Dynamic Heterogeneous Information Network Embedding in Hyperbolic Space

Dingyang Duan*, Daren Zha*, Xiao Yang** and Xiaobo Guo(✉)*

*Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
*School of Cyber Security,University of Chinese Academy of Sciences, Beijing, China
**Aerospace Internet of Things Technology Co.,Ltd, Beijing, China
**China Aerospace Times Electronics Co., Ltd, Beijing, China

***Abstract*—Heterogeneous information network (HIN) embedding, aiming to project HIN into a low-dimensional space, has attracted considerable research attention. Existing heterogeneous graph representation learning methods also take temporal evolution into consideration in Euclidean space which, however, underestimates the inherent complex and hierarchical properties in many real-world temporal networks, leading to sub-optimal embeddings. To explore these properties of a dynamic heterogeneous network, we propose a dynamic hyperbolic heterogeneous embedding(DyHHE) model that fully takes advantage of the hyperbolic geometry and structural heterogeneity. More specially, 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 DyHHE maps the temporal graph into hyperbolic space, and capture the structural heterogeneity and evolving behaviors by facilitating the proximity measurement. Experimental results on two real-world datasets demonstrate the superiority of DyHHE, as it consistently outperforms competing methods in link prediction task.**

***Index Terms*—Dynamic Graphs, Hyperbolic Space, Heterogeneous Information Network**

I. Introduction

Modeling data in the real world as heterogeneous information networks(HINs) can capture the internal relations of rich, complex data across various modalities. However, many real-world graphs are dynamic where graph structures constantly evolve over time. They are usually represented as sequence of graph snapshots at different time steps [1]. Examples include co-authorship networks where authors may periodically switch social network whose users may develop their multiple-type connections(follow, reply, retweet, etc) with others over time. The dynamics of a network and the structural heterogeneity provide abundant information for encoding nodes. So far, a number of HIN embedding methods have been proposed such as metapath2vec [2] and HAN [3]. These methods have overlooked a problem, that is, the formation of the neighbors is actually in order, it is related to time. There has been an ever-increasing amount of research on dynamic networks like DySAT [4] and HDGAN [5]. However, in the vast majority of these works, the space used for representing networks is Euclidean. In recent years, it has been suggested that complex networks may have underlying hyperbolic geometry and that hyperbolic space can better represent the structure of networks [6]. One fundamental property of hyperbolic space is that it expands exponentially and can be regarded as a smooth version of trees, abstracting the hierarchical organization. Despite the recent achievements in hyperbolic graph embedding, attempts on temporal heterogeneous networks are still scant. To fill this gap, in this work, we propose a novel dynamic hyperbolic heterogeneous embedding model, which fully takes advantage of the hyperbolic geometry and structural heterogeneity to capture the spatial dependency and temporal regularities of evolving networks via a recurrent learning paradigm. In summary, the main contributions are stated as follows:

- We propose a novel hyperbolic temporal graph embedding model on heterogeneous network, named DyHHE, to learn temporal regularities and implicitly hierarchical organization.
- We devise a hyperbolic structural network(HSN) module to preserve the HIN structure and semantic correlations in hyperbolic spaces based on the meta-path guided random walk. Then we apply a hyperbolic temporal network(HTN) module to effectively extract the diverse scope of historical information. To the best of our knowledge, this is the first study on dynamic heterogeneous network embedding in hyperbolic space.
- Experimental results on two real-world datasets demonstrate the superiority of DyHHE, as it consistently outperforms competing methods in link prediction task. The ablation study further gives insights into how each proposed component contributes to the success of the model.

II. Related Works

Recently, some methods have been proposed representation learning methods for HIN. Heterogeneous networks are usually characterized by meta-paths to find hidden relationships between nodes. Metapath2vec [2] obtains a corpus through random walks based on meta-paths, and uses skip-Gram for training. HAN [3] applies the attention mechanism to heterogeneous graphs through meta-path based neighbors. Real-world networks are not static, on the contrary, many networks are constantly changing, such as social networks. HDGAN [5] attempts to use the attention mechanism to take the

***This paper was withdrawn by the authors due to errors in the paper.
DOI reference number: 10.18293/SEKE2022-087

heterogeneity and dynamics of the network into account at the same time. DySAT [4] use the scaled dot-product form of attention to learn dynamic graph embedding. Most of the prevalent methods are built-in Euclidean space which, however, may underemphasize the intrinsic power-law distribution and hierarchical structure. Existing representation learning in hyperbolic space such as HGCN [11] and PoincareEmb [21] mainly focus in static graph. Despite the recent achievements in hyperbolic graph embedding, attempts on dynamic heterogeneous networks are still scant, which motivates us to explore hyperbolic geometry on dynamic heterogeneous networks.

III. Preliminaries

In this section, we first present the problem formulation of dynamic heterogeneous graph, then we introduce some fundamentals of hyperbolic geometry, which are essential in our proposed framework.

Definition 1. *Heterogeneous Information Network(HIN) [7].* An HIN is defined as a graph $G = (V, E)$, 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 $\phi(v) : V \rightarrow \mathcal{V}$ and $\psi(e) : E \rightarrow \mathcal{E}$ respectively. $\mathcal{V}$ and $\mathcal{E}$ denote the sets of node and relation types, where $|\mathcal{V}| + |\mathcal{E}| > 2$.

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

Definition 3. *Dynamic Heterogeneous Graph.* A heterogeneous temporal is defined as a graph $G =< V, E, A, T >$, from definition 1, we know that $|\mathcal{V}| + |\mathcal{E}| > 2$, where V represents the node type and E represents the edge type. A represents an event that sequence formed for each node's neighbors, and T is a time stamp. By definition 2, we can get the neighbor set of node i in the heterogeneous network. The neighbor formation sequence of node i refers to organizing the neighbors of the nodes in the network as a sequence based on the time of neighbor interaction events [8].

Then, we introduce some concepts of geometry to make this article more clear. A Riemannian manifold $\mathcal{M}$ is a space that generalizes the notion of a 2D surface to higher dimensions [9]. For each point $\mathbf{x} \in \mathcal{M}$, it associates with a **tangent space(Euclidean)** $\mathcal{T}_\mathbf{x}\mathcal{M}$ of the same dimensionality as $\mathcal{M}$. Intuitively, $\mathcal{T}_\mathbf{x}\mathcal{M}$ contains all possible directions in which one can pass through $\mathbf{x}$ tangentially (see Fig. 1). There are multiple models that can be used to represent hyperbolic space, each having different advantages. The Poincaré ball model is the best model for low dimensional visualizations of the embeddings. The Poincaré ball model with negative curvature $-c(c \geq 0)$ corresponds to the Riemannian manifold $(\mathbb{H}^{n,c}, g_\mathbb{H})$, where $\mathbb{H}^{n,c} = \{\mathbf{x} \in \mathbb{R}^n : c||\mathbf{x}||^2 \leq 1\}$ is an open n-dimensional ball. if $c = 0$, it degrades to Euclidean space, i.e., $\mathbb{H}^{n,c} = \mathbb{R}^n$. In addition, [9] shows how Euclidean and hyperbolic spaces can be continuously deformed into each other and provide a principled manner for basic operations (e.g., addition and multiplication) as well as essential functions (e.g., linear maps and softmax layer) in the context of neural networks and deep learning.

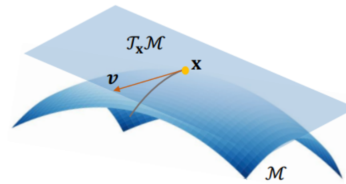

Fig. 1. The tangent space $\mathcal{T}_\mathbf{x}\mathcal{M}$ and a tangent vector v, along the given point $\mathbf{x}$ of a curve traveling through the manifold $\mathcal{M}$.

IV. Proposed Model

The overall framework of the proposed model DyHHE is illustrated in Fig. 2. DyHHE has two primary modules: hyperbolic structural module and hyperbolic temporal module, which benefits from the expressiveness of both hyperbolic embeddings and temporal evolutionary embeddings. As sketched in Fig. 2, DyHHE is a recurrent learning paradigm and falls into the prevalent discrete-time temporal graph architecture formulated by (1). More specifically, DyHHE can be summarized as two procedures: (a)Given the original input node feature, this procedure projects it into hyperbolic space, and preserve the structure by facilitating the proximity between the node $v \in V$ and its neighborhoods $c_\mathcal{V} \in C_\mathcal{V}(v)$ with type $\mathcal{V}$. We use meta-path guided random walks [2] to obtain heterogeneous neighborhoods of a node. (b)These sequences of node representations then feeds as input to the temporal recurrent module to capture the sequential patterns. Furthermore, we propose an attention mechanism based on the hyperbolic proximity to obtain the attentive hidden state. Owning to the superiorities of self-attention, this unit attending on multiple historical latent states to get a more informative hidden state. We elaborate on the details of each respective module in the following paragraphs.

$$H_t(\phi) = f_2(f_1(A_t, X_t), H_{t-1}) \tag{1}$$

A. Feature Map

Before going into the details of each module, we first introduce two bijection operations, the exponential map and the logarithmic map, for mapping between hyperbolic space and tangent space with a local reference point [10], [11], as presented below.

In this work, we use Poincaré model with constant curvature $c = 1$ as the hyperbolic space for entity embeddings [12]. In particular, we denote d-dimensional Poincaré centered at origin as $\mathbb{H}^{n,c} = \{\mathbf{x} \in \mathbb{R}^n : c||\mathbf{x}||^2 \leq 1\}$, where $|| \cdot ||$ is the Euclidean norm. The Poincaré model of hyperbolic space is equipped with Riemannian metric:

$$g_x^\mathbb{H} = \lambda_{\mathbf{x}'}^2 g^\mathbb{R} \tag{2}$$

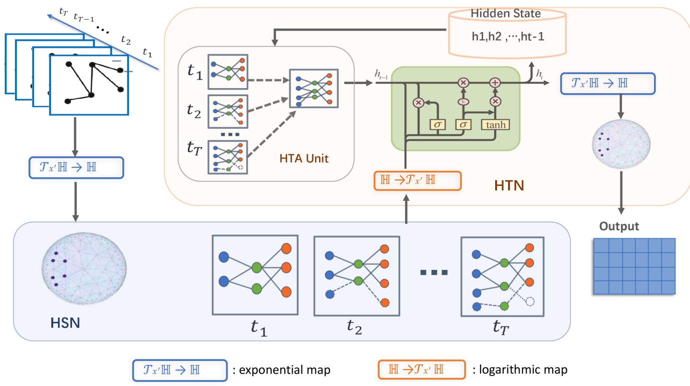

Fig. 2. Architecture of DyHHE.

where $\lambda_{\mathbf{x}'}^c := \frac{2}{1-c||x'||^2}$ and $g^{\mathbb{R}}$ denotes the Euclidean metric, i.e., $g^{\mathbb{R}} = \mathbb{I}$. The mobius addition $\oplus$ defined on Poincaré model with curvature c is given by:

$$u \oplus v := \frac{(1 + 2c\langle u, v\rangle + c||v||^2)u + (1 - c||u||^2)v}{1 + 2c\langle u, v\rangle + c^2||u||^2||v||^2}. \quad (3)$$

For each point $\mathbf{x}' \in \mathbb{H}^{d,c}$, the tangent space $\mathcal{T}_{\mathbf{x}'}\mathbb{H}^{d,c}$ is the Euclidean vector space containing all tangent vectors at $\mathbf{x}'$. For $\mathbf{x}' \in \mathbb{H}^{d,c}$, $a \in \mathcal{T}_{\mathbf{x}'}\mathbb{H}^{d,c}$, $b \in \mathbb{H}^{d,c}$, and $a \neq 0$, $b \neq \mathbf{x}'$. One can map vectors in $\mathcal{T}_{\mathbf{x}'}\mathbb{H}^{d,c}$ to vectors in $\mathbb{H}^{d,c}$ through exponential map $\exp_{\mathbf{x}'}^c(\cdot) : \mathcal{T}_{\mathbf{x}'}\mathbb{H}^{d,c} \rightarrow \mathbb{H}^{d,c}$ as follows:

$$\exp_{\mathbf{x}'}^c(a) = \mathbf{x}' \oplus^c (\tanh(\frac{\sqrt{c}\lambda_{\mathbf{x}'}^c||a||}{2})\frac{a}{\sqrt{c}||a||}) \quad (4)$$

Conversely, the logarithmic map $\log_{\mathbf{x}'}^c(\cdot) : \mathbb{H}^{d,c} \rightarrow \mathcal{T}_{\mathbf{x}'}\mathbb{H}^{d,c}$ maps vectors in $\mathbb{H}^{d,c}$ back to vectors in $\mathcal{T}_{\mathbf{x}'}$, in particular:

$$\log_{\mathbf{x}'}^c(b) := \frac{2}{\sqrt{c}\lambda_{\mathbf{x}'}^c} \operatorname{artanh}(\sqrt{c}||-\mathbf{x}' \oplus^c b||)\frac{-\mathbf{x}' \oplus^c b}{||-\mathbf{x}' \oplus^c b||} \quad (5)$$

Also, the hyperbolic distance between $u, v \in \mathbb{H}^{d,c}$ is:

$$d_c(u, v) = (2\sqrt{c}\operatorname{artanh}(\sqrt{c}|| - u \oplus^c v||)) \quad (6)$$

B. Hyperbolic Structural Network(HSN)

On a dynamic heterogeneous graph, various kinds of interactions are constantly being established over time, which can be regarded as a series of observed heterogeneous events. We aim to learn the representation of nodes to preserve the structure and semantic correlations in hyperbolic spaces for each snapshot. In each time step, HSN is employed to preserve the structure by facilitating the proximity between the node $v \in V$ and its neighborhoods $c_{\mathcal{V}} \in C_{\mathcal{V}(v)}$ with type $\mathcal{V}$, which leveraging promising properties of hyperbolic geometry.

The input of HSN is the node feature, whose norm could be out of the Poincaré ball defined in hyperbolic space. To make the node feature available in hyperbolic space, we use the exponential map to project the feature into the hyperbolic space, shown in (4). Specifically, let an Euclidean space vector $\mathbf{x}_i^E \in \mathbb{R}^d$ be the feature of node i, and then we regard it as the point in the tangent space $\mathcal{T}_{\mathbf{x}'}\mathbb{H}^{d,c}$ with the reference point $\mathbf{x}' \in \mathbb{H}^{d,c}$, using the exponential map to project it into hyperbolic space, obtaining $\mathbf{x}^{\mathcal{H}} \in \mathbb{H}^{d,c}$, which is defined as:

$$\mathbf{x}_i^{\mathcal{H}} = \exp_{\mathbf{x}'}^c(\mathbf{x}_i^R). \quad (7)$$

Then, We use meta-path guided random walks to obtain heterogeneous neighborhoods of a node [2]. Given an arbitrary meta-path $\mathcal{P} = \mathcal{V}_{v_1} \xrightarrow{\mathcal{E}_{e_1}} \mathcal{V}_{v_2} \xrightarrow{\mathcal{E}_{e_2}} ... \xrightarrow{\mathcal{E}_{e_{n-1}}} \mathcal{V}_{v_n}$, our goal is to learn the semantically meaningful embeddings for all nodes under the constraint of meta-path $\mathcal{P}$. The transition probability at step i is defined as follows:

$$p(v^{i+1}|v_{\mathcal{V}_{v_i}}^i, \mathcal{P}) = \begin{cases} \frac{1}{|N_{\mathcal{V}_{v_{i+1}}}(v_{\mathcal{V}_{v_i}}^i)|}, & (v^{i+1}, v_{\mathcal{V}_{v_i}}^i) \in E \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where $v_{\mathcal{V}_{v_i}}^i$ is node $v \in V$ with type $\mathcal{V}_{v_i}$, and $N_{\mathcal{V}_{v_{i+1}}}(v_{\mathcal{V}_{v_i}}^i)$ denotes the $\mathcal{V}_{v_{i+1}}$ type of neighborhood of node $v_{\mathcal{V}_{v_i}}^i$.The meta-path guided random walk strategy ensures that the semantic relationships between different types of nodes can be properly incorporated into HSN.

The premise of network embedding models is to preserve the proximity between a node and its neighborhood. Therefore, in hyperbolic space, we use distances in Poincaré model to measure their proximity, as given in (6). We use a probability to measure the node $c_{\mathcal{V}}$ is a neighborhood of node v as following:

$$p(v|c_{\mathcal{V}}; \Theta) = \sigma[-d(u, v)] \quad (9)$$

where $\sigma(\cdot) = \frac{1}{1+exp(-x)}$ is an activate function. According to the (9), the object of HSN module is to maximize the probability as followings:

$$argmax \sum_{v \in V} \sum_{c_{\mathcal{V}} \in C_{\mathcal{V}(v)}} \log p(v|c_{\mathcal{V}}; \Theta) \quad (10)$$

C. Hyperbolic Temporal Network(HTN)

Historical information plays an indispensable role in temporal graph modeling since it facilitates the model to learn the evolving patterns and regularities. Although the latest hidden state H_{t-1} obtained by the recurrent neural network already carries historical information before time t, some discriminate contents may still be under-emphasized due to the monotonic mechanism of RNNs that temporal dependencies are decreased along the time span [13]. Inspired by [14], we design the hyperbolic temporal attention(HTA) unit generalizes H_{t-1} to the latest time window w snapshots $H_{t-w}, ..., H_{t-1}$, attending on multiple historical latent states to get a more informative hidden state. These historical states in the state memory are concatenate together and feed as input to the HTA unit, which is performed in tangent space due to its computational efficiency. Owing to the superiorities of attention, this unit fuses the final hidden state by figuring out the importance each graph snapshots. The dataflow in the HTA unit is characterized by the following equations:

$$H_t^E = \log_{\mathbf{x}'}^c(H_t^{\mathcal{H}}) \tag{11}$$

$$H = concat(H_{t-w}^E, ..., H_{t-1}^E) \tag{12}$$

$$H_{t-1}^E = softmax(k^T \tanh(QH))H \tag{13}$$

$$H_{t-1}^{\mathcal{H}} = \exp_{\mathbf{x}}^c(H_{t-1}^E) \tag{14}$$

The learnable weight matrix Q and K are utilized to extract contextual information, where Q weights the node importance in each historical state and K determines the weights across the time windows.

Then, we use GRU, a variant of LSTM, as primary part of HTN to incorporate the current and historical node states, in view of the GRU is the newer generation of Recurrent Neural networks and is pretty similar to an LSTM. GRU gets rid of the cell state and used the hidden state to transfer information. It also only has two gates, a reset gate and update gate.

HTN unit receives the sequential node embedding $X_t^{\mathcal{H}}$ from HSN and the hidden state $H_{t-1}^{\mathcal{H}}$ which is obtained from HTA. As sketched in Fig. 2, the input representations of HTN unit are assumed to sufficiently capture local structural information as well as attentive hidden state. The dataflow in the HTN unit is characterized by the following equations:

$$X_t^E = \log_{\mathbf{x}'}^c(X_t^{\mathcal{H}}), \tag{15}$$

$$H_{t-1}^E = \log_{\mathbf{x}'}^c(H_{t-1}^{\mathcal{H}}), \tag{16}$$

$$H_t^E = GRU(X_t^E, H_{t-1}^E), \tag{17}$$

$$H_t^{\mathcal{H}} = \exp_{\mathbf{x}}^c(H_t^E). \tag{18}$$

The main part of the unit is GRU. As the GRU is built in tangent space, we use the logarithmic transformations to project the $X_t^{\mathcal{H}}$ and $H_{t-1}^{\mathcal{H}}$ into tangent space. After processing representation using GRU, we project the embedding back to hyperbolic space. As we can see, the final embedding $H_t^{\mathcal{H}}$ fuses structural heterogeneity, content, and temporal information.

V. Optimization

Uniting the above modules, we formulate the learning objective from two aspects: topological learning and temporal evolution, corresponding to the following hyperbolic structural loss and hyperbolic temporal loss.

A. Hyperbolic Structural Loss

We leverage the negative sampling proposed in [15], which basically samples a small number of negative objects to enhance the influence of positive objects. The hyperbolic structure loss $\mathcal{L}(\Theta)$ aims to minimize the proximity between v and its neighborhood $c_{\mathcal{V}}$ while maximize the proximity between v and its negative sampled node n. The objective equation (9) can be formulated as following:

$$\mathcal{L}(\Theta) = \log\sigma[-d(\mathbf{x}_v, \mathbf{x}_{c_{\mathcal{V}}})] + \sum_{m=1}^{M} \mathbb{E}_{n^m \sim P(n)}\{\log\sigma[d(\mathbf{x}_v, \mathbf{x}_{n^m})]\} \tag{19}$$

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.

B. Hyperbolic Temporal Loss

We build a hyperbolic temporal consistency consistency constraint $\mathcal{L}(t)$ on two consecutive time steps (G_t, G_{t-1})), which is defined as:

$$\mathcal{L}(t) = \sum_{t=1}^{T} \sigma[d(\mathbf{x}_v, \mathbf{x}_{n^m})] \tag{20}$$

where the t denotes the loss is with respect to time step t.

C. The Unified Loss

To enable the learned representations to capture structural evolution, our objective function set the final loss function as:

$$\mathcal{L} = \mathcal{L}(\Theta) + \lambda\mathcal{L}(t) \tag{21}$$

where $\lambda \in [0, 1]$ is the hyper-parameter to balance the temporal smoothness and structural regularity. The final $\mathcal{L}$ not only minimizing the hyperbolic distance of a node with its connected nodes and maximizing with the sampled negative neighbors, but also minimizing the distance between the same node over two consecutive timestamps. As the parameters of DyHHE live in a Poincaré model which has a Riemannian manifold structure, it should be noted, the back-propagated gradient is a Riemannian gradient. It makes no sense Euclidean gradient based optimization works in this manifold. Therefore, we optimize $\mathcal{L}$ via Riemannian stochastic gradient descent(RSGD) optimization method [16]. The gradient of their distance can be derived as:

$$\Delta_v(d(\mathbf{x}_v, \mathbf{x}_{n^m})) = \frac{4}{\beta\sqrt{\gamma^2 - 1}}\left(\frac{\|\mathbf{x}_v\|^2 - 2\langle\mathbf{x}_v, \mathbf{x}_{n^m}\rangle}{\alpha^2}\mathbf{x}_v - \frac{\mathbf{x}_v}{\alpha}\right) \tag{22}$$

where $\alpha = 1 - \|\mathbf{x}_v\|^2$, $\beta = 1 - \|\mathbf{x}_{n^m}\|^2$, $\gamma = 1 + \frac{2}{\alpha\beta}\|\mathbf{x}_v - \mathbf{x}_{n^m}\|^2$.

VI. Experiments and Analysis

In this section, we conduct extensive experiments with the aim of answering the following research questions:

- **RQ1** How does DyHHE perform.
- **RQ1** What does each component of DyHHE bring?

A. Datasets

To evaluate the effectiveness of our model, we conduct experiments on two datasets from real-world platforms. The datasets are summarized in Table I

- **DBLP** is a database of publications. Specifically, we collected the papers from four research areas 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). Timestamps denote the year of the publication.
- **MovieLens** [17] contains knowledge about movies. MovieLens users from the late 1990s to the early 2000s.

We extract a subset of MovieLens, which contains three types of nodes, i.e., actor(A), movie(M), and director(D) and two types of edges, i.e., actor-movie(act in) and director-movie(direct).

TABLE I
STATISTICS OF DATASETS

DBLP	A	P	V	AP	PV	Snapshots
	14475	14376	20	41794	14376	16
MovieLens	A	M	D	AM	MD	Snapshots
	11718	9160	3510	64051	9160	13

Data with power-law structure can be naturally modeled in hyperbolic spaces [6]. Therefore, we use two real-world HINs datasets which have been proved to conform to the power-law distribution of nodes [18], [19].

B. Baselines

We present comparisons against several static graph embedding methods to analyze the gains of using temporal information for link prediction. To ensure a fair comparision, we also conduct experiments on several heterogeneous network representation model to further demonstrate the superiority of the proposed model DyHHE. Moreover, we also compare to the hyperbolic embedding model, HGCN and PoincareEmb.

- **Node2vec** [20] is a static embedding method to generate vector representations of nodes on a graph. It learns low-dimensional representations for nodes in a graph through the use of random walks.
- **Metapath2vec** [2] is an HIN embedding method. It learns feature representations by capturing node pairs within w-hop heterogeneous neighborhood via meta-path guided random walks in the network.
- **HGCN** [11] is a static embedding method which leverages both the expressivenees of GCNs and hyperbolic geometry to learn node reoresentations for hierachical and scale-free graphs.
- **DySAT** [4] computes node representations by jointly employing self-attention layers along two dimensions: structural neighborhood and temporal dynamics.
- **HDGAN** [5] is based on three levels of attention, namely structural-level attention, semantic-level attention and time-level attention and attempts to use the attention mechanism to take the heterogeneity and dynamics of the network into account at the same time, so as to better learn network embedding.
- **PoincareEmb** [21] is a method that preserves proximities of node pairs linked by an edge via embedding network into a Poincaré ball.

For random walk based methods like Node2vec and Metapath2vec, we set neighborhood size to 5, walk length to 80, ignoring the temporal regularity. As for meta-path guided random walks like metapath2vec and PoincareEmb, we use meta-path ”A–P–V–P–A” in DBLP and ”A-M-D-M-A” in MovieLens. For dynamic homogeneous baselines Dysat, we treat events as homogeneous. The train/test ratio is set to 80%/20%.

C. Link Prediction Comparison(RQ1)

Link prediction is to predict the type $\mathcal{V}$ interaction at time step t, which can be used to test the generalization performance of a network embedding method. Given all temporal heterogeneous events before time step t and two nodes u and v. For each type of edge, we treat all events at time t as the positive link, and an equal number of negative examples in the training set are created by sampling the node pairs not interconnected. Subsequently, we split the chosen edges and negative samples into validation and test. In our experiments, we test the models regarding their ability of correctly classifying true and false edges by computing average precision (AP) and area under the ROC curve (AUC) scores. We uniformly train both the baselines and DyHHE by using early stopping based on the performance of the training set.

TABLE II
AUC SCORES OF LINK PREDICTION RESULT.

	DBLP		MovieLens	
Edge	**A-P**	**P-V**	**A-M**	**M-D**
Node2vec	85.32 ± 0.7	85.25 ± 0.8	81.27 ± 0.2	83.44 ± 1.1
Metapath2vec	87.46 ± 0.9	88.11 ± 0.9	82.3 ± 0.1	81.57 ± 1.4
HGCN	89.8 ± 1.2	90.27 ± 0.4	85.4 ± 0.2	85.17 ± 1.3
DySAT	90.66 ± 0.2	90.21 ± 0.4	87.32 ± 0.3	86.75 ± 0.9
HDGAN	87.42 ± 0.4	88.66 ± 0.6	85.78 ± 0.9	86.12 ± 0.7
PoincareEmb	87.85 ± 0.4	87.16 ± 0.3	86.71 ± 1.7	85.63 ± 0.9
DyHHE	$\mathbf{92.69 \pm 0.4}$	$\mathbf{93.19 \pm 0.3}$	$\mathbf{90.13 \pm 0.4}$	$\mathbf{90.77 \pm 0.3}$

TABLE III
AP SCORES OF LINK PREDICTION RESULT.

	DBLP		MovieLens	
Edge	**A-P**	**P-V**	**A-M**	**M-D**
Node2vec	86.94 ± 0.4	86.78 ± 0.6	83.17 ± 0.3	82.15 ± 0.9
Metapath2vec	87.83 ± 1.2	86.43 ± 0.7	81.77 ± 0.4	82.72 ± 0.3
HGCN	88.6 ± 0.7	89.33 ± 0.5	84.6 ± 1.2	84.45 ± 1.1
DySAT	90.37 ± 0.4	90.71 ± 0.3	85.72 ± 0.6	85.25 ± 0.3
HDGAN	89.61 ± 0.3	87.74 ± 0.6	86.33 ± 1.1	85.12 ± 0.9
PoincareEmb	88.15 ± 1.1	86.29 ± 0.2	84.12 ± 0.8	85.81 ± 0.7
DyHHE	$\mathbf{92.13 \pm 0.7}$	$\mathbf{93.49 \pm 0.5}$	$\mathbf{91.02 \pm 0.6}$	$\mathbf{89.43 \pm 0.8}$

We repeat each experiment five times and report the average value with the standard deviation on the test sets in Table II and Table III. It is observed our model achieves the best results and has a more than $4-6\%$ AUC and AP improvement comparing to the best baseline across all datasets. First of all, the Metapath2vec has a better performance than Node2vec which means the advantage of the proper consideration and accommodation of the network heterogeneity. Despite the existence of multiple types of nodes and edges in heterogeneous graph, Metapath2vec performs pooly compared with the dynamic methods DySAT and HDGAN, which confirms the importance of temporal regularity in dynamic graph modeling. Moreover, the performance gap between DyHHE and HDGAN suggests that the significantly benefit from hyperbolic geometry. It is

TABLE IV
ABLATION STUDY(AUC).

	DBLP		MovieLens	
edge	A-P	P-V	A-M	M-D
No Hyperbolic	89.03 ± 0.5	90.72 ± 0.4	85.34 ± 0.3	84.79 ± 0.5
No Temporal	89.18 ± 0.5	89.14 ± 0.6	89.26 ± 0.4	90.81 ± 0.6
Original	$\mathbf{92.69 \pm 0.4}$	$\mathbf{93.19 \pm 0.3}$	$\mathbf{92.13 \pm 0.4}$	$\mathbf{90.77 \pm 0.3}$

worth mentioning that HGCN and PoincareEmb also has not bad performance despite being agnostic to semantic relationships and temporal information in heterogeneous graph, which indicates further improvements to DyHHE on transforming embeddings from Euclidean space to Hyperbolic space.

D. Ablation Study (RQ2)

To investigate the superiority of the main components of our model, we conduct an ablation study by independently removing the hyperbolic geometry and temporal modules from DyHHE to create simpler architectures. And we compare DyHHE with different variants on DBLP and MovieLens datasets. When we remove the hyperbolic geometry and build the model in Euclidean space, the HSN and HTN units are converted to the corresponding Euclidean space. We show the variant models results in Table IV. From the results, we observe that in MovieLens the removal of hyperbolic geometry consistently deteriorates performance, while the DBLP only declines about 4%. One major explanation is that the MovieLens has a high-hyperbolicity, which indicates the dataset has a more evident hierarchical structure. And the hierarchical structure and tree-like data can naturally be represented and preserved by hyperbolic geometry. The effect of temporal module is also significant because of the performance degradation by removing the temporal block. This observation conforms to the nature of graph evolution since the behaviors usually have periodical patterns such as recurrent links or communities. In summary, DyHHE generates more appropriate embeddings for dynamic heterogeneous neetwork than comparative baselines, suggesting its ability to capture and incorporate the underlying structural and temporal information.

VII. CONCLUSIONS

In this work, we introduce a novel hyperbolic geometry-based node representation learning framework in dynamic heterogeneous networks in which there exists diverse types of nodes and links. To address the network heterogeneity and temporal evolution, we propose the DyHHE model. In general, DyHHE computes dynamic node representations through maximize proximity in consideration of multiple types of neighborhoods for a given node and follow the effective GRU framework by leveraging the superiority of hyperbolic graph neural network. Our experimental results on two real-world datasets indicate significant performance gains for DyHHE over several static and dynamic heterogeneous graph embedding baselines. An interesting future direction is generalizing our method to more challenging tasks.

REFERENCES

[1] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: Densification and shrinking diameters," *ACM transactions on Knowledge Discovery from Data (TKDD)*, vol. 1, no. 1, pp. 2–es, 2007.

[2] Y. Dong, N. V. Chawla, and A. Swami, "metapath2vec: Scalable representation learning for heterogeneous networks," in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 2017, pp. 135–144.

[3] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, "Heterogeneous graph attention network," in *The world wide web conference*, 2019, pp. 2022–2032.

[4] A. Sankar, Y. Wu, L. Gou, W. Zhang, and H. Yang, "Dysat: Deep neural representation learning on dynamic graphs via self-attention networks," in *Proceedings of the 13th International Conference on Web Search and Data Mining*, 2020, pp. 519–527.

[5] Q. Li, Y. Shang, X. Qiao, and W. Dai, "Heterogeneous dynamic graph attention network," in *2020 IEEE International Conference on Knowledge Graph (ICKG)*. IEEE, 2020, pp. 404–411.

[6] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguná, "Hyperbolic geometry of complex networks," *Physical Review E*, vol. 82, no. 3, p. 036106, 2010.

[7] C. Shi, Y. Li, J. Zhang, Y. Sun, and S. Y. Philip, "A survey of heterogeneous information network analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 1, pp. 17–37, 2016.

[8] Y. Zuo, G. Liu, H. Lin, J. Guo, X. Hu, and J. Wu, "Embedding temporal network via neighborhood formation," in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 2857–2866.

[9] O. Ganea, G. Bécigneul, and T. Hofmann, "Hyperbolic neural networks," *Advances in neural information processing systems*, vol. 31, 2018.

[10] Q. Liu, M. Nickel, and D. Kiela, "Hyperbolic graph neural networks," *arXiv preprint arXiv:1910.12892*, 2019.

[11] I. Chami, Z. Ying, C. Ré, and J. Leskovec, "Hyperbolic graph convolutional neural networks," *Advances in neural information processing systems*, vol. 32, pp. 4868–4879, 2019.

[12] M. Nickel and D. Kiela, "Poincaré embeddings for learning hierarchical representations," *Advances in neural information processing systems*, vol. 30, 2017.

[13] J. Liu, G. Wang, P. Hu, L.-Y. Duan, and A. C. Kot, "Global context-aware attention lstm networks for 3d action recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1647–1656.

[14] Q. Cui, S. Wu, Y. Huang, and L. Wang, "A hierarchical contextual attention-based network for sequential recommendation," *Neurocomputing*, vol. 358, pp. 141–149, 2019.

[15] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems*, vol. 26, 2013.

[16] S. Bonnabel, "Stochastic gradient descent on riemannian manifolds," *IEEE Transactions on Automatic Control*, vol. 58, no. 9, pp. 2217–2229, 2013.

[17] F. M. Harper and J. A. Konstan, "The movielens datasets: History and context," *Acm transactions on interactive intelligent systems (tiis)*, vol. 5, no. 4, pp. 1–19, 2015.

[18] X. Wang, Y. Zhang, and C. Shi, "Hyperbolic heterogeneous information network embedding," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 5337–5344.

[19] Y. Zhang, X. Wang, N. Liu, and C. Shi, "Embedding heterogeneous information network in hyperbolic spaces," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 16, no. 2, pp. 1–23, 2021.

[20] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 855–864.

[21] Z. Huang and N. Mamoulis, "Heterogeneous information network embedding for meta path based proximity," *arXiv preprint arXiv:1701.05291*, 2017.