Relation Structure-Aware Heterogeneous Information Network Embedding

Yuanfu Lu\textsuperscript{1}, Chuan Shi\textsuperscript{1,*}, Linmei Hu\textsuperscript{1}, Zhiyuan Liu\textsuperscript{2}
\textsuperscript{1}Beijing University of Posts and Telecommunications, Beijing, China
\textsuperscript{2}Tsinghua University, Beijing, China
\{luyuanfu, shichuan, hulinmei\}@bupt.edu.cn, liuzy@tsinghua.edu.cn

Abstract
Heterogeneous information network (HIN) embedding aims to embed multiple types of nodes into a low-dimensional space. Although most existing HIN embedding methods consider heterogeneous relations in HINs, they usually employ one single model for all relations without distinction, which inevitably restricts the capability of network embedding. In this paper, we take the structural characteristics of heterogeneous relations into consideration and propose a novel Relation structure-aware Heterogeneous Information Network Embedding model (RHINE). By exploring the real-world networks with thorough mathematical analysis, we present two structure-related measures which can consistently distinguish heterogeneous relations into two categories: Affiliation Relations (ARs) and Interaction Relations (IRs). To respect the distinctive characteristics of relations, in our RHINE, we propose different models specifically tailored to handle ARs and IRs, which can better capture the structures and semantics of the networks. At last, we combine and optimize these models in a unified and elegant manner. Extensive experiments on three real-world datasets demonstrate that our model significantly outperforms the state-of-the-art methods in various tasks, including node clustering, link prediction, and node classification.

1 Introduction
Network embedding has shed a light on the analysis of networks as it is effective to learn the latent features that encode the properties of a network (Cui et al. 2018; Cai, Zheng, and Chang 2018). Although the state-of-the-arts (Perozzi, Al-Rfou, and Skiena 2014; Grover and Leskovec 2016; Tang et al. 2015; Wang, Cui, and Zhu 2016) have achieved promising performance in many data mining tasks, most of them focus on homogeneous networks, which only contain one single type of nodes and edges. In reality, many networks are usually with multiple types of nodes and edges, widely known as heterogeneous information networks (HINs) (Sun et al. 2011; Shi et al. 2017). Taking the DBLP network for example, as shown in Figure 1(a), it contains four types of nodes: Author (A), Paper (P), Conference (C) and Term (T), and multiple types of relations: writing/written relations, publish/published relations, etc. In addition, there are composite relations represented by meta-paths (Sun et al. 2011) such as APA (co-author relation) and APC (authors write papers published in conferences), which are widely used to exploit rich semantics in HINs. Thus, compared to homogeneous networks, HINs fuse more information and contain richer semantics. Directly applying traditional homogeneous models to embed HINs will inevitably lead to reduced performance in downstream tasks.

To model the heterogeneity of networks, several attempts have been done on HIN embedding. For example, some models employ meta-path based random walk to generate node sequences for optimizing the similarity between nodes (Shang et al. 2016; Dong, Chawla, and Swami 2017; Fu, Lee, and Lei 2017) have achieved promising performance in many data mining tasks, most of them focus on homogeneous networks, which only contain one single type of nodes and edges. In reality, many networks are usually with multiple types of nodes and edges, widely known as heterogeneous information networks (HINs) (Sun et al. 2011; Shi et al. 2017). Taking the DBLP network for example, as shown in Figure 1(a), it contains four types of nodes: Author (A), Paper (P), Conference (C) and Term (T), and multiple types of relations: writing/written relations, and
two nodes close to each other, as illustrated in Figure 1(b).

However, various relations in an HIN have significantly different structural characteristics, which should be handled with different models. Let’s see a toy example in Figure 1(a). The relations in the network include atomic relations (e.g., AP and PC) and composite relations (e.g., APA and APC). Intuitively, AP relation and PC relation reveal rather different characteristics in structure. That is, some authors write some papers in the AP relation, which shows a peer-to-peer structure. While that many papers are published in one conference in the PC relation, which indicate peer-to-peer and one-centered-by-another structures respectively. The intuitive examples clearly illustrate that relations in an HIN indeed have different structural characteristics.

It is non-trivial to consider different structural characteristics of relations for HIN embedding, due to the following challenges: (1) How to distinguish the structural characteristics of relations in an HIN? Various relations (atomic relations or meta-paths) with different structures are involved in an HIN. Quantitative and explainable criteria are desired to explore the structural characteristics of relations and distinguish them. (2) How to capture the distinctive structural characteristics of different categories of relations? Since the various relations have different structures, modeling them with one single model may lead to some loss of information. We need to specifically design appropriate models which are able to capture their distinctive characteristics. (3) The different models for the differentiated relations should be easily and smoothly combined to ensure simple optimization in a unified manner.

In this paper, we present a novel model for HIN embedding, named Relation structure-aware HIN Embedding (RHINE). In specific, we first explore the structural characteristics of relations in HINs with thorough mathematical analysis, and present two structure-related measures which can consistently distinguish the various relations into two categories: Affiliation Relations (ARs) with one-centered-by-another structures and Interaction Relations (IRs) with peer-to-peer structures. In order to capture the distinctive structural characteristics of the relations, we then propose two specifically designed models. For ARs where the nodes share similar properties (Yang and Leskovec 2012), we calculate Euclidean distance as the proximity between nodes, so as to make the nodes directly close in the low-dimensional space. On the other hand, for IRs which bridge two compatible nodes, we model them as translations between the nodes. Since the two models are consistent in terms of mathematical form, they can be optimized in a unified and elegant way.

It is worthwhile to highlight our contributions as follows:

- To the best of our knowledge, we make the first attempt to explore the different structural characteristics of relations in HINs and present two structure-related criteria which can consistently distinguish heterogeneous relations into ARs and IRs.
- We propose a novel relation structure-aware HIN embedding model (RHINE), which fully respects the distinctive structural characteristics of ARs and IRs by exploiting appropriate models and combining them in a unified and elegant manner.
- We conduct comprehensive experiments to evaluate the performance of our model. Experimental results demonstrate that our model significantly outperforms state-of-the-art network embedding models in various tasks.

## 2 Related Work

Recently, network embedding has attracted considerable attention. Inspired by word2vec (Mikolov et al. 2013b), random walk based methods (Perozzi, Al-Rfou, and Skiena 2014; Grover and Leskovec 2016) have been proposed to learn representations of networks by the skip-gram model (Mikolov et al. 2013a). After that, several models are designed to better preserve network properties (Tang et al. 2015; Cao, Lu, and Xu 2015; Ou et al. 2016; Tu et al. 2016; Ribeiro, Saverese, and Figueiredo 2017). Besides, there are some deep neural network based models for network embedding (Yang and Leskovec 2012; Cao, Lu, and Xu 2016). However, all the aforementioned methods focus only on learning the representations of homogeneous networks.

Different from homogeneous networks, HINs consist of multiple types of nodes and edges. Several attempts have been done on HIN embedding and achieved promising performance in various tasks (Tang, Qu, and Mei 2015; Chang et al. 2015; Shang et al. 2016; Fu, Lee, and Lei 2017; Wang et al. 2018; Shi et al. 2018). PTE (Tang, Qu, and Mei 2015) decomposes an HIN to a set of bipartite networks and then performs network embedding individually. ESim (Shang et al. 2016) utilizes user-defined meta-paths as guidance to learn node embeddings. Metapath2vec (Dong, Chawla, and Swami 2017) combines meta-path based random walks and skip-gram model for HIN embedding. HIN2Vec (Fu, Lee, and Lei 2017) learns the embeddings of HINs by conducting multiple prediction training tasks jointly. HERec (Shi et al. 2018) filters the node sequences with type constraints and thus captures the semantics of HINs.

All the above-mentioned models deal with all relations in HINs with one single model, neglecting differentiated structures of relations. In this paper, we explore and distinguish the structural characteristics of relations with quantitative analysis. For relations with distinct structural characteristics, we propose to handle them with specifically designed models.

## 3 Preliminaries

In this section, we introduce some basic concepts and formalize the problem of HIN embedding.

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

**Definition 2.** **Meta-path.** A meta-path $m \in M$ is defined as a sequence of node types $t_{v_i}$ or edge types $t_{e_j}$ in the form
of \( t_{v_1} \xrightarrow{r_1} t_{v_2} \xrightarrow{r_2} \ldots \xrightarrow{r_{k+1}} t_{v_{k+1}} \) (abbreviated as \( t_{v_1} t_{v_2} \ldots t_{v_{k+1}} \)), which describes a composite relation between \( v_1 \) and \( v_{k+1} \).

**Definition 3. Node-Relation Triple.** In an HIN \( G \), relations \( R \) include atomic relations (e.g., links) and composite relations (e.g., meta-paths). A node-relation triple \( (u, r, v) \in P \) describes that two nodes \( u \) and \( v \) are connected by a relation \( r \in R \). Here \( P \) represents the set of all node-relation triples.

**Example 1.** For example, as shown in Figure 1(a), \( (a_1, R_{AP}, c_2) \) is a node-relation triple, meaning that \( a_1 \) writes a paper published in \( c_2 \).  

**Definition 4. Heterogeneous Information Network Embedding.** Given an HIN \( G = (V, E, T, \phi, \varphi) \), the goal of HIN embedding is to develop a mapping function \( f : V \rightarrow \mathbb{R}^d \) that projects each node \( v \in V \) to a low-dimensional vector in \( \mathbb{R}^d \), where \( d \ll |V| \).

### 4 Structural Characteristics of Relations

In this section, we first describe three real-world HINs and analyze the structural characteristics of relations in HINs. Then we present two structure-related measures which can consistently distinguish various relations quantitatively.

#### 4.1 Dataset Description

Before analyzing the structural characteristics of relations, we first briefly introduce three datasets used in this paper, including DBLP\(^1\), Yelp\(^2\) and AMIner\(^3\). The detailed statistics of these datasets are illustrated in Table 1.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Nodes</th>
<th>Number of Nodes</th>
<th>Relations</th>
<th>Number of Relations</th>
<th>Avg. Degree of ( t_u )</th>
<th>Avg. Degree of ( t_v )</th>
<th>Measures</th>
<th>Relation Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBLP</td>
<td>Term (T)</td>
<td>8,811</td>
<td>PC</td>
<td>14,376</td>
<td>1.0</td>
<td>718.8</td>
<td>718.8</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Paper (P)</td>
<td>14,376</td>
<td>APC</td>
<td>24,495</td>
<td>2.9</td>
<td>2089.7</td>
<td>1.0</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>Author (A)</td>
<td>14,475</td>
<td>AP</td>
<td>41,794</td>
<td>2.8</td>
<td>2.9</td>
<td>1.0</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>Conference (C)</td>
<td>20</td>
<td>PT</td>
<td>88,683</td>
<td>6.2</td>
<td>10.7</td>
<td>1.7</td>
<td>0.00007</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>APT</td>
<td>260,605</td>
<td>18.0</td>
<td>296</td>
<td>1.6</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BUB</td>
<td>528,332</td>
<td>405.3</td>
<td>405.3</td>
<td>1.0</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>UB</td>
<td>30,838</td>
<td>23.9</td>
<td>11.8</td>
<td>2.0</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BL</td>
<td>2,614</td>
<td>1.0</td>
<td>290.4</td>
<td>290.4</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BS</td>
<td>2,614</td>
<td>1.0</td>
<td>1307.0</td>
<td>1307.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BR</td>
<td>2,614</td>
<td>1.0</td>
<td>1307.0</td>
<td>1307.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PR</td>
<td>1,084,287</td>
<td>7.1</td>
<td>6.2</td>
<td>0.00004</td>
<td>AR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AP</td>
<td>127,623</td>
<td>1.0</td>
<td>1263.6</td>
<td>1264.6</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Paper (P)</td>
<td>127,623</td>
<td>PC</td>
<td>127,623</td>
<td>1.0</td>
<td>3515.6</td>
<td>1598.0</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Author (A)</td>
<td>164,472</td>
<td>APC</td>
<td>232,659</td>
<td>2.2</td>
<td>3515.6</td>
<td>1598.0</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Reference (R)</td>
<td>147,251</td>
<td>AP</td>
<td>355,072</td>
<td>2.2</td>
<td>3515.6</td>
<td>1598.0</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Conference (C)</td>
<td>101</td>
<td>PR</td>
<td>392,519</td>
<td>3.1</td>
<td>2.7</td>
<td>1.1</td>
<td>0.00002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>APR</td>
<td>1,084,287</td>
<td>7.1</td>
<td>7.1</td>
<td>1.1</td>
<td>0.00004</td>
</tr>
</tbody>
</table>

\(^1\)https://dblp.uni-trier.de  
\(^2\)https://www.yelp.com/dataset/  
\(^3\)https://www.aminer.cn/citation

#### 4.2 Affiliation Relations and Interaction Relations

In order to explore the structural characteristics of relations, we present mathematical analysis on the above datasets.

Since the degree of nodes can well reflect the structures of networks (Wasserman and Faust 1994), we define a degree-based measure \( D(r) \) to explore the distinction of various relations in an HIN. Specifically, we compare the average degrees of two types of nodes connected with the relation \( r \), via dividing the larger one by the smaller one \( (D(r) \geq 1) \). Formally, given a relation \( r \) with nodes \( u \) and \( v \) (i.e., node-relation triple \( (u, r, v) \)), \( t_u \) and \( t_v \) are the node types of \( u \) and \( v \), we define \( D(r) \) as follows:

\[
D(r) = \frac{\max[\bar{d}_{t_u}, \bar{d}_{t_v}]}{\min[\bar{d}_{t_u}, \bar{d}_{t_v}]},
\]

where \( \bar{d}_{t_u} \) and \( \bar{d}_{t_v} \) are the average degrees of nodes of the types \( t_u \) and \( t_v \), respectively.

A large value of \( D(r) \) indicates quite inequivalent structural roles of two types of nodes connected via the relation \( r \) (one-centered-by-another), while a small value of \( D(r) \) means compatible structural roles (peer-to-peer). In other words, relations with a large value of \( D(r) \) show much stronger affiliation relationships. Nodes connected via such relations share much more similar properties (Faust 1997). While relations with a small value of \( D(r) \) implicate much stronger interaction relationships. Therefore, we call the two categories of relations as Affiliation Relations (ARs) and Interaction Relations (IRs), respectively.

In order to better understand the structural difference between various relations, we take the DBLP network as an example. As shown in Table 1, for the relation \( PC \) with
$D(PC) = 718.8$, the average degree of nodes with type P is 1.0 while that of nodes with type C is 718.8. It shows that papers and conferences are structurally inequivalent. Papers are centered by conferences. While $D(AP) = 1.1$ indicates that authors and papers are compatible and peer-to-peer in structure. This is consistent with our common sense. Semantically, the relation $PC$ means that ‘papers are published in conferences’, indicating an affiliation relationship. Differently, $AP$ means that ‘authors write papers’, which explicitly describes an interaction relationship.

In fact, we can also define some other measures to capture the structural difference. For example, we compare the relations in terms of sparsity, which can be defined as:

$$S(r) = \frac{N_r}{N_t^u \times N_t^v},$$  \hspace{1cm} (2)

where $N_r$ represents the number of relation instances following $r$. $N_t^u$ and $N_t^v$ mean the number of nodes with type $t_u$ and $t_v$, respectively. The measure can also consistently distinguish the relations into two categories: ARs and IRs. The detailed statistics of all the relations in the three HINs are shown in Table 1.

Evidently, Affiliation Relations and Interaction Relations exhibit rather distinct characteristics: (1) ARs indicate one-centered-by-another structures, where the average degrees of the types of end nodes are extremely different. They imply an affiliation relationship between nodes. (2) IRs describe peer-to-peer structures, where the average degrees of the types of end nodes are compatible. They suggest an interaction relationship between nodes.

# 5 Relation Structure-Aware HIN Embedding

In this section, we present a novel Relation structure-aware HIN Embedding model (RHINE), which individually handles two categories of relations (ARs and IRs) with different models in order to preserve their distinct structural characteristics, as illustrated in Figure 1(c).

## 5.1 Basic Idea

Through our exploration with thorough mathematical analysis, we find that the heterogeneous relations can be typically divided into ARs and IRs with different structural characteristics. In order to respect their distinct characteristics, we need to specifically design different while appropriate models for the different categories of relations. For ARs, we propose to take Euclidean distance as a metric to measure the proximity of the connected nodes in the low-dimensional space. There are two motivations behind this: (1) First of all, ARs show affiliation structures between nodes, which indicate that nodes connected via such relations share similar properties. (Faust 1997; Yang and Leskovec 2012). Hence, nodes connected via ARs could be directly close to each other in the vector space, which is also consistent with the optimization of Euclidean distance (Danielsson 1980). (2) Additionally, one goal of HIN embedding is to preserve the high-order proximity. Euclidean distance can ensure that both first-order and second-order proximities are preserved as it meets the condition of the triangle inequality (Hsieh et al. 2017).

Different from ARs, IRs indicate strong interaction relationships between compatible nodes, which themselves contain important structural information of two nodes. Thus, we propose to explicitly model an IR as a translation between nodes in the low-dimensional vector space. Additionally, the translation based distance is consistent with the Euclidean distance in the mathematical form (Bordes et al. 2013). Therefore, they can be smoothly combined in a unified and elegant manner.

## 5.2 Different Models for ARs and IRs

In this subsection, we introduce two different models exploited in RHINE for ARs and IRs, respectively.

**Euclidean Distance for Affiliation Relations** Nodes connected via ARs share similar properties (Faust 1997), therefore nodes could be directly close to each other in the vector space. We take the Euclidean distance as the proximity measure of two nodes connected by an AR.

Formally, given an affiliation node-relation triple $(p, s, q) \in P_{AR}$ where $s \in R_{AR}$ is the relation between $p$ and $q$ with weight $w_{pq}$, the distance between $p$ and $q$ in the latent vector space is calculated as follows:

$$f(p, q) = w_{pq}\|X_p - X_q\|_2^2,$$  \hspace{1cm} (3)

in which $X_p \in \mathbb{R}^d$ and $X_q \in \mathbb{R}^d$ are the embedding vectors of $p$ and $q$, respectively. As $f(p, q)$ quantifies the distance between $p$ and $q$ in the low-dimensional vector space, we aim to minimize $f(p, q)$ to ensure that nodes connected by an AR should be close to each other. Hence, we define the margin-based loss (Bordes et al. 2013) function as follows:

$$L_{Eu AR} = \sum_{s \in R_{AR}} \sum_{(p, s, q) \in P_{AR}} \sum_{(p', s') \in R_{AR}} \sum_{(p, q) \in P_{AR}} \max[0, \gamma + f(p, q) - f(p', q')],$$  \hspace{1cm} (4)

where $\gamma > 0$ is a margin hyperparameter. $P_{AR}$ is the set of positive affiliation node-relation triples, while $P'_{AR}$ is the set of negative affiliation node-relation triples.

**Translation-based Distance for Interaction Relations** Interaction Relations demonstrate strong interactions between nodes with compatible structural roles. Thus, different from ARs, we explicitly model IRs as translations between nodes.

Formally, given an interaction node-relation triple $(u, r, v)$ where $r \in R_{IR}$ with weight $w_{uv}$, we define the score function as:

$$g(u, v) = w_{uv}\|X_u + Y_r - X_v\|,$$  \hspace{1cm} (5)

where $X_u$ and $X_v$ are the node embeddings of $u$ and $v$ respectively, and $Y_r$ is the embedding of the relation $r$. Intuitively, this score function penalizes deviation of $(X_u + Y_r)$ from the vector $X_v$.

For each interaction node-relation triple $(u, r, v) \in P_{IR}$, we define the margin-based loss function as follows:

$$L_{Tr IR} = \sum_{r \in R_{IR}} \sum_{(u, r, v) \in P_{IR}} \sum_{(u', v') \in P'_{IR}} \max[0, \gamma + g(u, v) - g(u', v')]$$  \hspace{1cm} (6)
where $P_{IR}$ is the set of positive interaction node-relation triples, while $P'_{IR}$ is the set of negative interaction node-relation triples.

### 5.3 A Unified Model for HIN Embedding

Finally, we smoothly combine the two models for different categories of relations by minimizing the following loss function:

$$L = L_{E_{uAR}} + L_{T_{vIR}}$$

(7)

$$= \sum_{s \in R_{AR}} \sum_{(p,s,q) \in P_{AR}} \sum_{(p',s',q') \in P'_{AR}} \max[0, \gamma + f(p, q) - f(p', q')]$$

$$+ \sum_{r \in R_{IR}} \sum_{(u,r,v) \in P_{IR}} \sum_{(u',r,v') \in P'_{IR}} \max[0, \gamma + g(u, v) - g(u', v')]$$

#### Sampling Strategy

As shown in Table 1, the distributions of ARs and IRs are quite unbalanced. What’s more, the proportion of relations are unbalanced within ARs and IRs. Traditional edge sampling may suffer from under-sampling for relations with a small amount or oversampling for relations with a large amount. To address the problems, we draw positive samples according to their probability distributions. As for negative samples, we follow previous work (Bordes et al. 2013) to construct a set of negative node-relation triples

$$P'_N = \{(u', r, v) \mid u' \in V \} \cup \{(u, r, v') \mid v' \in V \}$$

for the positive node-relation triple $(u, r, v)$, where either the head or tail is replaced by a random node, but not both at the same time.

### 6 Experiments

In this section, we conduct extensive experiments to demonstrate the effectiveness of our model RHINE.

#### 6.1 Datasets

As described in Subsection 4.1, we conduct experiments on three datasets, including DBLP, Yelp and AMiner. The statistics of them are summarized in Table 1.

#### 6.2 Baseline Methods

We compare our proposed model RHINE with six state-of-the-art network embedding methods.

- **DeepWalk** (Perozzi, Al-Rfou, and Skiena 2014) performs a random walk on networks and then learns low-dimensional node vectors via the skip-gram model.

- **LINE** (Tang et al. 2015) considers first-order and second-order proximities in networks. We denote the model that only uses first-order or second-order proximity as LINE-1st or LINE-2nd, respectively.

- **PTE** (Tang, Qu, and Mei 2015) decomposes an HIN to a set of bipartite networks and then learns the low-dimensional representation of the network.

- **ESim** (Shang et al. 2016) takes a given set of meta-paths as input to learn a low-dimensional vector space. For a fair comparison, we use the same meta-paths with equal weights in ESim and our model RHINE.

- **HIN2Vec** (Fu, Lee, and Lei 2017) learns the latent vectors of nodes and meta-paths in an HIN by conducting multiple prediction training tasks jointly.

- **Metapath2vec** (Dong, Chawla, and Swami 2017) leverages meta-path based random walks and skip-gram model to perform node embedding. We leverage the meta-paths APCPA, UBSBU and APCPA in DBLP, Yelp and AMiner respectively, which perform best in the evaluations.

#### Parameter Settings

For a fair comparison, we set the embedding dimension $d = 100$ and the size of negative samples $k = 3$ for all models. For DeepWalk, HIN2Vec and metapath2vec, we set the number of walks per node $w = 10$, the walk length $l = 100$ and the window size $\tau = 5$. For our model RHINE, the margin $\gamma$ is set to 1.

#### 6.3 Node Clustering

**Experimental Settings** We leverage K-means to cluster the nodes and evaluate the results in terms of normalized mutual information (NMI) (Shi et al. 2014).

**Results** As shown in Table 2, our model RHINE significantly outperforms all the compared methods. (1) Compared with the best competitors, the clustering performance of our model RHINE improves by 18.79%, 6.15% and 7.84% on DBLP, Yelp and AMiner, respectively. It demonstrates the effectiveness of our model RHINE by distinguishing the various relations with different structural characteristics in an HIN. In addition, it also validates that we utilize appropriate models for different categories of relations. (2) In all baseline methods, homogeneous network embedding models achieve the lowest performance, because they ignore the heterogeneity of relations and nodes. (3) RHINE significantly outperforms existing HIN embedding models (i.e., ESim, HIN2Vec and metapath2vec) on all datasets. We believe the reason is that our proposed RHINE with appropriate models for different categories of relations can better capture the structural and semantic information of HINs.

#### 6.4 Link Prediction

**Experimental Setting** We model the link prediction problem as a binary classification problem that aims to predict whether a link exists. In this task, we conduct co-author (A-A) and author-conference (A-C) link prediction for DBLP and AMiner. For Yelp, we predict user-business (U-B) links which indicate whether a user reviews a business. We first

<table>
<thead>
<tr>
<th>Methods</th>
<th>DBLP</th>
<th>Yelp</th>
<th>AMiner</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepWalk</td>
<td>0.3884</td>
<td>0.3043</td>
<td>0.5427</td>
</tr>
<tr>
<td>LINE-1st</td>
<td>0.2775</td>
<td>0.3103</td>
<td>0.3736</td>
</tr>
<tr>
<td>LINE-2nd</td>
<td>0.4675</td>
<td>0.3593</td>
<td>0.3862</td>
</tr>
<tr>
<td>PTE</td>
<td>0.3101</td>
<td>0.3527</td>
<td>0.4089</td>
</tr>
<tr>
<td>ESim</td>
<td>0.3449</td>
<td>0.2214</td>
<td>0.3409</td>
</tr>
<tr>
<td>HIN2Vec</td>
<td>0.4256</td>
<td>0.3657</td>
<td>0.3948</td>
</tr>
<tr>
<td>metapath2vec</td>
<td>0.6065</td>
<td>0.3507</td>
<td>0.5586</td>
</tr>
<tr>
<td>RHINE</td>
<td><strong>0.7204</strong></td>
<td><strong>0.3882</strong></td>
<td><strong>0.6024</strong></td>
</tr>
</tbody>
</table>

- RHINE
Table 3: Performance Evaluation of Link Prediction.

<table>
<thead>
<tr>
<th>Methods</th>
<th>DBLP (A-A)</th>
<th>DBLP (A-C)</th>
<th>Yelp (U-B)</th>
<th>AMiner (A-A)</th>
<th>AMiner (A-C)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC</td>
<td>F1</td>
<td>AUC</td>
<td>F1</td>
<td>AUC</td>
</tr>
<tr>
<td>DeepWalk</td>
<td>0.9131</td>
<td>0.8246</td>
<td>0.7634</td>
<td>0.7047</td>
<td>0.9122</td>
</tr>
<tr>
<td>LINE-1st</td>
<td>0.8264</td>
<td>0.7233</td>
<td>0.5335</td>
<td>0.6436</td>
<td>0.5084</td>
</tr>
<tr>
<td>LINE-2nd</td>
<td>0.7448</td>
<td>0.6741</td>
<td>0.8340</td>
<td>0.7396</td>
<td>0.7509</td>
</tr>
<tr>
<td>PTE</td>
<td>0.8853</td>
<td>0.8331</td>
<td>0.8843</td>
<td>0.7720</td>
<td>0.8061</td>
</tr>
<tr>
<td>ESim</td>
<td>0.9077</td>
<td>0.8129</td>
<td>0.7736</td>
<td>0.6795</td>
<td>0.6160</td>
</tr>
<tr>
<td>HIN2Vec</td>
<td>0.9160</td>
<td>0.8475</td>
<td>0.8966</td>
<td>0.7892</td>
<td>0.8653</td>
</tr>
<tr>
<td>metapath2vec</td>
<td>0.9153</td>
<td>0.8431</td>
<td>0.8987</td>
<td>0.8012</td>
<td>0.7818</td>
</tr>
<tr>
<td>RHINE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Performance Evaluation of Multi-class Classification.

<table>
<thead>
<tr>
<th>Methods</th>
<th>DBLP</th>
<th>Yelp</th>
<th>AMiner</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Macro-F1</td>
<td>Micro-F1</td>
<td>Macro-F1</td>
</tr>
<tr>
<td>DeepWalk</td>
<td>0.7475</td>
<td>0.7500</td>
<td>0.6723</td>
</tr>
<tr>
<td>LINE-1st</td>
<td>0.8091</td>
<td>0.8250</td>
<td>0.4872</td>
</tr>
<tr>
<td>LINE-2nd</td>
<td>0.7559</td>
<td>0.7500</td>
<td>0.5304</td>
</tr>
<tr>
<td>PTE</td>
<td>0.8852</td>
<td>0.8750</td>
<td>0.5389</td>
</tr>
<tr>
<td>ESim</td>
<td>0.8867</td>
<td>0.8750</td>
<td>0.6836</td>
</tr>
<tr>
<td>HIN2Vec</td>
<td>0.8631</td>
<td>0.8500</td>
<td>0.6075</td>
</tr>
<tr>
<td>metapath2vec</td>
<td>0.8976</td>
<td>0.9000</td>
<td>0.5337</td>
</tr>
<tr>
<td>RHINE</td>
<td>0.9344</td>
<td>0.9250</td>
<td>0.7132</td>
</tr>
</tbody>
</table>

randomly separate the original network into training network and testing network, where the training network contains 80% relations to be predicted (i.e., A-A, A-C and U-B) and the testing network contains the rest. Then, we train the embedding vectors on the training network and evaluate the prediction performance on the testing network.

Results The results of link prediction task are reported in Table 3 with respect to AUC and F1 score. It is clear that our model performs better than all baseline methods on three datasets. The reason behind the improvement is that our model based on Euclidean distance modeling relations can capture both the first-order and second-order proximities. In addition, our model RHINE distinguishes multiple types of relations into two categories in terms of their structural characteristics, and thus can learn better embeddings of nodes, which are beneficial for predicting complex relationships between two nodes.

6.5 Multi-Class Classification

Experimental Setting In this task, we employ the same labeled data used in the node clustering task. After learning the node vectors, we train a logistic classifier with 80% of the labeled nodes and test with the remaining data. We use Micro-F1 and Macro-F1 score as the metrics for evaluation (Dong, Chawla, and Swami 2017).

Results We summarize the results of classification in Table 4. As we can observe, (1) RHINE achieves better performance than all baseline methods on all datasets except Aminer. It improves the performance of node classification by about 4% on both DBLP and Yelp averagely. In terms of Aminer, the RHINE performs slightly worse than ESim, HIN2vec and metapath2vec. This may be caused by over-capturing the information of relations PR and APR (R represents references). Since an author may write a paper referring to various fields, these relations may introduce some noise. (2) Although ESim and HIN2Vec can model multiple types of relations in HINs, they fail to perform well in most cases. Our model RHINE achieves good performance due to the respect of distinct characteristics of various relations.

6.6 Comparison of Variant Models

In order to verify the effectiveness of distinguishing the structural characteristics of relations, we design three variant models based on RHINE as follows:

- **RHINE_{Eu}** leverages Euclidean distance to embed HINs without distinguishing the relations.
- **RHINE_{Tr}** models all nodes and relations in HINs with translation mechanism, which is just like TransE (Bordes et al. 2013).
- **RHINE_{Re}** leverages Euclidean distance to model IRs while translation mechanism for ARs, reversely.

We set the parameters of variant models as the same as those of our proposed model RHINE. The results of the three tasks are shown in Figure 2. It is evident that our model outperforms RHINE_{Eu} and RHINE_{Tr}, indicating that it is beneficial for learning the representations of nodes by distinguishing the heterogeneous relations. Besides, we find that
RHINE\textsubscript{T_r} achieves better performance than RHINE\textsubscript{E_r}. This is due to the fact that there are generally more peer-to-peer relationships (i.e., IRs) in the networks. Directly making all nodes close to each other leads to much loss of information. Compared with the reverse model RHINE\textsubscript{R_e}, RHINE also achieves better performance on all tasks, which implies that two models for ARs and IRs are well designed to capture their distinctive characteristics.

6.7 Visualization
To understand the representations of the networks intuitively, we visualize the vectors of nodes (i.e., papers) in DBLP learned with DeepWalk, metapath2vec and RHINE in Figure 3. As we can see, our model clearly clusters the paper nodes into four groups. It demonstrates that our model learns superior node embeddings by distinguishing the heterogeneous relations in HINs. In contrast, DeepWalk barely splits papers into different groups. Metapath2vec performs better than DeepWalk, but the boundary is blurry.

6.8 Parameter Analysis
In order to investigate the influences of different parameters in our model, we evaluate the RHINE in node clustering task. Specifically, we explore the sensitivity of two parameters, including the number of embedding dimensions and the number of negative samples. As shown in Figure 4(a), the performance of our model improves with the increase in the number of dimensions, and then tends to be stable once the dimension of the representation reaches around 100. Similarly, Figure 4(b) shows that as the number of negative ex-

Figure 2: Performance Evaluation of Variant Models.

Figure 3: Visualization of Node Embeddings.

Figure 4: Parameter Analysis.

amps increases, the performance of our model first grows and then becomes stable when the number reaches 3.

7 Conclusion
In this paper, we make the first attempt to explore and distinguish the structural characteristics of relations for HIN embedding. We present two structure-related measures which can consistently distinguish heterogeneous relations into two categories: affiliation relations and interaction relations. To respect the distinctive structures of relations, we propose a novel relation structure-aware HIN embedding model (RHINE), which individually handles these two categories of relations. Experimental results demonstrate that RHINE outperforms state-of-the-art baselines in various tasks. In the future, we will explore other possible measures to differentiate relations so that we can better capture the structural information of HINs. In addition, we will exploit deep neural network based models for different relations.
8 Acknowledgments

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

References


Cui, P.; Wang, X.; Pei, J.; and Zhu, W. 2018. A survey on network embedding. IEEE Transactions on Knowledge and Data Engineering.


Sun, Y.; Norick, B.; Han, J.; Yan, X.; Yu, P. S.; and Yu, X. 2013. Pathselclus: Integrating meta-path selection with user-guided object clustering in heterogeneous information networks. ACM Transactions on Knowledge Discovery from Data (TKDD) 7(3):11.


