2018/10/18应该值得一看

PME: Projected Metric Embedding on Heterogeneous Networks for Link Prediction

Hongxu Chen
The University of Queensland
Brisbane, QLD, Australia
hongxu.chen@uq.edu.au

Hongzhi Yin*
The University of Queensland
Brisbane, QLD, Australia
h.yin1@uq.edu.au

Weiqing Wang
Monash University
Melbourne, VIC, Australia
teresa.wang@monash.edu

Hao Wang
360 Search Lab
China
cashenry@126.com

Quoc Viet Hung Nguyen
Griffith University
Gold Coast, Australia
quocviethung1@gmail.com

Xue Li†
The University of Queensland
Brisbane, QLD, Australia
xueli@itee.uq.edu.au

采用dot product在低维空间度量近似度，仅能保留一阶相似度，保存全局相似度不足。

ABSTRACT

全局和局部网络结构都保存

对比同构信息网络多种类型的link, 关于关系的link分布严重倾斜

Heterogenous information network embedding aims to embed heterogenous information networks (HINs) into low dimensional spaces, in which each vertex is represented as a low-dimensional vector, and both global and local network structures in the original space are preserved. However, most of existing heterogenous information network embedding models adopt the dot product to measure the proximity in the low dimensional space, and thus they can only preserve the first-order proximity and are insufficient to capture the global structure. Compared with homogenous information networks, there are multiple types of links (i.e., multiple relations) in HINs, and the link distribution w.r.t relations is highly skewed.

To address the above challenging issues, we propose a novel heterogenous information network embedding model PME based on the metric learning to capture both first-order and second-order proximities in a unified way. To alleviate the potential geometrical inflexibility of existing metric learning approaches, we propose to build object and relation embeddings in separate object space and relation spaces rather than in a common space. Afterwards, we learn embeddings by firstly projecting vertices from object space to corresponding relation space and then calculate the proximity between projected vertices. To overcome the heavy skewness of the link distribution w.r.t relations and avoid "over-sampling" or "under-sampling" for each relation, we propose a novel loss-aware adaptive sampling approach for the model optimization. Extensive experiments have been conducted on a large-scale HIN dataset, and the experimental results show superiority of our proposed PME model in terms of prediction accuracy and scalability.

为了减轻metric learning 潜在的几何不灵活性，分别在对象空间和关系空间中，构建对象和关系的embedding. 而不是在同一个空间中。
基于metric learing提出PME,异构网络embedding模型，以统一的方式同时捕获一阶和二阶相似度

*This author is the corresponding author.
†The SSP at Nanjing University of Aeronautics and Astronautics, Nanjing, China.

KDD '18, August 19–23, 2018, London, United Kingdom

ACM ISBN 978-1-4503-5552-0/18/08…$15.00
https://doi.org/10.1145/3219819.3219986

将节点从对象空间 映射到 相应的关系空间，然后计算节点之间的相似度，以此来学习 embedding.
为了克服关于关系的link不平衡性，避免过渡采样和欠采样提出loss aware的采样方法。

CCS CONCEPTS

• **Information systems → Data mining**;

KEYWORDS

Heterogenous Network Embedding; Link Prediction;

ACM Reference Format:
Hongxu Chen, Hongzhi Yin, Weiqing Wang, Hao Wang, Quoc Viet Hung Nguyen, and Xue Li. 2018. PME: Projected Metric Embedding on Heterogeneous Networks for Link Prediction. In *KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19–23, 2018, London, United Kingdom.* ACM, London, United Kingdoms, 10 pages. https://doi.org/10.1145/3219819.3219986

在大数据时代，大规模的信息网络无所不在

1 INTRODUCTION

传统图表示有问题

In the era of Big Data, large-scale information networks are becoming ubiquitous in the real world, such as social networks, publication networks, E-commerce information networks and knowledge base graphs. Traditionally, an information network is represented as a graph $G =< V, E >$, where V is vertex set representing the nodes in a network, and E is an edge set representing the relationships among nodes. However, for large-scale information networks, the traditional graph-based representation poses a great challenge to numerous applications that search and mine information in them such as link prediction, node classification, clustering, and recommendation [33–38], due to the high computational complexity [8]. Recently, this motivates a lot of research interests [8] in network embedding techniques that aim to embed information networks into low dimensional vector spaces, in which every vertex is represented as a low-dimensional vector. A good embedding can preserve the proximity (i.e., similarity) between vertices in the original information network. Then, various search and mining tasks can be efficiently done in the embedded space with the help of off-the-shelf multidimensional indexing approaches and machine learning techniques for vector spaces.

对大量的应用提出了挑战。高的计算复杂度

刺激在网络嵌入技术方面的兴趣

现成的

While information network embedding has recently received a tremendous amount of research attention, most of them (e.g., LINE [26], DeepWalk [18], node2vec [11]) are focused on homogeneous network embedding that equally treats each type of nodes and each type of links. Heterogeneous information networks (HINs), such as publication networks [26], knowledge base graph [15] and E-commerce information networks, contain multiple types of nodes

很多关注同构网络的嵌入，

尽管信息网络嵌入已经收到大量的研究关注。
对等的看待每个类型的节点和每个类型链接。

本文与这篇文献很相似

它试图把观测节点对拉倒相同点，但是每个节点有很多邻居。这种内在的几何不灵活性导致有害的影响，当数据集规模较大时。使得所有的节点邻居在相同点。另一方面，一个对象有多面性，各种关系关注对象的不同面，具有不同的语义。

异构信息网络包含多种类型的节点，包含多种语义的多类型边。

dot product不行，所以弄 metric learning

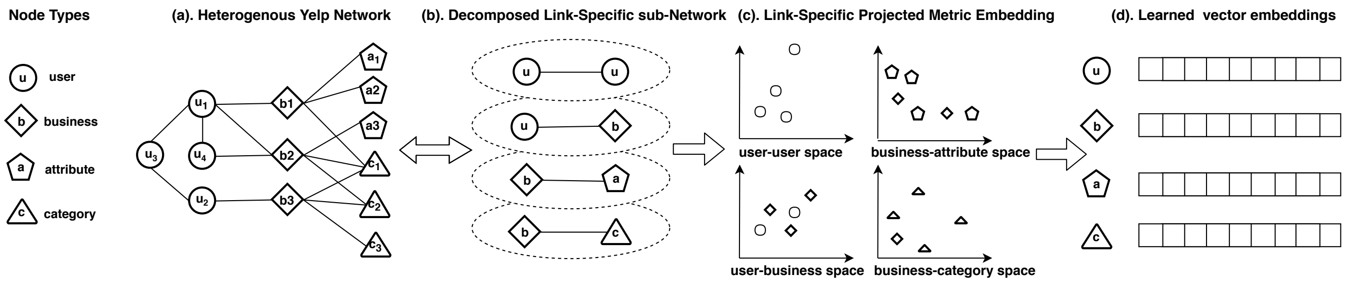


Figure 1: An illustrative example of a heterogenous social network (Yelp), and PME model architecture for embedding this heterogeneous network (a). A heterogenous Yelp network consists of four types of nodes (i.e., user, business, attribute, category), which are connected via 4 relations (i.e., user-user, user-business, business-attribute, business-category). (b). The heterogenous Yelp network can be decomposed into as 4 bipartite networks based on the 4 relations (i.e., user-user, user-business, business-attribute, and business-category). (c). Each bipartite network is projected to a relation specific semantic space in which the proximity of two vertices is measured by the Euclidian distance. (d). By joint learning multiple semantic-specific Euclidian spaces, the low dimensional vector representations for each vertex are learned.

compared to / with都可以 哦

and edges with diverse semantics. For example, in the Yelp platform, various types of objects (users, business, business attributes such as locations, and business categories) are inter-connected via multiple types of links (i.e., user-user friend relation, user-business interaction relation, business-attribute description relation, business-category classification relation). Compared to homogeneous network embedding, the proximity between objects in a HIN is not just a measure of closeness or distance, but it is also based on semantics. For example, in the HIN of Figure 1, vertex u_1 is close to both b_2 and u_3, but these relationships have different semantics. b_2 is a business visited by user u_1, while u_3 is a friend of u_1.

多种类型的link

对象之间的相似度不仅仅是紧密度或者距离的度量也是基于语义的

To model semantic-specific relationships, Huang and Mamoulis [13] introduced meta paths (i.e., a sequence of object types with edge types in between) in their heterogeneous information network embedding algorithm, and Xu et al. [32] introduced a harmonious embedding matrix to measure the proximity between nodes of different types in their proposed coupled heterogeneous network embedding method. However, both of them employed the dot product to compute the proximity between nodes in the low dimensional vector space. The dot product is not a metric based on distance learning, as it does not meet the condition of the triangle inequality that is the most crucial to the generalization of a learned metric [12]. The triangle inequality states that for any three objects, the sum of any two pairwise distance should be greater than or equal to the remaining pairwise distance. For example, vertices a and b are both close to vertex c. The triangle inequality implies a is also close to b. Therefore, these existing heterogeneous information network embedding approaches based on dot product proximity can only capture local structures represented by observed links in networks and preserve the first-order proximity (e.g., both a and b are close to c), but fail to capture the second-order proximity between vertices (e.g., a and b is also close) that is determined by the shared neighborhood structures of the vertices. It has been widely acknowledged that the first-order proximity is not sufficient for preserving the global network structures [26].

是这样吗 这个通过相似传递就可以呀

不能保存二阶相似度 由共享邻居结构决定

一阶相似对于保存全局结构不足。

In light of this, we propose a **Projected Metric Embedding** model (PME) for HIN embedding based on the metric or distance learning, to simultaneously preserve both first-order and second-order proximity in a unified and elegant way. Specifically, for each node in the HIN, we learn a low dimensional vector such that distances between pairs of nodes with observed links are smaller than those pairs of nodes without observed links in the latent space. However, directly applying the Euclidean distance as a metric will be problematic from both intuitive and mathematical perspectives. Mathematically, it is *geometrically restrictive* and also leads to an *ill-posed algebraic system* since it tries to fit each pair of linked nodes into the same point in the low dimensional space, but each node may have many neighbors. This intrinsic geometric inflexibility causes adverse repercussions when the dataset is large since it tries to force all of a node's neighbors onto the same point [28]. On the other hand, an object may have multiple aspects, and various relations focus on different aspects of objects and have different semantics in HINs.

基于metric learning，提出PME, 以统一的方式保存一阶和二阶相似度。

学习一个低维表示，可以观测到链接之间的节点距离小于没有观测到节点对。然而，直接采用欧几里得作为度量是有问题的，从直观和数学的角度。数学上，几何的约束严格，导致有训的代数系统。

To address the above issues, our PME introduces a relation-specific projection embedding matrix so that we model objects and relations in distinct spaces, i.e., one shared object space and multiple relation spaces (i.e., relation-specific object spaces), and performs proximity calculation via the Euclidean distance in the corresponding relation space. Hence, it is possible that some objects are far away from each other in the object space, but are close to each other in the corresponding relation spaces. This allows for a greater extent of geometric flexibility and modelling capability. The basic idea of PME is illustrated in Figure 1. For each observed link (v_i, v_j), vertices in the object space are first projected into r-relation space as v_i^r and v_j^r with operation matrix $\mathbf{M}_r$. The relation-specific projection can make vertices that actually hold the relation close with each other, and also get far away from those that do not hold the relation. As the number of unobserved links is huge in HINs, we adopt the bidirectional popularity-biased negative-sampling approach [33] to optimize our PME model, inspired by the good performance of the negative sampling-based optimization method in recent network embedding models such as LINE [26], PTE [25] and EOE [32].

由于负样本多，采取这种采样

为了建模指定语义的关系，huang等人介绍了meta path在异构网络嵌入中；Xu等人介绍Harmonious嵌入矩阵来度量不同类型节点之间的近似度。
然而，他们采用内积在低维空间中 计算 节点之间的相似度，内积不是基于距离学习的一个metric,它不满足三角不等式条件，这个三角不等式对于学习道德metric的泛化非常的关键。这是CML中描述的。
因此，基于内积的异构网络嵌入方法 仅能捕获局部结构，这些局部结构由观测到的link表示

为了处理上面的问题，PME介绍一种指定关系的projection embedding 矩阵，使得对象和关系在不同的空间中，一个对象空间，和多个关系空间，即关系指定对象空间。通过欧几里得距离在相应关系空间执行近似计算。因此，一些对象在对象空间很远，但是在关系空间很近。这赋予较大的结合灵活性和建模能力

对比同构信息网络，HIN中有多种乐行关系，观测到link关于关系的分布验证倾斜。对于模型的优化，提出了较大的挑战。在模型训练工程中，如果我们统一采样，执行随机梯度下降，那么大于一般的采样观测链接将来自user-user关系，导致训练模型不能保存user/attribute 和business/category关系。为了在联合训练多种类型关系时处理这个挑战，Tang的提出了对于每种关系，交替采样观测链接。因此每个关系收到相同数量的训练样本。这导致具有少数量观测链接的关系过渡采样，具有大量观测链接的关系采样不足。另外，保存节点之间相似度 在每个关系中的困难程度是不同的。因此对于每种关系，需要的训练样本不一样。而且，这个困难和需要的样本是动态变化的。

Compared with homogenous networks, there are multiple relations in a HIN and the distribution of observed links w.r.t. relations is heavily skewed, which poses a great challenge for the model optimization of PME. Take the Yelp dataset for example. More than half (54%) of observed links are user-user links, and second comes with user-business links, which takes up 34.3%. In contrast, business-attribute and business-category only take up 5.7% and 5.8%, respectively. During the model training, if we uniformly draw an observed link and perform stochastic gradient descent on the drawn case, just as done in the standard stochastic gradient descent algorithm, more than half of the sampled observed links would belong to the user-user relation, and it would lead to that the trained model may not be able to preserve the network structures of business-attribute and business-category relations. To address the challenge in the joint training of multiple relations, Tang et al. [25] proposed to alternatively sample observed links from each relation. Specifically, they first uniformly draw a relation, and then randomly sample an observed link from the drawn relation. Thus, each relation would receive the same number of training examples. It would result in that relations with a small number of observed links are over-sampled while those with a large number of links are under-sampled. Besides, the difficulty of preserving the proximity between pairs of vertices in each relation is different, therefore the required number of training examples for each relation should also be different. Moreover, both the difficulty of preserving the proximity and the required number of training examples for each relation are dynamically changing as the model parameters are updated during the model training process. To overcome the heavy skewness of the heterogeneous link distribution, we propose a novel loss-aware adaptive sampling approach to draw observed links for model optimization. The basic idea is that the relations with a larger loss should have a higher probability to be sampled, as they need more training examples to correct the current model parameters.

为了克服这个严重倾斜性，提出loss aware的采样方案。基本思想是loss的关系具有较大采样概率，由于他们需要更多的训练样本来矫正模型参数。

To summarize, we make the following contributions:

(1) We propose a novel heterogeneous information network embedding model called "PME", which suits arbitrary types of large-scale heterogeneous information networks. It learns a distance metric to preserve both the first-order and second-order proximities in a unified and elegant way, and introduces distinct latent spaces to model objects and relations to alleviate the potential geometrical inflexibility of existing metric learning approaches and scale to a larger number of links.
(2) To overcome the heavy skewness of the heterogeneous link distribution w.r.t relations, we propose a novel loss-aware adaptive sampling approach to draw training examples in the model optimization.
(3) We conduct extensive experiments to evaluate the performance of PME in terms of prediction accuracy and scalability on a large-scale HIN published by Yelp. The results show the superiority of our proposals by comparing with the state-of-the-art techniques.

The remainder of the paper is organized as follows. Section 2 reviews the related work, and Section 3 introduce the preliminaries. Section 4 details our proposed PME model. Section 5 reports the experimental setup and results, and Section 6 concludes the paper.

提出PME,适合任意类型的HIN.它学习距离度量保存一阶和二阶相似度。介绍不同的对象和关系空间现有metric learning内在的几何不灵活性。
为了处理严重倾斜性，提出loss aware的采样方法来进行模型优化。

2 RELATED WORK

We first introduce the related methods of general network embedding, and then discuss the recent works on heterogeneous network embedding.

原始上，是针对网络特征进行纬度约减的工具，如SVD,MDS, ISOMap 等，背后的思想是学习低维表示，保存网络特征的主体。

2.1 Network Embedding

Originally, graph or network embedding methods were proposed as tools of dimension reduction for network features, such as linear methods based on SVD [27], multi-dimensional scaling (MDS) [39], IsoMap [2] , Spectral clustering [17] and Laplacian Eigenmap [29]. The ideas behind those methods are to learn low dimensional latent factors that can preserve the majority of network features. However, these methods are not applicable for current large information networks because of their low efficiency and large computational complexity. Another graph embedding method called graph factorization [1] works out the low dimensional latent embeddings of a large graph through Matrix Factorization by utilizing network edges. It presents graphs as matrices where matrix elements correspond to edges between vertices. However, the graph factorization methods only preserve linkage information of directly linked nodes so it is insufficient for leaning the high-order proximity of a network. Moreover, representation learning on knowledge graphs is also related to our work. The representative methods such as [4] and Trans-family models (TransE [3], TransH [30], TransR[15]) have been shown effective for modelling knowledge bases. Our idea of building projection matrices for different relations is inspired by TransR but designed for different purposes (to alleviate geometric inflexibility when performing metric learning). Recently, With the advances in language modelling [16], skip-gram algorithm shows its superiority in modelling sentences by capturing the neighbour words concurrencies. Inspired by this idea, DeepWalk [18] was proposed to embed network structures by using local information obtained from truncated random walks as the equivalence of sentences. Along this line of research, node2vec [11] is another representative method. Besides, LINE [26] was proposed as an efficient network embedding method, has shown its robustness and effectiveness in dealing with large-scale information networks. Although it is proposed to be able to preserve both local and global proximity of the network vertices, it didn't consider the heterogeneity of complex information network.

由于低效率和计算复杂度，不能应用到大规模信息网络。
GF:利用网络边，进行矩阵分解，学习低维表示。仅能保存具有直接链接节点的linkage信息，不足以学习高阶相似性。
尽管可以保存网络结构的局部和全局相似性。

知识图谱上的表示学习与我们的工作相关。为不同关系构建projection矩阵受TransR启发，但是目的不同。

2.2 Heterogeneous Network Embedding

Different from homogenous networks, heterogeneous networks consist of different types of nodes and links. Although general network embedding methods might be applied by treating every node in the networks as the same type, it is still an interesting and challenging problem to develop more dedicated methods for modelling the heterogeneous types of nodes and links in a unified way.

针对节点分类的异构网络嵌入方法

A heterogeneous social network embedding algorithm [14] for classifying nodes was proposed by Yann et. al. They learn the representations of all types of nodes in a common vector space, and perform the inference in that space. In [5], a deep embedding method

最近，由于语言模型的进展，skip gram显示在建模句子方面的优越性，通过捕获邻居词汇之间的共现性。受此启发，DeepWalk 使用来自随机游走得到的局部信息作为句子等价体。提出嵌入网络结构。Line是一个有效的网络嵌入方法，已经显示健壮性和有效性在处理大规模网络方面

针对节点分类的异构网络嵌入方法。在统一的向量空间学习所有类型节点的表示，也在此空间执行推理。
在文献5中，深度嵌入的方法学习不同类型网络结构中的节点表示。他们使用CNN模型和全联通层学习图像和文本表示。然后把图像和文本表示 映射到一个共同空间。因此来自不同莫泰的数据之间的相似度可以直接度量。

类似于一般的网络嵌入，随机游走也被应用异构网络嵌入上。利用关于不同节点类型的预定义meta-path来指导随机游走学习网络结构。采用类似LIne的方法在低维空间保存近似度。

for heterogeneous network was proposed to learn the representations of nodes with different types of network structures. They use a CNN model and a fully connected layer to learn the embeddings of images and texts respectively, and then map the images and texts embeddings to a common space so that the similarities between data from different modalities can be directly measured.

Similar as general network embedding, the random walk process is also applied for heterogeneous networks embedding (i.e., metapath2vec [9]), which leverages the pre-defined meta-paths [23] w.r.t different node types to guide the random walk process to learn network structures. They adopt a similar strategy as LINE to preserve the proximity in the low dimensional space. Meta-path based methods also includes [21] [22] [23] [24].

PTE是Line在异构网络上的扩展。首先将异构网络分成多个二分图，然后独立执行Line.然后，不同网络节点学习是通过联合优化线性组合的目标函数。

PTE [25] was proposed as an extension of LINE to suite the heterogeneous networks. It first partitions a heterogeneous network into multiple bipartite graphs and performs network embedding individually by using LINE. Then, the representations of different network nodes can be learned by jointly optimizing the linearly combined objective function. PTE also addressed the sampling problem in heterogeneous network embedding by alternatively sampling positive edges from each type of edges. However, it is still problematic when various types of links are heavily unbalanced distributed. Moreover, PTE models vertices into a single space will make it difficult to distinguish the heterogeneity among different types of nodes and links. 在单一空间建模节点，不能区分易购性

A very recent work EOE [32] was proposed as a network embedding method for coupled heterogeneous network. The coupled heterogeneous network consists of two different but related homogeneous networks. For each homogeneous network, they adopt the same function as LINE to model the relationships between nodes. But, EOE is able to model both homogenous and heterogeneous network by using a harmonious embedding matrix to measure the closeness between nodes of different networks. Because the inter-network edges are able to provide the complementary information in the presence of intra-network edges, the learned embeddings of nodes also perform well on several tasks. However, it only models observed linkage information between heterogeneous nodes based on dot product of learned latent vectors and the second-proximity between network vertices cannot be preserved when the triangle inequality is violated. Also, EOE did not consider the comparabilities between different weights of network links, this will lead problems when optimizing the loss function. Besides, EOE can only suite the coupled heterogeneous network.

对于每个同构网络，采用LINE建模节点之间的关系。EOE可以建模同构和异构网络，使用har嵌入矩阵。度量不同信息网路节点之间的相近度。由于网络之间的边提供补充信息，学习的embedding在几个任务上表现很好。但是仅建模异构网络中可观察节点链接，基于隐藏向量之间的内积，二阶相似度不能保存

EOE没有考虑网络链接不同权重的兼容性优化loss是导致问题。EOE仅适用耦合异构网络

All above mentioned methods are either designed for specific tasks or have limitations on modelling multiple types of nodes. Our proposed **Projected Metric Embedding** using relation-specific projection matrices is versatile and more flexible to model arbitrary types of networks. Its intrinsic geometric flexibility is able to preserve first-order and second-order proximity naturally.

所述提及的方法要么是针对特定任务设计的，要么在建模多类型节点方面有局限性。PME使用关系指定的projection 矩阵，是多样化的，更加灵活在建模任意类型网络。它内在的几何灵活性能够自然的保存一阶和二阶近似度。

3 PRELIMINARIES

In this section, we first introduce preliminary concepts in heterogeneous information networks and then define the problem of heterogeneous information networks embedding.

PTE:处理采样问题，对一个类型的边，交替采样正样本。不同类型links严重不平衡分布时，还是存在问题。
PTE在单独空间中建模节点 是在不同类型节点和链接之间区分异构性变得困难。

V是不同类型节点的集合，E是不同类型链接的集合，R表示链接类型的集合，W是每个链接上的权重集合。

Definition 3.1. **A heterogeneous information network** is an information network with multiple types of objects and/or multiple types of links, formally defined as $G = (\mathcal{V}, \mathcal{E}, \mathcal{W}, \mathcal{R})$, where $\mathcal{V}$ is the union of different types of vertices, $\mathcal{E}$ is the union of different types of links, $\mathcal{R}$ denotes the link type set, and $\mathcal{W}$ is the union of the weight on each link. An edge $e \in \mathcal{E}$ is defined as: $e_{ijr} = (v_i, r, v_j), v_i, v_j \in \mathcal{V}, r \in \mathcal{R}$.

Problem 1. ***Projected Metric Embedding for Heterogeneous information Network***: *Given a heterogenous network G, the problem is to learn low-dimensional vector representations $X \in \mathbb{R}^{\|\mathcal{V}\| * d_v}$ for network nodes and low-dimensional latent representations $A \in \mathbb{R}^{\|\mathcal{R}\| * d_r * d_v}$ for heterogenous network relations, where d_v is the dimension of node embeddings, and $d_r * d_v$ is the dimension of relation-specific projection matrix.*

A是张量，一个|R|个二维矩阵，每个维度为d_v

Note that, the output of the problem consists of two parts: **a).** A low-dimensional Matrix X for node representations, with its i_{th} row representing the latent vector $\mathbf{v_i} \in \mathbb{R}^{d_v}$ for node v_i. **b).** A low-dimensional tensor A, with its r_{th} slice denoting the link-specific projection matrix $\mathbf{M_r} \in \mathbb{R}^{d_r * d_v}$ for link r, $r \in \mathcal{R}$.

with ...denoting

d_r取多少呢

4 PROJECTED METRIC EMBEDDING (PME)

In this section, we present the novel PME model for HINs and its optimization algorithm. Additionally, we also introduce a loss-aware adaptive positive sampling mechanism for optimization.

distingush区分 vt

4.1 The PME Model and Optimization

To address the key challenge of distinguishing the heterogeneity resulting from multiple types of vertices and relations in a HIN, we first project the latent representation $\mathbf{v_i}$ of a node v_i into relation-specific projection matrix $\mathbf{M_r}$. Then, the projected node embedding vector is defined as:

区分来自多类型点和关系的异构性是关键挑战

$$\mathbf{v_i^r} = \mathbf{M_r v_i} \quad \in \mathbb{R}^{d_r} \tag{1}$$

With the above defined link-specific projection, we now could perform the proximity calculation between two linked vertices in the corresponding relation space. For each observed link e_{ijr}, denoting vertex v_i and v_j are connected via a link r, the distance between v_i and v_j in the r-relation space is calculated as:

欧几里得距离满足三角不等式，保存一阶二阶

$$d_r(v_i, v_j) = \|\mathbf{M_r v_i} - \mathbf{M_r v_j}\|, r \in \mathcal{R} \tag{2}$$

The Euclidean distance is applied here to calculate the closeness between two nodes in specific relation space as Euclidian distance satisfy the triangle inequality and thus can preserve the first-order and second-order proximity naturally. At the same time, for the consideration of weighted edges in a HIN, we then define the following score function for an observed edge e_{ijr}:

采用欧几里得距离计算节点在特定关系空间中的紧密度

引入权重，只有观测link，对未观测link呢？

$$f_r(v_i, v_j) = w_{ij}\|\mathbf{M_r v_i} - \mathbf{M_r v_j}\|, r \in \mathcal{R} \tag{3}$$

With the defined score function for observed links in a given HIN, our idea is to keep vertices with links to be close to each other in certain relation space, and keep vertices without links far apart. We define the following margin-based loss function as objective for training:

$$L_r = \sum_{(v_i,v_j) \in D_r} \sum_{(v_i,v_k) \notin D_r} [m + f_r(v_i, v_j)^2 - f_r(v_i, v_k)^2]_+ \tag{4}$$

我们的idea:具有link的节点 在某关系空间相互靠近，没有link的节点远离。

where v_i and v_j is a pair of linked vertices, and v_k is a vertex not connected with v_i. D_r is the positive link set with relation type $r, r \in \mathcal{R}$. $[z]_+ = max(z, 0)$ is the standard hinge loss, r denotes a specific kind of link, and $m > 0$ is the safety margin size. The above loss function models one relation-specific network out of the entire given HIN. With respect to the whole heterogeneous network, the overall loss function is written as:

$$L = \sum_{r \in \mathcal{R}} \sum_{(v_i,v_j) \in D_r} \sum_{(v_i,v_k) \notin D_r} [m + f_r(v_i, v_j)^2 - f_r(v_i, v_k)^2]_+ \quad (5)$$

Then, the problem of learning embeddings of a heterogeneous network is turned to minimizing the following objective function:

$$\min_{\mathbf{v}_*, \mathbf{M}_*} \sum_{r \in \mathcal{R}} \sum_{(v_i,v_j) \in D_r} \sum_{(v_i,v_k) \notin D_r} [m + f_r(v_i, v_j)^2 - f_r(v_i, v_k)^2]_+$$
$$s.t. \quad \|\mathbf{v}_*\| <= 1 \quad and \quad \|\mathbf{M}_*\| <= 1 \quad (6)$$

We adopt the stochastic gradient algorithm for the model optimization. In each step, we sample a mini-batch of edges and update the embeddings.

4.2 Bidirectional Negative Sampling Strategy

直接最小化公式6计算开销大未观察的网络边的数量巨大，是观察网络节点和链接的三次方

However, directly minimizing Eqn. (6) is computationally expensive, as the number of unobserved network edges are huge and cubic to the number of observed network vertices and links. Inspired by the negative sampling techniques in [16] [26], instead of sampling all unobserved examples, we select some most likely negative examples for model optimization. For each sampled positive edge e_{ijr}, most related existing works on negative sampling [31] [19] [20] [1] [26] [25] generate the negative examples from only one side. Specifically, for example, given an unidirectional edge e_{ijr}, denoting a tripe (v_i, r, v_j) (i.e., two nodes v_i and v_j are connected via the relation r). Aforementioned negative samplers usually fix the vertex v_i and relation r, generating some negative vertices v_k (i.e., vertex v_k is never connected with v_i via r) according to a nosie distribution $P_n(v)$, and treat (v_i, r, v_k) as negative examples. This negative sampling strategy achieves good performance on most homogenous network embedding tasks [26] [16] [6] [11]; however, directly appling this negative sampling method on heterogeneous network will be insufficient and would lead to ineffective modelling results. For example, in the "business-category" sub-network in Yelp network, if we only sample negative nodes from category side, we cannot accurately learn latent vector representations for category nodes v_j, because only observed businesses are considered and thus the learned attributes vector v could not discriminate whether an unobserved business belongs to it.

从一边采负样本。

直接应用负采样的方法在异构网络将不足，导致无效的建模结果。

Thus, we follow our previous work [33] to draw negative samples from both sides of an edge. Specifically, for a sampled positive edge e_{ijr}, we first fix vertex v_i and edge type r, then generate K negative vertices v_k according to the widely adopted noise distribution [1] $P_n(v) \sim d_v^{0.75}$, where d_v is the degree of vertex v. Similarly, we then fix right side of e_{ijr}, and sample K negative vertex from the left side. Accordingly, the objective function can be refined as Eqn.(7).

$$O = \sum_{r \in \mathcal{R}} \sum_{(v_i,v_j) \in D_r} \Big(\sum_{k=1}^{K} E_{v_k} \sim p_n(v)[m + f_r(v_i, v_j)^2 - f_r(v_i, v_k)^2]_+$$
$$+ \sum_{k=1}^{K} E_{v_k} \sim p_n(v)[m + f_r(v_i, v_j)^2 - f_r(v_k, v_j)^2]_+\Big) \quad (7)$$

4.3 Loss-aware Adaptive Positive Sampling Strategy

Another challenging issue related to the model optimization is how to sample the positive examples since HINs contain multiple relation-specific sub-networks (i.e. the sub-networks extracted according to different link types), and the distribution of observed positive examples is heavily skewed and imbalanced. Table 1. shows the detailed statistics of the constructed Yelp heterogenous network. In every single state, the user-user relationships and the user-business interactions take the majority of all observed edges. If we adopt the uniform sampling to draw an observed edge and perform stochastic gradient descent algorithm, the most majority sampled observed edges would be user-user and user-business links. This sampling process will lead the trained model fail to preserve the the structure of business-attribute and business-category sub-networks. Besides, as the distribution of different types of links is quite different from different HINs, a fixed sampling mechanism is not able to fit all scenarios. Moreover, the efforts needed to preserve the network structure for different sub-network is different and will dynamically change while training. Thus, to build a versatile HIN embedding model, an adaptive sampling strategy for positive links is required.

We propose a novel loss-aware adaptive positive sampling strategy dedicated for heterogeneous networks. Intuitively, one can sample different types of positive examples from training set according to the training losses of individual sub-networks after each epoch during the training. As the distribution of various types of links in original training set is skewed and the difficulty of preserving the proximity between pairs of nodes in each relation is different, the convergence speed for each sub-network is different. Therefore, We can monitor the loss of each sub-network, if the loss of one particular sub-network is relatively high compared with the losses of other sub-networks, we adaptively increase the amount of positive samples for this kind of edges in next epoch. Otherwise, we decrease the amount of positive samples of this type of edges. Specifically, let $L = (l_1, l_2, l_3, ..., l_{\|\mathcal{R}\|})$ denote the sequence of the loss of each sub-network extracted from the complete heterogeneous network. One can simply calculate the sum of the losses $L_{sum} = \sum_{r \in \mathcal{R}} l_r$ and the percentage of each individual loss $\frac{l_r}{L_{sum}}$ after each training epoch. Then, draw a random value within the range of $[0, 1]$ to see which interval $[\sum_{j=0}^{r-1} \frac{l_j}{L_{sum}}, \sum_{j=0}^{r} \frac{l_j}{L_{sum}})$, the random value falls into. Obviously, this positive sampler will change accordingly while model parameters are updated because the parameter changes will lead the loss for each sub-network vary step by step. Thus, our proposed positive sampler is adaptive. As

看看落在那个区间

Table 1: Yelp network statistics

	No. of Edges				No. of Nodes			
State	User to User	User to Business	Business to Attributes	Business to Categories	Business	Users	Attributes	Categories
Complete	29,271,479	4,153,150	605,231	527,229	14,4072	1,029,432	81	1,191
NV	4,891,171	1,460,807	106,789	105,358	28,214	428,840	81	1,030
AZ	2,269,462	1,265,915	161,361	162,393	43,492	311,857	81	1,052
ON	465,204	500,812	120,241	84,491	24,507	92,997	66	777
WI	57,593	88,778	14,986	18,479	3,899	25,773	81	678
EDH	25,695	44,631	12,676	11,972	3,539	8,371	72	456

Algorithm 1 Training PME model

Input: A heterogeneous network $G(V, E, W, \mathcal{R})$, number of stochastic gradient steps, N, number of negative samples for each positive sample, K;

Output: Embeddings for network vertices and relation-specific projection matrix. (i.e., $\mathbf{v}$, M_r);

```
iter ← 0 :
while iter < N do
    if iter = 0 then
        Initialize the positive sampling probability as propor-
        tional to the original link distribution from G;
    else
        Sample M positive examples based on adaptive positive
        sampling strategy;
    End if
    For each sampled positive edge, sample K negative vertices
    from both sides of the edge;
    Compute gradients and update parameters;
    Censor the norm of v and projection matrix M_r;
    Compute relation-specific subgraph loss, and update the
    positive sampling probability;
    iter ← iter + 1 :
end
```

the loss for each sub-network is zero at the beginning of training, we initialize the positive sampling probability for each type of sub-network with proportional to their original edge distribution. The algorithm for optimizing our PME model is illustrated in Algorithm 1.

5 EXPERIMENTS

In this section, we first describe the experimental settings and then report the experimental results.

5.1 Dataset

We conduct our experiments on a large-scale and real-life dataset provided by Yelp Challenge[1] published in 2016. The dataset includes information about local business, user information, interactions between user and business (ratings, reviews), as well as friendship network among users. The original dataset contains the information in five states in the U.S, and we processed and extracted six (five individual state and one complete) large-scale heterogeneous social networks. Each network contains four different sub-networks, which are user-user, user-business, business-attribute, and business-category networks. Table 1. shows the detailed statistic of the extracted Yelp network. To make our experiments repeatable, we make our dataset and codes publicly available at our website[2].

Table 2: Statistics on AZ network

	u2u	u2b	b2a	b2c	|U|	|B|	|A|	|C|
Amount	1518610	961997	161361	162392	162345	43492	81	1052
Sparsity	0.99994	0.99986	0.95420	0.99645	-	-	-	-
Toal amount	2,804,360				206,970			

5.2 Evaluation Method

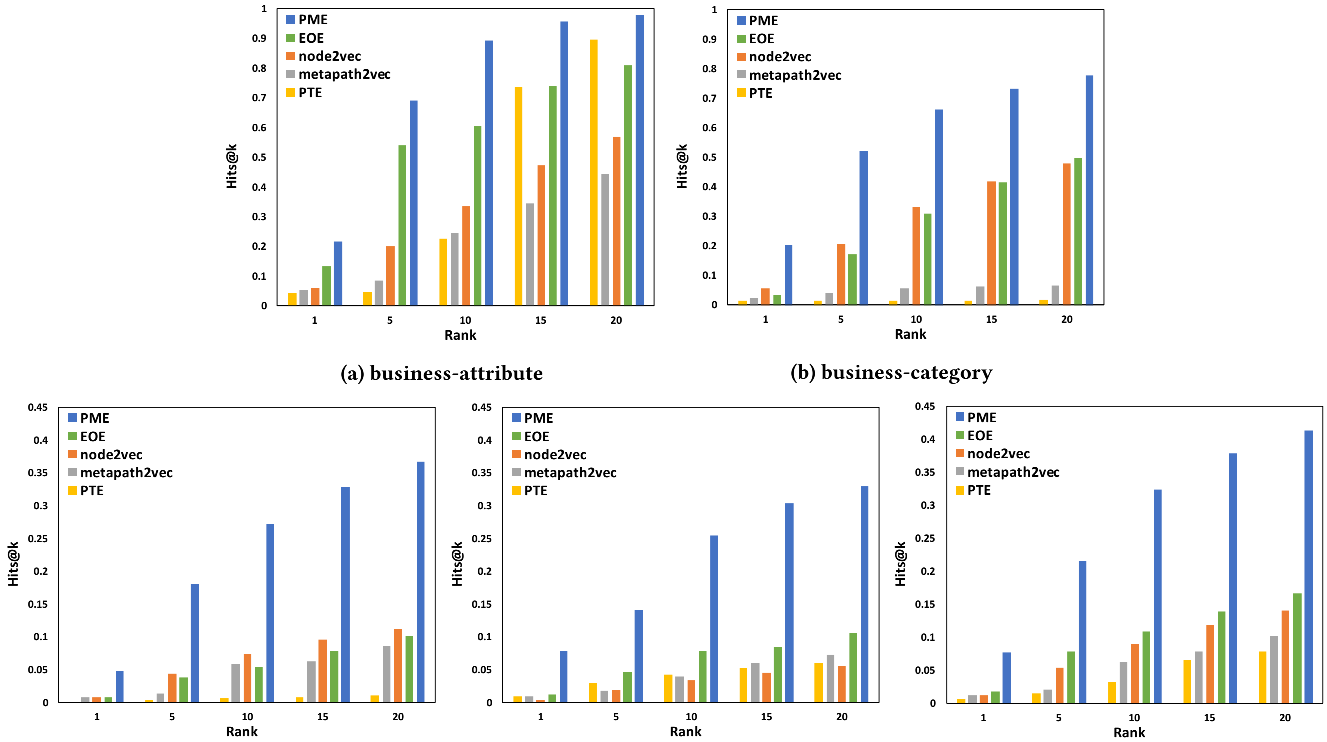
5.2.1 ***Evaluation of Prediction Accuracy.*** We perform this task on AZ (Arizona) state dataset. We further process this dataset by filtering out the nodes whose degree are less than 5. Then, we use 80-th percentile as the cut-off point so that the network linkage records before this point are used for training. In the training dataset, we choose the last 10% records as the validation data to tune the model parameters, including the dimension of latent feature vectors, margin, learning rate and the number of gradient steps. According to the above dividing strategies, we split the dataset D^+ into $D^+_{training}$, $D^+_{validate}$ and D^+_{test}. We summarise the detailed statistics of this dataset in Table 2.

To evaluate the embedding models, we employ the methodology and measurement $Hits@k$ which have been widely adopted by recommender system and knowledge graph communities [15] [7]. Specifically, for each linkage information (a triple consists of two vertices connected by a link) i.e., $e_{ijr} \in D^+_{test}$:

- We randomly choose 5000 items with which vertex v_i has been never connected by link type r to replace v_j and form 5000 negative examples.
- We compute a score for e_{ijr} as well as the 5000 negative examples by calculating their relative Euclidean distance by Equation (2).
- We form a ranked list by ordering these 5001 examples according to their distances to v_i. Let $rank(e_{ijr})$ denote the position of e_{ijr} in the ranking list.
- We form a top-k prediction list by picking the k top ranked examples from the list. If $rank(e_{ijr}) <= k$, we have a hit. Otherwise, we have a miss.

[1] https://www.yelp.com/dataset/challenge

[2] https://sites.google.com/view/hongxuchen



(a) business-attribute (b) business-category

(c) user-user (d) user-business (e) overall

Figure 2: Hit ratio@ top 20, 15, 10, 5, 1

The computation of Hits ratio proceeds as follows. We define $hit@k$ for a single test case as either 1, if the positive example e_{ijr} appears in the top-k results, or 0, if otherwise. The overall $Hits@k$ is defined by averaging over all test cases:

$$Hits@k = \frac{\#hit@k}{\|D^+_{test}\|}$$

where $\#hit@k$ denotes the number of hits in the test set, and $\|D^+_{test}\|$ is the number of all test cases. A good predictor should achieve higher $\#hit@k$. We can further divide D^+_{test} into four groups of triples according to the different edge types, and then analyze the performance of prediction models on each specific type of network edges. Besides Hits ratio, we also adopt the commonly used metric in information retrieval Mean Reciprocal Rank (MRR) to measure the prediction accuracy, and it is defined as follows:

$$MRR = \frac{1}{\|D^+_{test}\|} \sum_{e_{ijr} \in D^+_{test}} \frac{1}{rank(e_{ijr})}$$

MRR is an average of the reciprocal rank of a positive example over all sampled negative examples, and a good prediction model should have a bigger MRR value. In contrast to mean rank, MRR is less sensitive to outliers.

5.2.2 ***Binary Link Classification.*** Binary link classification tasks aim to predict whether the given links exist in the given networks. For this task, we split the original dataset into training, validation and testing dataset according to the same split strategy described in section 5.2.1, but we choose NV dataset to perform this task, which has a similar scale, but is geographically separate with AZ dataset. For each positive example (label as "1") in the testing set, we generate one negative example with label "0". Then, the performance of this task is evaluated with the widely used AUC metric [10].

5.3 Comparison Methods

We compare our proposed model with the following recent embedding methods for heterogeneous networks:

- **metapath2vec** [9] metapath2vec leverages predefined meta-path [23] guided random walks to construct the heterogeneous neighbourhood of a node and then applies a heterogeneous skip-gram model to perform node embedding. In our experiment, to include all types of nodes and links, we defined five different meta-paths: "ABA" (Attribute-Business-Attribute), "UBU" (User-Business-User), "CBC" (Category-Business-Category), "UBCBU" (User-Business-Category-Bus iness-User) and "UBABU" (User-Business-Attribute-Business-User) as the guidance of random walks.
- **node2vec** [11] This method diversifies the neighbourhood by using biased random walks over networks to produce paths of nodes. It also leverages the skip-gram architecture in word2vec [16] to model the network structure.
- **PTE** [25] PTE was further developed from LINE[26] , as an extension for heterogeneous network embedding. We construct four bipartite heterogeneous networks (user-user, user-business, bossiness-attributes, business-category) and restrain it as an unsupervised network embedding method.

- **EOE** [32] EOE learns embeddings for nodes in a coupled heterogenous network, and introduce a harmonious matrix to reconcile the heterogeneity between different types of nodes. However, EOE requires two inter-related homogenous networks, which has limitations when it is applied to general HINs embedding. Thus, we extend the EOE model by constructing bi-partite heterogenous networks and treating them as homogenous networks.

5.3.1 ***Parameter Settings.*** In the experiment, all the hyperparameters of both compared methods and our method are tuned to perform the best on the validation set. For our model, we set margin $m = 2$, learning rate $\alpha = 0.001$, batch size $B = 480$. To compare with all other methods, we set the common hyperparameters as follows, negative samples $N = 5$, embedding dimension $D = 128$. For random walk based methods node2vec[11] and metapath2vec [9], we set the number of walks per node $w = 1000$, walk length $l = 100$.

5.4 Experimental Results

In this section, we report our experimental results regarding social link prediction accuracy and binary link classification.

5.4.1 ***Social Link Predication Accuracy.*** In Figure 2, We present the prediction accuracy of all comparison methods in terms of $Hits@k$, where $k \in \{1, 5, 10, 15, 20\}$. Specifically, Figure 2 (a) - (d) show the individual prediction performance on each type of sub-network links (i.e., business-attribute, business-category, user-user, and user-business), and Figure 1 (e) shows the overall prediction accuracy on the whole test set that consists of all types of links.

It is clear that our proposed model consistently and significantly outperforms all compared methods in all types of network links prediction. Impressively, our model shows its superiority more significantly when the network is more sparse. For example, there are 162,345 users in our AZ dataset, which forms very sparse user-user (only 1,518,610 links, sparsity level 99.994%). Our model gains 3.6x, 35x, 4.26x, 3.28x times performance at $Hit@20$ compared with EOE, PTE, metapath2vec, node2vec, respectively as indicated in Figure 2(c). This reflects our model has good adaptability when dealing with data sparsity that is the nature of real-world HINs. The reason behind the superiority is that our PME model leverages a more geometrically flexible way to capture both the first-order and second-order proximity among nodes simultaneously. Thus, the weak relations in sparse network can be captured. Table 3 illustrates the prediction accuracy in terms of MRR metric, which is consistent with the performance in terms of $Hits@k$ in Figure 2.

We also noted that metapath2vec performs worse than node2vec in most experiments. We find the reason behind this is probably that the node2vec uses both BFS and DFS to traverse the network to generate node sequences, which is able to capture local and global network structure (higher-order proximity) at the same time. While, a key limitation of meta2path is that it treats the first-order proximity and the second-order relations as contributing equally to the learning. Moreover, the pre-defined meta-path for generating node sequences is also a key factor to the model performance. However, it is an interesting problem to select an appropriate meta-path based on different tasks and networks.

Table 3: Predication accuracy in terms of MRR

	PME	node2vec	PTE	EOE	metapath2vec
Overall	**0.1253**	0.0396	0.0181	0.0624	0.0098
user-user	**0.1249**	0.0314	0.0036	0.0260	0.0019
user-business	**0.0529**	0.0163	0.0219	0.0403	0.0089
business-attribute	**0.3701**	0.1539	0.1179	0.3059	0.0547
business-category	**0.3151**	0.1418	0.0321	0.2923	0.0435

Table 4: AUC scores on NV network

	PME	node2vec	PTE	EOE	metapath2vec
Overall	**0.9618**	0.8789	0.7494	0.8562	0.6232
user-user	**0.9672**	0.8909	0.6347	0.9033	0.5141
user-business	**0.9590**	0.8835	0.8615	0.9129	0.8179
business-attribute	**0.9376**	0.7522	0.8944	0.9201	0.5653
business-category	**0.9896**	0.9233	0.9652	0.9819	0.7725

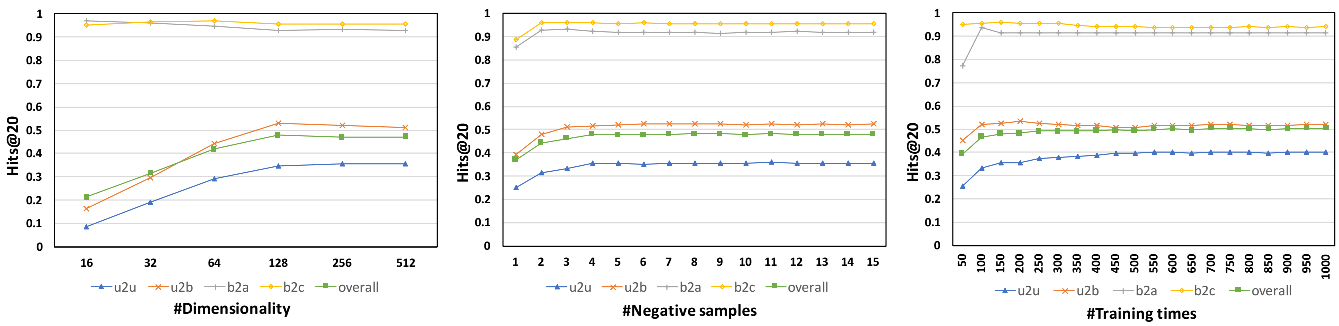
5.4.2 ***Binary Link Classification.*** Next, we introduce our experimental results on binary link classification task in Table 4, where we report the binary link classification results in terms of AUC metric of our PME model and different compared methods. Obviously, our model significantly improves the binary classification results consistently in all types of sub-networks.

We explore the reason behind the superiority of our proposed PME model. The superiority of our proposed PME model are two-folds. First, we deploy Euclidian distance as the metric to model the proximity in distinct relation-specific spaces, which preserves both the first-order and second-order proximity in a unified way, and the relation-specific space is helpful to represent the semantics of different relations. Other methods such as EOE that models the proximity between nodes by using dot product is not able to preserve the geometric properties of leant metric. Moreover, our PME model adopts a novel adaptive positive sampling and bidirectional negative sampling strategy while other models including EOE and PTE only consider replacing one side to draw samples. EOE employs gradient-based algorithms to perform the optimization and treats all unobserved links as negative examples. Although this solution empirically works well on small datasets, it has limited prediction accuracy because some of the missing links might be positive. Moreover, this solution cannot apply to large-scale HINs due to the huge number of unobserved links and the expensive computational cost.

5.5 Parameter Sensitivity Analysis

In this section, we investigate the sensitivity of different parameters in our model, including the number of embedding dimensions D, the number of negative samples N, the number of training times T (i.e., the number of epochs). We investigate how these parameters influence the performance of our proposed model by setting dimensions D to 32, 64, 128, 256 and 512, respectively; the number of epochs from 50 to 1000, and negative samples from 1 to 15.

Figure 3 (a) shows the results of prediction accuracy ($Hits@20$) w.r.t. the number of embedding dimensionality. From the results, we observe that the performance of our PME model improves with the increase of the number of dimensionality dramatically, and the performance becomes very stable when embedding dimensionality is going above 100. This implies our model is capable to capture the complex network structure among thousands of heterogeneous nodes and millions of links by only consuming such a low resource. Similar trends are also observed in figure 3 (b) and (c),



(a) Accuracy w.r.t. #dimensions (b) Accuracy w.r.t. #negative samples (c) Accuracy w.r.t. training steps

Figure 3: Parameter sensitivity

where in figure 3 (b) we can see that when the number of negative examples is larger than 4, the performance of our model archives a good and stable results. For training times, our model is starting converging after 200 epochs as shown in figure 3 (c).

5.6 Evaluation of Efficiency and Scalability

As heterogeneous networks are complex and contain such an impressive large number of nodes in the real world application scenario, it is necessary for a model being feasible to be applied in the large scale datasets. In this section, we investigate the scalability of our PME model optimized by the asynchronous stochastic gradient descent, which deploys multiple threads for parallel model optimization.Our experiments are conducted in a computer server with 64 cores and 1 Tb. memory. We run experiments with default settings (refers as in section 5.3.1) but different threads from 1 to 64. Figure 4 shows the speedup ratio w.r.t. the number of threads. The speedup ratio is quite close to linear, which shows that the optimization algorithm of the PME is quite scalable.

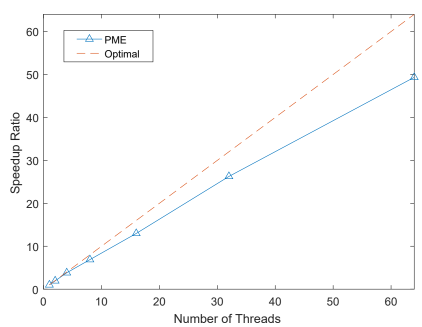
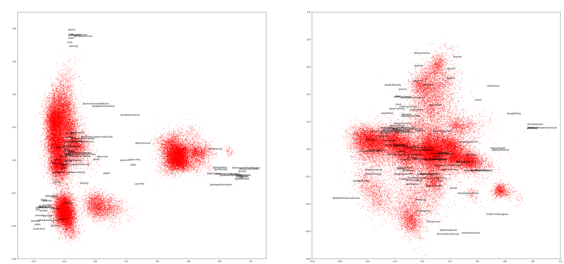


Figure 4: Scalability of PME

5.7 Case study: embedding visualization

Finally, for an intuitive understanding, we visualize the embedding vectors in a 2-dimensional space. Figure 5 (a) and (b) show the business-attribute relation space and business-category relation space respectively. From Figure 5 (a), we can see business nodes are clearly clustered into several groups and their distance to relevant attributes are revealed. This implies businesses are divided into groups based on their common attributes. In Figure 5 (b), we also observe our method successfully categorises businesses into more fine-grained clusters according to relevant categories because in our dataset, the number of categories is larger than attributes (i.e., 1052 categories and 81 attributes).



(a) business-attribute space (b) business-category space

Figure 5: visualization (zoom-in for a better readability)

6 CONCLUSION

In this work, we proposed a novel model PME to embed heterogenous information networks, which elegantly solves the challenging problem of modelling node and link heterogenities in elaborately designed relation-specific spaces. Besides, we apply Euclidian Distance as a metric to embed nodes proximities, which satisfies the crucial triangle inequality and preserves both the first-order and the second-order proximity at the same time. To optimize the PME model, we also introduce a novel loss-aware adaptive positive sampling strategy to overcome the heavy skewness of the heterogenous link distribution w.r.t. relations and further improve the model convergence speed. In addition, our model is versatile and suits arbitrary networks with no application limitations. Extensive experiments were conducted on a large-scale Yelp heterogenous network, and our PME model significantly outperforms the state-of-art heterogenous network embedding methods.

ACKNOWLEDGMENT

This work was supported by ARC Discovery Early Career Researcher Award (Grant No. DE160100308), ARC Discovery Project (Grant No. DP170103954 and Grant No. DP160104075) and New Staff Research Grant of The University of Queensland (Grant No.613134). It was also partially supported by National Natural Science Foundation of China (Grant No.61572335). The authors would also like to thank Rocky Chen for his patient discussions.

REFERENCES

[1] Amr Ahmed, Nino Shervashidze, Shravan Narayanamurthy, Vanja Josifovski, and Alexander J Smola. 2013. Distributed large-scale natural graph factorization. In *WWW*. 37–48.

[2] Mukund Balasubramanian and Eric L Schwartz. 2002. The isomap algorithm and topological stability. *Science* 295, 5552 (2002), 7–7.

[3] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *NIPS*. 2787–2795.

[4] Antoine Bordes, Jason Weston, Ronan Collobert, Yoshua Bengio, et al. 2011. Learning Structured Embeddings of Knowledge Bases.. In *AAAI*, Vol. 6. 6.

[5] Shiyu Chang, Wei Han, Jiliang Tang, Guo-Jun Qi, Charu C Aggarwal, and Thomas S Huang. 2015. Heterogeneous network embedding via deep architectures. In *KDD*. 119–128.

[6] Ting Chen, Yizhou Sun, Yue Shi, and Liangjie Hong. 2017. On Sampling Strategies for Neural Network-based Collaborative Filtering. In *KDD*. 767–776.

[7] Paolo Cremonesi, Yehuda Koren, and Roberto Turrin. 2010. Performance of recommender algorithms on top-n recommendation tasks. In *RecSys*. ACM, 39–46.

[8] Peng Cui, Xiao Wang, Jian Pei, and Wenwu Zhu. 2017. A Survey on Network Embedding. *CoRR* abs/1711.08752 (2017).

[9] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *KDD*. 135–144.

[10] Tom Fawcett. 2006. An introduction to ROC analysis. *Pattern recognition letters* 27, 8 (2006), 861–874.

[11] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *KDD*. 855–864.

[12] Cheng-Kang Hsieh, Longqi Yang, Yin Cui, Tsung-Yi Lin, Serge Belongie, and Deborah Estrin. 2017. Collaborative metric learning. In *WWW*. 193–201.

[13] Zhipeng Huang and Nikos Mamoulis. 2017. Heterogeneous Information Network Embedding for Meta Path based Proximity. *arXiv preprint arXiv:1701.05291* (2017).

[14] Yann Jacob, Ludovic Denoyer, and Patrick Gallinari. 2014. Learning latent representations of nodes for classifying in heterogeneous social networks. In *WSDM*. 373–382.

[15] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion.. In *AAAI*, Vol. 15. 2181–2187.

[16] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS*. 3111–3119.

[17] Andrew Y Ng, Michael I Jordan, and Yair Weiss. 2002. On spectral clustering: Analysis and an algorithm. In *NIPS*. 849–856.

[18] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *KDD*. 701–710.

[19] Steffen Rendle and Christoph Freudenthaler. 2014. Improving pairwise learning for item recommendation from implicit feedback. In *WSDM*. 273–282.

[20] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *uncertainty in artificial intelligence*. 452–461.

[21] Yizhou Sun, Rick Barber, Manish Gupta, Charu C Aggarwal, and Jiawei Han. 2011. Co-author relationship prediction in heterogeneous bibliographic networks. In *ASONAM*. IEEE, 121–128.

[22] Yizhou Sun, Jiawei Han, Charu C Aggarwal, and Nitesh V Chawla. 2012. When will it happen?: relationship prediction in heterogeneous information networks. In *WSDM*. 663–672.

[23] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. 2011. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *VLDB* 4, 11 (2011), 992–1003.

[24] Yizhou Sun, Brandon Norick, Jiawei Han, Xifeng Yan, Philip S Yu, and Xiao Yu. 2013. Pathselclus: Integrating meta-path selection with user-guided object clustering in heterogeneous information networks. *TKDD* 7, 3 (2013), 11.

[25] Jian Tang, Meng Qu, and Qiaozhu Mei. 2015. Pte: Predictive text embedding through large-scale heterogeneous text networks. In *KDD*. 1165–1174.

[26] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *WWW*. 1067–1077.

[27] Lei Tang and Huan Liu. 2009. Scalable learning of collective behavior based on sparse social dimensions. In *CIKM*. 1107–1116.

[28] Yi Tay, Anh Tuan Luu, and Siu Cheung Hui. 2017. Translational Recommender Networks. *CoRR* abs/1707.05176 (2017).

[29] Myo Thida, How-Lung Eng, and Paolo Remagnino. 2013. Laplacian eigenmap with temporal constraints for local abnormality detection in crowded scenes. *IEEE Transactions on Cybernetics* 43, 6 (2013), 2147–2156.

[30] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge Graph Embedding by Translating on Hyperplanes.. In *AAAI*, Vol. 14. 1112–1119.

[31] Min Xie, Hongzhi Yin, Hao Wang, Fanjiang Xu, Weitong Chen, and Sen Wang. 2016. Learning graph-based poi embedding for location-based recommendation. In *CIKM*. 15–24.

[32] Linchuan Xu, Xiaokai Wei, Jiannong Cao, and Philip S Yu. 2017. Embedding of Embedding (EOE): Joint Embedding for Coupled Heterogeneous Networks. In *WSDM*. 741–749.

[33] Hongzhi Yin, Hongxu Chen, Xiaoshuai Sun, Hao Wang, Yang Wang, and Quoc Viet Hung Nguyen. 2017. SPTF: A Scalable Probabilistic Tensor Factorization Model for Semantic-Aware Behavior Prediction. In *ICDM*. 585–594.

[34] Hongzhi Yin, Bin Cui, Yizhou Sun, Zhiting Hu, and Ling Chen. 2014. LCARS: A spatial item recommender system. *TOIS* (2014), 11.

[35] Hongzhi Yin, Bin Cui, Xiaofang Zhou, Weiqing Wang, Zi Huang, and Shazia Sadiq. 2016. Joint modeling of user check-in behaviors for real-time point-of-interest recommendation. *TOIS* (2016).

[36] Hongzhi Yin, Weiqing Wang, Hao Wang, Ling Chen, and Xiaofang Zhou. 2017. Spatial-Aware Hierarchical Collaborative Deep Learning for POI Recommendation. *TKDE* (2017), 2537–2551.

[37] Hongzhi Yin, Xiaofang Zhou, Bin Cui, Hao Wang, Kai Zheng, and Quoc Viet Hung Nguyen. 2016. Adapting to user interest drift for poi recommendation. *ICDE* 28, 10 (2016), 2566–2581.

[38] Hongzhi Yin, Lei Zou, Quoc Viet Hung Nguyen, Zi Huang, and Xiaofang Zhou. 2018. Joint eventpartner recommendation in event-based social networks. ICDE.

[39] Lubomir Zlatkov. 1978. Multidimensional Scaling (MDS). (1978).