

Hyperbolic Heterogeneous Information Network Embedding

Xiao Wang, Yiding Zhang, Chuan Shi*

School of Computer Science
Beijing University of Posts and Telecommunications
Beijing, China
{xiaowang, zyd, shichuan}@bupt.edu.cn

Abstract

Heterogeneous information network (HIN) embedding, aiming to project HIN into a low-dimensional space, has attracted considerable research attention. Most of the existing HIN embedding methods focus on preserving the inherent network structure and semantic correlations in Euclidean spaces. However, one fundamental problem is that whether the Euclidean spaces are the appropriate or intrinsic isometric spaces of HIN? Recent researches argue that the complex network may have the hyperbolic geometry underneath, because the underlying hyperbolic geometry can naturally reflect some properties of complex network, e.g., hierarchical and power-law structure. In this paper, we make the first effort toward HIN embedding in hyperbolic spaces. We analyze the structures of two real-world HINs and discover some properties, e.g., the power-law distribution, also exist in HIN. Therefore, we propose a novel hyperbolic heterogeneous information network embedding model. Specifically, to capture the structure and semantic relations between nodes, we employ the meta-path guided random walk to sample the sequences for each node. Then we exploit the distance in hyperbolic spaces as the proximity measurement. The hyperbolic distance is able to meet the triangle inequality and well preserve the transitivity in HIN. Our model enables the nodes and their neighborhoods have small hyperbolic distances. We further derive the effective optimization strategy to update the hyperbolic embeddings iteratively. The experimental results, in comparison with the state-of-the-art, demonstrate that our proposed model not only has superior performance on network reconstruction and link prediction tasks but also shows its ability of capture hierarchy structure in HIN via visualization.

Introduction

Heterogeneous information networks (HINs) are networks that consist of multiple types of nodes and edges. Modeling data in the real world as HINs can capture rich data semantics. For example, the bibliographic network can be modeled as an HIN with three types of nodes: author, paper, and venue (Fu, Lee, and Lei 2017). Further, the relations between nodes, such as author-paper (write), paper-venue (publish), have different types of edges. Recently, HIN embedding, aims at learning the node representations in the

low-dimensional spaces while preserving the structure and semantic information, has received increasing research attention. Benefited from HIN embedding, a variety of HIN based applications, e.g, recommendation (Shi et al. 2018; Hu et al. 2018) and link prediction (Chen et al. 2018), can be conducted and improved in the low-dimensional space.

So far, a number of HIN embedding methods have been proposed. Loosely speaking, there are the works based on random walk (Dong, Chawla, and Swami 2017; Fu, Lee, and Lei 2017), network partition based methods (Tang, Qu, and Mei 2015; Xu et al. 2017), and deep neural network based HIN embedding (Chang et al. 2015; Wang et al. 2018). In essence, because the structure and semantics are the two most important information in an HIN, most of them focus on how to effectively preserve the HIN structures and semantics in the low-dimensional spaces. However, another fundamental problem is that what are the appropriate or intrinsic underlying isometric spaces of HIN? Because the Euclidean spaces are the natural generalization of our intuition-friendly, and visual three-dimensional space, they have become the primary choice for the current HIN embedding methods. While more and more researches demonstrate that many types of complex data, e.g., social network, actually exhibit a highly non-Euclidean latent anatomy (Bronstein et al. 2017). This motivates us to rethink that whether the current chosen low-dimensional spaces for HIN embedding, i.e., Euclidean spaces, are optimal and are there other feasible non-Euclidean spaces?

Recently, hyperbolic spaces have gained momentum in the context of network science. Hyperbolic spaces are spaces of constant negative curvature (Cannon et al. 1997). A superiority of hyperbolic spaces is that they expand faster than Euclidean spaces (Nickel and Kiela 2017). For instance, considering a circle and a disk in a 2-dimensional hyperbolic space with constant curvature $K = -1$, the length of the circle and the area of the disk of hyperbolic radius r are given as $2\pi \sinh r$ and $2\pi(\cosh r - 1)$, respectively, and both of them grow as e^r with r . In a 2-dimensional Euclidean space, the length of a circle and the area of a disk of Euclidean radius r are given as $2\pi r$ and πr^2 , growing only linearly and quadratically with regard to r . For this reason, in hyperbolic spaces, it is easy to model complex data with low-dimensional embedding. Due to the characteristic of hyperbolic spaces, (Krioukov et al. 2010) assumes hyper-

*Corresponding author, Chuan Shi, shichuan@bupt.edu.cn.
Copyright © 2019, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

bolic spaces underlies complex network and find that data with power-law structure is suitable to be modeled in hyperbolic spaces. Because of these properties of hyperbolic spaces, some works begin to study the hyperbolic embeddings of different data. For instance, (Dhingra et al. 2018) embeds text in hyperbolic spaces. (Nickel and Kiela 2017) and (Ganea, Becigneul, and Hofmann 2018) learn the embeddings of homogeneous networks.

In this paper, we propose a novel hyperbolic heterogeneous information network embedding model (HHNE) which is able to preserve the structure and semantic information in hyperbolic spaces. We leverage the meta-path guided random walk to generate heterogeneous neighborhoods for each node to capture the structure and semantic relations in HIN. Then the proximity between nodes is measured by the distance in hyperbolic spaces. As the distance is defined in metric spaces, the proximity between nodes meets the triangle inequality and can well preserve the transitivity in HIN. Our model is able to maximize the proximity between the neighborhood nodes while minimizing the proximity between the negative sampled nodes. We further derive the effective optimization strategy to optimize hyperbolic embeddings iteratively.

We highlight the main contributions as follows:

- To our best knowledge, we are the first to study the problem of HIN embedding in hyperbolic spaces.
- We propose a novel HIN embedding model, named HHNE, to preserve the HIN structure and semantic correlations in hyperbolic spaces based on the meta-path guided random walk.
- We conduct extensive experiments to evaluate the performance of HHNE in terms of representation capacity and generalization ability on two real-world datasets. The results show the superiority of HHNE by comparing with the state-of-the-art techniques.

Related Work

Network Embedding

Recently, a significant amount of progress has been made toward network embedding (Cui et al. 2018). For instance, inspired by language modeling technique, DeepWalk (Perozzi, Al-Rfou, and Skiena 2014) regards the node sequences generated from random walks as “sentences”, and nodes as “words”, and then maximizes the co-occurrence probability among nodes. LINE (Tang et al. 2015) is able to efficiently learn the node embeddings while preserving both of the first-order and second-order structures. Node2vec (Grover and Leskovec 2016) is generalized from DeepWalk. It designs a parameterized random walk procedure to learn a mapping of nodes that maximizes the likelihood of preserving network neighborhoods of nodes. SDNE (Wang, Cui, and Zhu 2016) uses autoencoder to capture local and global network structures to learn user representation. Most of network embedding methods embed network into low-dimensional Euclidean spaces, while some researchers begin to embed network into low-dimensional hyperbolic spaces. (Nickel and Kiela 2017) embeds network into hyperbolic spaces to learn

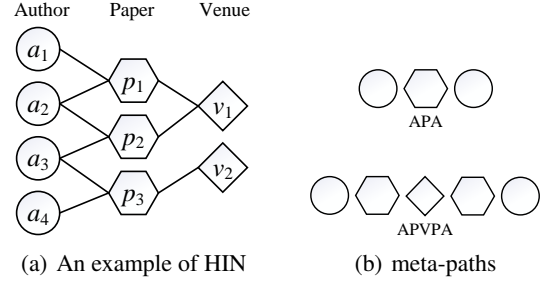


Figure 1: An illustrative example of a heterogeneous information network (DBLP) and meta-paths.

hierarchical feature representations of network. (Ganea, Becigneul, and Hofmann 2018) embeds the directed acyclic graphs into hyperbolic spaces to learn their hierarchical feature representations. However, these methods only focus on learning the representation of nodes in homogeneous networks and do not consider the heterogeneity of complex information network.

Heterogeneous Information Network Embedding

Recently, some methods have been proposed representation learning methods for HIN. Metapath2vec (Dong, Chawla, and Swami 2017) formalizes meta-paths based random walks to obtain heterogeneous neighborhoods of a node and leverages Skip-gram model to learn the network structure. HIN2vec (Fu, Lee, and Lei 2017) carries out multiple prediction tasks jointly to learn the presentation of nodes and meta-paths. PTE (Tang, Qu, and Mei 2015) partitions an HIN into a set of edgewise bipartite networks and then learn the feature representation jointly by using LINE. EOE (Xu et al. 2017) aims to embed the coupled HIN, which consists of two different but related homogeneous networks. It models each homogeneous network by using the same function as LINE. HNE (Chang et al. 2015) transfers different objects in HIN to unified feature representations and considers contents and topological structures in networks jointly for creating the embedding. SHINE (Wang et al. 2018) utilizes multiple deep autoencoders to extract users highly nonlinear representations while preserving the structure of original networks.

To sum up, all the above mentioned HIN embedding models project HIN into low-dimensional Euclidean spaces. Nevertheless, whether the Euclidean space is the most appropriate one is still an open question.

Preliminaries

Definitions in HIN

Definition 1. Heterogeneous Information Network (HIN) (Shi et al. 2017). An HIN is defined as a graph $G = (V, E, T, \phi, \psi)$, in which V and E are the sets of nodes and edges. Each node $v \in V$ and each edge $e \in E$ are associated with their mapping functions $\phi(v) : V \rightarrow T_V$ and $\psi(e) : E \rightarrow T_E$ respectively. T_V and T_E denote the sets of node and relation types, where $|T_V| + |T_E| > 2$.

Table 1: Statistics of datasets

DBLP	# A	# P	# V	# P-A	# P-V
	14475	14376	20	41794	14376
MovieLens	# A	# M	# D	# M-A	# M-D
	11718	9160	3510	64051	9160

Definition 2. Meta-path. Given an HIN $G = (V, E, T, \phi, \psi)$, a meta-path \mathcal{P} is a sequence of node types $t_{v_1}, t_{v_2}, \dots, t_{v_n}$ connected by edge types $t_{e_1}, t_{e_2}, \dots, t_{e_{n-1}}$: $\mathcal{P} = t_{v_1} \xrightarrow{t_{e_1}} t_{v_2} \xrightarrow{t_{e_2}} \dots \xrightarrow{t_{e_{n-1}}} t_{v_n}$. A meta-path instance consists of specific nodes and edges, e.g., $a_1 \xrightarrow{\text{write}} p_1 \xrightarrow{\text{publish}} v_1$.

For example, Figure 1(a) shows an academic network with authors (A), papers (P), venues (V) as nodes, wherein edges indicate the *write* (A-P), *publish* (P-V). The examples of meta-paths are shown in Figure 1(b). The meta-paths contain rich semantic information, such as *coauthor* (A-P-A), *authors attending the same venue* (A-P-V-P-A).

Definition 3. Heterogeneous Information Network Embedding. Given an HIN $G = (V, E, T, \phi, \psi)$, HIN embedding aims to project nodes into a latent low-dimensional representation space while preserving the original network structure and semantic correlations.

Relation Distribution in HIN

In order to find out whether there is some underlying correlations between HIN and hyperbolic spaces, we analyze some meta-paths in two real-world HINs. The basic statistics of the two HINs are shown in Table 1.

- **DBLP** is a bibliographic dataset in computer science. We use a subset of DBLP, i.e., DBLP-4-Area taken of (Sun et al. 2011) which contains three types of nodes, i.e., author (A), paper (P), venue (V) and two types of edges, i.e., author-paper (write), paper-venue (publish). The schema of DBLP network is shown in Figure 2(a).
- **MovieLens**¹ contains knowledge about movies (Cantador, Brusilovsky, and Kuflik 2011). We extract a subset of from MovieLens, which contains three types of nodes, i.e., actor (A), movie (M), director (D) and two types of edges, i.e., actor-movie (act in), director-movie (direct). The schema of MovieLens is shown in Figure 2(b).

As mentioned before, data with power-law structure can be naturally modeled in hyperbolic spaces. Therefore, we use two real-world HINs to check whether the power-law distribution of nodes also exists with different meta-paths. We calculate the distribution of nodes as follows: given a meta-path \mathcal{P} and a node v , we first calculate how many meta-path instances can be composed started with v following \mathcal{P} , and then calculate how many nodes have the same results. The two results are plotted as the horizontal axis and vertical axis, respectively. For DBLP dataset, we show the distribution of author-paper-author (A-P-A) relation and author-paper-venue (A-P-V) relation in Figure 3(a) and Figure 3(b),

¹<https://grouplens.org/datasets/hetrec-2011/>

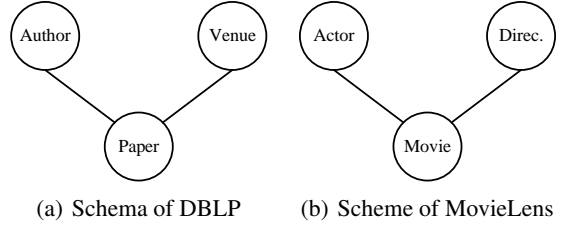


Figure 2: Schemas of two HINs.

respectively. For MovieLens dataset, we show the distribution of actor-movie-director (A-M-D) relation in Figure 3(c). We can see these distribution presents the power-law distribution. This fact implies that hyperbolic spaces may be the alternative spaces for embedding HINs. Please note that there exist a lot of meta-paths in an HIN, it makes sense that the nodes with some specific meta-paths may not always have power-law distribution, but as can be seen in the following experiments, the results are still very competitive. More fine-grained analysis on the meta-paths will leave for the future work.

Embedding HIN in Hyperbolic Spaces

Hyperbolic Geometry and Embedding

Hyperbolic geometry is a non-Euclidean geometry which is obtained by replacing Euclid's fifth geometric postulate (the parallel postulate). Hyperbolic geometry studies the spaces of constant negative curvature. A key property of hyperbolic spaces \mathbb{H} is that they expand faster than Euclidean spaces because Euclidean spaces \mathbb{R} expand polynomially while hyperbolic spaces \mathbb{H} expand exponentially. Specifically, each tile in Figure 4(a) (Tay, Tuan, and Hui 2018) is of the equal area in hyperbolic space but diminishes towards zero in Euclidean space towards the boundary. Because of this property, hyperbolic spaces can be thought of as a continuous version of trees. Specifically, as shown in Figure 4(b) (Nickel and Kiela 2017), consider a tree with branching factor b , the number of nodes at level l or no more than l hops from the root are $(b+1)b^{l-1}$ and $[(b+1)b^l - 2]/(b-1)$ respectively. The number of nodes grows exponentially with their distance to the root of the tree, which is similar to hyperbolic spaces as they expand exponentially. In hyperbolic spaces, the data with tree structure can be embedded into 2-dimensional hyperbolic spaces naturally. Given a node at level l , the node can be placed on a sphere in hyperbolic spaces with distance $d^{\mathbb{H}} \propto l$ to the origin of the sphere, and the branching factor b can be modeled by the constant curvature of hyperbolic spaces as $K = -\ln^2 b$. As mentioned above, the number of nodes of a tree grows exponentially with their distance to the root, and the nodes distribution of a tree follow the power-law distribution. Hence, the power-law distribution can naturally emerge as a direct consequence of the basic properties of hyperbolic geometry underlying the network. The data with power-law distribution is suitable to be modeled in hyperbolic spaces (Krioukov et al. 2010).

Hyperbolic spaces cannot be isometrically embedded into

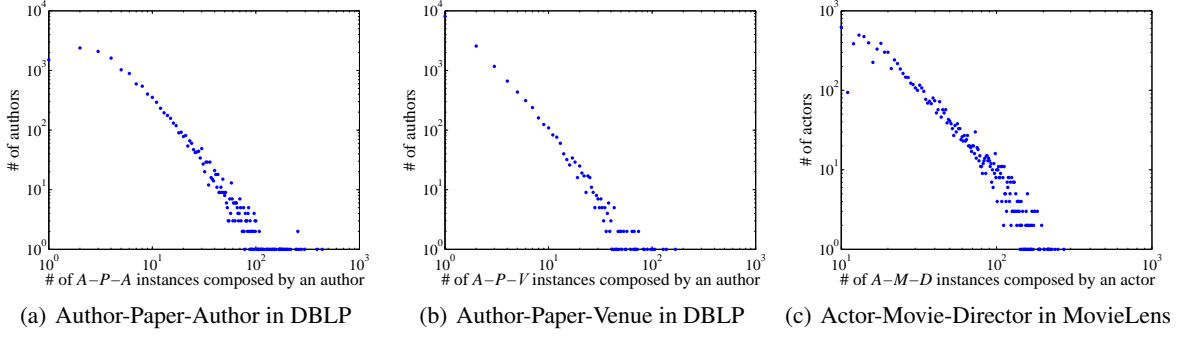


Figure 3: The distribution of some relations in two datasets.

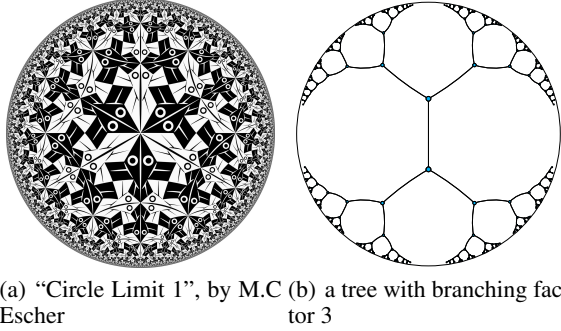


Figure 4: Two examples of Poincaré disk model of hyperbolic spaces.

Euclidean spaces, so they are difficult to envisage (Krioukov et al. 2010). Fortunately, there exist many equivalent models of hyperbolic spaces such as Poincaré disk (ball) model, Poincaré half-plane model, Beltrami-Klein model, hyperboloid model (Cannon et al. 1997). As Poincaré ball model is well-suited for gradient-based optimization, our work makes use of Poincaré ball model. Let $\mathbb{D}^d = \{x \in \mathbb{R}^d : \|x\| < 1\}$ be the *open* d -dimensional unit ball. The Poincaré ball model is defined by the manifold \mathbb{D}^d equipped with the following Riemannian metric tensor $g_x^{\mathbb{D}}$:

$$g_x^{\mathbb{D}} = \lambda_x^2 g^{\mathbb{E}} \quad \text{where } \lambda_x := \frac{2}{1 - \|x\|^2}, \quad (1)$$

where $x \in \mathbb{D}^d$, $g^{\mathbb{E}} = \mathbf{I}$ denotes the Euclidean metric tensor. It is worth noting that the Poincaré model is conformal, meaning that Euclidean angles between hyperbolic lines in the model are equal to their hyperbolic angles that makes it suitable for gradient-based method.

The HHNE Model

We aim to learn the representation of nodes to preserve the structure and semantic correlations in hyperbolic spaces. Given an HIN $G = (V, E, T, \phi, \psi)$ with $|T_V| > 1$, we are interested in learning the embeddings $\Theta = \{\theta_i\}_{i=1}^{|V|}$, $\theta_i \in \mathbb{D}^d$. We preserve the structure by facilitating the proximity between the node $v \in V$ and its neighborhoods $c_t \in C_t(v)$ with type t . We use meta-path guided random walks (Dong,

Chawla, and Swami 2017) to obtain heterogeneous neighborhoods of a node. In meta-path guided random walks, the nodes sequences are restrained by the node types which are defined by meta-paths. Specifically, given a meta-path

$\mathcal{P} = t_{v_1} \xrightarrow{t_{e_1}} t_{v_2} \xrightarrow{t_{e_2}} \dots \xrightarrow{t_{e_{n-1}}} t_{v_n}$, the transition probability at step i is defined as follows:

$$p(v^{i+1} | v_{t_{v_i}}^i, \mathcal{P}) = \begin{cases} \frac{1}{|N_{t_{v_{i+1}}}(v_{t_{v_i}}^i)|} & (v^{i+1}, v_{t_{v_i}}^i) \in E, \phi(v^{i+1}) = t_{v_{i+1}} \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where $v_{t_{v_i}}^i$ is node $v \in V$ with type t_{v_i} , and $N_{t_{v_{i+1}}}(v_{t_{v_i}}^i)$ denotes the $t_{v_{i+1}}$ type of neighborhood of node $v_{t_{v_i}}^i$.

The meta-path guided random walk strategy ensures that the semantic relationships between different types of nodes can be properly incorporated into HHNE.

In order to preserve the proximity between nodes and its neighborhoods in hyperbolic spaces, we use distances in Poincaré ball model to measure their proximity. Given nodes embeddings $\theta_i, \theta_j \in \mathbb{D}^d$, the distance in Poincaré ball is given by:

$$d_{\mathbb{D}}(\theta_i, \theta_j) = \cosh^{-1} \left(1 + 2 \frac{\|\theta_i - \theta_j\|^2}{(1 - \|\theta_i\|^2)(1 - \|\theta_j\|^2)} \right). \quad (3)$$

It is worth noting that as the Poincaré ball model is defined in metric spaces, the distance in Poincaré ball meet the triangle inequality and can well preserve the transitivity in HIN. Then, we use a probability to measure the node c_t is a neighborhood of node v as following:

$$p(v | c_t; \Theta) = \sigma[-d_{\mathbb{D}}(\theta_v, \theta_{c_t})],$$

where $\sigma(x) = \frac{1}{1 + \exp(-x)}$. Then the object of our model is to maximize the probability as following:

$$\arg \max_{\Theta} \sum_{v \in V} \sum_{c_t \in C_t(v)} \log p(v | c_t; \Theta). \quad (4)$$

To achieve efficient optimization, we leverage the negative sampling proposed in (Mikolov et al. 2013), which basically samples a small number of negative objects to enhance the influence of positive objects. In HHNE, for a given node v , we want to maximize the proximity between v and its neighborhood c_t while minimize the proximity between v

and its negative sampled node n . Therefore, the objective function Eq. 4 can be rewritten as following:

$$\begin{aligned} \mathcal{L}(\Theta) = & \log \sigma[-d_{\mathbb{D}}(\theta_{c_t}, \theta_v)] \\ & + \sum_{m=1}^M \mathbb{E}_{n^m \sim P(n)} \{\log \sigma[d_{\mathbb{D}}(\theta_{n^m}, \theta_v)]\}, \end{aligned} \quad (5)$$

where $P(n)$ is the pre-defined distribution from which a negative node n^m is drew from for M times. Our method builds the node frequency distribution by draw nodes regardless of their types.

Optimization

As the parameters of the model live in a Poincaré ball which has a Riemannian manifold structure, the back-propagated gradient is a Riemannian gradient. It means that the Euclidean gradient based optimization, such as $\theta_i \leftarrow \theta_i + \eta \nabla_{\theta_i}^E \mathcal{L}(\Theta)$, makes no sense as an operation in the Poincaré ball, because the addition operation is not defined in this manifold. Instead, we can optimize Eq. 5 via a Riemannian stochastic gradient descent (RSGD) optimization method (Bonnabel and others 2013). In particular, let $\mathcal{T}_{\theta_i} \mathbb{D}^d$ denote the tangent space of a node embedding $\theta_i \in \mathbb{D}^d$, and we can compute the Riemannian gradient $\nabla_{\theta_i}^R \mathcal{L}(\Theta) \in \mathcal{T}_{\theta_i} \mathbb{D}^d$ of $\mathcal{L}(\Theta)$. Using RSGD, parameter updates to maximize Eq. 5 are in the form:

$$\theta_i \leftarrow \exp_{\theta_i}(\eta \nabla_{\theta_i}^R \mathcal{L}(\Theta)), \quad (6)$$

where $\exp_{\theta_i}(\cdot)$ is exponential map in the Poincaré ball. In Euclidean spaces, the exponential map is familiar with us: given a node embedding $\theta_i \in \mathbb{D}^d$, $s = \eta \nabla_{\theta_i}^R \mathcal{L}(\Theta)$, the exponential map is given by $\exp_{\theta_i}(s) = \theta_i + s$. In our model, the exponential map is given by (Ganea, Becigneul, and Hofmann 2018):

$$\begin{aligned} \exp_{\theta_i}(s) = & \frac{\lambda_{\theta_i} \left(\cosh(\lambda_{\theta_i} \|s\|) + \langle \theta_i, \frac{s}{\|s\|} \rangle \sinh(\lambda_{\theta_i} \|s\|) \right)}{1 + (\lambda_{\theta_i} - 1) \cosh(\lambda_{\theta_i} \|s\|) + \lambda_{\theta_i} \langle \theta_i, \frac{s}{\|s\|} \rangle \sinh(\lambda_{\theta_i} \|s\|)} \theta_i \\ & + \frac{\frac{1}{\|s\|} \sinh(\lambda_{\theta_i} \|s\|)}{1 + (\lambda_{\theta_i} - 1) \cosh(\lambda_{\theta_i} \|s\|) + \lambda_{\theta_i} \langle \theta_i, \frac{s}{\|s\|} \rangle \sinh(\lambda_{\theta_i} \|s\|)} s. \end{aligned} \quad (7)$$

As the Poincaré ball model is a conformal model of hyperbolic spaces, i.e., $g_x^{\mathbb{D}} = \lambda_x^2 g_x^{\mathbb{E}}$, the Riemannian gradient ∇^R is obtained by rescaling the Euclidean gradient ∇^E by the inverse of the metric tensor, i.e., $\frac{1}{g_x^{\mathbb{D}}}$:

$$\nabla_{\theta_i}^R \mathcal{L} = \left(\frac{1}{\lambda_{\theta_i}} \right)^2 \nabla_{\theta_i}^E \mathcal{L}. \quad (8)$$

Furthermore, the gradients of Eq. 5 can be derived as follows:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \theta_{c_t}} = & \frac{4}{\alpha \sqrt{\gamma^2 - 1}} [\mathbb{I}_v[u^m] - \sigma(-d_{\mathbb{D}}(\theta_{c_t}, \theta_{u^m}))] \\ & \cdot \left[\frac{\theta_{c_t}}{\beta_m} - \frac{\|\theta_{c_t}\|^2 - 2\langle \theta_{c_t}, \theta_{u^m} \rangle + 1}{\beta_m^2} \theta_{u^m} \right], \end{aligned} \quad (9)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \theta_{c_t}} = & \sum_{m=0}^M \frac{4}{\beta_m \sqrt{\gamma^2 - 1}} [\mathbb{I}_v[u^m] - \sigma(-d_{\mathbb{D}}(\theta_{c_t}, \theta_{u^m}))] \\ & \cdot \left[\frac{\theta_{u^m}}{\alpha} - \frac{\|\theta_{u^m}\|^2 - 2\langle \theta_{c_t}, \theta_{u^m} \rangle + 1}{\alpha^2} \theta_{c_t} \right], \end{aligned} \quad (10)$$

where $\alpha = 1 - \|\theta_{c_t}\|^2$, $\beta_m = 1 - \|\theta_{u^m}\|^2$, $\gamma = 1 + \frac{2}{\alpha\beta} \|\theta_{c_t} - \theta_{u^m}\|^2$ and when $m = 0$, $u^0 = v$. $\mathbb{I}_v[u]$ is an indicator function to indicate whether u is v . Then, our model can be update by using Eq. 9-10 iteratively.

Complexity Analysis

The overall complexity of HHNE is $O(\tau \cdot l \cdot k \cdot n \cdot d \cdot |V|)$, where τ is the number of random walks, l is the length of random walks, k is the neighborhood size, n is the number of negative sampling, d is the embedding dimension, $|V|$ is the number of nodes in the network.

Experiments

Experiments Setup

Datasets The datasets have been introduced in the Preliminaries.

Baselines We compare our method with the following state-of-the-art methods: (1) the homogeneous network embedding methods, i.e., DeepWalk (Perozzi, Al-Rfou, and Skiena 2014), LINE (Tang et al. 2015) and node2vec (Grover and Leskovec 2016); (2) the heterogeneous information network embedding methods, i.e., metapath2vec (Dong, Chawla, and Swami 2017); (3) Moreover, the hyperbolic homogeneous network embedding methods, i.e., PoincaréEmb (Nickel and Kiela 2017).

Parameter Settings For random walk based methods DeepWalk, node2vec, metapath2vec and HHNE, we set neighborhood size as 5, walk length as 80, walks per node as 40. For LINE, metapath2vec, PoincaréEmb and HHNE, we set the number of negative samples as 10. For methods based on meta-path guided random walks, we use ‘‘APA’’ for relation ‘‘P-A’’ in network reconstruction and link prediction experiments in DBLP; ‘‘APVPA’’ for relation ‘‘P-V’’ in above experiments in DBLP; ‘‘AMDMA’’ for all relation in above experiments in MovieLens. In visualization experiment, in order to focus on analyzing the relation of ‘‘A’’ and ‘‘P’’, we use ‘‘APA’’.

Experimental Results

Network Reconstruction A good HIN embedding method should ensure that the learned embeddings can preserve the original HIN structure. The reconstruction error in relation to the embedding dimension is then a measure for the capacity of the model. More specifically, we use network embedding methods to learn feature representations. Then for each type of links in the HIN, we enumerate all pairs of objects that can be connected by such a link and calculate their proximity (Huang and Mamoulis 2017), i.e., the distance in Poincaré ball model for HHNE

Table 2: AUC scores for network reconstruction

Dataset	Edge	Dimension	Deepwalk	LINE(1st)	LINE(2nd)	node2vec	metapath2vec	PoincaréEmb	HHNE	
DBLP	P-A	2	0.6933	0.5286	0.6740	0.7107	0.6686	0.8251	0.9835	
		5	0.8034	0.5397	0.7379	0.8162	0.8261	0.8769	0.9838	
		10	0.9324	0.6740	0.7541	0.9418	0.9202	0.8921	0.9887	
		15	0.9666	0.7220	0.7868	0.9719	0.9500	0.8989	0.9898	
		20	0.9722	0.7457	0.7600	0.9809	0.9623	0.9024	0.9913	
		25	0.9794	0.7668	0.7621	0.9881	0.9690	0.9034	0.9930	
	P-V	2	0.7324	0.5182	0.6242	0.7595	0.7286	0.5718	0.8449	
		5	0.7906	0.5500	0.6349	0.8019	0.9072	0.5529	0.9984	
		10	0.8813	0.7070	0.6333	0.8922	0.9691	0.6271	0.9985	
		15	0.9353	0.7295	0.6343	0.9382	0.9840	0.6446	0.9985	
		20	0.9505	0.7369	0.6444	0.9524	0.9879	0.6600	0.9985	
		25	0.9558	0.7436	0.6440	0.9596	0.9899	0.6760	0.9985	
	MoiveLens	M-A	2	0.6320	0.5424	0.6378	0.6402	0.6404	0.5231	0.8832
			5	0.6763	0.5675	0.7047	0.6774	0.6578	0.5317	0.9168
			10	0.7610	0.6202	0.7739	0.7653	0.7231	0.5404	0.9211
15			0.8244	0.6593	0.7955	0.8304	0.7793	0.5479	0.9221	
20			0.8666	0.6925	0.8065	0.8742	0.8189	0.5522	0.9239	
M-D		25	0.8963	0.7251	0.8123	0.9035	0.8483	0.5545	0.9233	
		2	0.6626	0.5386	0.6016	0.6707	0.6589	0.6213	0.9952	
		5	0.7263	0.5839	0.6521	0.7283	0.7230	0.7266	0.9968	
		10	0.8246	0.6114	0.6969	0.8308	0.8063	0.7397	0.9975	
		15	0.8784	0.6421	0.7112	0.8867	0.8455	0.7378	0.9972	
	20	0.9117	0.6748	0.7503	0.9186	0.8656	0.7423	0.9982		
	25	0.9345	0.7012	0.7642	0.9402	0.8800	0.7437	0.9992		

and PoincaréEmb. Finally, we use the AUC (Fawcett 2006) to evaluate the performance of each embedding method. For example, for link type “write”, we calculate all pairs of authors and papers in DBLP and compute the proximity for each pair. Then using the links between authors and papers in real DBLP network as ground-truth, we compute the AUC value for each embedding method.

The results are shown in Table 2. As we can see, HHNE consistently performs the best in all the tested HINs. The results demonstrate that our proposed method can effectively preserve the original network structure and reconstruct the network, especially on the reconstruction of P-V and M-D edges. Also, please note that our method can achieve very promising results when the embedding dimension is very small. This suggests that regarding hyperbolic spaces underlying HIN is reasonable and hyperbolic spaces have strong ability of modeling network when the dimension of spaces is small.

Link Prediction Link prediction aims to infer the unknown links in an HIN given the observed HIN structure, which can be used to test the generalization performance of a network embedding method. We set our experiments similar to (Xu et al. 2017). For each type of edge, we remove 20% of edges randomly from the network while ensuring that the rest network structure is still connected. To generate negative samples, we randomly sample an equal number of node pairs from the network which have no edge connecting them. We split the chosen edges and negative samples into validation and test. In our experiments, we train the embeddings on the residual network, and use the validation data to tune the model parameters. We calculate proximity of all pair of nodes in the test. We still use AUC as the evaluation metric.

From the results in Table 3, HHNE outperforms the baselines upon all the dimensionality, especially in the low di-

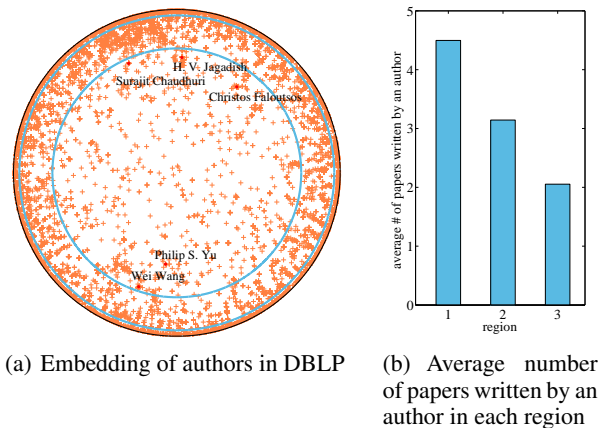


Figure 5: Visualization analysis of authors in DBLP dataset.

dimensionality. The results can demonstrate the generalization ability of HHNE. In DBLP dataset, the results of HHNE in 10 dimensionality exceed all the baselines in higher dimensionality results. In MovieLens dataset, HHNE with only 2 dimensionality surpasses baselines in all dimensionality. Besides, both of LINE(1st) and PoincaréEmb preserve proximities of node pairs linked by an edge, while LINE(1st) embed network into Euclidean spaces and Poincaré embed network into Hyperbolic spaces. PoincaréEmb performs better than LINE(1st) in most cases, especially in dimensionality lower than 10, suggesting the superiority of embedding network into hyperbolic spaces. But because HHNE can preserve high-order network structure and handle different types of nodes in HIN, HHNE is more effective than PoincaréEmb.

Visualization To evaluate whether our embeddings can reflect the latent hierarchy in HIN, we show the visualization in the two-dimensional embedding of authors in DBLP

Table 3: AUC scores for link prediction

Dataset	Edge	Dimension	Deepwalk	LINE(1st)	LINE(2nd)	node2vec	metapath2vec	PoincaréEmb	HHNE
DBLP	P-A	2	0.5813	0.5090	0.5909	0.6709	0.6536	0.6742	0.8777
		5	0.7370	0.5168	0.6351	0.7527	0.7294	0.7381	0.9041
		10	0.8250	0.5427	0.6510	0.8469	0.8279	0.7699	0.9111
		15	0.8664	0.5631	0.6582	0.8881	0.8606	0.7743	0.9111
		20	0.8807	0.5742	0.6644	0.9037	0.8740	0.7806	0.9106
		25	0.8878	0.5857	0.6782	0.9102	0.8803	0.7830	0.9117
	P-V	2	0.7075	0.5160	0.5121	0.7369	0.7059	0.8257	0.9331
		5	0.7197	0.5663	0.5216	0.7286	0.8516	0.8878	0.9409
		10	0.7292	0.5873	0.5332	0.7481	0.9248	0.9113	0.9619
		15	0.7325	0.5896	0.5425	0.7583	0.9414	0.9142	0.9625
		20	0.7522	0.5891	0.5492	0.7674	0.9504	0.9185	0.9620
MoiveLens	M-A	25	0.7640	0.5846	0.5512	0.7758	0.9536	0.9192	0.9612
		2	0.6278	0.5053	0.5712	0.6349	0.6168	0.5535	0.7715
		5	0.6353	0.5636	0.5874	0.6402	0.6212	0.5779	0.8255
		10	0.6680	0.5914	0.6361	0.6700	0.6332	0.5984	0.8312
		15	0.6791	0.6184	0.6442	0.6814	0.6382	0.5916	0.8319
		20	0.6868	0.6202	0.6596	0.6910	0.6453	0.5988	0.8318
		25	0.6890	0.6256	0.6700	0.6977	0.6508	0.5995	0.8309
	M-D	2	0.6258	0.5139	0.6501	0.6299	0.6191	0.5856	0.8520
		5	0.6482	0.5496	0.6607	0.6589	0.6332	0.6290	0.8967
		10	0.6976	0.5885	0.7499	0.7034	0.6687	0.6518	0.8984
		15	0.7163	0.6647	0.7756	0.7241	0.6702	0.6715	0.9007
		20	0.7324	0.6742	0.7982	0.7412	0.6746	0.6821	0.9000
		25	0.7446	0.6957	0.8051	0.7523	0.6712	0.6864	0.9018

dataset in Figure 5(a). Each dot in Figure 5(a) represents an author’s position in two-dimensional Poincaré ball (disk). The distances between nodes and origin of the disk can reflect the latent hierarchy of authors. In general, authors have high latent hierarchy if their distances to the origin are small. The distances between nodes and origin are defined as norms in Poincaré disk. The norm can be computed by:

$$\|x\|_{\mathbb{D}} := d_{\mathbb{D}}(0, x).$$

We split the authors in Poincaré disk into 3 regions by their norms. As the maximum norm of authors is slightly smaller than 6 in Figure 5(a), we split the authors who have norm between 0 and 2 into region 1, norm between 2 and 4 into region 2, norm between 4 and 6 into region 3. Therefore, the authors in region 1 have the highest hierarchy while the authors in region 3 have the lowest hierarchy. We mark the two boundaries of three regions by blue lines in Figure 5(a). In order to evaluate the influence of authors quantitatively, we calculate the average number of papers written by an author in each region. The results are shown in Figure 5(b). We can see that the average number of papers written by an author in a region decreases from region 1 to region 3. Moreover, we find that some high-impact researchers in computer science (i.e., Christos Faloutsos in Carnegie Mellon University, H. V. Jagadish in University of Michigan, Philip S. Yu in University of Illinois at Chicago, Surajit Chaudhuri in Microsoft Research, Wei Wang in University of California, Los Angeles), marked by red dot, are also in the region 1. These results demonstrate that the learned hyperbolic embeddings can reflect the latent hierarchy in HIN.

Parameters Sensitivity Due to the page limitation, we take the DBLP dataset as an example to conduct the parameter sensitivity analysis. In all parameters sensitivity experiments, we set dimensional as 5. Except for the parameter being tested, all other parameters are set as default. As can

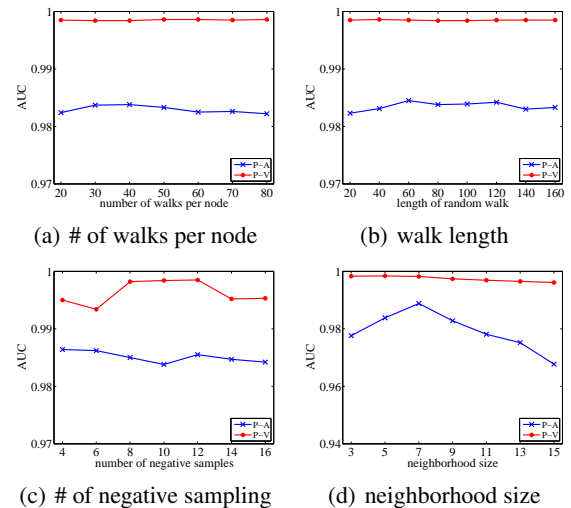


Figure 6: Parameter sensitivity in network reconstruction

be seen in Figure 6(a) and 6(b), the performance is relatively stable with respect to the number of walks per node and the walk length. Figure 6(c) suggests the negative size is relevant to the reconstruction. An appropriate negative size can make HHNE achieve better results. Figure 6(d) shows too large neighborhood size could make the results descent, so a smaller neighborhood size actually produces better embeddings for network reconstruction.

Conclusion

In this work, we study the problem of embedding HIN in hyperbolic spaces. We propose the HHNE method, which aims to maximize proximity in consideration of multiple types

of neighborhoods for a given node. We exploit the distance in hyperbolic spaces as the proximity measurements, which meets the triangle inequality and can well preserve the transitivity in HIN. To update the hyperbolic embeddings, we use a stochastic Riemannian optimization method. Extensive experiments demonstrate that HHNE is outperformed than some state-of-the-art network embedding methods, especially the dimension of embedding spaces is small and verify HHNE can discover the latent hierarchy in HIN.

Acknowledgement

This work is supported in part by the National Natural Science Foundation of China (No. 61772082, 61702296, 61806020, 61375058), the National Key Research and Development Program of China (2017YFB0803304), the Beijing Municipal Natural Science Foundation (4182043), and the CCF-Tencent Open Fund.

References

- Bonnabel, S., et al. 2013. Stochastic gradient descent on riemannian manifolds. *IEEE Trans. Automat. Contr.* 58(9):2217–2229.
- Bronstein, M. M.; Bruna, J.; LeCun, Y.; Szlam, A.; and Vandergheynst, P. 2017. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine* 34(4):18–42.
- Cannon, J. W.; Floyd, W. J.; Kenyon, R.; Parry, W. R.; et al. 1997. Hyperbolic geometry. *Flavors of geometry* 31:59–115.
- Cantador, I.; Brusilovsky, P.; and Kuflik, T. 2011. 2nd workshop on information heterogeneity and fusion in recommender systems (hetrec 2011). In *RecSys*.
- Chang, S.; Han, W.; Tang, J.; Qi, G.-J.; Aggarwal, C. C.; and Huang, T. S. 2015. Heterogeneous network embedding via deep architectures. In *SIGKDD*, 119–128.
- Chen, H.; Yin, H.; Wang, W.; Wang, H.; Nguyen, Q. V. H.; and Li, X. 2018. Pme: Projected metric embedding on heterogeneous networks for link prediction. In *SIGKDD*, 1177–1186.
- Cui, P.; Wang, X.; Pei, J.; and Zhu, W. 2018. A survey on network embedding. *IEEE Transactions on Knowledge and Data Engineering*.
- Dhingra, B.; Shallue, C.; Norouzi, M.; Dai, A.; and Dahl, G. 2018. Embedding text in hyperbolic spaces. In *TextGraphs*, 59–69.
- Dong, Y.; Chawla, N. V.; and Swami, A. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *SIGKDD*, 135–144.
- Fawcett, T. 2006. An introduction to roc analysis. *Pattern recognition letters* 27(8):861–874.
- Fu, T.-y.; Lee, W.-C.; and Lei, Z. 2017. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. In *CIKM*, 1797–1806.
- Ganea, O.; Becigneul, G.; and Hofmann, T. 2018. Hyperbolic entailment cones for learning hierarchical embeddings. In *ICML*, 1646–1655.
- Grover, A., and Leskovec, J. 2016. node2vec: Scalable feature learning for networks. In *SIGKDD*, 855–864.
- Hu, B.; Shi, C.; Zhao, W. X.; and Yu, P. S. 2018. Leveraging meta-path based context for top-n recommendation with a neural co-attention model. In *SIGKDD*, 1531–1540.
- Huang, Z., and Mamoulis, N. 2017. Heterogeneous information network embedding for meta path based proximity. *arXiv preprint arXiv:1701.05291*.
- Krioukov, D.; Papadopoulos, F.; Kitsak, M.; Vahdat, A.; and Boguná, M. 2010. Hyperbolic geometry of complex networks. *Physical Review E* 82(3):036106.
- Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS*, 3111–3119.
- Nickel, M., and Kiela, D. 2017. Poincaré embeddings for learning hierarchical representations. In *NIPS*, 6338–6347.
- Perozzi, B.; Al-Rfou, R.; and Skiena, S. 2014. Deepwalk: Online learning of social representations. In *SIGKDD*, 701–710.
- Shi, C.; Li, Y.; Zhang, J.; Sun, Y.; and Philip, S. Y. 2017. A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering* 29(1):17–37.
- Shi, C.; Hu, B.; Zhao, X.; and Yu, P. 2018. Heterogeneous information network embedding for recommendation. *IEEE Transactions on Knowledge and Data Engineering*.
- Sun, Y.; Han, J.; Yan, X.; Yu, P. S.; and Wu, T. 2011. Paths: Meta path-based top-k similarity search in heterogeneous information networks. In *VLDB*, 992–1003.
- Tang, J.; Qu, M.; Wang, M.; Zhang, M.; Yan, J.; and Mei, Q. 2015. Line: Large-scale information network embedding. In *WWW*, 1067–1077.
- Tang, J.; Qu, M.; and Mei, Q. 2015. Pte: Predictive text embedding through large-scale heterogeneous text networks. In *SIGKDD*, 1165–1174.
- Tay, Y.; Tuan, L. A.; and Hui, S. C. 2018. Hyperbolic representation learning for fast and efficient neural question answering. In *WSDM*, 583–591.
- Wang, H.; Zhang, F.; Hou, M.; Xie, X.; Guo, M.; and Liu, Q. 2018. Shine: signed heterogeneous information network embedding for sentiment link prediction. In *WSDM*, 592–600.
- Wang, D.; Cui, P.; and Zhu, W. 2016. Structural deep network embedding. In *SIGKDD*, 1225–1234.
- Xu, L.; Wei, X.; Cao, J.; and Yu, P. S. 2017. Embedding of embedding (eoe): Joint embedding for coupled heterogeneous networks. In *WSDM*, 741–749. ACM.