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

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. The existing graph embedding methods cannot preserve the asymmetric transitivity well, which is a critical property of directed graphs. Asymmetric transitivity depicts the correlation among directed edges, that is, if there is a directed path from $u$ to $v$, then there is likely a directed edge from $u$ to $v$. Asymmetric transitivity can help in capturing structures of graphs and recovering from partially observed graphs. To tackle this challenge, we propose the idea of preserving asymmetric transitivity by approximating high-order proximity which are based on asymmetric transitivity. In particular, we develop a novel graph embedding algorithm, High-Order Proximity preserved Embedding (HOPE for short), which is scalable to preserve high-order proximities of large scale graphs and capable of capturing the asymmetric transitivity. More specifically, we first derive a general formulation that cover multiple popular high-order proximity measurements, then propose a scalable embedding algorithm to approximate the high-order proximity measurements based on their general formulation. Moreover, we provide a theoretical upper bound on the RMSE (Root Mean Squared Error) of the approximation. Our empirical experiments on a synthetic dataset and three real-world datasets demonstrate that HOPE can approximate the high-order proximities significantly better than the state-of-art algorithms and outperform the state-of-art algorithms in tasks of reconstruction, link prediction and vertex recommendation.

Categories and Subject Descriptors
H.4 [Information Systems Applications]: Miscellaneous; D.2.8 [Software Engineering]: Metrics—complexity measures, performance measures

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WOODESTOCK ’97 El Paso, Texas USA
Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...$15.00.

Figure 1: Asymmetric Transitivity in Real Data. Given a pair of vertexes $(u, v)$, the horizontal axis is the Number of 2-hop paths from $u$ to $v$. For the red (or black) curve with circle (or star) marker, the vertical axis represents the cumulative forward (or backward) transiting probability, that is the connection probability from $u$ (or $v$) to $v$ (or $u$), when the number of 2-hop paths is less than corresponding horizontal axis value. The left figure is the statistics from the social network of Tencent Weibo, and the right figure is the statistics from the social network of Twitter. As the two red curves increases monotonically, we can claim that the more paths from $u$ to $v$ there are, the more probable it is that there exists an edge from $u$ to $v$, which reflects the transitivity assumption. The forward connection probability is much larger than the backward connection probability, which reflects the asymmetry assumption. In summary, asymmetric transitivity can be widely observed in real datasets.

Keywords
asymmetric transitivity, directed graph, high-order proximity, graph embedding

1. INTRODUCTION
Nowadays, more and more applications are based on larger and larger networks and graphs. It is well recognized that graph data is sophisticated and challenging. To process graph data effectively, the first critical challenge is graph data representation, that is, how to represent graphs properly so that advanced analytic tasks, such as pattern discovery, analysis, and prediction, can be conducted efficiently in both time and space. However, graph representation in general remains an open problem.

Recently, the graph embedding paradigm [35] is proposed to represent vertices of a graph in a low-dimensional vector space while the structures (i.e. edges and other high-order
structures) of the graph can be reconstructed in the vector space. With proper graph embedding, we can easily apply classic vector-based machine learning techniques to process graph data. Meanwhile, it effectively facilitates the parallelization of graph analysis by decoupling highly correlated vertices into independent vectors. However, most of the existing graph embedding methods target on undirected graphs, a natural and popularly used representation of data in many applications, such as social networks and webpage networks, are largely untouched.

Can we straightforwardly apply undirected graph embedding methods on directed graphs? The answer is no due to a fundamentally different characteristic of directed graphs: asymmetric transitivity.

Transitivity is a common characteristic of undirected and directed graphs [29] (see Figure 1). In undirected graphs, if there is an edge between vertices u and v, and one between w and v, then it is likely that u and v are connected by an edge. Transitivity plays a key role in graph inference and analysis tasks, such as calculating similarities between nodes [16, 14] and measuring the importance of nodes. Transitivity is symmetric in undirected graphs. However, transitivity is asymmetric in directed graphs. If there is a directed link from u to w and a directed link from w to v, there is likely a directed link from u to v; but not from v to u. More generally, if there is a directed path from u to v, there is likely a directed link from u to v, but not from v to u.

Preserving the symmetric transitivity in undirected graphs in an embedding space is natural and straightforward, as most of the distance metrics defined in the vector space possess the property of symmetric transitivity. However, how to preserve the asymmetric transitivity of directed graphs in a vector space is much more challenging.

Recently, a paucity of studies [37, 20, 3, 2, 24] focus on directed graph embedding. In order to reflect the asymmetric property in vector space, these methods design asymmetric metrics on the embedding vectors. Unfortunately, those asymmetric metrics are not fully transitive, which severely limit the capability of the learned embedding space in reflecting the structure of graphs as well as supporting graph inference and analysis.

In this paper, we tackle the challenging problem of asymmetric transitivity preserving graph embedding. Our major idea is that we learn two embedding vectors, source vector and target vector, for each node to capture asymmetric edges, as illustrated in Figure 2. Then, for a directed link from v1 to v2 without the reverse link from v2 to v1, we can assign similar values to v1’s source vector and v2’s target vector, and assign very different values to v1’s target vector and v2’s source vector. In this way, the feasibility of asymmetry transitivity is preserved. The key challenge is how well the asymmetric transitivity of directed graphs can be preserved in the embedding learning.

From the learning perspective, a good way is to let the learned embedding vectors directly approximate a target measure reflecting asymmetric transitivity of graphs. From
the graph embedding perspective, the property of asymmetric transitivity leads to the assumption that the more and the shorter paths from \( v_i \) to \( v_j \), the more similar should be \( v_i \)'s source vector and \( v_j \)'s target vector. This assumption coincides with high-order proximities of nodes in graphs, such as Katz [16] and rooted pagerank [30]. What is more, these high-order proximities are defined to be asymmetric in directed graphs. Thus, in its place, we propose to use high-order proximities of nodes as the target measure, resulting in a novel directed graph embedding algorithm, High Order Proximity preserved Embedding (HOPE).

In this model, we theoretically derive a general form covering multiple high-order proximities. Interestingly, the general form is consistent with the formulation of generalized SVD. Based on this, we propose a scalable embedding algorithm for large-scale graphs by avoiding the time-consuming computation of high-order proximities. Moreover, we derive a theoretical upper bound on the approximation error of HOPE, which is used to estimate the embedding quality in theory and determine embedding dimensions automatically.

To verify the advantages of our algorithm, we conduct experiments on both synthetic data and 3 real datasets. The experiments show that our method consistently and significantly outperforms the state-of-the-art baselines in several tasks, including high-order proximity approximation, graph reconstruction, link prediction and vertex recommendation.

The main contributions of this paper are as follows:

- We propose a high-order proximity preserved embedding (HOPE) method to solve the challenging problem of asymmetric transitivity in directed graph embedding.
- We derive a general form covering multiple commonly used high-order proximities, enabling the scalable solution of HOPE with generalized SVD.
- We provide an upper bound on the approximation error of HOPE.
- Extensive experiments are conducted to verify the usefulness and generality of the learned embedding in various applications.

The rest of the paper is organized as follows. In Section 2, we review the related work. We develop our method in Section 3 and report the experimental results in Section 4. We conclude the paper in Section 5.

2. RELATED WORK

2.1 Graph Embedding

Graph embedding technology has been widely studied in the fields of dimensionality reduction [15, 27, 33, 1, 8], natural language processing [18], network analysis [11] and so on.

For dimensionality reduction, adjacency matrices of graphs are constructed from the feature similarity (distance) between samples [35]. And the graph embedding algorithms aim to preserve the feature similarity in the embedded latent space. For example, Laplacian Eigenmaps (LE) [1] aims to learn the low-dimensional representation to expand the manifold where data lie. Locality Preserving Projections (LPP) [8] is a linearization variant of LE which learns a linear projection from feature space to embedding space. Besides, there are many other graph embedding algorithms for dimensionality reduction, including non-linear [33, 27], linear [15, 6], kernelized [28] and tensorized [36] algorithms. All of these algorithms are based on undirected graphs derived from symmetric similarities. Thus, they cannot preserve asymmetric transitivity.

In the field of natural language processing, the graph of words is often used to learn the representation of words [19, 18, 23]. Mikolov et. al. [19] propose to utilize the context of words to learn representation, which has been proved equivalent to factorizing word-context matrix [18]. Pennington et. al. [23] exploit a word-word co-occurrence matrix. In network analysis, Hof et. al. [11] first propose to learn latent space representation of vertexes in graph and the probability of a relation depends on the distance between vertexes in the latent space, and they apply it to link prediction problem [10]. Handcock et. al. [7] propose to apply the latent space approach to clustering in graph. And Zhu et. al. [37] propose to address the classification problem in graph with graph embedding model. While early graph embedding works focus on modeling the observed first-order relationship (i.e. edges in graph) between vertexes, some recent works try to model the directed higher order relationships between vertexes in sparse graphs [24, 2]. GraRep [2] is related to our work. But, it cannot fully capture transitivity. Moreover, it is not scalable for large-scale graph.

2.2 Directed Graph

Theoretically, any type (undirected and directed) of graph can be represented as directed graph. So, modeling directed graph is a critical problem for graph analysis. Holland et. al. [12] propose the p1 distribution model to capture the structural properties in directed graph, including the attractiveness and expansiveness of vertexes and the reciprocation of edges. Besides these properties, Wang et. al. [34] take the group information of vertexes into consideration. Recently, some works adopt graph embedding [4, 26, 25, 21] to model directed graphs. Chen et. al. [4] learn the embedding vectors in Euclidean space with locality preserved. Perrault-Joncas et. al. [26, 25] and Mousazadeh et. al. [21] learn the embedding vectors based on Laplacian type operators and preserve the asymmetry property of edges in a vector field. However, all of these methods cannot preserve asymmetry property in embedding vector space.

3. HIGH-ORDER PROXIMITY PRESERVED EMBEDDING

In this section, we will derive how to preserve high-order proximities in the embedding space. Before introducing the detailed derivation, we will clarify the symbols and definition that will be used.

3.1 Notations

We define a directed graph as \( G = (V, E) \). \( V \) is the vertex set. Let \( V = \{v_1, \ldots, v_i, \ldots, v_N\} \) where \( N \) is the number of vertexes. \( E \) is the directed edge set. \( e_{ij} = (v_i, v_j) \in E \) represents a directed edge from \( v_i \) to \( v_j \). The adjacency matrix is denoted as \( A \). We define a high-order proximity matrix as \( S \), where \( S_{ij} \) is the proximity between \( v_i \) and \( v_j \). And \( U = [U^*, U^]\) is the embedding matrix, where the \( i \)-th row, \( u_i \), is the embedding vector of \( v_i \). \( U^*, U^\in\mathbb{R}^{N \times K} \) are
the source embedding vectors and target embedding vectors respectively, where $K$ is the embedding dimensions. For any matrix $B$, the lowercase symbol $b_{i}$ represents the $i$-th row of $B$, and $B_{ij}$ represents the element at $i$-th row and $j$-th column.

### 3.2 Problem Definition

In this paper, we focus on the directed graph embedding problem. It aims to represent the vertexes in the numerical vector space, where asymmetric transitivity is preserved. As high-order proximities are derived from asymmetric transitivity, we propose to preserve the asymmetric transitivity by approximating high-order proximity. Formally, we adopt the L2-norm below as the loss function which need to be minimized:

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

As Figure 2 shows, the embedding vectors can well preserve the asymmetric transitivity.

### 3.3 High order proximities

Many high-order proximity measurements in graph can reflect the asymmetric transitivity. Moreover, we found that many of them share a general formulation which will facilitate the approximation of these proximities, that is:

$$S = M_y^{-1} \cdot M_t
$$

where $M_y$ and $M_t$ are both polynomial of matrices. Below, we will introduce some popular proximity measurements and transform them into this formulation.

**Katz Index** [16]. This index is an ensemble of all paths, which is a weighted summation over the path set between two vertexes. The weight of a path is a exponential function of its length. Actually, the formula of Katz index can also be written as a recurrent formula:

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

where $\beta$ is a decay parameter. It determines how fast the weight of a path decay when the length of path grows. $\beta$ should be properly set to preserve the series converging. Actually, $\beta$ must be smaller than the spectral radius of adjacency matrix.

Then, we can get that:

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

where $I$ is an identity matrix.

For Katz index,

$$M_y = I - \beta \cdot A
$$

$$M_t = \beta \cdot A
$$

**Rooted PageRank (RPR)**. $S^{RPR}_{ij}$ is the probability that a random walk from node $v_i$ will locate at $v_j$ in the steady state. Consider one step of a random walk from $v_i$, the random walker randomly jumps to one of the neighbor of current node with probability $\alpha$, and jumps back to $v_i$ with probability $\alpha$. The formula is:

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

, where $\alpha \in [0,1)$ is the probability to randomly walk to a neighbor, and $P$ is the probability transition matrix satisfying that $\sum_{j=1}^{N} P_{ij} = 1$.

Then, we can get:

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

So, for rooted PageRank,

$$M_y = I - \alpha P
$$

$$M_t = (1 - \alpha) \cdot I
$$

**Common Neighbors (CN)**. $S_{ij}^{CN}$ counts the number of vertexes connecting to both $v_i$ and $v_j$. For directed graph, $S_{ij}^{CN}$ is the number of vertexes which is the target of an edge from $v_i$ and the source of an edge to $v_j$. Formally,

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

So, for common neighbors,

$$M_y = I
$$

$$M_t = A^2
$$

**Adamic-Adar (AA)**. Adamic-Adar is a variant of common neighbors. Unlike common neighbors, Adamic-Adar assigns each neighbor a weight that is the reciprocal of the degree of the neighbor. This means that the more vertexes one vertex connected to, the less important it is on evaluating the proximity between a pair of vertex. Formally,

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

where $D$ is a diagonal matrix,

$$D_{ii} = 1/ \sum_{j} (A_{ij} + A_{ji})
$$

So, for Adamic-Adar,

$$M_y = I
$$

$$M_t = A \cdot D \cdot A
$$

We list the corresponding formula of $M_y$ and $M_t$ of each proximity measurement in Table 1. Note that $M_y$ and $M_t$ are both the polynomial of adjacency matrix or its variants. The above proximity measurements can be classified into two types, i.e. global proximity and local proximity. Global proximities, i.e. Katz index and rooted PageRank, are derived from a recurrent formula, which make the proximity can preserve global asymmetric transitivity. And local proximities, i.e. common neighbors and Adamic-Adar, have no recurrent structure and just preserve the asymmetric transitivity in the local structure, which we call it local asymmetric transitivity. Intuitively, $M_y$ is highly related to global asymmetric transitivity. $M_t$ has a formulation like $I - \alpha \cdot B$, where $\alpha$ is a parameter and $B$ is a transition matrix. The larger the alpha is, the easier the observed relationship (the edges of the graph) is to be transited in the graph. When
\( \alpha = 0 \), the observed relationship can just transit in a subgraph whose range is limited by the order of \( M_i \).

### 3.4 Approximation of High-Order Proximity

The objective in Equation (1) aims to find an optimal rank-\( K \) approximation of the proximity matrix \( S \). 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. Formally, if

\[
S = \sum_{i=1}^{N} \sigma_i v_i^T v_i^\top \tag{18}
\]

where \( \{\sigma_1, \sigma_2, \ldots, \sigma_N\} \) is the singular values sorted in decreasing order, \( v_i^T \) and \( v_i^\top \) are corresponding singular vectors of \( \sigma_i \),

then, we can get the optimal embedding vectors as:

\[
U^* = [\sqrt{\sigma_1} v_1^\top, \ldots, \sqrt{\sigma_K} v_K^\top] \tag{19}
\]

\[
U^T = [\sqrt{\sigma_1} v_1^\top, \ldots, \sqrt{\sigma_K} v_K^\top] \tag{20}
\]

However, this solution need to calculate the proximity matrix \( S \). Even for sparse adjacency matrix, the time complexity of matrix inversion is up to \( O(N^3) \). And the matrix polynomial operation on adjacency matrix will make \( S \) much denser than adjacency matrix \( A \) which will also make the SVD on \( S \) very expensive. Thus, the solution is not feasible for large scale graphs.

As the calculation of \( S \) is the efficiency bottleneck and \( S \) is just the intermediate product in our problem, we propose a novel algorithm to avoid the calculation of \( S \) and learn the embedding vectors directly. As many proximity measurements have the general formulation in Equation (2), we transform the original SVD problem into a generalized SVD problem [22] for proximity measurements with the general formulation.

According to [22], it is easy to derive the following theorem:

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

\[
M_g^{-1} \cdot M_I = V^* \Sigma V^\top
\]

where \( V^T \) and \( V^\top \) are two orthogonal matrices,

\[
\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_N)
\]

Then, there exists a nonsingular matrix \( X \) and two diagonal matrices, i.e. \( \Sigma^d \) and \( \Sigma^g \), satisfying that

\[
V^\top M_g^{-1} X = \Sigma^d
\]

\[
V^\top M_g X = \Sigma^g
\]

, where

\[
\Sigma^d = \text{diag}(\sigma_1^d, \sigma_2^d, \ldots, \sigma_N^d)
\]

\[
\Sigma^g = \text{diag}(\sigma_1^g, \sigma_2^g, \ldots, \sigma_N^g)
\]

\[
\sigma_1^d \geq \sigma_2^d \geq \ldots \geq \sigma_N^d \geq 0
\]

\[
0 \leq \sigma_1^g \leq \sigma_2^g \leq \ldots \leq \sigma_N^g
\]

\[
\forall i \quad \sigma_i^d + \sigma_i^g = 1
\]

, and

\[
\sigma_i = \frac{\sigma_i^d}{\sigma_i^g} \tag{21}
\]

**Algorithm 1** High-order Proximity preserved Embedding

**Require:** adjacency matrix \( A \), embedding dimension \( K \), parameters of high-order proximity measurement\( \theta \).

**Ensure:** embedding source vectors \( U^* \) and target vectors \( U^T \).

1: calculate \( M_g \) and \( M_I \).
2: perform JDGSVD with \( M_g \) and \( M_I \) and obtain the generalized singular values \( \{\sigma_1^d, \ldots, \sigma_K^d\} \) and \( \{\sigma_1^g, \ldots, \sigma_K^g\} \), and the corresponding singular vectors, \( \{v_1^d, \ldots, v_K^d\} \) and \( \{v_1^g, \ldots, v_K^g\} \).
3: calculate singular values \( \{\sigma_1, \ldots, \sigma_K\} \) according to Equation (21).
4: calculate embedding matrices \( U^* \) and \( U^T \) according to Equation (19) and (20).

As the generalized SVD can also derive the results of SVD, we can still use Equation (19) (20) to get the embedding vectors.

Complete generalized SVD will also achieve \( O(N^3) \) time complexity, which is not feasible for large scale graphs. As we only need the largest \( K \) singular values and corresponding singular vectors, we adopt a partial generalized SVD algorithm [9], which we call it JDGSVD. This is an iterative Jacobi-Davidson type algorithm which is very scalable when \( K \ll N \). Algorithm 1 lists the steps of our algorithm.

### 3.4.1 Complexity Analysis

In this algorithm, we do not explicitly perform the polynomial operation on adjacency matrix in \( M_g \) and \( M_I \), whose time complexity is up to \( O(N^3) \). Because, in JDGSVD, we only need to multiply \( M_g \) and \( M_I \) with some thin matrices whose size is \( N \times K \). If we change the multiplication order, and first perform multiplication between adjacency matrix and thin matrix, the time complexity of this operation will reduce to \( O(m \cdot K) \), where \( m \) is the number of non-zero elements in adjacency matrix (i.e. the number of edges in the graph) and much smaller than \( N^2 \) in sparse graph. And, the total time complexity of JDGSVD is \( O(m \cdot K^2 \cdot L) \), where \( L \) is the iteration number. We can see that the time complexity is just linear with respect to the volume of data (i.e. the number of edges), which means that it is scalable for large scale graphs.

### 3.4.2 Approximation Error

Finally, we give the error bound of our algorithm:

**Theorem 2.** Given the proximity matrix, \( S \), of a directed graph, and the embedding vectors, \( U^* \) and \( U^T \), learned by HOPPE. Then the approximation error is

\[
\| S - U^* \cdot U^\top \|_F \leq \sum_{i=K+1}^{N} \sigma_i^d
\]

, and the relative approximation error is:

\[
\frac{\| S - U^* \cdot U^\top \|_F}{\| S \|_F} = \frac{\sum_{i=K+1}^{N} \sigma_i^2}{\sum_{i=1}^{N} \sigma_i^2} \tag{22}
\]
where \( \{\sigma_i\} \) are the singular values of \( S \) in descend order.

Prove:

\[
\|S - U^* \cdot U^T\|_F^2 = \| \sum_{i=1}^N \sigma_i \cdot v_i^T \cdot v_i - \sum_{j=1}^K \sigma_j \cdot v_j^T \cdot v_j \|_F^2 \\
= \| \sum_{i=K+1}^N \sigma_i \cdot v_i^T \cdot v_i \|_F^2 \\
= \sum_{i=K+1}^N \sigma_i^2
\]

and we have:

\[
\|S\| = \| \sum_{i=1}^N \sigma_i \cdot v_i^T \cdot v_i \| = \sum_{i=1}^N \sigma_i^2
\]

So, the relative approximation error is:

\[
\frac{\|S - U^* \cdot U^T\|_F^2}{\|S\|} = \frac{\sum_{i=K+1}^N \sigma_i^2}{\sum_{i=1}^N \sigma_i^2}
\]

If \( S \) is low-rank, then the singular values \( \{\sigma_{K+1}, \cdots, \sigma_N\} \) will be close to zero and the error will be very small. That is, the lower the rank of \( S \) is, the smaller the error is.

4. EXPERIMENTS

We conduct experiments on a synthetic data and three real-world datasets. We will first introduce the experiment setting, then show the results and analysis on the two types of datasets respectively.

4.1 Experiment Setting

4.1.1 Datasets

We use four datasets, whose statistics are summarized in Table 2.

Table 2: Statistics of datasets. \(|V|\) denotes the number of vertexes and \(|E|\) denotes the number of edges.

<table>
<thead>
<tr>
<th></th>
<th>Syn</th>
<th>Cora</th>
<th>SN-Twitter</th>
<th>SN-TWeibo</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
<td>V</td>
<td>)</td>
<td>10,000</td>
<td>23166</td>
</tr>
<tr>
<td>(</td>
<td>E</td>
<td>)</td>
<td>144,555</td>
<td>91500</td>
</tr>
</tbody>
</table>

- Synthetic Data (Syn): We generate the synthetic data by the forest fire model [17]. The model can generate powerlaw graphs. We can observe asymmetric transitivity in the generated graphs. As the time complexity of computation of some high-order proximities (e.g., Katz, RPR) is too high, we generate small-scaled synthetic data to allow for the computation of high-order proximities, so that we can evaluate the accuracy of high-order proximity approximation. We randomly generate ten synthetic datasets.
- Cora\(^1\) [31]: This is a citation network of academic papers. The vertexes are academic papers and the directed edges are the citation relationship between papers.
- Twitter Social Network\(^2\) (SN-Twitter) [5]: This dataset is a subnetwork of Twitter. The vertexes are users of Twitter, and the directed edges are following relationships between users.
- Tencent Weibo Social Network\(^3\) (SN-TWeibo): This dataset contains a subnetwork of the social network in Tencent Weibo\(^4\), a Twitter-style social platform in China. The vertexes are users and the directed edges are following relationship between users.

Due to high time complexity of calculating high-order proximity measurements, we first evaluate the high-order proximity approximation error on two small datasets, i.e. Synthetic Data and Cora, where it is feasible to evaluate high-order proximity. Furthermore, we conduct other application experiments on two large datasets, i.e. SN-Twitter and SN-TWeibo, to evaluate the performance of our algorithm in real large-scale graphs. For Synthetic Data and Cora, where it is feasible to calculate proximity approximation error on two small datasets, i.e. Katz, RPR, we prove the following theorems.

Due to high time complexity of calculating high-order proximity measurements, we first evaluate the high-order proximity approximation error on two small datasets, i.e. Synthetic Data and Cora, where it is feasible to evaluate high-order proximity. Furthermore, we conduct other application experiments on two large datasets, i.e. SN-Twitter and SN-TWeibo, to evaluate the performance of our algorithm in real large-scale graphs. For Synthetic Data and Cora, where it is feasible to calculate proximity approximation error on two small datasets, i.e. Katz, RPR, we prove the following theorems.

4.1.2 Baseline Methods

- LINE [32]: This algorithm preserves the first-order and second-order proximity between vertexes. but it only can preserve symmetric second-order proximity when applied to directed graph. We use vertex vectors as source vectors and context vectors as target vectors. We use LINE1 to represent LINE preserving first-order proximity and LINE2 to represent LINE preserving second-order proximity.
- DeepWalk [24]: this algorithm first randomly walks on the graph, and assumes a pair of vertexes similar if they are close in the random path. Then, the embedding is learned to preserve these pairwise similarities in the embedding.
- PPE (Partial Proximity Embedding) [30]: This algorithm first selects a small subset of vertexes as landmarks, and learns the embedding by approximating the proximity between vertexes and landmarks.
- Common Neighbors: We rank the links by the number of common neighbors. We use it for link prediction and vertex recommendation.
- Adamic-Adar: We rank the links by the Adamic-Adar values. We use it for link prediction and vertex recommendation.

In graph reconstruction, link prediction and vertex recommendation experiments, we adopt Katz as the target high-order proximity for HOPE and PPE methods, as Katz has shown superior performance in related tasks in previous works, to approximate for HOPE and PPE. For PPE, we sample 1000 landmarks. For LINE and DeepWalk, we search the parameters grid search.

\(^1\)http://konect.uni-koblenz.de/networks/subelj_cora
\(^2\)http://konect.uni-koblenz.de/networks/munmun_twitter_social
\(^3\)http://www.kddcup2012.org/c/kddcup2012-track1/data
\(^4\)http://t.qq.com/
4.1.3 Evaluation Metrics

In the experiments, we adopt RMSE, Precision@k and MAP (Mean Average Precision) as the evaluation metrics.

RMSE is used to evaluate the approximation error of the proximity approximation algorithms, including HOPE and PPE. The formula of RMSE in our problem is:

\[
RMSE = \sqrt{\frac{\|S - U^*U^T\|^2}{N^2}}
\]

NRMSE (Normalized RMSE) is used to evaluate the relative error of the proximity approximation algorithms. The formula of NRMSE in our problem is:

\[
NRMSE = \frac{\sqrt{\|S - U^*U^T\|^2}}{\sqrt{\|S\|^2_F N^2}} = \sqrt{\frac{\|S - U^*U^T\|^2}{\|S\|^2_F}}
\]

Precision@k is used to evaluate the performance of link prediction, which measures the prediction precision of top k links. The formula of Precision@k is:

\[
Precision@k = \frac{|\{(i,j) | (i,j) \in E_p \cap E_o\}|}{|E_p|}
\]

where \(E_p\) is the set of top k predicted links, \(E_o\) is the set of observed links and \(|\cdot|\) represents the size of set.

MAP is used to evaluate the performance of vertex recommendation, which measures the rank accuracy of recommended vertex list. The formula of MAP@k is:

\[
AP@k(i) = \frac{\sum_{j=1}^{k} P_j(i) \cdot \delta_j(i)}{\sum_{j=1}^{k} \delta_j(i)}
\]

\[
MAP@k = \frac{\sum_{v \in V_G} AP(v)}{|V_G|}
\]

4.2 High-order Proximity Approximation

As we preserve asymmetric transitivity by approximating high-order proximity, the error of approximation can evaluate how well we preserve asymmetric transitivity.

We evaluate the approximation error on Synthetic Data and Cora. Besides our algorithms, PPE can also approximate high-order proximity. Here, we compare HOPE with PPE on four proximity measurements, including Katz, RPR, Common Neighbors and Adamic-Adar. Figure 3 shows the results on Synthetic Data, and second row is the results on Cora. For Katz, \(\beta = 0.1\); for RPR, \(\alpha = 0.5\).

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, \(\alpha = 0.5\).
4.3 Graph Reconstruction

As the representation of a graph, embedding vectors are expected to well reconstruct the graph. We evaluate the reconstruction ability on training sets of SN-TWeibo and SN-Twitter. We reconstruct the graph edges based on the reconstructed proximity between vertexes. For graph embedding algorithms (i.e., HOPE, PPE, Line and DeepWalk), we use the inner product between embedding vectors to reconstruct the proximity matrix. For Common Neighbors and Adamic-Adar, we directly calculate the proximity matrix. We rank the pairs of vertexes according to their corresponding reconstructed proximity. Then, we calculate the ratio of real links in top \( k \) pairs of vertexes as the reconstruction precision. As the number of possible pairs of vertexes (\( N \cdot (N - 1) \)) is too large in SN-TWeibo and SN-Twitter, we randomly sample about 0.1% pairs of vertexes for evaluation.

Figure 5 shows the Precision@\( k \) with different \( k \). HOPE significantly outperforms baselines. As mentioned above, PPE just approximates a sub-block of proximity matrix. So, it works poorer than HOPE. Comparing with Line2 which directly reconstructs the directed links, HOPE still achieves much better performance. It may because HOPE uses high-order proximity as weight of directed edges, but LINE2 assigns equivalent weight for each directed edges. The edges in densely connected vertex clusters will obtain higher weights. Thus, these edges will be preserved first. Moreover, the number of these edges is larger than that in sparsely connected vertex clusters. Thus, HOPE may reconstruct more edges than Line2.

4.4 Link Prediction

As asymmetric transitivity captures the correlation among edges, we can use it to predict missing edges in graphs. On the other hand, the performance of HOPE on link prediction can also reflect how well the asymmetric transitivity is preserved. This experiment is conducted on SN-TWeibo and SN-Twitter. We train the embedding vectors on training graphs, and evaluate the prediction performance on testing graphs. We randomly sample about 0.1% pairs of vertexes for evaluation. Then, we rank them according to the inner product between embedding vectors (for HOPE, LINE1, LINE2 and PPE) or the calculated proximity (for Common Neighbors and Adamic-Adar), and evaluate the prediction precision in top \( k \) pairs of vertexes.

Figure 6 shows the precision@\( k \) of link prediction with different \( k \). Our algorithm, HOPE, outperforms the baselines significantly. Compared to Section 4.3, the performance of PPE significantly decreases in prediction task. This is mainly because PPE only trains on partial proximity matrix, which is easy to overfit. Besides, we can see that all the curves converges to a point when \( k \) is large. This is because almost all the real edges have been correctly predicted by all the algorithms.

4.5 Vertex Recommendation

The setting of training procedure in this experiment is the same as link prediction. But, we evaluate the performance of algorithms from the vertex view, i.e. vertex recommendation performance. We randomly sample 1000 vertexes. For each vertex \( v_i \), we randomly hide 20% outgoing links as groundtruth. Then we derive the top 100 vertexes with the highest proximity with \( v_i \) as the candidates that \( v_i \) will possible point to. After that, we use MAP@10, MAP@50 and MAP@100 to evaluate the quality of recommendation. Table 3 shows the MAPs of different algorithms. HOPE outperforms the baseline algorithms. Especially, the MAP@10 of HOPE achieves at least 88.5% improvement and the MAP@50 of HOPE achieves at least 95% improvement.
Table 3: MAP of vertex recommendation on SN-TWeibo and SN-Twitter. For each vertex, the recommended vertex list is ranked according to the predicted proximity between vertexes. For embedding algorithms, we calculate the predicted proximity by performing inner product between embedding vectors.

<table>
<thead>
<tr>
<th>Method</th>
<th>SN-TWebio MAP@10</th>
<th>SN-TWebio MAP@50</th>
<th>SN-TWebio MAP@100</th>
<th>SN-Twitter MAP@10</th>
<th>SN-Twitter MAP@50</th>
<th>SN-Twitter MAP@100</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOPE</td>
<td>0.2295</td>
<td>0.1869</td>
<td>0.169</td>
<td>0.1000</td>
<td>0.0881</td>
<td>0.0766</td>
</tr>
<tr>
<td>PPE</td>
<td>0.0928</td>
<td>0.0845</td>
<td>0.077</td>
<td>0.0061</td>
<td>0.0077</td>
<td>0.0081</td>
</tr>
<tr>
<td>LINE1</td>
<td>0</td>
<td>0</td>
<td>0.005</td>
<td>0.0209</td>
<td>0.0221</td>
<td>0.0221</td>
</tr>
<tr>
<td>LINE2</td>
<td>0.051</td>
<td>0.051</td>
<td>0.048</td>
<td>0.0044</td>
<td>0.0043</td>
<td>0.0035</td>
</tr>
<tr>
<td>DeepWalk</td>
<td>0.0635</td>
<td>0.0583</td>
<td>0.004</td>
<td>0.0006</td>
<td>0.0008</td>
<td>0.001</td>
</tr>
<tr>
<td>Common Neighbors</td>
<td>0.1217</td>
<td>0.1031</td>
<td>0.155</td>
<td>0.0394</td>
<td>0.0379</td>
<td>0.0369</td>
</tr>
<tr>
<td>Adamic-Adar</td>
<td>0.1173</td>
<td>0.0990</td>
<td>0.156</td>
<td>0.0455</td>
<td>0.0442</td>
<td>0.0423</td>
</tr>
</tbody>
</table>

81.2% improvement over all baseline algorithms. In most settings, the algorithms preserving directed high-order relationship, including HOPE, Common Neighbors, Adamic-Adar and PPE, outperforms LINE2 which only preserves directed first-order relationship. This proves that preserving directed high-order relationship is critical to capture the structure of directed graphs.

5. CONCLUSION

In this paper, we aim to preserve asymmetric transitivity in directed graphs, and propose to preserve asymmetric transitivity by approximating high-order proximities. We propose a scalable approximation algorithm, called High-Order Proximity preserved Embedding (HOPE). In this algorithm, we first derive a general formulation of a class of high-order proximity measurements, then apply generalized SVD to the general formulation, whose time complexity is linear with the size of graph. The empirical study demonstrates the superiority of asymmetric transitivity and our proposed algorithm, HOPE. Our future direction is to develop a nonlinear model to better capture the complex structure of directed graphs.

6. ACKNOWLEDGEMENTS

This work was supported by National Program on Key Basic Research Project, No. 2015CB352300; National Natural Science Foundation of China, No. 61370022, No. 61531006, No. 61472444 and No. 61210008. Thanks for the research fund of Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology.

7. REFERENCES


