The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)

Deep Structural Knowledge Exploitation and Synergy for Estimating Node Importance Value on Heterogeneous Information Networks

Yankai Chen[1], Yixiang Fang[2*], Qiongyan Wang[3], Xin Cao[4], Irwin King[1]

[1] The Chinese University of Hong Kong, Hong Kong
[2] The Chinese University of Hong Kong, ShenZhen
[3] University of Copenhagen, Denmark
[4] The University of New South Wales, Australia
{ykchen,king}@cse.cuhk.edu.hk, fangyixiang@cse.cuhk.edu.cn, qiongyanwang010@gmail.com, xin.cao@unsw.edu.au

Abstract

Node importance estimation problem has been studied conventionally with homogeneous network topology analysis. To deal with network heterogeneity, a few recent methods employ graph neural models to automatically learn diverse sources of information. However, the major concern revolves around that their full adaptive learning process may lead to insufficient information exploration, thereby formulating the problem as the *isolated* node value prediction with underperformance and less interpretability. In this work, we propose a novel learning framework: SKES. Different from previous automatic learning designs, SKES exploits heterogeneous structural knowledge to enrich the informativeness of node representations. Based on a *sufficiently uninformative* reference, SKES estimates the importance value for any input node, by *quantifying* its disparity against the reference. This establishes an interpretable node importance computation paradigm. Furthermore, SKES dives deep into the understanding that "*nodes with similar characteristics are prone to have similar importance values*" whilst guaranteeing that such informativeness disparity between any different nodes is orderly reflected by the embedding distance of their associated latent features. Extensive experiments on three widely-evaluated benchmarks demonstrate the performance superiority of SKES over several recent competing methods.

Introduction

Estimating node importance, as one of the classic problems in network science, founds various downstream applications, such as recommender systems, web information search and retrieval, query disambiguation, and resource allocation optimization (Zhang and Zhu 2019; Park et al. 2020; Zheng et al. 2021; Yang et al. 2022; Zhang et al. 2022a; Hu et al. 2020a, 2022, 2021b; Song, Zhang, and King 2023c; Chen et al. 2021; Fang et al. 2017; He et al. 2023b,a). Traditional approaches revolve around the analyses of *network topologies*, e.g., closeness centrality (Nieminen 1974), degree analysis (Nieminen 1974), and PageRank methodologies (Page et al. 1999; Haveliwala 2003).

With the proliferation of heterogeneous information in graph data, conventional methods focusing on topology analysis may thus fail to capture such diverse semantic knowledge embedded. Heterogeneous Information Networks (HINs), which are essentially prevalent in various domains including bibliographic information networks, social media, and knowledge graphs, usually are composed of multiple typed nodes and edges. Notably, **network heterogeneity leads to the variations of semantics and values in interpreting node importance**. This makes the studied problem more challenging on HINs than that on homogeneous counterparts. We illustrate this by an HIN example of DBLP network in Figure 1(a). The important values of *authors*, *papers*, *venues*, and *topics* are indicated by their h-index values, citation numbers, venue rank, and popularity (e.g., numbers of Web pages in Google), respectively. Despite the network connection, adjacent nodes may have different influences on the importance value of the target node. For example, *authors* and *topics* have different facets of contributions to the importance values of *papers*. Moreover, while the scholar's h-index values are often in [0, 350][1], the maximum paper citation number could be over 300,000[2]. This shows that the importance value heterogeneity is essentially influenced by semantic heterogeneity. Consequently, it becomes evident that both variations are inherently difficult to be simply analyzed from homogeneous network topologies, necessitating the exigency for new methodologies.

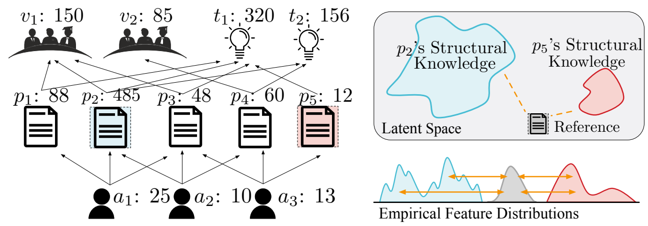

Figure 1: An HIN example and SKES methodology.

Related Works. Among a few recent attempts, methods with *Graph Neural Networks* (GNNs) have emerged as a promising direction (Park et al. 2019). Due to good knowledge mining ability from high-order topologies, GNNs can produce semantic enrichment to vectorized node represen-

*The corresponding author.

[1] https://www.webometrics.info/en/hlargerthan100

[2] https://www.genscript.com/top-100-most-cited-publications.html

tations benefiting downstream tasks (Chen et al. 2022a,b, 2023a; Yang et al. 2023a,b; Zhang et al. 2023b; Chen et al. 2020). By incorporating specific designs for HINs, GNN-based methods show the potential in dealing with information heterogeneity (Liang et al. 2023; Fu and King 2023), especially for node importance estimation. For instance, GENI (Park et al. 2019) applies GNN and attention mechanism to aggregate the structure information for node importance estimation. MULTIIMPORT (Park et al. 2020) improves GENI by using a variety of external input signals. RGTN (Huang et al. 2021) utilizes both the network structure information and nodes' attributes for estimating the importance of nodes. However, all these works focus on solving the *importance-based ranking problem*, without inferring the specific importance values. Recent work HIVEN (Huang et al. 2022) considers the value heterogeneity of node importance in HINs by learning both local and global node information. Nevertheless, the primary concern lies in that HIVEN purely relies on GNNs for automatic information aggregation but ignores the explicit structural knowledge mining on HINs, making the model underperforming and less interpretable in importance calculation.

Our Contribution. We push forward the investigation of node importance estimation over HINs by introducing a novel learning framework, namely SKES (Deep Structural Knowledge Exploitation and Synergy). SKES makes the assumption that each node corresponds to a unique *high-dimensional feature distribution* reflecting its essential characteristics and knowledge to determine the node importance. However, such feature distribution is unknown and agnostic that can only be sampled and observed by certain *empirical feature representations*. These empirical representations are expected to be as much informative with heterogeneous knowledge as possible, so that the importance of each node within the underlying HIN can thus be estimated from its associated feature representations. Then SKES transforms the importance regression problem into quantifying the *informativeness* of these empirical node feature representations. Underpinned by Optimal Transport Theory (Villani et al. 2009), our formulation eventually provides an effectual and interpretable learning paradigm with theoretical guarantees.

Specifically, SKES consists of three progressive modules. For each node, (1) *Structural Priori Knowledge Exploitation* focuses on mining the intrinsic intra- and inter-node information, i.e., *centrality* and *similarity*, from the given HIN, providing the comprehensive coverage of structural knowledge with diversity and heterogeneity. Then our (2) *Synergetic Representation of Feature Distribution* module learns to empirically represent the node's unique, complicated, and high-dimensional feature distribution with adaptive heterogeneous knowledge synergy. Lastly, (3) we manually create a random feature distribution as the reference, which functions as the "coordinate origin" in the embedding space. Due to the randomness, this reference is *sufficiently uninformative*. Then our *Node Importance Value Estimation* module quantifies the informativeness of the input node by measuring its distance against the reference in the latent sapce, and transforms such measurement for node importance estimation. Furthermore, anchoring on this reference, the estimated importance values obey *triangle inequality* such that the informativeness gap between different node pairs can also be captured. This produces a fine-grained importance learning framework, which is different from previous methods that formulate the problem as the importance value prediction for isolated nodes. We provide a high-level illustration in Figure 1(b) and summarize our principal contributions as:

- To the best of our knowledge, we are the first to formulate the HIN node importance estimation problem via quantifying node feature informativeness with Optimal Transport methodology, providing a novel and interpretable perspective to the related community.
- We propose SKES model with three effective modules that operate progressively from structural knowledge exploitation and synergy to node importance estimation.
- We conduct model evaluation on three real-world benchmarks. Experimental results demonstrate the performance superiority of our model against competing methods as well as the effectiveness of each module contained therein.

Preliminaries and Problem Formulation

Definition 1: Heterogeneous Information Network (HIN). It is a directed graph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ with a node type mapping function $\phi : \mathcal{V} \rightarrow \mathcal{A}$ and an edge type mapping function $\psi : \mathcal{E} \rightarrow \mathcal{R}$, where $\mathcal{A}$ is a set of node types and $\mathcal{R}$ is a set of edge types satisfying $|\mathcal{A}| + |\mathcal{R}| > 2$.

Definition 2: Metapaths. A metapath $\mathcal{P}$ is with the form $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \cdots \xrightarrow{R_H} A_{H+1}$ that defines on the node and edge types, i.e., $A_i \in \mathcal{A}$, $R_i \in \mathcal{R}$. We omit the edge types if they are unique between two connected node types, e.g., $A_1A_2 \cdots A_{H+1}$. We call a path between nodes v_1 to v_{H+1} a *path instance* of $\mathcal{P}$, if $\forall i$, the node v_i and edge $e_i = (v_i, v_{i+1})$ satisfy that $\phi(v_i) = A_i$ and $\psi(e_i) = R_i$.

Definition 3: 1-Wasserstein Distance. While Optimal Transport (OT) is the problem of moving one distribution of mass, e.g., P, to another, e.g., Q, as efficiently as possible, 1-Wasserstein distance is the derived minimum distribution distance that is defined by the following formulation:

$$W(P, Q) = \inf_{f \in TP(P,Q)} \int \|\boldsymbol{x} - f(\boldsymbol{x})\|_1 dP(\boldsymbol{x}), \quad (1)$$

where the infimum is over $TP(P, Q)$ that denotes all transport plans. If a minimizer f^* exists, it is thus the solution to compute $W(P, Q)$. For common one-dimensional distributions, there is a closed-form solution to compute the optimal plan f^* as $f^*(x) := F_P^{-1}(F_Q(x))$; F is the cumulative distribution function (CDF) associated with the underlying distribution. 1-Wasserstein distance satisfies *positive-definiteness*, *symmetry*, and *triangle inequality* (Nietert et al. 2022; Korotin, Selikhanovych, and Burnaev 2023; Chen et al. 2023c,b; Naderializadeh et al. 2021).

<u>**Problem Definition**</u>. *Given an HIN $\mathcal{H}$ =($\mathcal{V}$, $\mathcal{E}$) and the importance values for a subset of nodes $\mathcal{V}' \subset \mathcal{V}$ of some given types $\mathcal{A}' \subseteq \mathcal{A}$, we aim to learn a mapping function $g(\cdot) : \mathcal{V} \rightarrow \mathbb{R}$ that estimates the importance value of every node of the given types $\mathcal{A}'$ in $\mathcal{H}$.*

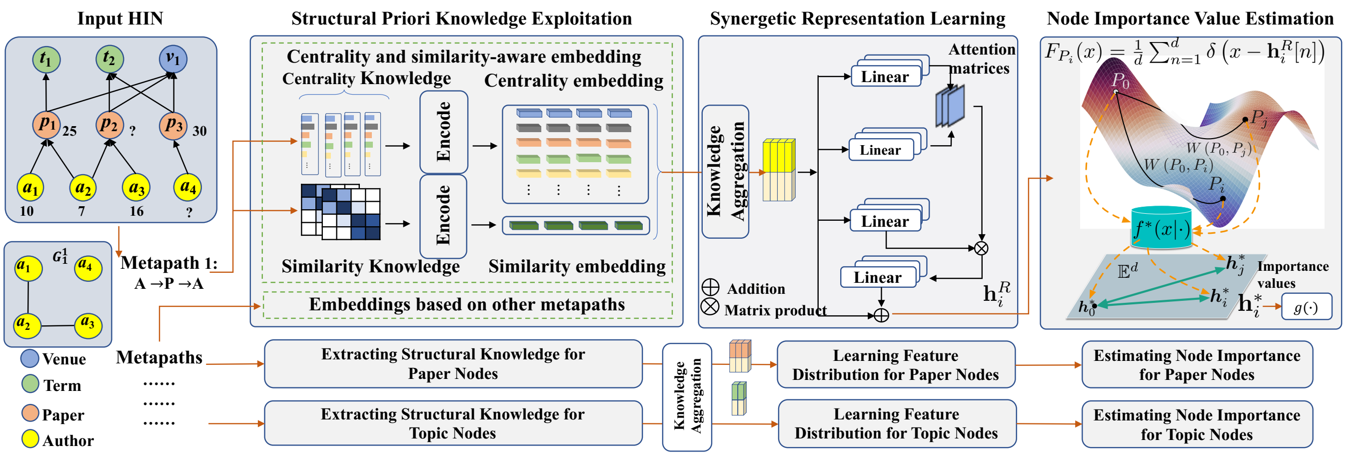

Figure 2: The framework of our proposed model (best view in color).

SKES Methodology

Overview

We now formally introduce our SKES model (Deep Structural Knowledge Exploitation and Synergy). To implement the mapping function $g(\cdot)$, the general framework of SKES to estimate the node importance value is: *given a random distribution P_0 representing the sufficiently uninformative reference, for each node $v_i \in \mathcal{V}$, we first learn to represent the high-dimensional feature distribution P_i for v_i, then the mapping function $g(\cdot)$ can be implemented as $\forall v_i \in \mathcal{V}$, $g(v_i) = g(W(P_0, P_i))$.* As depicted in Figure 2, SKES comprises three progressively-operated modules.

Structural Priori Knowledge Exploitation

Acquiring informative structural knowledge is critical to estimating node importance. To achieve this in HINs, we adopt the metapath-based methodologies (Sun et al. 2011) to first obtain the underlying sub-networks as follows.

1) Metapath-induced Sub-network Construction. There are limited but meaningful *metapaths* in HINs that describe the meta information of HINs (Fu et al. 2020; Fu and King 2024). For the k-th metapath, the induced sub-network is denoted as $G_k^j = (\mathcal{V}_k^j, \mathcal{E}_k^j)$, such that $\mathcal{V}_k^j$ contains all j-th typed nodes and $\mathcal{E}_k^j$ contains all the edges between nodes in $\mathcal{V}_k^j$, i.e., two nodes are linked in this sub-network if there is a path instance of the metapath between them.

2) Priori Centrality and Similarity Embedding. After obtaining the induced sub-networks, we propose to extract node pointwise *centrality* and pairwise *similarity*, which are two essential network properties that reveal the *intra-* and *inter*-node priori knowledge. Specifically, centrality measures generally are either defined based on network properties (Nieminen 1974; Hirsch 2005; Egghe et al. 2006; Negre et al. 2018; Page et al. 1999; Dorogovtsev, Goltsev, and Mendes 2006), or designed based on the shortest paths (Shaw 1954; Marchiori and Latora 2000; Sabidussi 1966). To fully capture the diverse information of the underlying structures, we pre-process these popular centrality measures to improve knowledge coverage. Detailed formulations are listed in Appendix. Then each centrality value is vectorized into 128-dimension via a two-layer perceptron. We denote the embedding calculated from the l-th (l ranges from 1 to L) centrality value of node v_i by $\mathbf{c}_{i,k}^{(l)}$.

While centrality reflects the property of a given node oneself, *similarity* reveals the contrastive node information compared to others. In this work, SKES embeds the knowledge from the *attribute-* and *topology-aware* similarity. Specifically, for each edge in the induced sub-network G_k^j, we assign the edge weight by the cosine similarity between the attributes of its two end nodes. Then we transform the similarity matrix as the transition probabilities to compute the node embeddings by adopting node2vec (Grover and Leskovec 2016). To embed topology-aware similarity, we directly follow PathSim (Sun et al. 2011) to calculate the similarity of each node pair in G_k^j and then take an analogous embedding procedure via node2vec. Subsequently, we take the summation of two corresponding embeddings, denoted by $\mathbf{f}_{i,k}^{att}$ and $\mathbf{f}_{i,k}^{top}$ of the node v_i, and finally have the similarity knowledge embedding as $\mathbf{c}_{i,k}^{(L+1)} = \mathbf{f}_{i,k}^{att} + \mathbf{f}_{i,k}^{top}$.

To summarize, centrality and similarity knowledge provides cohesive and complementary views of underlying networks, as each one of these measures usually represents a specific perspective of structural information. As we will show in experiments, they substantially boost SKES performance and stabilize the model training.

Synergetic Representation of Feature Distribution

1) Heterogeneous Knowledge Aggregation. Since the priori centrality and similarity knowledge represents the node information from different perspectives, we propose to synergetically fuse them for later representing the unknown high-dimensional feature distribution. Concretely, for each node v_i, let $\alpha_{i,k}^{(l)}$ denote the weighting coefficient of the l-th centrality embedding and $\alpha_{i,k}^{(L+1)}$ represent the coefficient of v_i's similarity embedding. We first derive the following coefficient calculations with l ranging from 1 to $L+1$:

$$\alpha_{i,k}^{(l)} = \frac{1}{|G_k^{\phi(v_i)}|} \sum_{v_{i'} \in G_k^{\phi(v_i)}} \mathbf{W}_1 \tanh(\mathbf{W}_1' \mathbf{c}_{i',k}^{(l)} + \mathbf{b}_1), \tag{2}$$

where $G_k^{\phi(v_i)}$ denotes the induced sub-network. $\mathbf{W}_1$, $\mathbf{W}_1'$, and $\mathbf{b}_1$ are learnable parameters. These coefficients are further normalized with the softmax function, i.e., $\widehat{\alpha}_{i,k}^{(l)} = \exp(\alpha_{i,k}^{(l)}) / \sum_{l'=1}^{L+1} \exp(\alpha_{i,k}^{(l')})$. $\widehat{\alpha}_{i,k}^{(l)}$ attentively contributes to the k-th metapath derived knowledge embedding $\mathbf{e}_{i,k}$ as:

$$\mathbf{e}_{i,k} = \sum_{l=1}^{L+1} \widehat{\alpha}_{i,k}^{(l)} \mathbf{c}_{i,k}^{(l)}. \tag{3}$$

Besides, for each node v_i, since different metapaths lead to different sub-network extraction, we further adaptively fuse embeddings from these different sub-networks containing v_i as well. Similarly, the coefficient $\tau_{i,k}$ of v_i induced by the k-th metapath is defined as:

$$\tau_{i,k} = \frac{1}{|G_k^{\phi(v_i)}|} \sum_{v_{i'} \in G_k^{\phi(v_i)}} \mathbf{W}_2 \tanh(\mathbf{W}_2' \mathbf{e}_{i',k} + \mathbf{b}_2), \tag{4}$$

where $\mathbf{W}_2$, $\mathbf{W}_2'$, and $\mathbf{b}_2$ are learnable parameters. Similarly, Eqn. (4) is further normalized across all other related coefficients as, $\widehat{\tau}_{i,k} = \exp(\tau_{i,k}) / \sum_{k'=1}^{N_{\phi(v_i)}} \exp(\tau_{i,k'})$. $N_{\phi(v_i)}$ denotes the number of metapaths starting from node type $\phi(v_i)$. We derive the aggregated embedding of $\mathbf{e}_i$ as follows:

$$\mathbf{e}_i = \sum_{k=1}^{N_{\phi(v_i)}} \widehat{\tau}_{i,k} \mathbf{e}_{i,k}. \tag{5}$$

Given v_i's initial feature $\mathbf{e}_i'$ and the learned knowledge embedding $\mathbf{e}_i$, we obtain the aggregated representation $\mathbf{x}_i$ by concatenation (denoted as $||$), i.e., $\mathbf{x}_i = \mathbf{e}_i' \mid\mid \mathbf{e}_i$.

2) Empirical Representation of Feature Distribution. Intuitively, $\mathbf{x}_i$ contains both the initial node features and refined structural knowledge. We then adaptively learn the empirical representations that are informative to represent the *unknown high-dimensional node feature distributions*.

We achieve this by leveraging the self-attention mechanism (Vaswani et al. 2017; Chen et al. 2022c). Specifically, we extract hidden features from the input $\mathbf{x}_i$ via implementing M attention heads in each layer. We denote the hidden feature of node v_i learned by the r-th layer as $\mathbf{h}_i^{(r)}$. We follow the conventional computation protocol to firstly obtain d-dimensional *query*, *key* and *value* variables:

$$\mathbf{q}_{i,m}^{(r)} = \mathbf{W}_{qry}^m \mathbf{h}_{i,m}^{(r)},\ \mathbf{k}_{i,m}^{(r)} = \mathbf{W}_{key}^m \mathbf{h}_{i,m}^{(r)},\ \mathbf{v}_{i,m}^{(r)} = \mathbf{W}_{val}^m \mathbf{h}_{i,m}^{(r)}, \tag{6}$$

where $\mathbf{q}_{i,m}^{(r)}$, $\mathbf{k}_{i,m}^{(r)}$, and $\mathbf{v}_{i,m}^{(r)}$ denote the m-th query, key, and value vectors of v_i at the r-th layer. $\mathbf{W}_{qry}^m$, $\mathbf{W}_{key}^m$, and $\mathbf{W}_{val}^m$ are learnable weights. Then, the attentive coefficient between nodes v_j and v_i is calculated as:

$$S_m^{(r)}(v_j, v_i) = \frac{\exp(\mathbf{q}_{i,m}^{(r)} \mathbf{W}_{\psi(e_{j,i})} (\mathbf{k}_{i,m}^{(r)})^\top \frac{\mu^{\psi(e_{j,i})}}{\sqrt{d}})}{\sum\limits_{v_{j'} \in \mathcal{N}(v_i)} \exp(\mathbf{q}_{i,m}^{(r)} \mathbf{W}_{\psi(e_{j',i})} (\mathbf{k}_{j',m}^{(r)})^\top \frac{\mu^{\psi(e_{j',i})}}{\sqrt{d}})}, \tag{7}$$

where $e_{j,i}$ denotes the edge from v_j to v_i, and $\mathbf{W}_{\psi(e_{j,i})}$ represents the learnable weight matrix of edge type $\psi(e_{j,i})$. $\mu^{\psi(e_{j,i})}$ is the learnable magnitude for type $\psi(e_{j,i})$ and $\mathcal{N}(i)$ is the neighbor set of node v_i. Then the embedding $\mathbf{v}_{i,m}^{(r)}$ is updated via aggregating adjacent information as follows:

$$\tilde{\mathbf{v}}_{i,m}^{(r)} = \sum_{v_j \in \mathcal{N}(v_i)} S_m^{(r)}(v_j, v_i) \mathbf{v}_{i,m}^{(r)}. \tag{8}$$

Let $||$ denote the concatenation and $\mathbf{W}_{out}$ is a learnable weight matrix. We finally complete the target feature representation by iteratively updating from $r=1$ to $R-1$:

$$\mathbf{h}_{i,m}^{(r+1)} = \mathbf{h}_{i,m}^{(r)} + \mathbf{W}_{out} \cdot \left(||_{m=1}^{M} \tilde{\mathbf{v}}_{i,m}^{(r)} \right). \tag{9}$$

The output of Eqn. (9), i.e., $\mathbf{h}_i^R$ for brevity, is expected to be empirically representative for the *unknown node feature distribution* with heterogeneous knowledge synergy. We explain our implementation to estimate node importance via mensurating the empirical distribution distances as follows.

Node Importance Value Estimation

As we mentioned earlier, notation P_i denotes the feature distribution associated with node v_i. Since we use $\mathbf{h}_i^R$ to represent P_i, which is discrete, then its empirical CDF can be defined as follows:

$$F_{P_i}(x) = \frac{1}{d} \sum_{n=1}^{d} \delta \left(x - \mathbf{h}_i^R[n] \right), \tag{10}$$

where $\delta(\cdot)$ returns 1 if the input is zero and 0 otherwise[3]. $\mathbf{h}_i^R[n]$ is the n-th element. To explicitly measure the distribution distance, we propose to compare P_i with a fixed *random reference* that functions as the "origin" in the embedding space. Specifically, we introduce a reference distribution P_0 with associated feature representation $\mathbf{h}_0 \in \mathbb{R}^d$, elements of which are uniformly sampled. Then the distribution distance between P_i and P_0 can be explicitly measured by 1-Wasserstein distance, i.e., $W(P_0, P_i)$.

As we have introduced in Preliminaries, $W(P_0, P_i)$ is computed via implementing the optimal transport plan $f^*(x) := F_{P_i}^{-1}(F_{P_0}(x))$. Based on the empirical CDFs of P_0 and P_i, we can quantitatively interpreted $f^*(x)$ as:

$$f^* \left(x | \mathbf{h}_i^R \right) = \operatorname{argmin}_{x' \in \mathbf{h}_i^R} \left(F_{P_i}(x') = \gamma \right),\ \gamma = F_{P_0}(x). \tag{11}$$

Moreover, let $\pi(x'|\mathbf{h}_i^R)$ denote the ranking of each input x' in the ascending sorting of elements in $\mathbf{h}_i^R$. We can further achieve the following algorithmic implementation:

$$f^* \left(x | \mathbf{h}_i^R \right) = \operatorname{argmin}_{x' \in \mathbf{h}_i^R} \left(\pi(x'|\mathbf{h}_i^R) = \pi(x|\mathbf{h}_0) \right). \tag{12}$$

Please notice that, the indicator $\pi(\cdot)$ can be actually preprocessed via "argsort" to $\mathbf{h}_i^R$ and "sort" to $\mathbf{h}_0$, which *essentially permutes and encodes* $\mathbf{h}_i^R$ *via referring to* $\mathbf{h}_0$. Therefore, the resultant representation is denoted as $\mathbf{h}_i^*$:

$$\mathbf{h}_i^* = ||_{n=1}^{d} \left(f^*(\mathbf{h}_0[n] | \mathbf{h}_i^R) - \mathbf{h}_0[n] \right). \tag{13}$$

$\mathbf{h}_i^* \in \mathbb{R}^d$ presents several desirable geometric properties, as it can efficiently reflect the 1-Wasserstein distance between distributions P_0 and P_i as follows:

$$\|\mathbf{h}_i^*\|_1 \propto W(P_0, P_i) \text{ and } \left\|\mathbf{h}_i^* - \mathbf{h}_j^*\right\|_1 \propto W(P_i, P_j). \tag{14}$$

[3]Dirac delta function with $\int \delta(x) dx = 1$ for continuous inputs.

Dataset	# Nodes	# Edges	# Node types	# Edge types	Target node	Meaning	# Node with Importance	Range
MUSIC10K	22,986	80,272	4	8	Artist	Familiarity	4,214 (18.3%)	[0, 1]
					Song	Hotness	4,411 (19.1%)	[0, 1]
TMDB5K	76,926	359,780	7	12	Movie	Popularity	4,802 (6.2%)	[-7.89, 6.77]
					Director	Box office	1,159 (1.5%)	[0.021, 10.55]
DBLP	249,903	2,428,250	4	6	Author	H-index	101,958 (40.8%)	[0, 159]
					Paper	Citations	100,000 (40.0%)	[0, 34191]

Table 1: Statistics of three datasets (MUSIC10K, TMDB5K and DBLP).

We attach the proof of Eqn. 14 in Appendix. $\mathbf{h}_i^*$ naturally inherit the relative order of the distance ranking with the theoretical guarantees. Based on $\mathbf{h}_i^*$, we finally implement the importance estimation function $g(\cdot)$:

$$g(v_i) = \boldsymbol{\lambda} \cdot \mathbf{h}_i^*, \tag{15}$$

where $\boldsymbol{\lambda}$ is a learnable vector to provide better regression capability. Obviously, for any node v_i, its importance value $g(v_i)$ is correlated with its distribution distance to the reference P_0. We showcase and analyze its performance superiority over competing methods in Experimental Evaluation.

Training Objective

We adopt the common regression loss with mean squared error between estimated and ground-truth importance values:

$$\mathcal{L}_{mse} = \frac{1}{|\mathcal{A}'|}\sum_{j\in\mathcal{A}'}\frac{1}{|\mathcal{V}^j|}\sum_{v_i\in\mathcal{V}^j}(g(v_i) - y_{v_i})^2, \tag{16}$$

where y_{v_i} denotes the ground-truth importance value of node v_i. Then the complete training objective is defined as:

$$\mathcal{L} = \mathcal{L}_{mse} + \mu\|\Delta\|_2^2. \tag{17}$$

$\|\Delta\|_2^2$ is the $L2$-regularizer of trainable embeddings and variables to avoid over-fitting with the hyperparameter μ.

Experiments

We evaluate SKES with the research questions (RQs) as:

- **RQ1**: How does SKES compare to state-of-the-art methods on the tasks of *node importance value estimation* and *important node ranking*?
- **RQ2**: How does our proposed design of knowledge synergy and measurement contribute to SKES performance?
- **RQ3**: How does other proposed components of SKES influence the model performance?

Experiment Setups

1) Benchmarks. We include three widely evaluated real-world HIN datasets, namely MUSIC10K, TMDB5K, and DBLP. Dataset statistics are reported in Table 1 and detailed data descriptions are attached in Appendix.

2) Evaluation Metrics. For the node importance estimation task, three metrics are applied for performance evaluation, including mean absolute error (MAE), root mean square error (RMSE), and normalized root mean square error (NRMSE). The lower value of these metrics, the better the model performance. On the importance-based ranking task, we use normalized discounted cumulative gain (NDCG) and Spearman correlation coefficient (SPEARMAN), where a higher value indicates better performance.

3) Experimental Settings. In line with prior work (Park et al. 2019; Huang et al. 2022), we perform five-fold cross validation for testing and report the average performance. For each fold, 80% and 20% of nodes with ground-truth importance values are used for training and testing. 15% of training nodes is used for validating. We select symmetric metapaths with lengths less than four. We implement SKES using Python 3.8 and PyTorch 1.8.0 on a Linux machine with 4 Nvidia A100 GPUs and 4 Intel Core i7-8700 CPUs. Following HIVEN (Huang et al. 2022), node features are initialized from textual contents via sentence-BERT (Reimers and Gurevych 2019). We set the learning rate as 10^{-3} and train the model via Adam optimizer. Metapaths and hyperparameter settings are reported in Appendix.

4) Competing Models. We include three groups of existing models: (1) traditional network analytic methods, i.e., PageRank (PR) (Page et al. 1999) and personalized PageRank (PPR) (Haveliwala 2003); (2) machine learning methods, i.e., linear regression (LR) and random forest (RF); (3) neural network based models, i.e., GAT (Veličković et al. 2017), HGT (Hu et al. 2020b), GENI (Park et al. 2019), Multiimport (MULTI) (Park et al. 2020), RGTN (Huang et al. 2021), and HIVEN (Huang et al. 2022). Detail descriptions are referred in Appendix.

Experimental Evaluation (RQ1)

1) Task of Importance Value Estimation. As shown in Table 2, we observe that: **(1)** Compared to HGT, GENI, and MULTI that mainly learn the graph heterogeneity, HIVEN considers the degree centrality that showcases its usefulness for estimating the node importance. **(2)** SKES further outperforms the baseline methods with performance gains over 2.75%, 4.76%, and 1.20% in MAE, RMSE, and NRMSE on three datasets. This demonstrates its effectiveness of structural knowledge exploitation and synergy for accurate estimation of importance values. **(3)** For the target *"Director"* type of TMDB5K, we observe a performance gap against HIVEN with -2.19% of MAE. One explanation is due to the data scarcity issue of *"Director"* for model training, i.e., 1.5% shown in Table 1, as SKES may be under-trained to produce sporadic inaccurate estimations. These *"outliers"* may significantly influence MAE that considers the average absolute difference between the predicted and actual values. On the contrary, RMSE and NRMSE are less sensitive to these outliers, providing a more objective evaluation by considering the magnitude of errors and reflecting the influence of outliers less prominently. **(4)** Additionally, we conduct the Wilcoxon signed-rank tests and the results show that all

Method	MUSIC10K						TMDB5K						DBLP					
	Song			Artist			Movie			Director			Paper			Author		
	M	R	N	M	R	N	M	R	N	M	R	N	M	R	N	M	R	N
PR	0.460	0.489	0.633	0.540	0.563	0.596	2.517	2.771	0.448	0.499	1.005	0.155	2.293	3.244	0.352	1.728	1.894	0.467
PPR	0.460	0.490	0.633	0.540	0.563	0.596	2.512	2.771	0.448	0.499	1.005	0.155	2.293	3.243	0.352	1.173	1.894	0.467
LR	0.137	0.165	0.208	0.126	0.164	0.173	0.672	0.851	0.138	0.541	0.812	0.125	1.061	1.319	0.146	0.628	0.7657	0.204
RF	0.122	0.152	0.213	0.110	0.143	0.151	0.774	0.954	0.154	0.500	0.857	0.132	1.057	1.312	0.145	0.555	0.674	0.180
NN	0.127	0.155	0.200	0.111	0.143	0.151	0.720	0.889	0.144	0.402	0.801	0.124	1.042	1.297	0.141	1.133	1.402	0.153
GAT	0.126	0.156	0.205	0.109	0.140	0.149	0.635	0.808	0.131	0.415	0.753	0.116	0.991	1.240	0.143	0.492	0.606	0.153
HGT	0.129	0.160	0.207	0.118	0.145	0.154	0.581	0.764	0.126	0.352	0.681	0.105	0.996	1.248	0.135	0.514	0.629	0.158
GENI	0.131	0.158	0.200	0.121	0.155	0.158	0.594	0.748	0.121	0.347	0.677	0.104	1.006	1.259	0.145	0.493	0.607	0.154
MULTI	0.147	0.185	0.234	0.147	0.186	0.190	0.982	1.166	0.189	0.479	0.764	0.117	2.061	2.451	0.268	1.316	1.467	0.371
RGTN	0.123	0.155	0.216	0.111	0.143	0.153	0.624	0.798	0.182	0.316	0.557	0.093	0.990	1.240	0.136	0.496	0.613	0.155
HIVEN	0.122	0.152	0.209	0.102	0.132	0.144	0.523	**0.664**	0.108	**0.268**	0.539	0.084	1.024	1.283	0.141	0.507	0.620	0.151
SKES	**0.106**	**0.137**	**0.177**	**0.099**	**0.126**	**0.133**	**0.509**	0.667	**0.106**	0.274	**0.507**	**0.083**	**0.940**	**1.179**	**0.130**	**0.462**	**0.571**	**0.143**
Gain	15.09%	10.94%	12.99%	3.03%	4.76%	8.27%	2.75%	-0.45%	1.89%	-2.19%	6.31%	1.20%	5.32%	5.17%	3.85%	6.49%	6.13%	5.59%

Table 2: Quantitative comparison on the importance value estimation task. Bold and underlined digits are the best and second-best metric values (M, R, and N denote MAE, RMSE, and NRMSE, respectively).

Method	MUSIC10K				TMDB5K				DBLP			
	Song		Artist		Movie		Director		Paper		Author	
	SP	NDCG	SP	NDCG	SP	NDCG	SP	NDCG	SP	NDCG	SP	NDCG
PR	0.013	0.596	0.176	0.743	0.548	0.775	0.182	0.473	-0.104	0.331	0.443	0.916
PPR	-0.020	0.581	0.188	0.732	0.707	0.846	0.195	0.489	0.051	0.333	0.453	0.913
LR	0.226	0.701	-0.037	0.645	0.669	0.858	0.393	0.672	0.312	0.538	0.2445	0.676
RF	0.461	0.797	0.441	0.783	0.590	0.854	0.333	0.484	0.325	0.615	0.396	0.759
NN	0.383	0.774	0.431	0.820	0.657	0.850	0.414	0.613	0.352	0.583	-0.002	0.399
GAT	0.408	0.786	0.481	0.830	0.728	0.867	0.660	0.794	0.401	0.597	0.491	0.922
HGT	0.342	0.753	0.448	0.810	0.758	0.892	0.301	0.463	0.426	0.644	0.458	0.857
GENI	0.402	0.793	0.485	0.784	0.753	0.895	0.678	0.851	0.412	0.602	0.491	0.923
MULTI	0.467	0.808	0.500	0.871	0.728	0.867	0.660	0.704	0.364	0.596	0.452	0.918
RGTN	0.414	0.787	0.486	0.853	0.682	0.901	0.623	0.822	0.438	0.643	0.488	0.907
HIVEN	0.480	0.814	0.544	0.885	0.793	0.910	**0.701**	**0.862**	0.404	0.612	0.459	0.913
SKES	**0.565**	**0.865**	**0.602**	**0.894**	**0.823**	**0.942**	0.680	0.847	**0.483**	**0.674**	**0.589**	**0.925**
Gain	17.77%	6.27%	10.66%	1.02%	3.78%	3.52%	-3.00%	-1.74%	10.27%	4.66%	19.96%	0.22%

Table 3: Quantitative comparison on the importance-based node ranking task (SP denotes SPEARMAN).

SPEARMAN	NDCG	Training Time/epoch (ms)
0.680→ 0.755 (+11.03%)	0.847→0.902 (+6.49%)	745→1,219 (+63.62%)

Table 4: Evaluation with the marginal ranking loss.

the improvements that SKES has achieved are statistically significant with at least 95% confidence level.

2) Task of Important Node Ranking. We present evaluation results for important node ranking in Table 3 with twofold discussions: **(1)** Our model generally presents performance superiority over all baselines with 3.78%∼19.96% and 0.22%∼6.27% of metric improvements, respectively. This is intuitive as our model SKES, performing well for essential importance value estimation task, thus naturally inherits to present a good capability for ranking. **(2)** For the type *"Director"* of TMDB5K, SKES obtains the second-best performance. Since our model incorporates more heterogeneous information, this implies that it may need more training data for better knowledge learning and fusion.

One solution could be to supplement the original regression objective, i.e., MSE loss of Eqn. (16), by adding *triplet ranking objective*, e.g., marginal ranking loss, as follows:

$$\mathcal{L}_{mrl} = \max\left\{0, m+\big(g(v_i)-g(v_+)\big)-\big(g(v_i)-g(v_-)\big)\right\}, \quad (18)$$

where m is the margin. v_+ and v_- are two nodes that the importance value gap between v_+ and v_i is larger than the value gap between v_- and v_i. The general idea of Eqn. (18) is to reduce the disparity between v_i and its positive pair, i.e., v_+, against its negative counterpart v_-. We conduct experiments on this case with results shown in Table 4. The observation indicates that supplementing the ranking regularization of Eqn. (18) will boost the performance as expected. With the performance increasing over 6.49%, our model also surpasses HIVEN shown in Table 3. On the other hand, the training time cost inevitably increases, which is a practical trade-off to consider. Hence, in this paper, we mainly report the results based on our original objective function and leave the further design of learning frameworks as future work.

Study of Knowledge Synergy and Measurement (RQ2)

1) Synergy of Structural Knowledge. To validate the contribution of structural knowledge, we randomly disable the knowledge proportion, i.e., $\mathbf{c}_{i,k}^{(l)}$, from 0% (intact) to -80%.

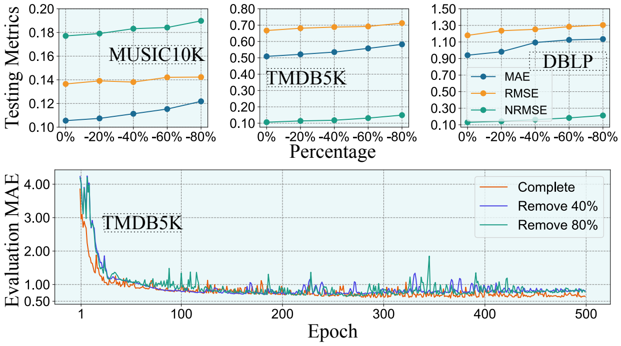

Figure 3: (1) Decreasing contribution of structural knowledge; (2) curves of evaluation MAE (best view in color).

Variant	Movie			Director		
	MSE	RMSE	NRMSE	MSE	RMSE	NRMSE
`w/o WD`	0.540	0.692	0.120	0.284	0.523	0.092
`w/o` λ	0.524	0.675	0.116	0.279	0.519	0.092
SKES	**0.509**	**0.667**	**0.106**	**0.274**	**0.507**	**0.083**

Table 5: Study of node importance estimation.

Our observations from Figure 3 are twofold: (1) From the upper-row figures, we notice remarkable performance degradations across three datasets, where DBLP presents a more pronounced performance perturbation on MAE and RMSE curves. This demonstrates the efficacy of our knowledge synergy mechanism in identifying node importance, especially in larger HINs with diverse and complex structural information. (2) We plot MAE curves for cases of *keeping intact*, *removing 40%*, and *removing 80%* prior knowledge in the first 500 training epochs on "*Movie*" of TMDB5K. As shown in the lower-row figure, we observe that SKES produces much less bursting perturbations than the curves of *removing 40%* and *removing 80%*. This implies that our implementation to adaptively fuse heterogeneous knowledge can also help to stabilize the model performance.

2) Mechanism of Node Importance Estimation. To evaluate the effectiveness of our proposed importance estimation method, we design two variants on TMDB5K. (1) Firstly, we replace our original design with 1-Wasserstein distance by simply utilizing a two-layer of MLP for node importance estimation. We denote the variant as `w/o WD`. (2) Secondly, we retain the measurement of 1-Wasserstein distance but remove $\boldsymbol{\lambda}$ in Eqn. (15) for regression, denoted as `w/o` λ. Results in Table 5 not only justify that integrating Optimal Transport theory for importance estimation achieves better performance (comparing `w/o WD` to `w/o` λ), but also prove the simplicity yet effectiveness of non-linear value regression with learnable $\boldsymbol{\lambda}$ (comparing SKES to `w/o` λ). We further visualize the absolute value gap between the ground truth and estimated values of "*Movie*" nodes. To provide a readable visualization in Figure 4, the curves are based on averaged values with the rolling window as 10 nodes. As the curve gets closer to the bottom, the value estimation will be more accurate. The curves in Figure 4 indicate that SKES consistently performs better than those two variants.

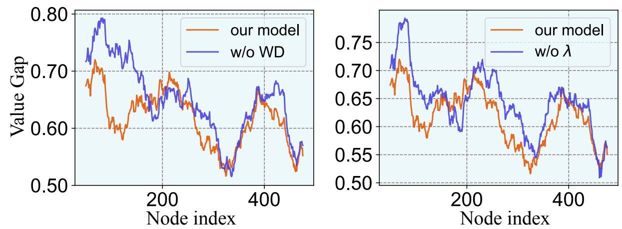

Figure 4: Absolute importance value gap (best view in color)

Variant	Movie			Director		
	MSE	RMSE	NRMSE	MSE	RMSE	NRMSE
`w/o NH`	0.551	0.722	0.113	0.320	0.573	0.098
`w/o ATT`	0.544	0.696	0.118	0.319	0.549	0.096
SKES	**0.509**	**0.667**	**0.106**	**0.274**	**0.507**	**0.083**

Table 6: Ablation study of other module designs.

Ablation Study (RQ3)

1) Network Heterogeneity Learning. We adopt metapath-based methodology to learn heterogeneity. To validate its usefulness, we propose a variant `w/o NH` via directly treating input networks as homogeneous. From Table 6, we observe over 8.33% (MSE of "*Movie*" type) performance decay on TMDB5K. This demonstrates the necessity of explicitly distinguishing the heterogeneity of node and edge types for the node importance estimation task.

2) Self-attention in Synergetic Representation Learning. We create a variant `w/o ATT` by disabling all designs in Eqn's. (6-9) and simply using $\mathbf{e}_i$ to replace $\mathbf{h}_i^R$ for importance estimation. The empirical results between `w/o ATT` and SKES in Table 6 clearly demonstrate that the self-attention mechanism also work for our model to attentively adjust the contributions for different sources of structural prior knowledge in model learning.

Conclusion and Future Work

We propose a novel framework SKES for estimating HIN node importance. SKES effectively leverages structural knowledge to harness information synergy, providing a robust measurement of node importance. Our empirical model evaluation on three public benchmarks demonstrates the performance superiority of SKES against competing baselines. As for future work, we identify two promising directions: (1) It is worth investigating other Learning paradigm (Zhang et al. 2022b, 2023c,a; Song, Zhang, and King 2023b,a; He et al. 2023c) to further improve the quality of learned node embeddings from heterogeneous information. (2) We also plan to incorporate *language/vision* modeling (Qiu et al. 2022, 2021; Chen et al. 2022d; Li et al. 2022; Sun et al. 2022; Li et al. 2020; Hu et al. 2021a, 2023; Zhu et al. 2023) as practical HINs may contain multi-modal information.

Acknowledgments

The work described here was partially supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (CUHK 14222922, RGC GRF No. 2151185) and (RGC Research Impact Fund R5034-18; CUHK 2410021). Yixiang Fang was supported in part by NSFC (Grant 62102341), Guangdong Talent Program (Grant 2021QN02X826), Shenzhen Science and Technology Program (Grants JCYJ20220530143602006 and ZDSYS 20211021111415025), and Shenzhen Science and Technology Program and Guangdong Key Lab of Mathematical Foundations for Artificial Intelligence.

References

Chen, Y.; Fang, Y.; Zhang, Y.; and King, I. 2023a. Bipartite Graph Convolutional Hashing for Effective and Efficient Top-N Search in Hamming Space. In *WWW*, 3164–3172.

Chen, Y.; Guo, H.; Zhang, Y.; Ma, C.; Tang, R.; Li, J.; and King, I. 2022a. Learning binarized graph representations with multi-faceted quantization reinforcement for top-k recommendation. In *SIGKDD*, 168–178.

Chen, Y.; Truong, T.; Shen, X.; Wang, M.; Li, J.; Chan, J.; and King, I. 2023b. Topological representation learning for e-commerce shopping behaviors. In *MLG-KDD*.

Chen, Y.; Wu, Y.; Ma, S.; and King, I. 2020. A Literature Review of Recent Graph Embedding Techniques for Biomedical Data. In *ICONIP*, 21–29.

Chen, Y.; Yang, M.; Zhang, Y.; Zhao, M.; Meng, Z.; Hao, J.; and King, I. 2022b. Modeling scale-free graphs with hyperbolic geometry for knowledge-aware recommendation. In *WSDM*, 94–102.

Chen, Y.; Yang, Y.; Wang, Y.; Bai, J.; Song, X.; and King, I. 2022c. Attentive knowledge-aware graph convolutional networks with collaborative guidance for personalized recommendation. In *ICDE*, 299–311. IEEE.

Chen, Y.; Zhang, J.; Fang, Y.; Cao, X.; and King, I. 2021. Efficient community search over large directed graphs: An augmented index-based approach. In *IJCAI*, 3544–3550.

Chen, Y.; Zhang, Y.; Guo, H.; Tang, R.; and King, I. 2022d. An Effective Post-training Embedding Binarization Approach for Fast Online Top-K Passage Matching. In *AACL*, 102–108.

Chen, Y.; Zhang, Y.; Yang, M.; Song, Z.; Ma, C.; and King, I. 2023c. WSFE: Wasserstein Sub-graph Feature Encoder for Effective User Segmentation in Collaborative Filtering. In *SIGIR*, 2521–2525.

Dorogovtsev, S. N.; Goltsev, A. V.; and Mendes, J. F. F. 2006. K-core organization of complex networks. *Physical review letters*, 96(4): 040601.

Egghe, L.; et al. 2006. An improvement of the h-index: The g-index. *ISSI newsletter*, 2(1): 8–9.

Fang, Y.; Cheng, R.; Chen, Y.; Luo, S.; and Hu, J. 2017. Effective and efficient attributed community search. *The VLDB Journal*, 26: 803–828.

Fu, X.; and King, I. 2023. FedHGN: A Federated Framework for Heterogeneous Graph Neural Networks. In *IJCAI*, 3705–3713.

Fu, X.; and King, I. 2024. MECCH: Metapath Context Convolution-based Heterogeneous Graph Neural Networks. *Neural Networks*, 170: 266–275.

Fu, X.; Zhang, J.; Meng, Z.; and King, I. 2020. Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding. In *WWW*, 2331–2341.

Grover, A.; and Leskovec, J. 2016. node2vec: Scalable feature learning for networks. In *SIGKDD*, 855–864.

Haveliwala, T. H. 2003. Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. *TKDE*, 15(4): 784–796.

He, B.; He, X.; Zhang, R.; Zhang, Y.; Tang, R.; and Ma, C. 2023a. Dynamic Embedding Size Search with Minimum Regret for Streaming Recommender System. In *CIKM*, 741–750.

He, B.; He, X.; Zhang, Y.; Tang, R.; and Ma, C. 2023b. Dynamically Expandable Graph Convolution for Streaming Recommendation. In *WWW*, 1457–1467.

He, B.; Sun, Z.; Liu, J.; Zhang, S.; Chen, X.; and Ma, C. 2023c. Offline imitation learning with variational counterfactual reasoning. *arXiv preprint arXiv:2310.04706*.

Hirsch, J. E. 2005. An index to quantify an individual's scientific research output. *PNAS*, 102(46): 16569–16572.

Hu, X.; Chen, J.; Li, X.; Guo, Y.; Wen, L.; Yu, P. S.; and Guo, Z. 2023. Do Large Language Models Know about Facts? *arXiv preprint arXiv:2310.05177*.

Hu, X.; Guo, Z.; Wu, G.; Liu, A.; Wen, L.; and Yu, P. S. 2022. CHEF: A Pilot Chinese Dataset for Evidence-Based Fact-Checking. In *NAACL-HLT*, 3362–3376.

Hu, X.; Wen, L.; Xu, Y.; Zhang, C.; and Yu, P. S. 2020a. SelfORE: Self-supervised Relational Feature Learning for Open Relation Extraction. In *EMNLP*, 3673–3682.

Hu, X.; Zhang, C.; Ma, F.; Liu, C.; Wen, L.; and Yu, P. S. 2021a. Semi-supervised Relation Extraction via Incremental Meta Self-Training. In *EMNLP*, 487–496.

Hu, X.; Zhang, C.; Yang, Y.; Li, X.; Lin, L.; Wen, L.; and Yu, P. S. 2021b. Gradient Imitation Reinforcement Learning for Low Resource Relation Extraction. In *EMNLP*, 2737–2746.

Hu, Z.; Dong, Y.; Wang, K.; and Sun, Y. 2020b. Heterogeneous graph transformer. In *WWW*, 2704–2710.

Huang, C.; Fang, Y.; Lin, X.; Cao, X.; Zhang, W.; and Orlowska, M. 2022. Estimating Node Importance Values in Heterogeneous Information Networks. In *ICDE*, 846–858.

Huang, H.; Sun, L.; Du, B.; Liu, C.; Lv, W.; and Xiong, H. 2021. Representation Learning on Knowledge Graphs for Node Importance Estimation. In *SIGKDD*, 646–655.

Korotin, A.; Selikhanovych, D.; and Burnaev, E. 2023. Neural Optimal Transport. In *ICLR*. OpenReview.net.

Li, J.; Li, Z.; Ge, T.; King, I.; and Lyu, M. R. 2022. Text Revision by On-the-Fly Representation Optimization. In *AAAI*, 10956–10964.

Li, J.; Li, Z.; Mou, L.; Jiang, X.; Lyu, M.; and King, I. 2020. Unsupervised text generation by learning from search. In *NeurIPS*, volume 33, 10820–10831.

Liang, L.; Hu, X.; Xu, Z.; Song, Z.; and King, I. 2023. Predicting Global Label Relationship Matrix for Graph Neural Networks under Heterophily. In *NeurIPS*.

Marchiori, M.; and Latora, V. 2000. Harmony in the small-world. *Physica A: Statistical Mechanics and its Applications*, 285(3-4): 539–546.

Naderializadeh, N.; Comer, J. F.; Andrews, R.; Hoffmann, H.; and Kolouri, S. 2021. Pooling by sliced-Wasserstein embedding. volume 34, 3389–3400.

Negre, C. F.; Morzan, U. N.; Hendrickson, H. P.; Pal, R.; Lisi, G. P.; Loria, J. P.; Rivalta, I.; Ho, J.; and Batista, V. S. 2018. Eigenvector centrality for characterization of protein allosteric pathways. *PNAS*, 115(52): E12201–E12208.

Nieminen, J. 1974. On the centrality in a graph. *Scandinavian journal of psychology*, 15(1): 332–336.

Nietert, S.; Goldfeld, Z.; Sadhu, R.; and Kato, K. 2022. Statistical, robustness, and computational guarantees for sliced wasserstein distances. *NeurIPS*, 35: 28179–28193.

Page, L.; Brin, S.; Motwani, R.; and Winograd, T. 1999. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

Park, N.; Kan, A.; Dong, X. L.; Zhao, T.; and Faloutsos, C. 2019. Estimating node importance in knowledge graphs using graph neural networks. In *SIGKDD*, 596–606.

Park, N.; Kan, A.; Dong, X. L.; Zhao, T.; and Faloutsos, C. 2020. Multiimport: Inferring node importance in a knowledge graph from multiple input signals. In *SIGKDD*, 503–512.

Qiu, Z.; Su, Q.; Ou, Z.; Yu, J.; and Chen, C. 2021. Unsupervised Hashing with Contrastive Information Bottleneck. In *IJCAI*, 959–965.

Qiu, Z.; Su, Q.; Yu, J.; and Si, S. 2022. Efficient Document Retrieval by End-to-End Refining and Quantizing BERT Embedding with Contrastive Product Quantization. In *EMNLP*, 853–863.

Reimers, N.; and Gurevych, I. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Sabidussi, G. 1966. The centrality index of a graph. *Psychometrika*, 31(4): 581–603.

Shaw, M. E. 1954. Group structure and the behavior of individuals in small groups. *The Journal of psychology*, 38(1): 139–149.

Song, Z.; Zhang, Y.; and King, I. 2023a. No Change, No Gain: Empowering Graph Neural Networks with Expected Model Change Maximization for Active Learning. In *NeurIPS*.

Song, Z.; Zhang, Y.; and King, I. 2023b. Optimal Block-wise Asymmetric Graph Construction for Graph-based Semi-supervised Learning. In *NeurIPS*.

Song, Z.; Zhang, Y.; and King, I. 2023c. Towards Fair Financial Services for All: A Temporal GNN Approach for Individual Fairness on Transaction Networks. In *CIKM*, 2331–2341. ACM.

Sun, X.; Ge, T.; Ma, S.; Li, J.; Wei, F.; and Wang, H. 2022. A Unified Strategy for Multilingual Grammatical Error Correction with Pre-trained Cross-Lingual Language Model. In Raedt, L. D., ed., *IJCAI*, 4367–4374. Main Track.

Sun, Y.; Han, J.; Yan, X.; Yu, P. S.; and Wu, T. 2011. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *VLDB*, 4(11): 992–1003.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *NeurIPS*, 30.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.

Villani, C.; et al. 2009. *Optimal transport: old and new*, volume 338. Springer.

Yang, M.; Li, Z.; Zhou, M.; Liu, J.; and King, I. 2022. Hicf: Hyperbolic informative collaborative filtering. In *SIGKDD*, 2212–2221.

Yang, M.; Zhou, M.; Pan, L.; and King, I. 2023a. κHGCN: Tree-likeness Modeling via Continuous and Discrete Curvature Learning. In *SIGKDD*, 2965–2977.

Yang, M.; Zhou, M.; Ying, R.; Chen, Y.; and King, I. 2023b. Hyperbolic Representation Learning: Revisiting and Advancing. *ICML*.

Zhang, X.; Chen, Y.; Gao, C.; Liao, Q.; Zhao, S.; and King, I. 2022a. Knowledge-aware Neural Networks with Personalized Feature Referencing for Cold-start Recommendation. *arXiv preprint arXiv:2209.13973*.

Zhang, Y.; Chen, Y.; Song, Z.; and King, I. 2023a. Contrastive cross-scale graph knowledge synergy. In *SIGKDD*, 3422–3433.

Zhang, Y.; and Zhu, H. 2019. Doc2hash: Learning discrete latent variables for documents retrieval. In *NAACL*.

Zhang, Y.; Zhu, H.; Chen, Y.; Song, Z.; Koniusz, P.; King, I.; et al. 2023b. Mitigating the Popularity Bias of Graph Collaborative Filtering: A Dimensional Collapse Perspective. In *NeurIPS*.

Zhang, Y.; Zhu, H.; Song, Z.; Koniusz, P.; and King, I. 2022b. COSTA: covariance-preserving feature augmentation for graph contrastive learning. In *SIGKDD*.

Zhang, Y.; Zhu, H.; Song, Z.; Koniusz, P.; and King, I. 2023c. Spectral feature augmentation for graph contrastive learning and beyond. In *AAAI*.

Zheng, Y.; Zhang, X.; Chen, S.; Zhang, X.; Yang, X.; and Wang, D. 2021. When Convolutional Network Meets Temporal Heterogeneous Graphs: An Effective Community Detection Method. *IEEE TKDE*.

Zhu, X.; Zhang, R.; He, B.; Guo, Z.; Zeng, Z.; Qin, Z.; Zhang, S.; and Gao, P. 2023. Pointclip v2: Prompting clip and gpt for powerful 3d open-world learning. In *ICCV*, 2639–2650.