matrix $\tilde{\mathbf{A}}_+$ contains only positive values in the semi-row normalized matrix $\tilde{\mathbf{A}}$. The negative semi-row normalized matrix $\tilde{\mathbf{A}}_-$ contains absolute values of negative elements in $\tilde{\mathbf{A}}$. In other words, $\tilde{\mathbf{A}} = \tilde{\mathbf{A}}_+ - \tilde{\mathbf{A}}_-$. ■

Based on Definitions 3 and 4, Equation (1) is represented as the following equation:

$$\begin{aligned}
\mathbf{r}^+ &= (1-c)\left(\tilde{\mathbf{A}}_+^\top \mathbf{r}^+ + \tilde{\mathbf{A}}_-^\top \mathbf{r}^-\right) + c\mathbf{q} \\
\mathbf{r}^- &= (1-c)\left(\tilde{\mathbf{A}}_-^\top \mathbf{r}^+ + \tilde{\mathbf{A}}_+^\top \mathbf{r}^-\right)
\end{aligned} \quad (2)$$

where $\mathbf{q}$ is a vector whose sth element is 1, and all other elements are 0.

C. Balance Attenuation Factors

The signed surfer measures trust and distrust of nodes w.r.t. a seed node according to edge relationships as discussed in Section II-A. Our model strongly supports balance theory describing the four cases between nodes as shown in Figure 2(b). However, the naive balance theory would not hold for explaining behaviors of people in real-world signed networks, since unbalanced relationships frequently appear (e.g., the enemy of my friend could be my friend) due to the uncertainty in trusting the friendship of enemies.

We reflect the uncertainty of the friendship of an enemy into our ranking model by introducing stochastic parameters, β and γ, called *balance attenuation factors*. Note that we assume the friendship of a friendly user is reliable. β is a parameter for the uncertainty of "the enemy of my enemy is my friend", and γ is for "the friend of my enemy is my enemy." We first explain β using the fourth case (enemy's enemy) in Figure 2(b). Suppose a surfer with a positive sign starts at node s toward node t and encounters two consecutive negative edges. Based on balance theory, her sign becomes negative at the intermediate node m and positive at node t. However, some people might think that the enemy of my enemy is my enemy. In this case,

Algorithm 1: Normalization phase of SRWR

Input: signed adjacency matrix: $\mathbf{A}$

Output: positive semi-row normalized matrix: $\tilde{\mathbf{A}}_+$, and negative semi-row normalized matrix: $\tilde{\mathbf{A}}_-$

1: compute out-degree matrix $\mathbf{D}$ of $|\mathbf{A}|$, $\mathbf{D}_{ii} = \sum_j |\mathbf{A}|_{ij}$
2: compute semi-row normalized matrix, $\tilde{\mathbf{A}} = \mathbf{D}^{-1}\mathbf{A}$.
3: split $\tilde{\mathbf{A}}$ into $\tilde{\mathbf{A}}_+$ and $\tilde{\mathbf{A}}_-$ such that $\tilde{\mathbf{A}} = \tilde{\mathbf{A}}_+ - \tilde{\mathbf{A}}_-$
4: **return** $\tilde{\mathbf{A}}_+$ and $\tilde{\mathbf{A}}_-$

Algorithm 2: Iteration phase of SRWR

Input: positive semi-row normalized matrix: $\tilde{\mathbf{A}}_+$, and negative semi-row normalized matrix: $\tilde{\mathbf{A}}_-$, and seed node: s, restart probability: c, balance attenuation factors: β and γ, and error tolerance: ϵ.

Output: positive SRWR score vector: $\mathbf{r}^+$ and negative SRWR score vector: $\mathbf{r}^-$

1: set the starting vector $\mathbf{q}$ from the seed node s
2: set $\mathbf{r}^+ = \mathbf{q}$, $\mathbf{r}^- = \mathbf{0}$, and $\mathbf{r}' = [\mathbf{r}^+; \mathbf{r}^-]^\top$
3: **repeat**
4: $\quad \mathbf{r}^+ \leftarrow (1-c)(\tilde{\mathbf{A}}_+^\top \mathbf{r}^+ + \beta \tilde{\mathbf{A}}_-^\top \mathbf{r}^- + (1-\gamma)\tilde{\mathbf{A}}_+^\top \mathbf{r}^-) + c\mathbf{q}$
5: $\quad \mathbf{r}^- \leftarrow (1-c)(\tilde{\mathbf{A}}_-^\top \mathbf{r}^+ + \gamma \tilde{\mathbf{A}}_+^\top \mathbf{r}^- + (1-\beta)\tilde{\mathbf{A}}_-^\top \mathbf{r}^-)$
6: $\quad$ concatenate $\mathbf{r}^+$ and $\mathbf{r}^-$ into $\mathbf{r} = [\mathbf{r}^+; \mathbf{r}^-]^\top$
7: $\quad$ compute the error between $\mathbf{r}$ and $\mathbf{r}'$, $\delta = \|\mathbf{r} - \mathbf{r}'\|$
8: $\quad$ update $\mathbf{r}' \leftarrow \mathbf{r}$ for the next iteration
9: **until** $\delta < \epsilon$
10: **return** $\mathbf{r}^+$ and $\mathbf{r}^-$

her sign will be negative at nodes m and t. To consider this uncertainty, we introduce a parameter β so that if the negative surfer at node m encounters a negative edge, her sign becomes positive with probability β or negative with $1-\beta$ at node t. The other parameter γ is also interpreted similarly to β. When the negative surfer at node m encounters a positive edge, her sign will be negative with probability γ or positive with $1-\gamma$ at node t (e.g., the third case in Figure 2(b)). SRWR with the balance attenuation factors is represented as follows:

$$\begin{aligned} \mathbf{r}^+ &= (1-c)\left(\tilde{\mathbf{A}}_+^\top \mathbf{r}^+ + \beta \tilde{\mathbf{A}}_-^\top \mathbf{r}^- + (1-\gamma)\tilde{\mathbf{A}}_+^\top \mathbf{r}^-\right) + c\mathbf{q} \\ \mathbf{r}^- &= (1-c)\left(\tilde{\mathbf{A}}_-^\top \mathbf{r}^+ + \gamma \tilde{\mathbf{A}}_+^\top \mathbf{r}^- + (1-\beta)\tilde{\mathbf{A}}_-^\top \mathbf{r}^-\right) \end{aligned} \tag{3}$$

Note that the uncertainty of a friend's friendship could be considered by adding other factors similarly to the proposed approach, but in this work, we only reflect the uncertainty of an enemy's friendship for simplicity.

D. Algorithm for Signed Random Walk with Restart

We present an iterative algorithm for computing SRWR scores efficiently and accurately based on Equation (3).

Normalization phase (Algorithm 1). Our proposed algorithm first computes the out-degree diagonal matrix $\mathbf{D}$ of $|\mathbf{A}|$, which is the absolute adjacency matrix of $\mathbf{A}$ (line 1). Then, the algorithm computes the semi-row normalized matrix $\tilde{\mathbf{A}}$ using $\mathbf{D}$ (line 2). We split $\tilde{\mathbf{A}}$ into two matrices: the positive semi-row normalized matrix ($\tilde{\mathbf{A}}_+$) and the negative semi-row normalized matrix ($\tilde{\mathbf{A}}_-$) (line 3) satisfying $\tilde{\mathbf{A}} = \tilde{\mathbf{A}}_+ - \tilde{\mathbf{A}}_-$.

Iteration phase (Algorithm 2). Our algorithm computes the SRWR score vectors $\mathbf{r}^+$ and $\mathbf{r}^-$ for the seed node s with the balance attenuation factors (β and γ) in the iteration phase. We set $\mathbf{q}$ to sth unit vector, and initialize $\mathbf{r}^+$ to $\mathbf{q}$ and $\mathbf{r}^-$ to $\mathbf{0}$ (lines 1 and 2). Our algorithm iteratively computes Equation (3) (lines 4 and 5). We concatenate $\mathbf{r}^+$ and $\mathbf{r}^-$ vertically (line 6) into $\mathbf{r}$. We then compute the error δ between $\mathbf{r}$ and $\mathbf{r}'$ which is the result in the previous iteration (line 7). We update $\mathbf{r}$ into $\mathbf{r}'$ for the next iteration (line 8). The iteration stops when the error δ is smaller than a threshold ϵ (line 9).

TABLE II: Dataset statistics.

Dataset	Node	Edge	+ Edge	− Edge
Epinions[1]	131,828	841,372	85.3%	14.7%
Slashdot[2]	79,120	515,397	76.1%	23.9%
Wikipedia[3]	7,118	103,675	78.4%	21.6%

[1] http://www.trustlet.org/wiki/Extended_Epinions_dataset
[2] http://dai-labor.de/IRML/datasets
[3] http://snap.stanford.edu/data/wiki-Vote.html

III. Experiment

A. Experimental Settings

Experiments are performed on a PC with Intel(R) Core(TM) i5-4590 CPU @ 3.30GHz and 8GB memory. The signed networks used in our experiments are summarized in Table II. We use all datasets in the sign prediction task (Section III-B). We use the Slashdot dataset in the troll identification task (Section III-C) since there is a troll list only in the dataset.

Methods. We compare our proposed model with Random Walk with Restart (RWR) [3], Modified Random Walk with Restart (M-RWR) [12], Modified Personalized SALSA (M-PSALSA) [10], Personalized Signed spectral Rank (PSR) [7], Personalized Negative Rank (PNR) [7], and Troll-Trust Model (TR-TR) [14]. Note that RWR is computed on the absolute adjacency matrices of signed networks. Since most existing methods compute global trust and distrust rankings in the context of PageRank, we set a starting vector of each method as in line 1 of Algorithm 2 to obtain personalized rankings.

Parameters. We set the restarting probability c to 0.15 for all random walk based methods including our method. We set other parameters which give the best performances as follows:

- Sign prediction task: In our model, we set $\beta = 0.5, \gamma = 0.9$ in the Epinions dataset, $\beta = 0.6, \gamma = 0.9$ in the Slashdot dataset, and $\beta = 0.1, \gamma = 0.6$ in the Wikipedia dataset. We set $\beta = 0.5, \lambda_1 = 1.0$ in TR-TR.
- Troll identification task: In our model, we set $\beta = 0.1, \gamma = 1.0$. We set $\beta = 0.5, \lambda_1 = 1.0$ in TR-TR.

B. Sign Prediction Task

We evaluate ranking models on the sign prediction task defined as follows: given a signed network containing missed signs of edges connected from a node, predict those signs. We randomly select $5,000$ seed nodes for the experiment and choose 20% edges of positive and negative links of each node as a test set. Then, we remove each selected edge $(s \rightarrow t)$, and predict the edge's sign based on personalized ranking scores w.r.t. node s. We compute $\mathbf{r}^d = \mathbf{r}^+ - \mathbf{r}^-$ whose values range from -1 to 1. If $\mathbf{r}^d_t$ is greater than 0, then we predict the sign of the edge $(s \rightarrow t)$ as positive. Otherwise, it is considered as negative. We measure the prediction accuracy by comparing $\mathbf{r}^d_t$ with the true sign of the edge.

Results. We compare the performance of SRWR, M-RWR, M-PSALSA, TR-TR, and PSR on the sign prediction task. As

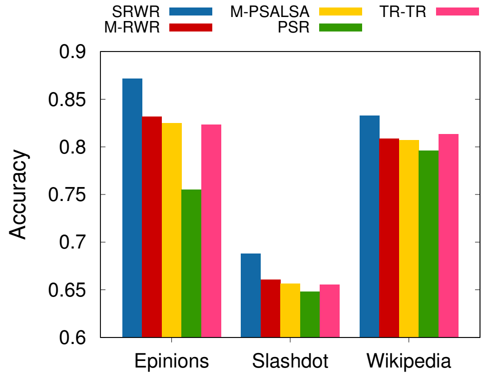
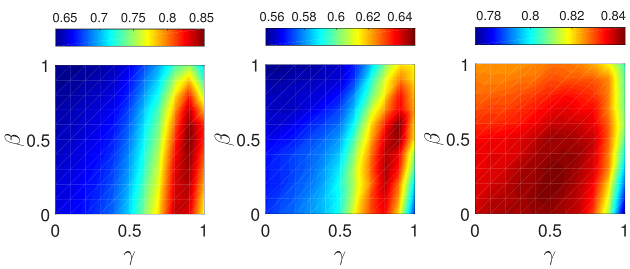


Fig. 4: The accuracy of SRWR in the sign prediction task on different signed networks. As shown in the results, our proposed SIGNED RANDOM WALK WITH RESTART provides the best performance in terms of accuracy.



(a) Epinions (b) Slashdot (c) Wikipedia

Fig. 5: Accuracy maps according to β and γ where color indicates a degree of accuracy. The Epinions and the Slashdot datasets present similar tendencies while the Wikipedia dataset shows a different result from those of the two datasets.

shown in Figure 4, SRWR is the most accurate in predicting signs for all the datasets. This implies that SRWR provides more effective personalized rankings than other methods.

Balance attenuation factors. We adjust the balance attenuation factors of SRWR, and evaluate the sign prediction task to examine how well balance theory explains signed networks. In this experiment, we use the top-100 highest degree nodes as a test set for each network. The Epinions and the Slashdot datasets show similar results where larger values of β and γ achieve high accuracy as shown in Figures 5(a) and 5(b). Unlike these two datasets, the accuracy is high when β is small in the Wikipedia network as shown in Figure 5(c). This implies that ”an enemy of my enemy is my friend” would not be correct in the network, which means balance theory does not apply well to the Wikipedia dataset. The reason is that the Wikipedia network represents votes between users to elect administrators; thus, the dataset is different from the Epinions and the Slashdot networks which are general social networks. Another observation is that the ideal balance theory does not apply to real-world signed networks because the accuracy is not the best over all datasets when $\beta = 1$ and $\gamma = 1$.

C. Troll Identification Task

We evaluate ranking models on the task of identifying trolls who behave abnormally or cause normal users to be irritated. The task is defined as follows: given a signed network and

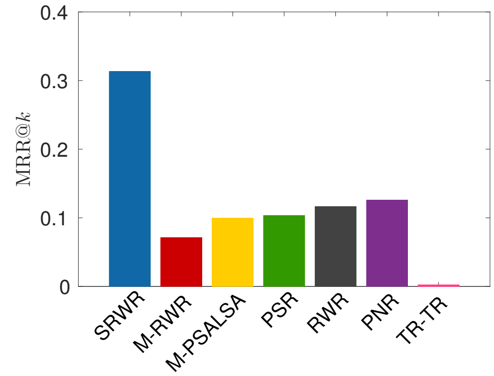


Fig. 6: MRR@k of SRWR ($k = 2,000$). The measure indicates how trolls are ranked high in a personalized distrust ranking. The MRR value of SRWR is the highest.

a user, identify trolls using a personalized ranking w.r.t. the user. As in the previous work [7], we also use the enemies of a user called *No-More-Trolls* in the Slashdot dataset as trolls. The user is an administrative account created for the purpose of collecting a troll list. There are 96 trolls in the list. Since we focus on personalized ranking in signed networks, we consider the following case: trolls are likely to be enemies of each normal user. Hence, trolls would be ranked high in a personalized distrust ranking ($\mathbf{r}^-$) w.r.t. the user.

The task consists of identifying trolls based on personalized rankings for each normal user in the Slashdot network after excluding edges adjacent to *No-More-Trolls*. For each user with the minimum out-degree 1, we search trolls within the top-k distrust ranking, and measure Mean Average Precision (MAP@k), Normalized Discount Cumulated Gain (NDCG@k), Precision@k, Recall@k, and Mean Reciprocal Rank (MRR@k). Since there are no user-graded relevance scores for the troll list, we set those scores to 1 for NDCG.

Results. SRWR significantly outperforms other ranking models for the troll identification task as shown in Figures 1 and 6. More trolls are identified by our proposed model within the top-k ranking according to MAP@k shown in Figure 1(a). This observation is also consistent with the results in Figures 1(c) and 1(d), indicating that SRWR achieves higher Precision@k and Recall@k than other methods. SRWR provides $4\times$ better performance than PNR, the second best one, in terms of Precision@k. Moreover, the ranking of a top ranked troll from our proposed model is higher than that of other ranking models because MRR@k of our model is the highest among other competitors as shown in Figure 6. Many trolls tend to be ranked high in our distrust ranking because SRWR achieves better MAP@k and NDCG@k than other ranking models as presented in Figures 1(a) and 1(b).

We list the top-10 personalized rankings for a user called *”freejung”* in Table III. According to the result, more trolls are ranked high in the distrust ranking from SRWR, indicating that our model is more sensitive in identifying trolls than other models. Also, the query user is ranked high in the distrust rankings from M-RWR and M-PSALSA while the user is ranked low in the distrust ranking from our model. The query

TABLE III: Troll prediction results of SRWR and other models w.r.t. a normal user *"freejung"*. For each model, we show top 10 trusted and distrusted nodes. Red-colored users are trolls, a blue-colored user is a query user, and the black-colored are normal users. Note that SRWR shows the best result: in SRWR, the query user is ranked 1st in the trust ranking, and many trolls are ranked high in the distrust ranking. M-RWR and M-PSALSA provide inferior results since they rank the query user high in the distrust ranking, although the query user is the most trusted user for this task. PSR and TR-TR are not satisfactory either: they do not contain many trolls in their top distrusted users.

	SRWR (proposed)		**M-RWR**		**M-PSALSA**		**PSR**		**TR-TR**	
Rank	**Trust Ranking**	**Distrust Ranking**	**Trust Ranking**	**Distrust Ranking**	**Trust Ranking**	**Distrust Ranking**	**Trust Ranking**	**Distrust Ranking**	**Trust Ranking**	**Distrust Ranking**
1	**freejung**	**Twirlip+o**	**freejung**	**freejung**	CleverNic	**freejung**	**freejung**	manifest3	**freejung**	inTheLoo
2	CmdrTaco	**Klerck**	CmdrTaco	TheJesusC	CmdrTaco	**Klerck**	CmdrTaco	rpiquepa	daoine	(TK14)Des
3	TomorrowP	CmdrTaco	CleverNic	Fnkmaster	Bruce+Per	**CmderTaco**	TomorrowP	JonKatz	Jamie+Zaw	westbake
4	Gryll	**%24%24%24**	FortKnox	Professor	John+Carm	spinlocke	Gryll	johnnyb	KshGoddes	2forshow
5	CleverNic	JonKatz	TomorrowP	rqqrtnb	**%24%24%24**	JonKatz	autosentr	**TrollBurg**	shadowspa	43Percent
6	FortKnox	CleverNic	gleam	dubba-dum	kfg	twitter	CleverNic	HanzoSan	turg	ABeowulfC
7	autosentr	HanzoSan	Gryll	drhairsto	NewYorkCo	StarManta	meowsquea	kalka	ryanr	abigsmurf
8	meowsquea	ekrout	autosentr	howcoome	**freejung**	tomstdeni	FortKnox	p00p	slothdog	AdiBean
9	Ethelred+	**CmderTaco**	quadong	khuber	AKAImBatm	Doc+Ruby	Ethelred+	fimbulvet	TheIndivi	airjrdn
10	SolemnDra	manifest3	meowsquea	Skapare	FortKnox	**stratjakt**	SolemnDra	HBergeron	avitzur	alewar

user should trust himself; thus, the user should be ranked high in a trust ranking, whereas the user must be positioned low in a distrust ranking. This implies our model is more desirable than other models for personalized rankings in signed networks.

IV. Related Works

In this section, we review related works, which are categorized into two parts: 1) ranking in unsigned networks, and 2) ranking in signed networks.

Ranking in unsigned networks. There are various global ranking measures based on link structure and random walk, e.g., PageRank (PR) [11], HITS [6], and SALSA [8]. Furthermore, personalized ranking methods are explored in terms of relevance such as Personalized PageRank (PPR) [3], Personalized SALSA (PSALSA) [1]. Among these measures, RWR has received much interests and has been applied to many applications [5], [13], [4]. Note that these methods are not applicable to signed graphs because they assume only positive edges; on the contrary, our model works on signed networks.

Ranking in signed networks. Wu et al. [14] proposed Troll-Trust model (TR-TR) which is a variant of PageRank. In the algorithm, the trustworthiness of individual data is modeled as a probability that represents the underlying ranking values. Kunegis et al. [7] presented Signed spectral Ranking (SR) by extending PageRank to the case of negative edge. Shahriari et al. [12] suggested Modified PageRank (MPR), which computes PageRank in a positive subgraph and a negative subgraph separately, and subtracts negative PageRank scores from positive ones. These models cannot explain complex relationships between negative and positive edges; in contrast, our model has the ability to account for the relationships.

V. Conclusion

We propose Signed Random Walk with Restart, a novel model which provides personalized trust or distrust rankings in signed networks. In our model, we introduce a signed random surfer so that she considers negative edges by changing her sign for surfing on signed networks. Consequently, our model provides personalized trust or distrust rankings reflecting signed edges based on balance theory. We experimentally show that SRWR achieves the best accuracy (up to 87%) for sign prediction, and predicts trolls $4\times$ more accurately than other ranking models.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF-2015R1C1A2A01055739 and NRF-2015K1A3A1A14021055). U Kang is the corresponding author.

References

[1] B. Bahmani, A. Chowdhury, and A. Goel. Fast incremental and personalized pagerank. *Proceedings of the VLDB Endowment*, 4(3):173–184, 2010.

[2] D. Cartwright and F. Harary. Structural balance: a generalization of heider's theory. *Psychological review*, 63(5):277, 1956.

[3] T. H. Haveliwala. Topic-sensitive pagerank. In *Proceedings of the 11th international conference on World Wide Web*, pages 517–526. ACM, 2002.

[4] J. Jung, K. Shin, L. Sael, and U. Kang. Random walk with restart on large graphs using block elimination. *ACM Transactions on Database Systems (TODS)*, 41(2):12, 2016.

[5] U. Kang, H. Tong, and J. Sun. Fast random walk graph kernel. In *Proceedings of the Twelfth SIAM International Conference on Data Mining, Anaheim, California, USA, April 26-28, 2012.*, pages 828–838, 2012.

[6] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

[7] J. Kunegis, A. Lommatzsch, and C. Bauckhage. The slashdot zoo: mining a social network with negative edges. In *Proceedings of the 18th international conference on World wide web*, pages 741–750. ACM, 2009.

[8] R. Lempel and S. Moran. Salsa: the stochastic approach for link-structure analysis. *ACM Transactions on Information Systems (TOIS)*, 19(2):131–160, 2001.

[9] 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*, pages 641–650. ACM, 2010.

[10] A. Y. Ng, A. X. Zheng, and M. I. Jordan. Stable algorithms for link analysis. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 258–266. ACM, 2001.

[11] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: bringing order to the web. 1999.

[12] M. Shahriari and M. Jalili. Ranking nodes in signed social networks. *Social Network Analysis and Mining*, 4(1):1–12, 2014.

[13] K. Shin, J. Jung, S. Lee, and U. Kang. Bear: Block elimination approach for random walk with restart on large graphs. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, pages 1571–1585. ACM, 2015.

[14] Z. Wu, C. C. Aggarwal, and J. Sun. The troll-trust model for ranking in signed networks. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 447–456. ACM, 2016.

[15] B. Yang, W. K. Cheung, and J. Liu. Community mining from signed social networks. *Knowledge and Data Engineering, IEEE Transactions on*, 19(10):1333–1348, 2007.