Topic-aware Heterogeneous Graph Neural Network for Link Prediction

Siyong Xu[1], Cheng Yang[1], Chuan Shi[1], Yuan Fang[2], Yuxin Guo[1], Tianchi Yang[1]
Luhao Zhang[3], Maodi Hu[3]
[1]Beijing University of Posts and Telecommunications
[2]Singapore Management University, [3]Meituan
{xusiyong, yangcheng, shichuan, gyxx, yangtianchi}@bupt.edu.cn
{yfang}@smu.edu.sg, {zhangluhao, humaodi}@meituan.com

ABSTRACT

Heterogeneous graphs (HGs), consisting of multiple types of nodes and links, can characterize a variety of real-world complex systems. Recently, heterogeneous graph neural networks (HGNNs), as a powerful graph embedding method to aggregate heterogeneous structure and attribute information, has earned a lot of attention. Despite the ability of HGNNs in capturing rich semantics which reveal different aspects of nodes, they still stay at a coarse-grained level which simply exploits structural characteristics. In fact, rich unstructured text content of nodes also carries latent but more fine-grained semantics arising from multi-facet topic-aware factors, which fundamentally manifest why nodes of different types would connect and form a specific heterogeneous structure. However, little effort has been devoted to factorizing them.

In this paper, we propose a **T**opic-aware **H**eterogeneous **G**raph **N**eural **N**etwork, named THGNN, to hierarchically mine topic-aware semantics for learning multi-facet node representations for link prediction in HGs. Specifically, our model mainly applies an alternating two-step aggregation mechanism including intra-metapath decomposition and inter-metapath mergence, which can distinctively aggregate rich heterogeneous information according to the inferential topic-aware factors and preserve hierarchical semantics. Furthermore, a topic prior guidance module is also designed to keep the quality of multi-facet topic-aware embeddings relying on the global knowledge from unstructured text content in HGs. It helps to simultaneously improve both performance and interpretability. Experimental results on three real-world HGs demonstrate that our proposed model can effectively outperform the state-of-the-art methods in the link prediction task, and show the potential interpretability of learnt multi-facet topic-aware representations.

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**.

KEYWORDS

Heterogeneous graph, graph neural networks, representation learning, link prediction

ACM Reference Format:
Siyong Xu[1], Cheng Yang[1], Chuan Shi[1], Yuan Fang[2], Yuxin Guo[1], Tianchi Yang[1], Luhao Zhang[3], Maodi Hu[3], [1]Beijing University of Posts and Telecommunications, [2]Singapore Management University, [3]Meituan, {xusiyong, yangcheng, shichuan, gyxx, yangtianchi}@bupt.edu.cn, {yfang}@smu.edu.sg, {zhangluhao, humaodi}@meituan.com . 2021. Topic-aware Heterogeneous Graph Neural Network for Link Prediction. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management (CIKM '21), November 1–5, 2021, Virtual Event, QLD, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3459637.3482485

1 INTRODUCTION

Heterogeneous graphs (HGs), which are composed of different types of nodes and relations, also known as heterogeneous information network, exist in a variety of real-world scenarios, ranging from bibliographic networks [4, 32], social networks [33] to recommendation systems [23]. For example, as shown in Fig.1 (a), an academic network has multiple types of nodes (author, paper, conference, and term) and edges defined by their relations (e.g., author-paper, paper-conference). Due to its heterogeneity in both graph structure and node attributes, HGs often carry immensely rich and diverse semantics. Thus, much effort has been devoted to heterogeneous graph embedding to map HGs into a low-dimension vector space [5, 7, 22] for downstream tasks. Among the tasks in HGs, link prediction is a fundamental and important problem that estimates the existence probability of a link between two nodes, which serves as the basis in many data mining tasks like recommendation[23, 27].

Recently, graph neural networks (GNNs), as a powerful family of deep representation learning method to combine both structures and node features for graph data, has attained considerable success [17, 26]. Inspired by the well-designed mechanisms in GNNs for homogeneous graphs, heterogeneous graph neural networks (HGNNs) [6, 8, 28] have also attracted a lot of attention in recent years. One major line of HGNNs defines and leverages metapaths [24] to preserve semantics and model heterogeneous structure [8, 28], since different metapaths are able to reveal different aspects of target nodes from a global perspective. For example, in an academic network shown in Fig. 1(a), Author-Paper-Author (APA) and Author-Paper-Conference-Paper-Author (APCPA) are metapaths describing two different relations among authors. The APA metapath associates two co-authors, while the APCPA metapath associates two

CIKM '21, November 1–5, 2021, Virtual Event, QLD, Australia

ACM ISBN 978-1-4503-8446-9/21/11…$15.00
https://doi.org/10.1145/3459637.3482485

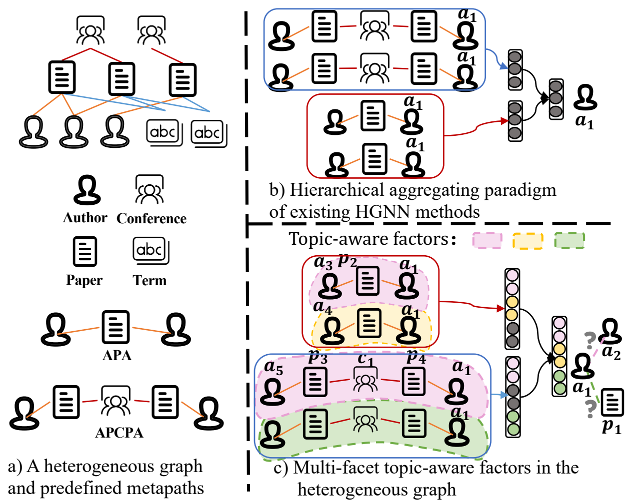


Figure 1: An illustration for heterogeneous graph (node types, metapaths) and comparison between our model and previous HGNNs.

authors who published papers in the same conference. Specifically, a classical paradigm of HGNNs is to adopt a hierarchical aggregation at node and semantic level, to fuse information from different metapaths, which can be intuitively shown in Fig. 1(b).

In spite of the informativeness of heterogeneous structure like metapaths, rich unstructured text content carried by nodes such as paper abstracts, descriptions or reviews, are also pervasive in HGs. Furthermore, text content is often a mixture of semantics arising from multi-facet topic-aware factors, which fundamentally manifest why nodes of different types would connect and form a specific heterogeneous structure. Such topic-aware semantics are more fine-grained than structural semantics for link prediction. For example, an author node a_1 in Fig. 1(c), may have multi-facet research interest on different topics, which can be reflected in a variety of local heterogeneous context. Along her APA-based context, she works with author a_3 due to their mutual interest on the topic of "Graph Mining" (shown in red), while builds a co-author relation with author a_4 stemming from the shared aim of "NLP Application" (shown in yellow). Similarly, along a_1's APCPA-based context, it can be inferred that both a_1's and a_5's papers are accepted by the same conference also related to "Graph Mining" topic (also shown in red). If we do not consider such fine-grained semantics through identifying the latent multi-facet topic-aware factors, and simply fuse derived confounding features, it will inevitably limit the performance of node representations for link prediction.

Recently, there are some attempts at disentangled learning to identify the latent explanatory factors behind the data with some promising results. Most prior efforts related to disentangled learning are devoted mainly in the field of image representation learning [9, 11, 16]. In order to tackle the non-Euclidean graph data, there are also a couple of works that explore the potential factors of edge formation between a pair of nodes for homogeneous graphs [18, 19, 31]. Although disentangled representation learning has been adopted in HGNNs [30], it only focuses on a coarse-grained and also local level aiming to automatically factorize structural semantics and avoid metapath selection only from neighbor nodes, but cannot further recognize and unveil more fined-grained semantics underlying the bare node connections.

Given the above limitations in current approaches, in this paper, we exploit both heterogeneous structures and unstructured text content in HGs. Specifically, we look deeper into identifying the potential but fundamental topic-aware factors based on the rich structural semantics in HGs, in order to learn multi-facet topic-aware representations for nodes while preserving such hierarchical semantics for link prediction. However, it imposes several challenges and thus can not be a straightforward extension of existing solutions. Firstly, HGs usually contain complex interactions and diverse attribute information among nodes, but there are no explicit labels indicating the latent and subtle topic-aware factors. This poses difficulties to distinguish the mixed information and factorize the feature vector into multi-facet topic-aware components. Secondly, after identifying potential topic-aware factors, an appropriate mechanism is required to combine both structural and topic-aware semantics. Thirdly, to keep up with the structural semantics at the global level, it is also important to preserve the global characteristics of topic-aware semantics and maintain the quality of multi-facet topic-aware embeddings, thus simultaneously improving both performance and interpretability.

To address the above challenges, we introduce a new model for link prediction, **T**opic-aware **H**eterogeneous **G**raph **N**eural **N**etwork (THGNN), that aims to further mine fine-grained topic-aware semantics based on structural semantics for multi-facet topic-aware representations learning in HGs. More precisely, THGNN applies a multi-facet transformation matrix to project the features of different types of nodes into the multiple topic-aware subspaces. For the first two challenges, THGNN applies an alternating two-step aggregation mechanism including intra-metapath decomposition and inter-metapath mergence, in order to learn multi-facet topic-aware embeddings for each target node. Specifically, the main goal of intra-metapath decomposition step is to infer the topic-aware distribution of metapath-based contexts and aggregate the context information according to the distribution to form multi-facet representations, so as to capture the fine-grained topic-aware semantics. On the other hand, inter-metapath mergence step adopts a multi-facet attention mechanism to fuse different metapaths for the final multi-facet embeddings, thus preserving both structural and topic-aware semantics for link prediction. For the last challenge, we introduce another module named topic prior guidance, which leverages topic modeling to obtain global statistical knowledge from unstructured textual content and helps to guide the context aggregation. In this way, it serves as a regularizer to encourage the inferential topic-aware subspaces to be more orthogonal and improve the interpretability of learnt multi-facet topic-aware representations.

To summarize, this work makes the following main contributions:

- From a fresh point of view, we propose to identify multi-facet topic-aware factors underlying the bare connections between associated nodes, making full use of both heterogeneous structures and unstructured text content for link prediction in HGs.

- We introduce a novel model THGNN for link prediction, which can distinctively aggregates rich heterogeneous information according to the inferred multi-facet topic-aware factors, so as to generate multi-facet topic-aware representations preserving both structural and topic-aware semantics. Besides, the topic prior guidance module is further designed to leverage the global knowledge from unstructured textual content, further keeping the quality of multi-facet topic-aware embeddings.
- Experiments on real heterogeneous graph datasets demonstrate that our proposed model significantly outperforms state-of-the-art methods in link prediction task, and also show the potential interpretability of our learnt multi-facet topic-aware embeddings.

2 RELATED WORK

In this section, we will review the most related work to ours, including graph neural networks and heterogeneous graph embedding.

2.1 Graph Neural Networks

Graph neural networks (GNNs) are a set of methods that apply deep learning to arbitrary graph-structured data. They generally fall into two categories: spectral and spatial methods. Based on spectral graph theory, Joan Bruna et al. [3] proposed a way to perform convolution in the spectral domain using the Fourier basis of a given graph. Kipf et al. [17] proposed a spectral approach, named Graph Convolutional Network, which simplified GNNs by involving the first-order approximation of the spectral graph convolution. A major limitation of traditional graph convolutional methods is that the filters are learned on the entire graph Laplacian, which lacks feasibility and scalability when graph is changed. Therefore, spatial approaches are proposed to define convolution operations directly on the graph, operating on groups of spatially close neighbors. Hamilton et al. [10] proposed GraphSage to summarize neighborhood information and a variety of solutions have been designed based on more advanced spatial-based filters. Inspired by Transformer [25], GAT [26] incorporated node features to measure the relative importance of neighbors via a multi-head self-attention mechanism.

Motivated by disentangled representation learning [1] successfully applied in the field of computer vision [9, 12, 16], Ma et al. [19] firstly proposed to learn disentangled representations in graph-structured data, employing neighborhood routing to automatically discover the independent latent factors of edges connecting a given node to its neighbors. As a follow up, IPGDN [18] improved disentangled GCNs by enforcing independence between the latent representations. Unlike prior efforts on decoding only a single attribute for a neighboring node, FactorGCN [31] enabled multi-relation disentangling, producing block-wise interpretable node features by analyzing the global-level topological semantics. Despite the results of the previous works on discovering latent factors in graphs, they are either built for homogeneous graphs, or designed for specific applications such as recommendation systems [15, 29].

2.2 Heterogeneous Graph Embedding

Heterogeneous graph embedding (i.e., HGE), aiming to learn a function that maps input space into a lower-dimension space while preserving the heterogeneous structures and semantics, has drawn considerable attentions and been applied to various scenarios in recent years [6, 13, 14, 34]. One of the classical paradigms is to leverage metapath for semantic-preserving embeddings. Metapath2vec [5] and HERec [23] generated random walks guided by selected metapaths and adopted skip-gram to perform HGE. HIN2vec [7] carries out multiple prediction training tasks and learns latent embeddings of nodes and metapaths simultaneously. Recent studies have attempted to extend GNNs for modeling HGs. HAN [28] utilized metapaths to convert a heterogeneous graph to multiple metapath-based homogeneous graphs, but adopted hierarchical attention mechanism (including both of the node-level and semantic-level attentions) to aggregate information. MAGNN [8] improved HAN by considering node information along the metapaths instead of only two end nodes, and thus exploited more comprehensive information in HGs.

Recently, disentangled representation learning is also applied in HGNNs. DisenHAN [30] was proposed to identify the major aspect of the relation between node pairs and propagate corresponding information semantically so as to automatically extract metapaths. Although mining such rich structural semantics in HGs is able to reflect multiple aspects of nodes in existing methods, there is still a room for improvement by exploiting the multi-facet topic-aware factors hidden by pervasive unstructured text content from a gloval perspective in HGs. These fine-grained topic-aware semantics can essentially bring insight into node connections and heterogeneous structure formation, which suggests us to identify the hidden multi-facet factors and model HGs more distinctively.

3 PRELIMINARY

In this section, we formalize some important definitions related to heterogeneous graphs.

Definition 3.1. **Heterogeneous Graph [24].** A heterogeneous graph is defined as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with an object type mapping function $\varphi : \mathcal{V} \rightarrow \mathcal{A}$ and a link type mapping function $\psi : \mathcal{E} \rightarrow \mathcal{R}$. $\mathcal{A}$ and $\mathcal{R}$ denote the sets of predefined object types and link types, where$|\mathcal{A}| + |\mathcal{R}| > 2$.

Definition 3.2. **Metapath [24].** A metapath M is a path denoted in the form of $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \dots \xrightarrow{R_l} A_{l+1}$, which defines a composite relation $R = R_1 \circ R_2 \circ \cdots \circ R_l$ between objects A_1 and A_{l+1}, where $\circ$ denotes the composition operator on relations.

Definition 3.3. **Metapath Instance.** Given a metapath M of a heterogeneous graph, a metapath instance m of M is defined as a node sequence in the graph matching the sequence of types in M. A metapath instance connecting node u and v where node u is a target node is denoted as m_u.

Definition 3.4. **Metapath-based Context.** Given a metapath instance m_u, the metapath-based context c of the target node u is defined as the node sequence in the rest of metapath instance without node u, denoted as c_u=$m_u \backslash \{u\}$, in which there exists nodes that contain text content. More specifically, the set of metapath M-based contexts with a target node u is denoted as C_u^M. The text-related node sequence in c_u denoted as c_u^{text}.

Example. As shown in Figure 1(a), we construct a HG to model DBLP. It consists of multiple types of objects (Author (A), Paper (P), Term (T), Conference (C)) and relations (P-A, P-T and P-C).

Two authors can be connected based on multiple metapaths, e.g., Author-Paper-Author (APA) and Author-Paper-Conference-Paper-Author (APCPA). Given a metapath APA, a_3-p_2-a_1 is a metapath instance related to the target node a_1. The author a_1 connects with other authors via different metapaths, and then node sequences (a_3, p_2) (APA-based) and (a_5, p_3, c_1, p_4) (APCPA-based) constitute the context of a_1, different metapath-based contexts may reveal different topic-aware semantics as paper nodes in metapath-based contexts carry rich text content.

4 METHODOLOGY

To guide information aggregation for nodes with multi-facet factors by the hidden topic-aware semantics underlying HGs, we propose a unified topic-aware heterogeneous graph neural network (THGNN). In this section, we will give a detailed description about the architecture of THGNN, whose three basic components are: multi-facet projection, multi-facet heterogeneous graph neural network, topic prior guidance. The multi-facet projection component serves as a preprocessing stage to facilitate the aggregation process of THGNN. The key building block "multi-facet heterogeneous graph neural network" consists of two steps: intra-metapath decomposition and inter-metapath mergence, to learn multi-facet topic-aware embeddings for nodes with multi-facet factors iteratively. A topic prior guidance is also introduced to guide context aggregation, further keeping the quality of multi-facet topic-aware embedding. Figure 2 illustrates the overall topic-aware embedding generation.

4.1 Multi-facet Projection

Due to the heterogeneity of nodes in HGs, different types of nodes and edges have different attributes, which are usually located in totally different feature spaces. With the goal of mining potential topic-aware subspaces in a HG, we need to project different types of node features into the same shared latent vector subspaces that indicates multi-facet topic-aware semantics.

Therefore, assuming that there exists K potential topic-aware subspaces in the HG, for each type of nodes we apply K type-specific linear transformation by projecting feature vectors into K latent topic-aware subspaces before feeding node vectors. As shown in Fig. 2 (a), for a node $u \in \mathcal{V}$ of type $\varphi(u) \in \mathcal{A}$, the multi-facet projection process can be shown as follows:

$$\mathbf{h}_{u,k} = \mathbf{P}_k^{\varphi(u)} \cdot \mathbf{x}_u, \tag{1}$$

where $k = 1, 2, \cdots, K$. $\mathbf{x}_u \in \mathbb{R}^{d_{\varphi(u)}}$ is the original feature vector of node u and $\mathbf{h}_{u,k} \in \mathbb{R}^{\frac{D}{K}}$ is the projected feature in the k^{th} latent topic-aware subspace. $\mathbf{P}_k^{\varphi(u)} \in \mathbb{R}^{\frac{D}{K} \times d_{\varphi(u)}}$ is the k^{th} training weight matrix for nodes of type $\varphi(u)$.

After multi-facet projection, all nodes features share the same D dimension with K components, which is convenient for the multi-facet heterogeneous graph neural network in Section 4.2.

Sampling Strategy Given a node u with multi-facet factors in the HG, we firstly need to sample some metapath-based contexts via different metapaths. For the sake of identifying multi-facet topic-aware factors, we employ a sampling strategy to pay more attention to those metapath-based contexts containing more than one text-related node with high topic consistency. The sampling process is as follows:

$$p_{c_u} = \frac{\sum_{v_s, v_{s+1} \in c_u^{text}} \cos(\boldsymbol{\lambda}_{v_s}^{c_u}, \boldsymbol{\lambda}_{v_{s+1}}^{c_u})}{\sum_{c'_u \in C_u^M} \sum_{v_s, v_{s+1} \in c'^{text}_u} \cos(\boldsymbol{\lambda}_{v_s}^{c'_u}, \boldsymbol{\lambda}_{v_{s+1}}^{c'_u})}, \tag{2}$$

where p_{c_u} is the sampling probability of metapath M-based context c_u, and $\boldsymbol{\lambda}_{v_s}^{c_u}$ is the pre-calculated topic distribution of text content carried by node v_s in context c_u through topic model LDA.

4.2 Multi-facet Heterogeneous Graph Neural Network

In the ensuing discussion, we shall zoom into the key building block of THGNN, which is composed of two steps: intra-metapath decomposition to capture topic-aware semantics and inter-metapath mergence to preserve structural semantics. The aim of intra-metapath decomposition step is to infer topic-aware distribution of the metapath M-based context in C_u^M preliminarily and aggregate the context information based on the inferred distribution to form the multi-facet representation. Inter-metapath mergence step aims to fuse different metapaths to generate the final multi-facet embedding preserving structural and topic-aware semantics at current iteration. The above two steps are conducted alternatively, which can be summarized as follows:

$$\mathbf{y}_u = g(\{\mathbf{h}_{u,k} | 1 \le k \le K\}, \{\mathbf{h}_{u,k}^{c_u} | 1 \le k \le K, c_u \in C_u^{M_i}, M_i \in \mathcal{M}\}), \tag{3}$$

where $g(\cdot)$ is the aggregation function to learn the final multi-facet topic-aware representation $\mathbf{y}_u$ of u, and $\mathbf{h}_{u,k}^{c_u}$ indicates embedding of the metapath-based context c_u in the k^{th} topic-aware subspace. $\mathcal{M}$ is the set of selected metapaths. As assuming that there exist K latent topic-aware subspaces, we would like to let $\mathbf{y}_u$ be composed of K topic-aware components $\mathbf{y}_u = [\mathbf{z}_{u,1}, \mathbf{z}_{u,2}, \cdots, \mathbf{z}_{u,K}]$, where $\mathbf{z}_{u,k} \in \mathbb{R}^{\frac{D}{K}}$ is able to characterize the k^{th} topic-aware factor of u.

4.2.1 Intra-metapath Decomposition. As shown in Fig. 2 (b)-1), given a metapath $M_i \in \mathcal{M}$, the goal of this step is to infer which topic that each sampled metapath-based context pertain to. Recall that we expect $\mathbf{h}_{u,k}^{c_u}$ to capture the k^{th} facet topic-aware of target node u by leveraging the current metapath-based context. The context encoder can be shown in the following manner:

$$\mathbf{h}_{u,k}^{c_u} = f(\{\mathbf{h}_{v,k}, \forall v \in c_u\}), \tag{4}$$

where $c_u \in C_u^{M_i}$, and we choose the average pooling operation in experiments for $f(\cdot)$ in consideration of simplicity and efficiency.

After encoding the metapath-based context into multi-facet representations in topic-aware subspaces, we adopt a cosine similarity between multi-facet representations of target node and its current metapath-based context, to infer the most related topic-aware subspace both of them share. The inferential topic distribution is given by:

$$p_{k|c_u} = \frac{\exp(\cos(\mathbf{h}_{u,k}, \mathbf{h}_{u,k}^{c_u})}{\sum_{k'=1}^{K} \exp(\cos(\mathbf{h}_{u,k'}, \mathbf{h}_{u,k'}^{c_u})}, \tag{5}$$

where $k = 1, 2, \cdots, K$ and $c_u \in C_u^{M_i}$. $C_u^{M_i}$ indicates a set of sampled metapath M_i-based contexts of u. Naturally, the probability $p_{k|c_u}$ should be high if the target node u is related to the current metapath-based context c_u in the k^{th} topic-aware subspace. Meanwhile, it

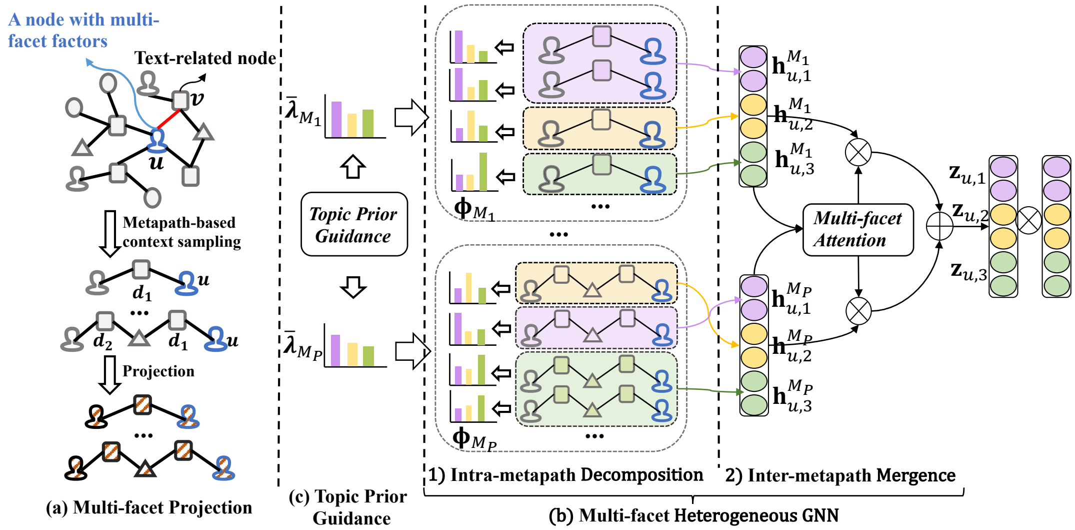


Figure 2: The overall framework of the proposed model THGNN.

also serves as the importance weight of the current metapath-based context information when propagating it to the target node in the k^{th} topic-aware subspace. Therefore, we can obtain the metapath-specific multi-facet topic-aware representation of target node u:

$$\mathbf{h}_{u,k}^{M_i} = L2\_Norm(\sum_{c \in C_u^{M_i}} p_{k|c_u} \cdot \mathbf{h}_{u,k}^{c_u}), \tag{6}$$

where $k = 1, 2, \cdots, K$ and $M_i \in \mathcal{M}$.

In summary, given the multi-facet projected feature vectors of the target node u, and the set of selected metapaths $\mathcal{M} = (M_1, M_2, \cdots, M_P)$ for u, the intra-metapath decomposition generates P groups of metapath-specific multi-facet topic-aware representations for u, denoted as $\{[\mathbf{h}_{u,1}^{M_1}, \mathbf{h}_{u,2}^{M_1}, \cdots, \mathbf{h}_{u,K}^{M_1}], [\mathbf{h}_{u,1}^{M_2}, \mathbf{h}_{u,2}^{M_2}, \cdots, \mathbf{h}_{u,K}^{M_2}], \cdots, [\mathbf{h}_{u,1}^{M_P}, \mathbf{h}_{u,2}^{M_P}, \cdots, \mathbf{h}_{u,K}^{M_P}]\}$.

As the topic-aware distributions and multi-facet topic-aware representations derived from the intra-metapath decomposition step aim to be as diverse as possible, the inter-metapath mergence step, is designed to find overall importance of multi-facet information conditioned on certain structure semantic.

4.2.2 Inter-metapath Mergence. Since every node in a HG contains multiple types of structural semantic information which can be revealed by metapaths, we propose a multi-facet attention to learn the importance of different metapaths and merge them to generate the final multi-facet topic-aware representation for the target node u, which is shown at Fig. 2(b)-2).

To learn the importance of given metapath M_i, firstly we transform the metapath-specific single-facet representations with a weight matrix $\mathbf{W}$, then we summarize metapath M_i in the k^{th} topic-aware subspace by averaging the transformed metapath-specific single-facet representations for target nodes in a batch:

$$\mathbf{s}_k^{M_i} = \frac{1}{|B|} \sum_{u \in B} \tanh(\mathbf{W} \cdot \sigma(\mathbf{h}_{u,k}^{M_i})), \tag{7}$$

where $k = 1, 2, \cdots, K$, $\mathbf{W} \in \mathbb{R}^{\frac{D}{K} \times \frac{D}{K}}$ is learnable parameter and $|B|$ is size of batch.

Then we use multi-facet attention vectors $(\mathbf{q}_1^T, \mathbf{q}_2^T, \cdots, \mathbf{q}_K^T)$ to measure the relative importance of the metapath M_i in each topic-aware subspace. Since different metapaths describe various semantics behind a HG at structural level, the overall importance of different metapaths should be shared in all topic-aware subspaces, and thus we sum up multi-facet relative importance as the overall importance for the metapath M_i:

$$e^{M_i} = \sum_{k=1}^{K} \mathbf{q}_k^T \cdot \mathbf{s}_k^{M_i}, \tag{8}$$

$$\beta^{M_i} = \frac{\exp(e^{M_i})}{\sum_{M \in \mathcal{M}} \exp(e^{M})}, \tag{9}$$

where $\mathbf{q}_k \in \mathbb{R}^{\frac{D}{K}}$, and β^{M_i} can be interpreted as the contribution of the metapath M_i.

With the computed coefficients $(\beta^{M_1}, \beta^{M_2}, \cdots, \beta^{M_P})$, we can merge all the metapath-specific multi-facet representations to obtain final multi-facet topic-aware representation in current iteration:

$$\hat{\mathbf{h}}_{u,k} = \sum_{M_i \in \mathcal{M}} \beta^{M_i} \cdot \mathbf{h}_{u,k}^{M_i}, \tag{10}$$

where $k = 1, 2, \cdots, K$.

The above two steps are performed together as an iteration and the current output $\{\hat{\mathbf{h}}_{u,k}, k = 1, 2, \cdots, K\}$ will in turn serve as guidance of the next topic-aware distribution inference, which can be also interpreted as updated K topic-aware cluster centers for different types of metapath-based contexts of the target node u. After T iterations, THGNN outputs K final enriched representations for $\mathbf{y}_u$ in Eq.(3) going through an activation function: $\{\mathbf{z}_{u,k} = \sigma(\hat{\mathbf{h}}_{u,k}^{(T)}) | k = 1, \cdots, K\}$(for a text-related node v, let $\mathbf{z}_{v,k} = \sigma(\mathbf{h}_{v,k})$ without any iteration), and also final inferential topic-aware distribution of all the sampled metapath-based contexts, denoted as $\{\mathbf{\Phi}_u^{M_i}, M_i \in \mathcal{M}\}$.

Here $\Phi_u^{M_i} \in \mathbb{R}^{|C_u^{M_i}| \times K}$ can be interpreted as a soft cluster assignment matrix of metapath M_i-based contexts, and each row of $\Phi_u^{M_i}$ contains K elements calculated through Eq.(5) at the last iteration.

4.3 Topic Prior Guidance

In spite of the iterative process in multi-facet heterogeneous graph neural network inferring topic-aware distribution for metapath-based context and generating chunked representations for the target nodes, there are still two points needing to go further: (1) due to the heterogeneity of metapath-based context, there might be still confounding among different topic-aware subspaces, even leading to an extreme collapse case; (2) the intra-metapath decomposition step is performed essentially as a kind of local clustering, making it difficult to capture global knowledge from K potential topic-aware subspaces we assume. The first suggests the inferential topic-aware subspaces to be more independent and each one to maintain a convenient size. Another means global prior knowledge is necessary to keep the globality of topic-aware subspaces. To meet the two requirements, we introduce another module named topic prior guidance to encourage the inferential topic-aware subspaces to be more orthogonal and improve interpretability.

As shown in Fig.2 (c), inspired by the orthogonality loss term for graph clustering [2], the added module will serve as a regularizer, ingeniously leveraging topic model to obtain global statistical knowledge from unstructured textual content, to guide context aggregation among the HG, which has the following form:

$$\mathcal{L}_T = \frac{1}{|\mathcal{M}|} \sum_{M_i \in \mathcal{M}} \parallel \frac{\Phi_{M_i}^T \Phi_{M_i}}{\parallel \Phi_{M_i}^T \Phi_{M_i} \parallel_F} - \frac{\overline{\boldsymbol{\lambda}}_{M_i}}{\parallel \overline{\boldsymbol{\lambda}}_{M_i} \parallel_F} \parallel_F, \tag{11}$$

$$\overline{\boldsymbol{\lambda}}_{M_i} = \frac{1}{|C^{M_i}|} \sum_{c \in C^{M_i}} \boldsymbol{\lambda}_{d_1}^c, \tag{12}$$

where each row of Φ^{M_i} indicates the inferential topic-aware distribution of metapath M_i-based context from Eq.(5), while $\boldsymbol{\lambda}_{d_1}^c$ means the pre-calculated topic distribution of the first document node d_1 along the single metapath-based context c obtained from LDA. Note that in experiments, we concatenate all the sampled metapath M_i-based contexts of target nodes in the current mini-batch to form the matrix Φ_{M_i}, instead of a single target node, due to the sparsity of data and so as to save computation. Similarly, C^{M_i} also indicates all the sampled metapath M_i-based contexts of target nodes in the current mini-batch. In this way, $\overline{\boldsymbol{\lambda}}_{M_i}$ plays a prior guiding role in context aggregation by the hidden topic-aware subspaces underlying the HG.

4.4 Model Training

After applying above described three basic components, we obtain multi-facet topic-aware representations for nodes, the forward propagation process is shown in Algorithm 1. We estimate the similarity for a training pair (u, v) through inner product in multi-facet topic-aware subspaces, and then we sum all of them as the final matching score for link prediction:

$$s_{uv} = \sum_{k=1}^{K} \mathbf{z}_{u,k}^T \cdot \mathbf{z}_{v,k}, \tag{13}$$

We construct the loss function of graph reconstruction using mini-batch gradient in two major learning paradigms:

Algorithm 1 The forward propagation of THGNN.

Require:
The heterogeneous graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$,
node types $\mathcal{A}$=$\{A_1, A_2, \cdots, A_{|\mathcal{A}|}\}$,
the type of nodes with multi-facet factors A,
metapaths $\mathcal{M}$=$\{M_1, M_2, \cdots, M_{\mathcal{M}}\}$,
node features $\{\mathbf{x}_v, \forall v \in \mathcal{V}\}$,
the number of multi-facet topic-aware factors K

Ensure:
$\{\mathbf{z}_{v,k} | 1 \leq k \leq K, \forall v \in \mathcal{V}\}$;

```
for node type A_i ∈ 𝒜 do
    for k = 1, 2, ··· , K do
        Multi-facet projection h_{v,k} = P_k^{A_i} · x_v, ∀v ∈ 𝒱_{A_i};
    end for
end for
for T iterations do
    for metapath M_i ∈ ℳ do
        for u ∈ 𝒱_A do
            for k = 1, 2, ··· , K do
                Calculate h_{u,k}^{c_u} for all c_u ∈ C_u^{M_i} with Eq. 4;
                Infer the topic-aware distribution p_{k|c_u} with Eq. 5;
                Obtain h_{u,k}^{M_i} with Eq. 6;
            end for
        end for
    end for
    Calculate β^{M_i} of metapath M_i ∈ ℳ with Eq. 7 8, 9;
    Obtain ĥ_{u,k}, h_{u,k} ← ĥ_{u,k}, k = 1, 2, ...K with Eq. 10;
end for
return {z_{v,k} = σ(h_{v,k}) | k = 1, ··· , K, ∀v ∈ 𝒱}
```

- Only one node with multi-facet factors in training pair (u, v): assuming that u is the only node with multi-facet factors , then

$$\mathcal{L}_{G,\mathcal{B}} = -\sum_{(u,v) \in \mathcal{B}^+} \log \sigma(s_{uv}) - \sum_{(u,v') \in \mathcal{B}^-} \log \sigma(-s_{uv'}), \tag{14}$$

  where $\mathcal{B}$ =$\{\mathcal{B}^+ \cup \mathcal{B}^-\}$ denotes the training pairs involving the observed edges $\mathcal{B}^+$ and unobserved edges $\mathcal{B}^-$.
- Both are nodes with multi-facet factors in training pair (u, v):

$$\mathcal{L}_{G,\mathcal{B}} = -\sum_{(u,v) \in \mathcal{B}^+} \log \sigma(\frac{1}{K} s_{uv}) - \sum_{(u',v') \in \mathcal{B}^-} \log \sigma(-\frac{1}{K} s_{u'v'}), \tag{15}$$

  where $\mathcal{B}$ =$\{\mathcal{B}^+ \cup \mathcal{B}^-\}$. Here we take the average of the final matching score because THGNN outputs the final multi-facet representations after l2 normalization operation.

The overall training loss which combines the loss of graph reconstruction and regularizer term can be rewritten as:

$$\mathcal{L}_{\mathcal{B}} = \mathcal{L}_{G,\mathcal{B}} + \gamma \mathcal{L}_T. \tag{16}$$

Table 1: Description of datasets.

Dataset	Relation (A-B)	#A	#B	#A-B
DBLP	Paper-Author	11,248	11,569	32,534
	Paper-Conference	11,248	16	11,248
	Paper-Term	11,248	2,463	65,818
YELP	Business-User	2,203	1,430	27,793
	Business-City	2,203	66	2,157
	Business-Category	2,203	347	9,787
Amazon	Movie-User	1,754	2,476	34,042
	Movie-Brand	1,754	293	734
	Movie-Category	1,754	95	4,913

5 EXPERIMENTS

In this section, we evaluate the effectiveness of THGNN on three real-world graph datasets, namely, DBLP, YELP and Amazon, whose nodes carry a wealth of text content. A comparative evaluation against a wide variety of baselines is performed in the link prediction task to evaluate the model. Besides, we also visualize the quality of the learned multi-facet topic-aware embeddings to demonstrate the discovered topic-aware semantics. The experiments aim to address the following questions:

- Q1: How does THGNN perform in link prediction task compared with state-of-the-art methods?
- Q2: How does topic prior guidance affect the result of THGNN?
- Q3: Can THGNN capture multi-facet topic-aware semantics?

5.1 Experimental Setup

Datasets The detailed descriptions of the HGs used in our experiments are shown in Table 1.

- **DBLP**: We extract a subset of DBLP which contains 11,569 authors (A), 11,248 papers (P), 16 conferences (C), 2,463 terms (T) between year 2000 to 2010, to build an academic graph. The contents of papers consist of abstracts and titles and are transformed into bag-of-word representations of keywords for author features. Here we employ the metapath set {APA, APCPA, APTPA} to perform experiments. For DBLP dataset, we split training and test data sequentially with a split year 2008 and predict both author-paper links and co-author hyper-links. Due to the temporality of DBLP dataset, we train the topic model LDA only using papers before split year. We also randomly sample disconnected node pairs of the given form as negative instances.
- **YELP**: We extract a subset of YELP which contains 1,430 users (U), 2,230 businesses (B), 66 cities (Ci), 347 categories (Ca), to build a review graph. Business nodes carry text content from some useful reviews. Here we employ the metapath set {UBU, UBCiBU, UBCaBU} to perform experiments. We randomly hide 25% of user-business links as the ground truth positives, and randomly sample disconnected node pairs of the given form as negative instances. All the text content carried by business nodes is used to train the topic model LDA. The ground truth serves as our test set of YELP dataset.
- **Amazon**: We extract a subset of Amazon which contains 2,476 users (U), 1,754 movies (M), 293 brands (Br), 95 categories (Ca), to build a review graph. Movie nodes carry text content from some useful reviews and descriptions. Here we employ the metapath set {UMU, UMBrMU, UMCaMU} to perform experiments. All the text content carried by movie nodes is used to train the topic model. Similar to YELP dataset, we randomly hide 25% of user-movie links as the ground truth positives.

Baselines We compare THGNN against three categories of graph embedding methods: random walk-based, GNN-based, HGNN-based. **DeepWalk** [21] is a traditional random walk-based homogeneous method employing the skip-gram model [20]. Here we ignore the heterogeneity of nodes and perform DeepWalk on the whole heterogeneous graphs. **Mp2vec** [5] is a random walk-based heterogeneous method, which generates node embeddings by feeding metapath-based random walks and also employs the skip-gram model. Here we test all the metapaths for metapath2vec and report the best performance. **HERec** [23] is a random walk-based heterogeneous method which designs a type-constrained strategy to o filter the node sequence and utilizes skip-gram to embed the HGs. **GraphSage** [10] is a classical GNNs which leverages sampler and aggregator to encode homogeneous graph. **GAT** [26] is a homogeneous GNN method which adopts multi-head additive attention on neighbors. **DisenGCN** [19] is a homogeneous disentangled GNN method which designs a neighborhood routing mechanism and embedding propagation to disentangle latent factors underlying edges between nodes and their neighbors. **HAN** [28] is a HGNN-based method which extracts metapath-based homogeneous graphs and adopts hierarchical attentions to aggregate neighobor information via different metapaths. **DisenHAN** [30] is a HGNN-based method which can iteratively identify the major aspect of meta relations and aggregate corresponding aspect features from each meta relation for target nodes. **MAGNN** [8] is a HGNN-based method which also adopts hierarchical attentions but to aggregate metapath instances leveraging node content features. **THGNN**$_{\backslash MA}$ is a variant of THGNN, which removes the multi-facet attention and employs the simple average strategy on all metapaths.

Parameter Settings For all the GNN-based and HGNN-based methods including our model, we search the dimension of output node embeddings in {64, 128, 256} and report the best performance, and we unify them in an inductive version. For our method, the number of metapath-based context samples is 30 and the number of iterations is 3 in default setting. We fix the dimension of output node embedding for each task and search the number of latent topic-aware factors in {4, 8, 16, 32}. To prevent overfitting, we employ dropout where the ratio is tuned among $\{0.0, 0.1, \cdots, 0.5\}$. We optimize THGNN with Adam optimizer by setting the learning rate to $1e\text{-}7 \sim 1e\text{-}4$. We use AUC and average precision (AP) to evaluate performance of models. For all methods, we run 5 times with the same partition and report the average results. We will release the codes after the paper is accepted.

5.2 Link Prediction (Q1)

In link prediction, the goal is to predict potential or missing links connecting pairwise nodes in a graph. It is a widely used task to evaluate the quality of learnt node representations, and it is also a demanding task in real world scenarios with multi-facet characteristic. The prediction performance of all models on the three datasets is summarized in Table 2. Analyzing such performance comparison, we have the following observations:

Table 2: **The performance comparison of link prediction. The underlined means the best performance in baselines.**

	DBLP (A-P)		DBLP (A-A)		YELP (U-B)		Amazon (U-M)	
Models	AUC	AP	AUC	AP	AUC	AP	AUC	AP
DeepWalk	0.7754	0.7782	0.7488	0.7568	0.6200	0.6281	0.7684	0.7840
Mp2vec	0.7213	0.7243	0.7600	0.7484	0.6792	0.6872	0.7812	0.7834
HERec	0.8186	0.7927	0.7565	0.7871	0.8383	0.8215	0.8085	0.8083
GraphSage	0.8218	0.8341	0.8524	0.8817	0.8874	0.8810	0.8236	0.8327
GAT	0.8322	0.8367	0.8579	0.8805	0.8915	0.8847	0.8329	0.8343
DisenGCN	0.8301	0.8352	0.8692	0.8936	0.8957	0.8838	<u>0.8435</u>	0.8390
HAN	0.8535	0.8719	0.8744	<u>0.9046</u>	0.9013	0.8722	0.8418	0.8414
DisenHAN	0.8496	0.8654	<u>0.8868</u>	0.8947	0.8976	0.8843	0.8426	<u>0.8419</u>
MAGNN	<u>0.8543</u>	<u>0.8727</u>	0.8671	0.9035	<u>0.9123</u>	<u>0.8926</u>	0.8378	0.8278
$\text{THGNN}_{\backslash MA}$	0.8546	0.8683	0.8712	0.8939	0.9154	0.8945	0.8578	0.8502
THGNN	**0.8807**	**0.9000**	**0.8996**	**0.9174**	**0.9243**	**0.9037**	**0.8634**	**0.8515**

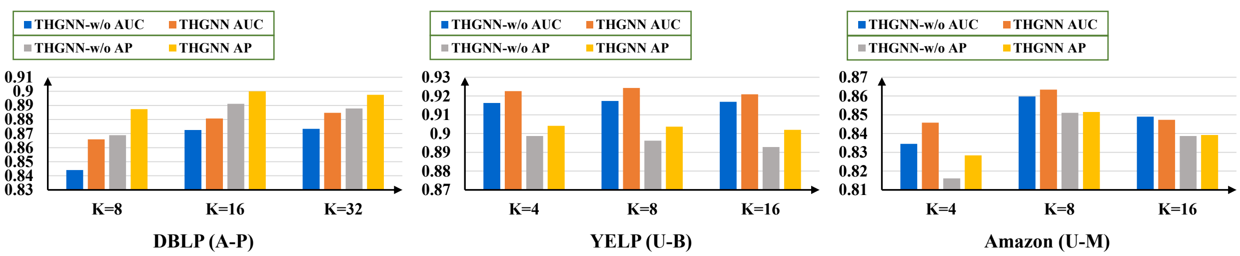


Figure 3: Effect of the topic prior guidance module. THGNN-w/o means THGNN without topic prior guidance.

It is clear that our model consistently performs better than all baselines on three datasets. Compared to the best performance of baselines, THGNN achieves improvements in terms of both AUC (up to 1.32% $\sim$ 3.09%) and AP (up to 1.14% $\sim$ 3.13%), which indicates the effectiveness of the delicate designs for factorizing multi-facet topic-aware semantics in THGNN. The results of its variant also achieve competitive performance compared to baselines, it is still significantly worse than the complete THGNN. It indicates the effectiveness of both structural and topic-aware semantics.

Comparing across baselines, we can observe that GNNs often obtain better performance against random walk-based methods since they combine the structures and features information in different ingenious ways. HGNNs further outperform GNNs by better capturing HGs' complex structures and rich semantic information.

DisenGCN substantially outperforms other homogeneous GNNs in most cases. It is reasonable since it is capable of dynamically recognizing and disentangling the latent factors that cause edges through a neighbor routing mechanism. The improvements of DisenHAN is not obvious. Possible reasons are that the main goal of DisenHAN is to automatically factorize structural semantics, which further confirms the necessity and benefits of mining fine-grained topic-aware semantics in HGs.

5.3 Analysis of Topic Prior Guidance (Q2)

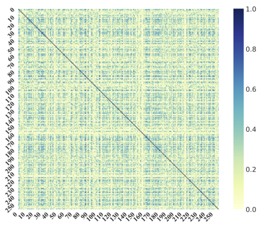
To investigate the effect of topic prior guidance module on improving performance, we test THGNN and THGNN without topic prior guidance (THGNN-w/o) with topic-aware factors in {4, 8, 16, 32} on DBLP (A-P), YELP (U-B) and Amazon (U-M). As shown in Fig. 3, a better performance on AUC and AP is substantially coupled with topic prior guidance especially when K is set to be relatively small. This is because a small K with higher dimension for each factor easily leads to collapse and the topic prior guidance can help to relieve and keep the quality of multi-facet topic-aware embeddings. We also notice that the improvements on Amazon when $K = 8$ and $K = 16$ is much less than that on others. This might suggest that, the inferential process is performed well enough and THGNN still achieves competitive performance against baselines without topic prior guidance since users' preference on movies are diverse.

5.4 Multi-facet Embedding Visualization (Q3)

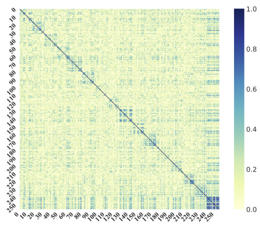
To analyze the quality of the multi-facet topic-aware embeddings learned by THGNN intuitively, we visualize the correlation analysis between the elements of multi-facet node embeddings learned in DBLP (A-P) task and Amazon (U-M) task. The number of multi-facet topic-aware factors is set to $K = 16$ in DBLP and $K = 8$ in Amazon. Fig. 4 (a), Fig. 4 (b) and Fig. 4 (c, d) show the results derived from MAGNN, THGNN without topic prior guidance and THGNN respectively. The latent features learned by MAGNN are still hiddenly entangled. It can be explained that the multi-head attention mechanism [25] adopted by MAGNN has little ability to capture more fine-grained multi-facet semantics, and its main function is to reflect in stabilizing the training process. According to Fig. 4 (c, d) we can also observe on both DBLP and Amazon dataset that THGNN is able to extract independent multi-facet representations especially on Amazon dataset, as the correlation plot exhibits K clear diagonal blocks. Comparing the quality of Fig. 4 (b) and Fig. 4 (c), we can further make a conclusion that the topic prior guidance plays an important role on giving more orthogonality and keeping the quality of multi-facet topic-aware embeddings, and thus THGNN achieves a better performance in the link prediction task.

Table 3: Case study of the author u198's top 2 papers by the probability in her three most concerned factors on DBLP (A-P).

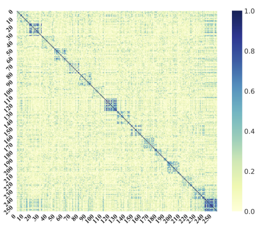
Factor	Text Content (Title) of Top 2 Papers	Probability
9	BibFinder/StatMiner: Effectively Mining and Using Coverage and Overlap Statistics in Data Intergration.	0.1327
	Improving text collection selection with coverage and overlap statistics.	0.1621
14	When is Temporal Planning Really Temporal?	0.1404
	Parallelizing State Space Plans Online.	0.1480
16	Query Processing over Incomplete Autonomous Databases.	0.2782
	Answering Imprecise Queries over Autonomous Web Databases.	0.2632



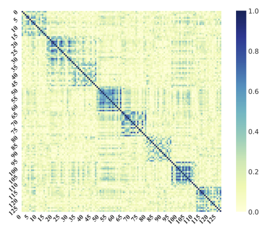
(a) MAGNN DBLP (A-P).



(b) THGNN-w/o DBLP (A-P).

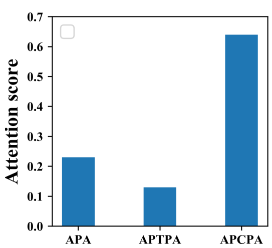


(c) THGNN DBLP (A-P).

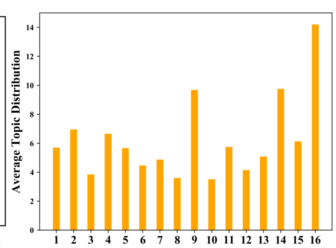


(d) THGNN Amazon (U-M).

Figure 4: The magnitude of the correlations between the elements of the 256-dimensional representations for DBLP (A-P), and 128-dimensional representations for Amazon (U-B).



(a) Attention scores of metapaths.

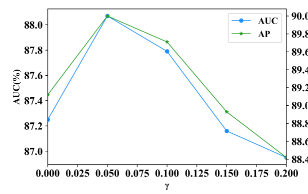


(b) Average topic distribution.

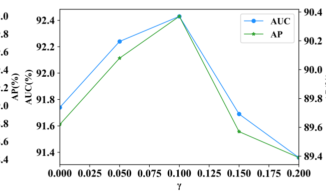
Figure 5: Attention values on metapaths and average topic-aware distribution of APA-based context of u198 on DBLP (A-P).

5.5 Semantics Analysis: Case Study

In order to investigate the semantics in both structural level and topic-aware level, we further present a case study for a deeper understanding. We randomly select an author $u198$ from DBLP (A-P) task, and the attention score of metapaths as well as her average topic distribution of APA-based context are shown in Fig. 5. Fig. 5 (a) shows the metapath APCPA is given the largest weight which means that $u198$ considers the APCPA as the most critical metapath when she connects with a paper node. It suggests that a conference may represent a research topic. From Fig. 5 (b) we can find that



(a) DBLP (A-P).



(b) YELP (U-B).

Figure 6: Analysis of parameter γ.

$u198$ mainly focus on topic-aware factor 9, 14 and 16. For each one, we visualize the title of top 2 papers in her metapath-based context as well as their probability. As shown in Table 3, for example, factor 9 is probably related to "Overlap statistics" topic, factor 16 may indicate the "Database" research field.

5.6 Parameter Study

Finally, we perform a sensitivity analysis of the parameter γ on DBLP and YELP datasets to explore how topic prior guidance influences model performance. The results are shown in Fig. 6. As the topic prior coefficient increases, the performance increases initially as the guidance is beneficial to the multi-facet topic-aware representation learning. However, it will drop quickly if γ is larger than 0.1 on both DBLP and YELP because too sharp topic-aware distribution inference which cannot maintain topic relevance will negatively affect the model.

6 CONCLUSIONS

In this paper, we proposed to identify and reason multi-facet topic-aware factors underlying the bare connections between associated nodes, by taking advantage of both heterogeneous structures and unstructured text content in HGs. Specifically, we proposed a novel framework THGNN for link prediction to distinctively aggregate rich heterogeneous information according to the inferential multi-facet topic-aware factors, so as to generate multi-facet topic-aware representations preserving both structural and topic-aware semantics. Moreover, we incorporated a topic prior guidance module, which aims to leverage global knowledge from unstructured text content, to further improving the quality of multi-facet topic-aware embeddings. Experiments show that the learned multi-facet topic-aware node embeddings are more predictive and interpretable.

7 ACKNOWLEDGMENTS

This research is supported in part by the National Natural Science Foundation of China (No. U20B2045, 61772082, 62002029), the Fundamental Research Funds for the Central Universities 2020RC23, and Meituan.

REFERENCES

[1] Yoshua Bengio, Aaron Courville, and Pascal Vincent. 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence* 35, 8 (2013), 1798–1828.

[2] Filippo Maria Bianchi, Daniele Grattarola, and Cesare Alippi. 2020. Spectral clustering with graph neural networks for graph pooling. In *ICML*. PMLR, 874–883.

[3] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. 2013. Spectral networks and locally connected networks on graphs. *arXiv preprint arXiv:1312.6203* (2013).

[4] Ting Chen and Yizhou Sun. 2017. Task-guided and path-augmented heterogeneous network embedding for author identification. In *WSDM*. 295–304.

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

[6] Shaohua Fan, Junxiong Zhu, Xiaotian Han, Chuan Shi, Linmei Hu, Biyu Ma, and Yongliang Li. 2019. Metapath-guided heterogeneous graph neural network for intent recommendation. In *KDD*. 2478–2486.

[7] Tao-yang Fu, Wang-Chien Lee, and Zhen Lei. 2017. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. In *CIKM*. 1797–1806.

[8] Xinyu Fu, Jiani Zhang, Ziqiao Meng, and Irwin King. 2020. MAGNN: metapath aggregated graph neural network for heterogeneous graph embedding. In *WWW*. 2331–2341.

[9] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. 2018. Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728* (2018).

[10] William L Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *arXiv preprint arXiv:1706.02216* (2017).

[11] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. 2016. beta-vae: Learning basic visual concepts with a constrained variational framework. (2016).

[12] Jun-Ting Hsieh, Bingbin Liu, De-An Huang, Li Fei-Fei, and Juan Carlos Niebles. 2018. Learning to decompose and disentangle representations for video prediction. *arXiv preprint arXiv:1806.04166* (2018).

[13] Binbin Hu, Chuan Shi, Wayne Xin Zhao, and Philip S Yu. 2018. Leveraging meta-path based context for top-n recommendation with a neural co-attention model. In *KDD*. 1531–1540.

[14] Binbin Hu, Zhiqiang Zhang, Chuan Shi, Jun Zhou, Xiaolong Li, and Yuan Qi. 2019. Cash-out user detection based on attributed heterogeneous information network with a hierarchical attention mechanism. In *AAAI*. 946–953.

[15] Linmei Hu, Siyong Xu, Chen Li, Cheng Yang, Chuan Shi, Nan Duan, Xing Xie, and Ming Zhou. 2020. Graph neural news recommendation with unsupervised preference disentanglement. In *ACL*. 4255–4264.

[16] Hyunjik Kim and Andriy Mnih. 2018. Disentangling by factorising. In *ICML*. PMLR, 2649–2658.

[17] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[18] Yanbei Liu, Xiao Wang, Shu Wu, and Zhitao Xiao. 2020. Independence promoted graph disentangled networks. In *AAAI*. 4916–4923.

[19] Jianxin Ma, Peng Cui, Kun Kuang, Xin Wang, and Wenwu Zhu. 2019. Disentangled graph convolutional networks. In *ICML*. PMLR, 4212–4221.

[20] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

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

[22] Jingbo Shang, Meng Qu, Jialu Liu, Lance M Kaplan, Jiawei Han, and Jian Peng. 2016. Meta-path guided embedding for similarity search in large-scale heterogeneous information networks. *arXiv preprint arXiv:1610.09769* (2016).

[23] Chuan Shi, Binbin Hu, Wayne Xin Zhao, and S Yu Philip. 2018. Heterogeneous information network embedding for recommendation. *IEEE Transactions on Knowledge and Data Engineering* 31, 2 (2018), 357–370.

[24] 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. *Proceedings of the VLDB Endowment* 4, 11 (2011), 992–1003.

[25] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *arXiv preprint arXiv:1706.03762* (2017).

[26] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

[27] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural graph collaborative filtering. In *Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval*. 165–174.

[28] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous graph attention network. In *WWW*. 2022–2032.

[29] Xiang Wang, Hongye Jin, An Zhang, Xiangnan He, Tong Xu, and Tat-Seng Chua. 2020. Disentangled Graph Collaborative Filtering. In *SIGIR*. 1001–1010.

[30] Yifan Wang, Suyao Tang, Yuntong Lei, Weiping Song, Sheng Wang, and Ming Zhang. 2020. DisenHAN: Disentangled Heterogeneous Graph Attention Network for Recommendation. In *CIKM*. 1605–1614.

[31] Yiding Yang, Zunlei Feng, Mingli Song, and Xinchao Wang. 2020. Factorizable graph convolutional networks. *arXiv preprint arXiv:2010.05421* (2020).

[32] Chuxu Zhang, Chao Huang, Lu Yu, Xiangliang Zhang, and Nitesh V Chawla. 2018. Camel: Content-aware and meta-path augmented metric learning for author identification. In *WWW*. 709–718.

[33] Jiawei Zhang, Congying Xia, Chenwei Zhang, Limeng Cui, Yanjie Fu, and S Yu Philip. 2017. BL-MNE: emerging heterogeneous social network embedding through broad learning with aligned autoencoder. In *ICDM*. IEEE, 605–614.

[34] Yizhou Zhang, Yun Xiong, Xiangnan Kong, Shanshan Li, Jinhong Mi, and Yangyong Zhu. 2018. Deep collective classification in heterogeneous information networks. In *WWW*. 399–408.