Supervised and Extended Restart in Random Walks for Ranking and Link Prediction in Networks (Supplementary Document)

Woojeong Jin* Jinhong Jung* U Kang*

Abstract

In this supplementary document, we provide proofs of lemmas and theorems, information of experiments in the main paper, and additional experiments.

1 Proofs

The equation of Random Walk with Extended Restart is as follows:

$$\mathbf{r} = \tilde{\mathbf{A}}^\top(\mathbf{I} - diag(\mathbf{c}))\mathbf{r} + \left(\mathbf{c}^\top\mathbf{r}\right)\mathbf{q} \tag{1.1}$$

where $\tilde{\mathbf{A}}$ is a row-normalized matrix, $\mathbf{c}$ is a restart vector whose i-th entry is c_i, $diag(\mathbf{c})$ is a matrix whose $diag(\mathbf{c})_{ii} = c_i$ and other entries are 0, and $\mathbf{q}$ is a vector whose s-th element is 1, and all other elements are 0.

1.1 Proof of Lemma 3.1.

Proof. Note that $\mathbf{c}^\top\mathbf{r}$ is a scalar; thus, $(\mathbf{c}^\top\mathbf{r})\mathbf{q} = \mathbf{q}(\mathbf{c}^\top\mathbf{r}) = (\mathbf{q}\mathbf{c}^\top)\mathbf{r}$. Hence, equation (1.1) is represented as follows:

$$\begin{aligned}\mathbf{r} &= \tilde{\mathbf{A}}^\top(\mathbf{I} - diag(\mathbf{c}))\mathbf{r} + \left(\mathbf{q}\mathbf{c}^\top\right)\mathbf{r} \\ &= \tilde{\mathbf{A}}^\top(\mathbf{I} - diag(\mathbf{c}))\mathbf{r} + \left(\mathbf{q}\mathbf{c}^\top\right)\mathbf{r} - \mathbf{q} + \mathbf{q}\end{aligned}$$

Since $\mathbf{r}$ is a probability vector and $\mathbf{1}^\top\mathbf{r} = 1$, the above equation is written in the following equation:

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

where $\mathbf{B}$ is $\tilde{\mathbf{A}}^\top(\mathbf{I} - diag(\mathbf{c})) + \mathbf{q}(\mathbf{c} - \mathbf{1})^\top$. Finally, $\mathbf{r}$ is represented in the following closed form:

$$\mathbf{r} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{q}$$

Note that $\mathbf{I} - \mathbf{B}$ is invertible (see the following Lemma 1.1). ■

Table 1: Table of symbols.

Symbol	Definition
$\mathcal{G}$	input graph
n	number of nodes in $\mathcal{G}$
m	number of edges in $\mathcal{G}$
s	query node (= seed node)
c	restart probability
$\mathbf{c}$	$(n \times 1)$ restart probability vector
$\mathbf{r}$	$(n \times 1)$ relevance vector
$\mathbf{o}$	$(n \times 1)$ origin vector
$\mathbf{A}$	$(n \times n)$ adjacency matrix of $\mathcal{G}$
$\tilde{\mathbf{A}}$	$(n \times n)$ row-normalized matrix of $\mathcal{G}$
$\mathbf{A}(i,:)$	$(1 \times n)$ i-th row of a matrix $\mathbf{A}$
$\mathbf{A}(:,j)$	$(n \times 1)$ j-th column of a matrix $\mathbf{A}$
$\mathbf{J}^{ij}$	$(n \times n)$ single-entry matrix whose (i,j) entry is 1
$\mathbf{q}$	$(n \times 1)$ starting vector
P, N	set of positive and negative nodes
$\circ$	Hadamard product

Lemma 1.1. *Suppose* $\mathbf{B}$ *is* $\tilde{\mathbf{A}}^\top(\mathbf{I} - diag(\mathbf{c})) + \mathbf{q}(\mathbf{c} - \mathbf{1})^\top$. *Then,* $\mathbf{M} = \mathbf{I} - \mathbf{B}$ *is invertible.*

Proof. Let $\mathbf{C}$ be $\mathbf{I} - diag(\mathbf{c})$, and $\mathbf{X}$ be $\mathbf{I} - \tilde{\mathbf{A}}^\top\mathbf{C}$. Then, $\mathbf{M} = \mathbf{I} - \mathbf{B}$ is represented as follows:

$$\begin{aligned}\mathbf{M} &= \mathbf{I} - \tilde{\mathbf{A}}^\top\mathbf{C} - \mathbf{q}(\mathbf{c} - \mathbf{1})^\top \\ &= \mathbf{X} + \mathbf{q}(\mathbf{1} - \mathbf{c})^\top\end{aligned}$$

Note that $(\tilde{\mathbf{A}}^\top\mathbf{C})_{ij} \geq 0$, $1 \leq i,j \leq n$ and the largest eigenvalue $\lambda_{max}(\tilde{\mathbf{A}}^\top\mathbf{C}) \leq 1$ since $\tilde{\mathbf{A}}$ is stochastic and $\mathbf{C}$ is sub-stochastic. In other words, $\mathbf{X} = \mathbf{I} - \tilde{\mathbf{A}}^\top\mathbf{C}$ is M-matrix [4]; thus, $\mathbf{X}^{-1}$ exists and all entries of $\mathbf{X}^{-1}$ are nonnegative. By Sherman-Morrison lemma [7], $(\mathbf{X} + \mathbf{q}(\mathbf{1} - \mathbf{c})^\top)^{-1}$ is represented as follows:

$$\left(\mathbf{X} + \mathbf{q}(\mathbf{1} - \mathbf{c})^\top\right)^{-1} = \mathbf{X}^{-1} - \frac{\mathbf{X}^{-1}\mathbf{q}(\mathbf{1} - \mathbf{c})^\top\mathbf{X}^{-1}}{1 + (\mathbf{1} - \mathbf{c})^\top\mathbf{X}^{-1}\mathbf{q}}$$

Since all entries of $\mathbf{X}^{-1}$, $\mathbf{1} - \mathbf{c}$ and $\mathbf{q}$ are nonnegative, $(\mathbf{1} - \mathbf{c})^\top\mathbf{X}^{-1}\mathbf{q} \geq 0$; thus, $1 + (\mathbf{1} - \mathbf{c})^\top\mathbf{X}^{-1}\mathbf{q} \geq 1$. Hence, the right side of the above equation exists, i.e.,

*Seoul National University, Email: woojung211@snu.ac.kr; jinhongjung@snu.ac.kr; ukang@snu.ac.kr

$\left(\mathbf{X}+\mathbf{q}(\mathbf{1}-\mathbf{c})^\top\right)^{-1} = \mathbf{M}^{-1}$ exists. $\mathbf{I}-\mathbf{B}^\top$ is also invertible, which is proved similarly to the proof of $\mathbf{I}-\mathbf{B}$. ■

1.2 Proof of Theorem 3.1.

Proof. Equation (1.1) is represented as follows:

$$\begin{aligned}\mathbf{r} &= \tilde{\mathbf{A}}^\top(\mathbf{I}-diag(\mathbf{c}))\mathbf{r}+\left(\mathbf{q}\mathbf{c}^\top\right)\mathbf{r}\\ &= \left(\tilde{\mathbf{A}}^\top(\mathbf{I}-diag(\mathbf{c}))+\left(\mathbf{q}\mathbf{c}^\top\right)\right)\mathbf{r}=\mathbf{G}\mathbf{r}\end{aligned}$$

Note that $\mathbf{G}$ is a column stochastic matrix. Moreover, $\mathbf{G}$ is irreducible since $\tilde{\mathbf{A}}$ is irreducible, and aperiodic due to the self-loop at node s by restart. Hence, $\mathbf{r}$ is the eigenvector corresponding to the principal eigenvalue of $\mathbf{G}$, and the power iteration for $\mathbf{r}$ converges [5].

1.3 Proof of Theorem 3.2.

Proof. We assume that the number of edges is greater than that of nodes for simplicity. The iterative algorithm for computing RWER scores takes $O(Tm)$ time since each iteration requires the sparse matrix-vector multiplication which takes $O(m)$ time.

1.4 Proof of Lemma 3.2.

Proof. $\mathbf{M}(x,:)^\top$ is a column vector which is the transpose of the x-th row of the matrix $\mathbf{M}$. In other words, $\mathbf{M}(x,:)=\mathbf{e}_x^\top\mathbf{M}\Leftrightarrow\mathbf{M}(x,:)^\top=\mathbf{M}^\top\mathbf{e}_x$ where $\mathbf{e}_x$ is an $n\times 1$ vector whose x-th element is 1 and the others are 0. Then, $\tilde{\mathbf{r}}$ is represented as follows:

$$\begin{aligned}\tilde{\mathbf{r}} &= \sum_{x\in P,y\in N}\frac{\partial h(\delta_{yx})}{\partial\delta_{yx}}\left(\mathbf{M}(y,:)-\mathbf{M}(x,:)\right)^\top\\ &= \sum_{x\in P,y\in N}\frac{\partial h(\delta_{yx})}{\partial\delta_{yx}}\left(\mathbf{M}^\top\mathbf{e}_y-\mathbf{M}^\top\mathbf{e}_x\right)\\ &=\mathbf{M}^\top\sum_{x\in P,y\in N}\frac{\partial h(\delta_{yx})}{\partial\delta_{yx}}\left(\mathbf{e}_y-\mathbf{e}_x\right)=\mathbf{M}^\top\tilde{\mathbf{p}}\end{aligned}$$

where $\tilde{\mathbf{p}} = \sum_{x\in P,y\in N}\frac{\partial h(\delta_{yx})}{\partial\delta_{yx}}(\mathbf{e}_y-\mathbf{e}_x)$. Also, $\tilde{\mathbf{r}}$ is represented as the following linear system:

$$\tilde{\mathbf{r}}=\mathbf{M}^\top\tilde{\mathbf{p}}=(\mathbf{I}-\mathbf{B})^{-\top}\tilde{\mathbf{p}}\Leftrightarrow(\mathbf{I}-\mathbf{B}^\top)\tilde{\mathbf{r}}=\tilde{\mathbf{p}}$$

where $\mathbf{B}=\tilde{\mathbf{A}}^\top(\mathbf{I}-diag(\mathbf{c}))+\mathbf{q}(\mathbf{c}-\mathbf{1})^\top$.

■

1.5 Proof of Lemma 3.3.

Proof. Note that solving a sparse linear system $\mathbf{A}\mathbf{x}=\mathbf{b}$ with an iterative method such as GMRES [6] requires $O(T|\mathbf{A}|)$ time where T is the number of iterations, and $|\mathbf{A}|$ is the number of non-zeros of $\mathbf{A}$. Hence, it takes $O(T|\mathbf{I}-\mathbf{B}^\top|) = O(Tm)$ time to solve the linear system $(\mathbf{I}-\mathbf{B}^\top)\tilde{\mathbf{r}}=\tilde{\mathbf{p}}$ by the iterative method where the number of non-zeros of $\mathbf{I}-\mathbf{B}^\top$ is bounded by $O(m)$. In addition, setting $\tilde{\mathbf{p}}$ takes $O(|P||N|)$ time. Thus, the overall time complexity for computing $\sum_{x,y}\frac{\partial h(\delta_{yx})}{\partial\delta_{yx}}(\mathbf{M}(y,:)-\mathbf{M}(x,:))^\top$ is $O(Tm+|P||N|)$. ■

1.6 Proof of Theorem 3.3.

Proof. The computation of $\mathbf{r}$ takes $O(T'm)$ time according to Theorem 3.2. The computation of $\tilde{\mathbf{r}} = \sum_{x,y}\frac{\partial h(\delta_{yx})}{\partial\delta_{yx}}(\mathbf{M}(y,:)-\mathbf{M}(x,:))^\top$ takes $O(T''m+|P||N|)$ time according to Lemma 3.3. Besides, the computation of $\frac{\partial F(\mathbf{c})}{\partial\mathbf{c}}$ takes additional $O(m)$ time due to the sparse matrix-vector multiplication. Hence, SuRe takes $O(T_1(T_2m+|P||N|))$ time where T_1 is the number of iterations of the gradient descent procedure in order that the restart vector converges and $T_2=T'+T''$. ■

2 Experiments

Datasets. We experiment on various real-world network datasets including social networks, collaboration networks, citation networks, and political networks. The Wikipedia dataset is a hyperlink network. The Epinions and Slashdot datasets are signed networks. Epinions is an online review website where users show their views toward each other with positive and negative signs. Slashdot is a technology-related news website, in which users can rate each other positively or negatively. The HepPh and HepTh datasets are collaboration networks where nodes are authors, and edges are collaboration relationships time-stamped from May 15 1992 to August 14 1996 and from October 1 1993 to December 10 1999, respectively. The Polblogs dataset is a political network made up of liberal and conservative blogs.

Parameters. We set the origin vector $\mathbf{o}$ in SuRe and restart probability c in RWR, SRW, and QUINT to the ones that give the best performance. Also, in SuRe, SRW, and QUINT, we set $\lambda = 10^{-2}$ among $\{1,10^{-1},10^{-2},10^{-3}\}$, $b = 10^{-2}$ among $\{10^{-1},10^{-2},10^{-3},10^{-4}\}$, and $\eta = 10^{-2}$ where the selected parameters give the best performance as well. For SRW, we use each node's degree and the number of common neighbors as features.

2.1 Experimental Setup for Ranking. In Polbogs dataset, a node represents a blog, and an edge between nodes indicates a hyperlink between blogs. In the dataset, each node has a label which is either *liberal* or *conservative*. Among nodes connected from the query node (i.e., neighbors), we choose nodes having the same political position to the query node as positive nodes, and nodes having the opposite propensity to the query node as negative nodes. Note that the numbers of positive nodes and negative nodes do not exceed their query node's degree. We sample 115 nodes whose

Table 2: AUC result. We compare SuRe with other baselines including supervised methods SRW and QUINT for link prediction. SuRe provides the best link prediction accuracy.

Dataset	**SuRe**	RWR	SRW	QUINT
HepPh	**0.95470**	0.93597	0.94405	0.93612
HepTh	**0.96005**	0.94002	0.94845	0.94203

degrees are greater than 4 as query nodes to perform this experiment. We use all nodes except neighbors from a query node as test nodes. In this experiment, we aim to boost ranks of nodes having the same political position as the query node.

In Epinions and Slashdot datasets which are signed networks, we choose nodes connected from the seed positively as positive nodes, and nodes connected negatively as negative nodes. According to balance theory [2], friend of my friend $(+,+)$ is my friend, friend of my enemy $(+,-)$ is my enemy, and enemy of my friend $(-,+)$ is my enemy. Based on the theory, we use them as test nodes, and the goal is to rank $(+,+)$ nodes from the seed higher than $(+,-)$ and $(-,+)$ nodes. In this experiment, we do not include $(-,-)$ nodes since it can be interpreted differently depending on *strong* or *weak* version of balance theory [2, 3].

2.2 Experimental Setup for Link Prediction. In the link prediction task, we aim to predict future links from a query node to other nodes based on relevance scores. Here, we focus on predicting links to nodes that are 2-hops away from the query node since most of new edges are created closing a triangle [1]. The experimental setting is as follows:

- We consider time-stamp information in networks to construct training and test datasets as in [1]. We select nodes whose degrees are greater than 30 as query nodes. For each query node s, let $t_{s,\min}$ and $t_{s,\max}$ denote the minimum and maximum time-stamp of 1-hop neighbors of node s, respectively. Suppose $t_{s,\min} < t_{s,1} < t_{s,2} < t_{s,\max}$. We select links $(s,\, v)$ created between $t_{s,2}$ and $t_{s,\max}$ as test data where node v is 2-hop neighbors from node s before the links are created. We choose the query node's 1-hop neighbors created between $t_{s,1}$ and $t_{s,2}$ as positive nodes. We sample the same number of negative nodes as that of positive ones, which are 3-hops or more from node s. We exploit other links excepts the test links as training data. We select $t_{s,1}$ and $t_{s,2}$ such that $t_{s,1} = t_{s,\min} + 0.3L$ and $t_{s,2} = t_{s,\min} + 0.7L$, respectively, where $L = t_{s,\max} - t_{s,\min}$ is the total time length.

Table 3: Dataset statistics. The query nodes are used for the ranking and the link prediction tasks.

Dataset	**# Nodes**	**# Edges**	**# Queries**
Cora[1]	23,116	91,500	154
DBLP[1]	12,591	49,728	162

[1] http://konect.uni-koblenz.de

2.3 AUC Result for Link Prediction. Table 2 compares the link prediction performance of SuRe with that of RWR and other supervised methods SRW and QUINT in terms of AUC. Specifically, SuRe gets 1.2% relative improvement over the second best supervised method SRW on the HepTh dataset.

3 Additional Experiments

In this section, we present the results of additional experiments to answer the following questions:

- **Q1. Ranking performance.** Does our proposed method SuRe provide better relevances scores for ranking compared to other methods in non time-stamped datasets?

3.1 Link Prediction Performance. We examine the link prediction performance of our proposed method SuRe compared to other methods in non time-stamped datasets; The Cora and DBLP datasets are citation networks where nodes are papers, and edges indicate citation relationships.

Experimental Setup for Link Prediction. We also focus on predicting links to nodes that are 2-hops away from the query node. The experimental setting is as follows:

- We keep a fraction (80%) of all edges as a snapshot of a network, and evaluate on a test set composed of the remaining edges. Since there is no label information in the datasets, we choose positive and negative nodes as follows: from the snapshot, we select all nodes connected from the query node as positive nodes; we sample the equal number of nodes that is not connected from the query node as negative nodes. Note that the numbers of positive nodes and negative nodes do not exceed their query node's degrees. We select nodes whose degrees are greater than 4 as query nodes.

Result. Figures 1 and 2, and Table 4 show the link prediction performance in terms of MAP,

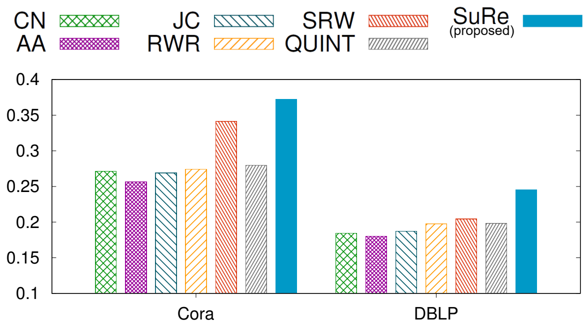

Figure 1: Link prediction performance in terms of MAP (Mean Average Precision). SuRe obtains up to 20% relative improvement (in DBLP) compared to the best competitor SRW.

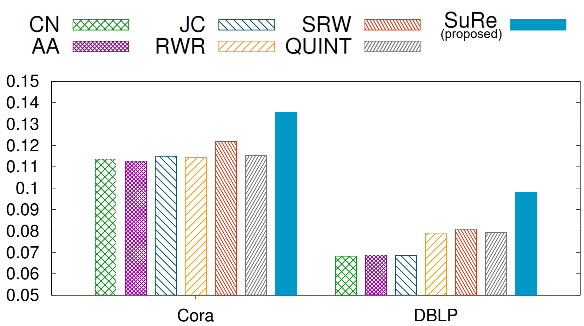

Figure 2: Link prediction performance in terms of Precision@20. SuRe achieves 21% improvement on the best competitor in the DBLP dataset.

Precision@20, and AUC, respectively. As shown in the results, our proposed method SuRe outperforms other competitors including SRW and QUINT which are the state-of-the-art methods for link prediction. In the DBLP dataset, compared to the best competitor SRW, SuRe achieves 20% improvement in terms of MAP (Figure 1), and 21% improvement in terms of Precision@20 (Figure 2). Table 2 compares the link prediction performance of SuRe with that of RWR and other supervised methods SRW and QUINT in terms of AUC. Note that SuRe provides the best prediction over all datasets. Specifically, SuRe gets 6% relative improvement over the second best supervised method SRW on the DBLP dataset.

References

[1] L. Backstrom and J. Leskovec. Supervised random walks: predicting and recommending links in social networks. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 635–644. ACM, 2011.

Table 4: AUC result. We compare SuRe with other baselines including supervised methods SRW and QUINT for link prediction. SuRe provides the best link prediction accuracy.

Dataset	**SuRe**	RWR	SRW	QUINT
Cora	**0.72274**	0.64708	0.69112	0.64914
DBLP	**0.73786**	0.68570	0.69917	0.68658

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

[3] J. A. Davis. Clustering and structural balance in graphs. *Human relations*, 20(2):181–187, 1967.

[4] T. Fujimoto and R. R. Ranade. Two characterizations of inverse-positive matrices: the hawkins-simon condition and the le chatelier-braun principle. *Electronic Journal of Linear Algebra*, 11(1):6, 2004.

[5] A. N. Langville and C. D. Meyer. *Google's PageRank and beyond: The science of search engine rankings.* Princeton University Press, 2011.

[6] Y. Saad and M. H. Schultz. Gmres: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on scientific and statistical computing*, 7(3):856–869, 1986.

[7] J. Sherman and W. J. Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21(1):124–127, 1950.