Asymmetric Transitivity Preserving Graph Embedding

Mingdong Ou
Tsinghua University
oumingdong@gmail.com

Peng Cui
Tsinghua University
cuip@tsinghua.edu.cn

Jian Pei
Simon Fraser University
jpei@cs.sfu.ca

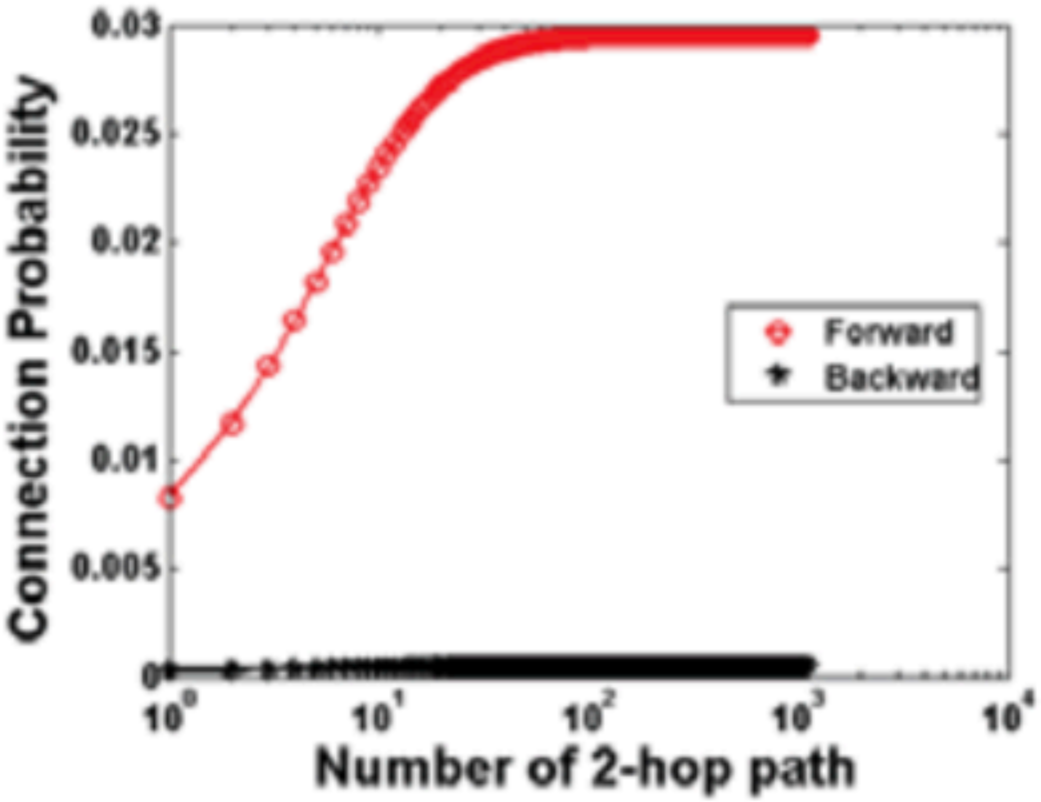
Ziwei Zhang
Tsinghua University
zhangziwei1993@126.com

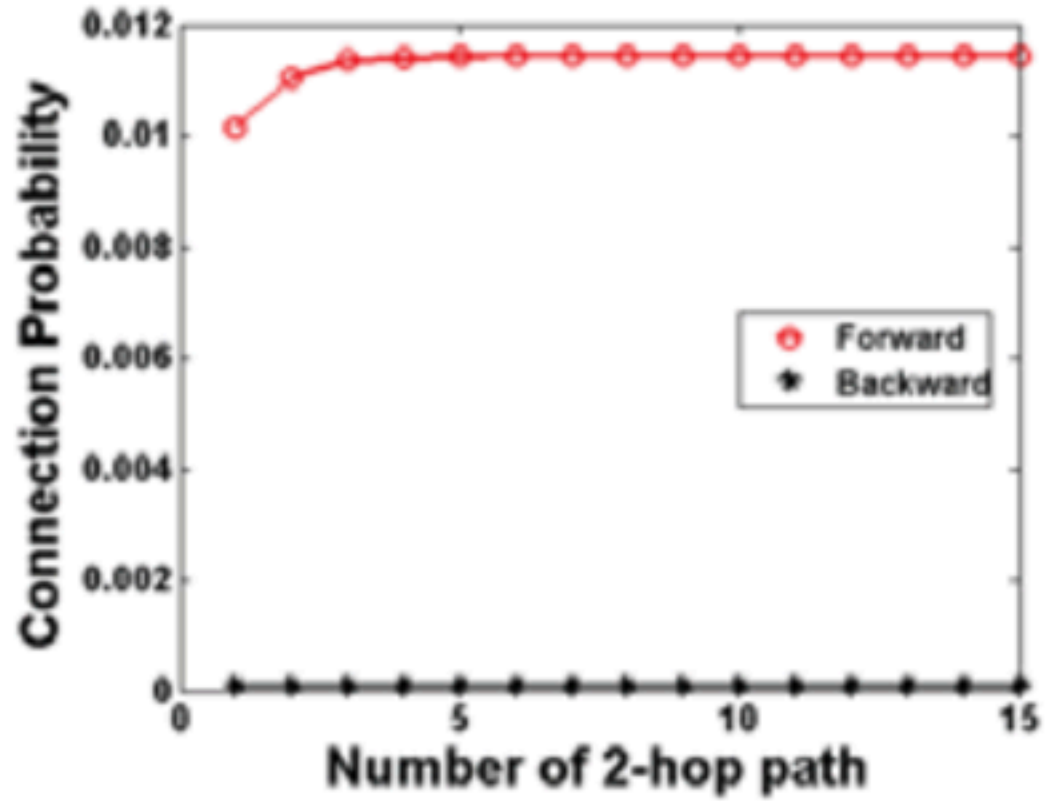
Wenwu Zhu
Tsinghua University
wwzhu@tsinghua.edu.cn

ABSTRACT

Graph embedding algorithms: embed a graph into a vector space where the structure and the inherent properties of the graph are preserved.

This paper propose the idea of preserving asymmetric transitivity by approximating high-order proximity which are based on asymmetric transitivity.

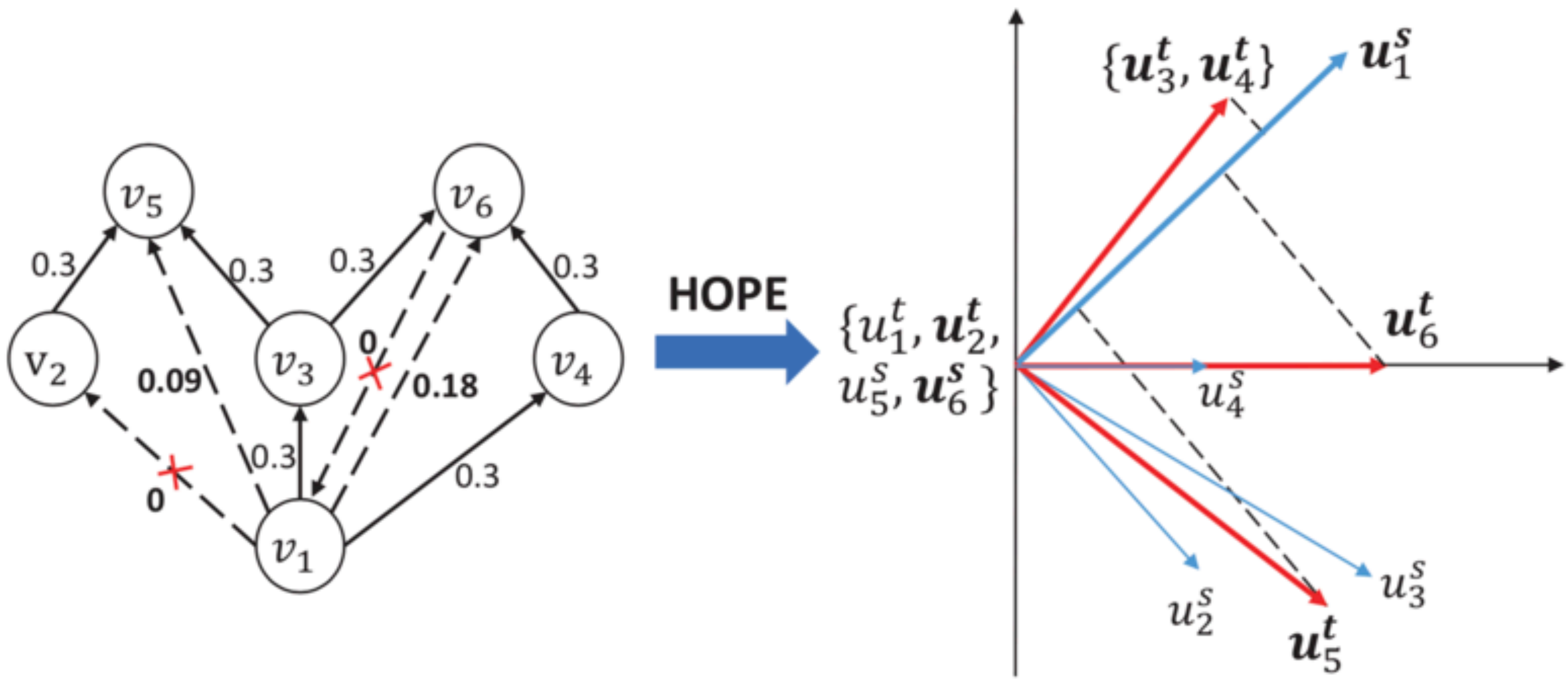




Transitivity assumption: the more paths from u to v there are, the more probable it is that there exists an edge from u to v.

Asymmetry assumption: the forward connection probability is much larger than the backward connection probability,

Problem Definition



$$\min \|\mathbf{S} - \mathbf{U}^s \cdot \mathbf{U}^{t\top}\|_F^2$$

High order proximities

Many of them share a general formulation which will facilitate the approximation of these proximities, that is:

$$\mathbf{S} = \mathbf{M}_g^{-1} \cdot \mathbf{M}_l$$

Katz Index

$$\mathbf{S}^{Katz} = \sum_{l=1}^{\infty} \beta \cdot \mathbf{A}^l = \beta \cdot \mathbf{A} \cdot \mathbf{S}^{Katz} + \beta \cdot \mathbf{A}$$

$$\mathbf{S}^{Katz} = (\mathbf{I} - \beta \cdot \mathbf{A})^{-1} \cdot \beta \cdot \mathbf{A}$$

Rooted PageRank (RPR)

$$\mathbf{S}^{RPR} = \alpha \cdot \mathbf{S}_{ij}^{RPR} \cdot \mathbf{P} + (1 - \alpha) \cdot \mathbf{I}$$

$$\mathbf{S}^{RPR} = (1 - \alpha) \cdot (\mathbf{I} - \alpha \mathbf{P})^{-1}$$

Common Neighbors(CN)

$$\mathbf{S}^{CN} = \mathbf{A}^2$$

Adamic-Adar (AA)

$$\mathbf{S}^{AA} = \mathbf{A} \cdot \mathbf{D} \cdot \mathbf{A}$$

Table 1: General Formulation for High-order Proximity Measurements

Proximity Measurement	$\mathbf{M}_g$	$\mathbf{M}_l$
Katz	$\mathbf{I} - \beta \cdot \mathbf{A}$	$\beta \cdot \mathbf{A}$
Personalized Pagerank	$\mathbf{I} - \alpha\mathbf{P}$	$(1-\alpha) \cdot \mathbf{I}$
Common neighbors	$\mathbf{I}$	$\mathbf{A}^2$
Adamic-Adar	$\mathbf{I}$	$\mathbf{A} \cdot \mathbf{D} \cdot \mathbf{A}$

According to [13], the solution is to perform SVD (Singular Value Decomposition) on S and use the largest K singular value and corresponding singular vectors to construct the optimal embedding vectors.

$$\mathbf{S} = \sum_{i=1}^{N} \sigma_i \mathbf{v}_i^s {\mathbf{v}_i^t}^\top$$

the optimal embedding vectors as

$$\mathbf{U}^s = [\sqrt{\sigma_1} \cdot \mathbf{v}_1^s, \cdots, \sqrt{\sigma_K} \cdot \mathbf{v}_K^s]$$

$$\mathbf{U}^t = [\sqrt{\sigma_1} \cdot \mathbf{v}_1^t, \cdots, \sqrt{\sigma_K} \cdot \mathbf{v}_K^t]$$

. [13] J. Hopcroft and R. Kannan. Foundations of data science. 2014.

Theorem 1. If we have the singular value decomposition of the general formulation

$$\mathbf{M}_g^{-1} \cdot \mathbf{M}_l = \mathbf{V}^s \Sigma \mathbf{V}^{t\top}$$

V_t and V_s are two orthogonal matrices, $\Sigma = diag(\sigma_1, \sigma_2, \cdots, \sigma_N)$

Then, there exists a nonsingular

$$\mathbf{V}^{t\top} \mathbf{M}_l^\top \mathbf{X} = \Sigma^l$$
$$\mathbf{V}^{s\top} \mathbf{M}_g^\top \mathbf{X} = \Sigma^g$$

diagonal matrices, i.e. Σ_l and Σ_g, satisfying that :

Where

$$\Sigma^l = diag(\sigma_1^l, \sigma_2^l, \cdots, \sigma_N^l)$$
$$\Sigma^g = diag(\sigma_1^g, \sigma_2^g, \cdots, \sigma_N^g)$$
$$\sigma_1^l \geq \sigma_2^l \geq \cdots \geq \sigma_K^l \geq 0$$
$$0 \leq \sigma_1^g \leq \sigma_2^g \leq \cdots \leq \sigma_K^g$$
$$\forall i \qquad {\sigma_i^l}^2 + {\sigma_i^g}^2 = 1$$

. M. Hochstenbach. A jacobi–davidson type method for the generalized singular value problem. Linear Algebra and its Applications, 431(3):471–487, 2009.

Algorithm 1 High-order Proximity preserved Embedding

Require: adjacency matrix $\mathbf{A}$, embedding dimension K, parameters of high-order proximity measurementθ.

Ensure: embedding source vectors $\mathbf{U}^s$ and target vectors $\mathbf{U}^t$.

1: calculate $\mathbf{M}_g$ and $\mathbf{M}_l$.

2: perform JDGSVD with $\mathbf{M}_g$ and $\mathbf{M}_l$, and obtain the generalized singular values $\{\sigma_1^l, \cdots, \sigma_K^l\}$ and $\{\sigma_1^g, \cdots, \sigma_K^g\}$, and the corresponding singular vectors, $\{\mathbf{v}_1^s, \cdots, \mathbf{v}_K^s\}$ and $\{\mathbf{v}_1^t, \cdots, \mathbf{v}_K^t\}$.

3: calculate singular values $\{\sigma_1, \cdots, \sigma_K\}$ according to Equation (21).

4: calculate embedding matrices $\mathbf{U}^s$ and $\mathbf{U}^t$ according to Equation (19) and (20).

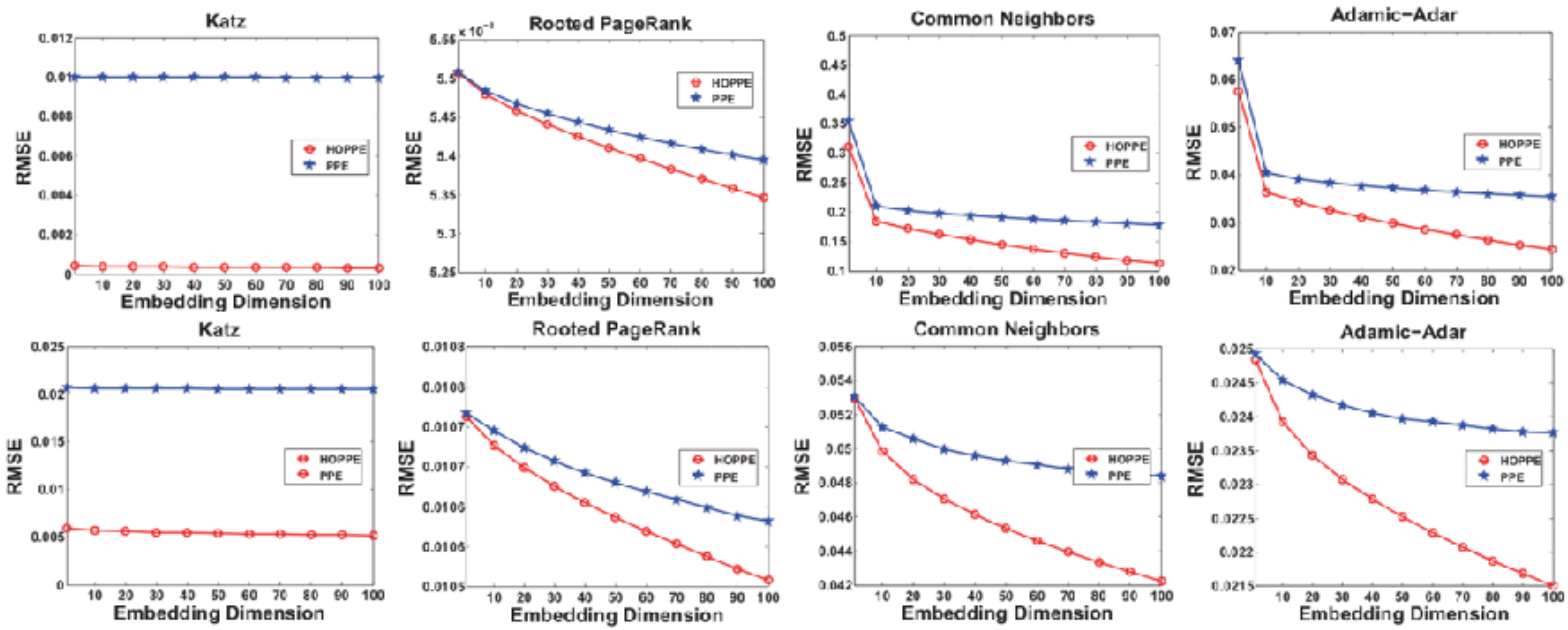
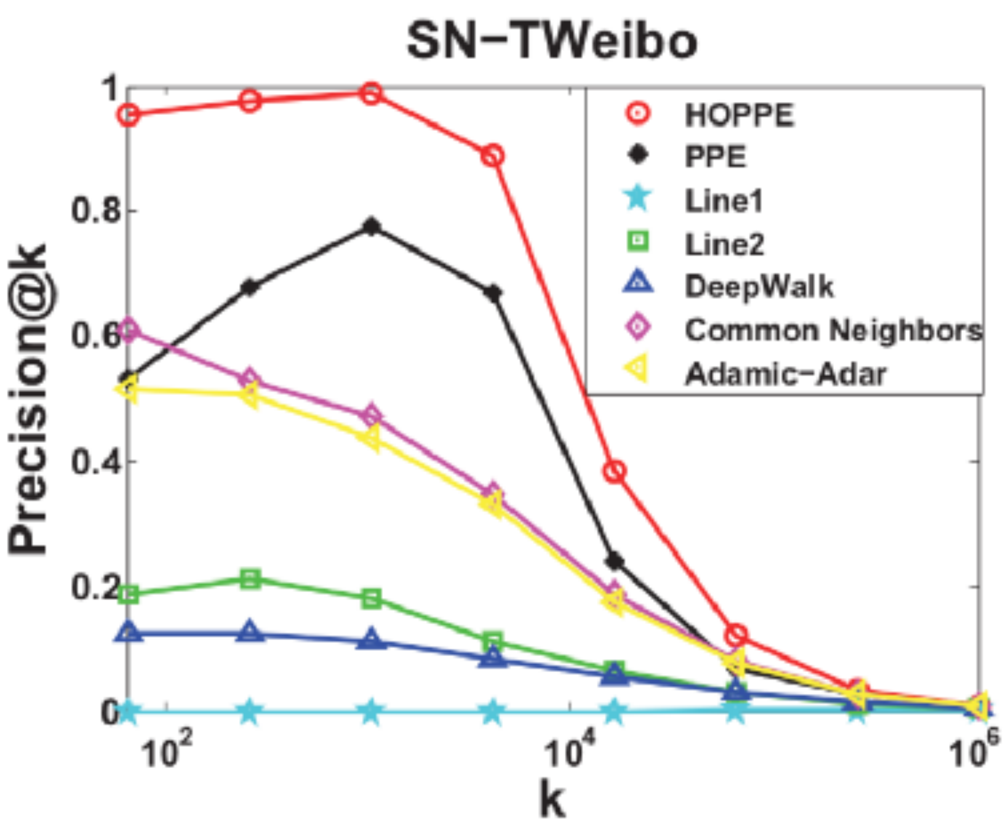
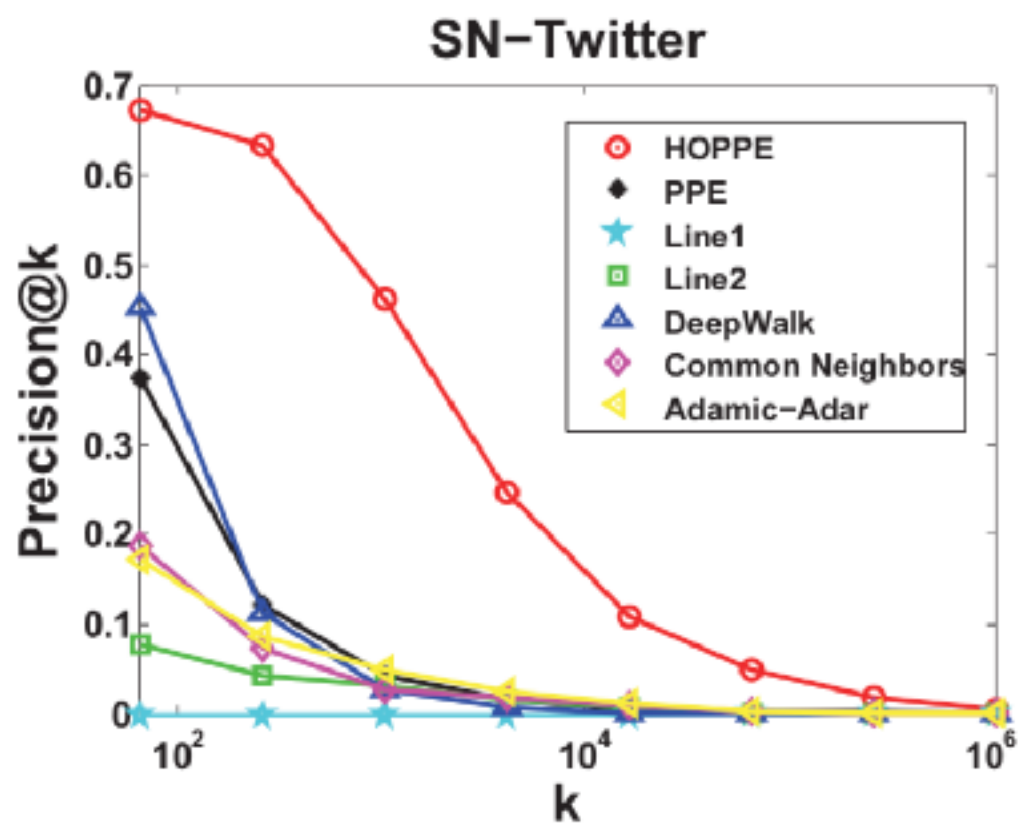


Figure 3: Error of proximity approximation. We evaluate the errors of HOPE and PPE in approximating four proximity measurements, including Katz, RPR, Common Neighbors and Adamic-Adar. First row is the results on Synthetic Data, and second row is the results on Cora. For Katz, $\beta = 0.1$; for RPR, $\alpha = 0.5$.

Graph Reconstruction





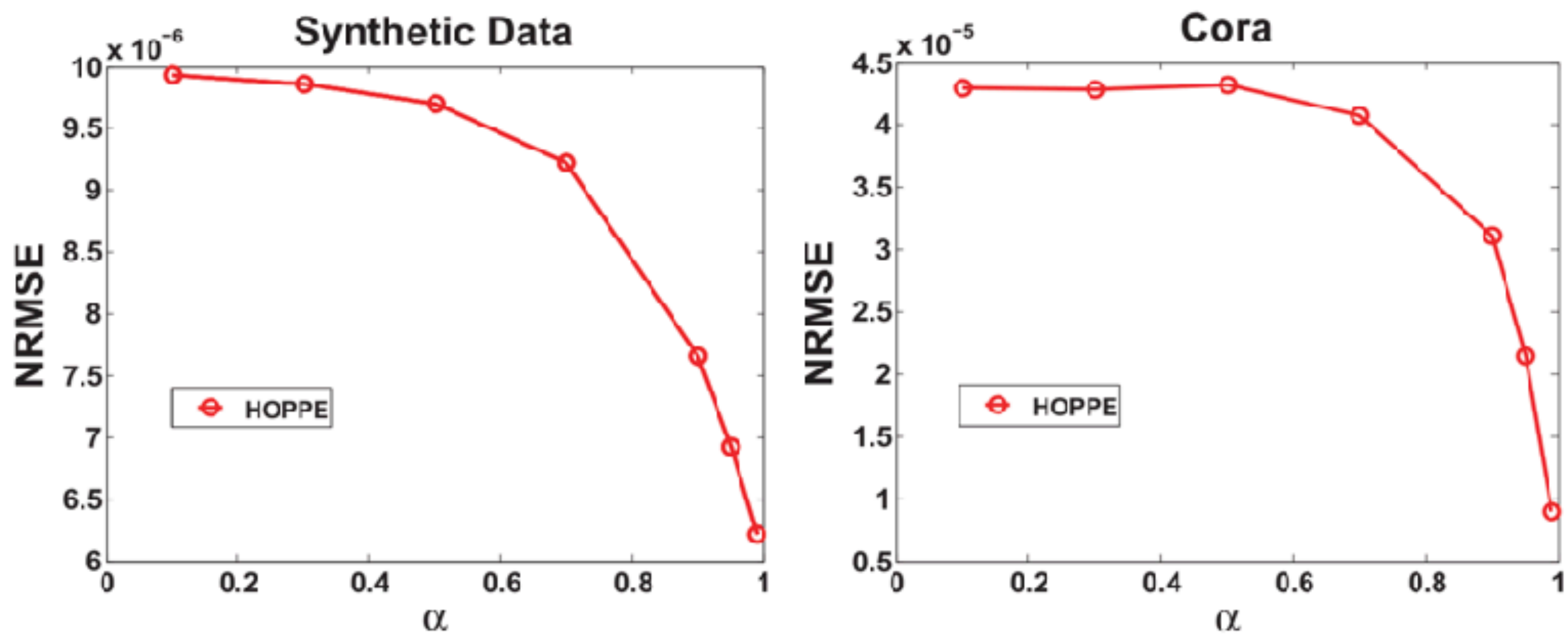


Figure 4: Correlation between relative approximation error and the rank of proximity matrix. The parameter α of Rooted PageRank is highly related to the rank of proximity matrix. Here, we use α to simulate the rank of proximity matrix. The embedding dimension is 100.

Vertex Recommendation