Heterogeneous Graph Learning for Explainable Recommendation over Academic Networks

Xiangtai Chen
School of Software
Dalian University of Technology
Dalian, China
chenxiangtai@outlook.com

Tao Tang
School of Engineering, IT and Physical Sciences
Federation University Australia
Ballarat, Australia
tau.tang@outlook.com

Jing Ren
School of Engineering, IT and Physical Sciences
Federation University Australia
Ballarat, Australia
ch.yum@outlook.com

Ivan Lee✉
STEM
University of South Australia
Adelaide, Australia
ivan.lee@unisa.edu.au

Honglong Chen
College of Control Science and Engineering
China University of Petroleum
Qingdao, China
chenhl@upc.edu.cn

Feng Xia
School of Engineering, IT and Physical Sciences
Federation University Australia
Ballarat, Australia
f.xia@ieee.org

ABSTRACT

With the explosive growth of new graduates with research degrees every year, unprecedented challenges arise for early-career researchers to find a job at a suitable institution. This study aims to understand the behavior of academic job transition and hence recommend suitable institutions for PhD graduates. Specifically, we design a deep learning model to predict the career move of early-career researchers and provide suggestions. The design is built on top of scholarly/academic networks, which contains abundant information about scientific collaboration among scholars and institutions. We construct a heterogeneous scholarly network to facilitate the exploring of the behavior of career moves and the recommendation of institutions for scholars. We devise an unsupervised learning model called HAI (Heterogeneous graph Attention InfoMax) which aggregates attention mechanism and mutual information for institution recommendation. Moreover, we propose scholar attention and meta-path attention to discover the hidden relationships between several meta-paths. With these mechanisms, HAI provides ordered recommendations with explainability. We evaluate HAI upon a real-world dataset against baseline methods. Experimental results verify the effectiveness and efficiency of our approach.

CCS CONCEPTS

• **Information systems → Information retrieval**; • **Computing methodologies → Knowledge representation and reasoning**; • **Applied computing →** *Education.*

ACM ISBN 978-1-4503-9187-0/21/12…$15.00
https://doi.org/10.1145/3498851.3498926

KEYWORDS

academic social networks, recommender systems, explainability, graph learning, heterogeneous networks

ACM Reference Format:
Xiangtai Chen, Tao Tang, Jing Ren, Ivan Lee✉, Honglong Chen, and Feng Xia. 2021. Heterogeneous Graph Learning for Explainable Recommendation over Academic Networks. In *IEEE/WIC/ACM International Conference on Web Intelligence (WI-IAT '21 Companion), December 14–17, 2021, ESSENDON, VIC, Australia.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3498851.3498926

1 INTRODUCTION

Recent years have witnessed the rapid growth of academic information (i.e., big scholarly data) due to a large number of research works carried out by academia and industry [34]. Meanwhile, academic social networks are continuously expanded. It is well recognized that different types of academic social relationship have intrinsically-different effect among researchers, which form a complex force that influence the dynamics of academic social networks. From the UNESCO Science Report[1], the number of full-time equivalent researchers grew by 21% from 2007 to 2013, accounting for 0.1% of the global population. Nowadays, the number of doctoral graduates is expanding every year, which results in input inflation of academic researchers. It is worse for early-career researchers that industry might be unable to absorb such numerous researchers, while academic institutions are generally adopting the Tenure-Track on recruiting researchers. With such fierce competition, doctoral graduates are facing a dilemma when choosing institutions.

However, a survey about PhD degrees in nature reveals that despite facing a love–hurt relationship, doctoral students are as committed as ever on pursuing research careers [31]. The first choice of which academic institution to join has a great impact on the future academic career of doctoral graduates, further choosing a suitable institution could contribute to future academic success. But how can the doctoral students find a research position? The

[1] https://en.unesco.org/node/252277

Nature survey [1] indicates that doctoral students are largely finding their career advice online. Just one-third credited advice from a supervisor as a reason for their career choice. With the rapid growth of social media, researchers can use such platforms liked Twitter to expand their social contacts and find jobs [1]. Although some platforms such as LinkedIn aim to build a career network for job hunting, early-career researchers would like to find an academic position through friendship relationships and collaboration relationships on Twitter.

Utilizing social media to find an academic job is time-consuming and heavily relies on manually posting, which greatly limits its widespread usage. An ideal solution is to design a method to automatically discover the hidden collaboration information, colleague information, academic research direction, etc. from academic social networks (ASNs). Nevertheless, accurately distinguishing social relationships is difficult, especially in a real-world network. In the real data, the vast majority of the patterns of career moves are hidden in the collaboration information, which is not currently revealed in most social networks.

Currently, the problem of finding an optimal academic institution for early-career researchers is not well addressed, and existing social networks are not sufficient to address the issues about finding academic institutions on a large scale.

In this work, we aim to tackle the above problem. First, we construct an undirected heterogeneous scholarly network (HSN) with various types of nodes and edges, where nodes include scholars, institutions, papers, while edges include "works-with" and "writes". Furthermore, we design a model called HAI which works on automatically discovering collaboration information and colleague information for institution recommendation for junior scholars. We adopt attention mechanism on the meta-path neighbors connected by two meta-paths: Author-Paper-Author (*APA*) and Author-Institution-Author (*AIA*). The scholar mutual information is adopted to maximize the similarity of local features and global features. Generally, there has no label to indicate which academic institution the scholar will join in the future for a recommendation system. To address the problem of lack of labels in the real-world data, we maximize the scholar mutual information as the objective function and apply unsupervised learning to learn the information of scholars. Moreover, the experimental results show that our model performs well in terms of the AUC metric and HR metric.

The primary contributions of this work are as follows:

- We devise a novel unsupervised learning algorithm to learn low-dimension features of scholars by taking advantage of the attention mechanism and mutual information of scholars.
- We proposed an explainable recommendation model (i.e., HAI) which can recommend suitable academic institutions for early-career researchers while providing explanations.
- Extensive experiments have been conducted upon a real-world dataset. The results verify the superiority of HAI against state-of-the-art baseline methods.

The rest of this paper is organized as follows: Section 2 gives a discussion of the related works. Section 3 presents the primary preliminaries of heterogeneous graph learning and defines the problem we are dealing with. Section 4 proposes the method including scholar attention, meta-path attention, and scholar mutual information to recommend institutions for junior scholars. Section 5 shows the data analysis, experimental results, and case studies. In the end, Section 6 concludes this study and discusses the future directions.

2 RELATED WORK

Meta-path, as an important characteristic of a heterogeneous network, is regarded as a useful tool for heterogeneous network embedding [33]. Many researchers design algorithms based on meta-path and heterogeneous neighborhoods generated by random walks [2, 4, 32]. By exploring meta-paths, node-level attentions are present to learn heterogeneous relations [11, 27]. Random walks on heterogeneous structures walk slowly on meta-path, while attention on HAI runs much faster than random walks. Furthermore, neural networks are proposed to embed heterogeneous networks. Generative or discriminative adversarial networks based framework [8, 12] works on complex neural networks to learn the node distribution. Qu et al. [12] study learning curricula for node embedding in heterogeneous star networks and propose an approach based on deep reinforcement learning for this problem. Zhang et al. [38] propose a heterogeneous graph neural network model which uses two modules to aggregate feature information of sampled neighboring nodes. Wang et al. [26] represent scholars to vector for lifetime collaborator prediction.

Network embedding has shown great power in the analysis of homogeneous networks [7]. HAI is different from the existing studies in heterogeneous network embedding. The previous works mainly focus on meta-path embedding or employ attention mechanism on labeled data for supervised learning, while HAI aggregates attention mechanism and mutual information on HSN for unsupervised learning.

In the early age of recommendation systems, Collaborative Filtering [16] and Matrix Factorization are two of the most widely used algorithms in the industry. Wan et al. [25] recommend citations with network representation learning by Collaborative Filtering and [23] utilize deep Matrix Factorization for Trust-Aware recommendation in social networks. Zhao et al. [39] introduce Matrix Factorization and Factorization Machine to learn similarities generated by each meta-path for the recommendation.

Nowadays, various kinds of embedding information are integrated for item recommendation [18]. Meta-path contexts [9] are leveraged for TOP K recommendation. Wang et al. [29] assume there exist some common characteristics under different meta-paths for each user or item and propose a unified embedding model. Zhao et al. [40] construct a heterogeneous co-occurrence network in a recommendation-oriented heterogeneous network. Ying et al. [35] develop a data-efficient graph convolutional network for web-scale recommender systems.

More recently, embedding information is applied to mine the relationships hidden in ASNs. In [24], academic collaboration networks are built for academic relationship mining. [13] and [30] design metrics to evaluate the academic potentials of scholars and academia units respectively. Guo et al. [5] make predictions on graduate employment with bias, but not providing suggestions for graduate employment. In [28], attractive communities are detected in ASNs and Yu et. al [36] optimize academic teams when the outlier member is leaving. However, none of the previous works focus on helping scholars to deal with the career problem. For tackling the problem, we design a novel model HAI for academic institution recommendation.

3 PRELIMINARIES

In this section, we introduce the input to our framework. The input HSN is a kind of Heterogeneous Information Network (HIN) with various types of nodes and various types of edges and has much more complicated structural information than homogeneous networks.

3.1 Heterogeneous Graph

Definition 3.1 (Heterogeneous Graph). The input graph can be denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A}, \mathcal{X})$, where $\mathcal{V}$ denotes a set of nodes with different types, $\mathcal{E}$ denotes a set of edges with different types, $\mathcal{A}$ denotes the set of adjacency matrices for each type of edge, and $\mathcal{X}$ denotes the set of matrices of nodes features for each type of node.

Each node $v \in \mathcal{V}$ and each edge $e \in \mathcal{E}$ are associated with type mapping functions $\tau(v) : \mathcal{V} \rightarrow \mathcal{T}^v$ and $\phi(e) : \mathcal{E} \rightarrow \mathcal{T}^e$, where $\mathcal{T}^v$ and $\mathcal{T}^e$ represent the set of node types and edge types respectively. The graph becomes homogeneous when $|\mathcal{T}^v| = |\mathcal{T}^e| = 1$, in contrast the graph is considered as heterogeneous when $|\mathcal{T}^v| + |\mathcal{T}^e| > 1$. The whole nodes' information in the graph is stored in a set of node feature matrices $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n\}$, where $n = |\mathcal{T}^v|$; $\mathbf{X}_n \in \mathbb{R}^{N_n \times D_n}$ is the feature matrix of all nodes with type $\mathcal{T}_n^v$; N_n is the number of nodes in type $\mathcal{T}_n^v$; and D_n is the dimension of each node's embedding. The edge information is present as a set of adjacency matrices $\mathcal{A} = \{\mathbf{A}_1, \mathbf{A}_2, \ldots \mathbf{A}_m\}$, where $m = |\mathcal{T}^e|$.

3.2 Meta-Path

A remarkable difference between heterogeneous network and homogeneous network is the connection information among various types of nodes. The edge $e = (s, d)$ connected from source node s to destination node d of a different type is denoted as $\langle \tau(s), \phi(e), \tau(d) \rangle$, and the edge is considered as meta-relation.

Definition 3.2 (Meta-Path). Meta-path denoted as ρ is a sequence of those meta-relations, which can be defined in the form of sequence $v_1 \xrightarrow{\phi(e_1)} v_2 \xrightarrow{\phi(e_2)} \ldots \xrightarrow{\phi(e_l)} v_{l+1}$ connected by l meta-relations.

3.3 Meta-Path Neighbors

Definition 3.3 (Meta-Path Neighbors). For a node v and meta-path ρ in heterogeneous network, the nodes which linked with node v through meta-path ρ are defined as meta-path neighbors denoted by $\mathcal{N}_v^\rho$.

The node v itself is also included in the set of meta-path neighbors because of attention operations.

Example 3.4. As shown in Figure 1, the meta-path neighbors of a_1 through meta-path Author-Paper-Author include a_3, a_4, and a_1 itself. Those meta-path neighbors on meta-path Author-Paper-Author are authors that a_1 has collaborated with. Similarly, the meta-path neighbors of a_1 through meta-path Author-Institution-Author contain the authors that work with the same institution of a_1, which are a_2, a_1 from Figure 1.

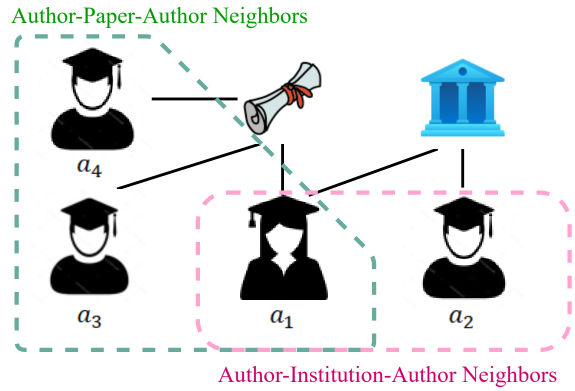


Figure 1: An instance of meta-path neighbors. The scholars in the area bounded by the green dotted line are Author-Paper-Author neighbors of a_1. The scholars in the area bounded by the red dotted line are Author-Institution-Author neighbors of a_1.

3.4 Problem Formulation

The proposed model HAI is denoted as function μ. Each scholar has an ordered preference list on institutions in the form of $Pre(a) = (i_1, i_2, \ldots i_s)$, where $s = |\mathcal{V}^i|$. For a specific scholar $a_i \in \mathcal{V}^a$, we generate a preference list $Pre(a_i) = (i_1, i_2, i_3, i_4 \ldots)$ for a_i in $\mathcal{V}^a$. Assuming the institution i_3 is where the scholar a_i works with, the preference list $Pre(a_i)$ indicates that a_i may prefer i_1 to i_2 and prefers staying in i_3 instead of moving to i_4. The problem we addressed is to recommend institutions $\mathcal{I}_i$ for the scholar a_i from the preference list $Pre(a_i)$.

The institution recommendation problem in this research is defined as follows:

Problem 1. *Given a preference list $Pre(a) = \mu(a) \in \mathcal{V}^i$ for scholar a, the target is to recommend institutions $\mathcal{I} \in Pre(a)$ for any $a \in \mathcal{V}^a$.*

4 DESIGN OF HAI

In this section, we propose a novel unsupervised heterogeneous graph learning model HAI for institution recommendation. From the inspiration of HAN [27], the model we proposed utilizes scholar attention to learn the attention score of meta-path neighbors and meta-path attention to learn the node embeddings from the meta-paths. Despite HAN has good performance on node classification, it could not be used for the institution recommendation because HAN is a semi-supervised learning algorithm that needs labels. Hence, we import mutual information mechanism to fill this gap.

4.1 Overall Framework

Figure 2 illustrates the overall framework of the model HAI. The black lines in the figure are second-order links connected by meta-paths. Two kinds of meta-path play a significant role in calculating scholar attentions. Figure 2(a) shows the process of scholar attention on a specific scholar based on two meta-path neighbors $\mathcal{N}^{\rho_P}$ and $\mathcal{N}^{\rho_I}$. Figure 2(b) shows the process of the meta-path attention which aggregates different kinds of scholar attentions. More about scholar attention, meta-path attention ,and mutual information are detailed in the following sections.

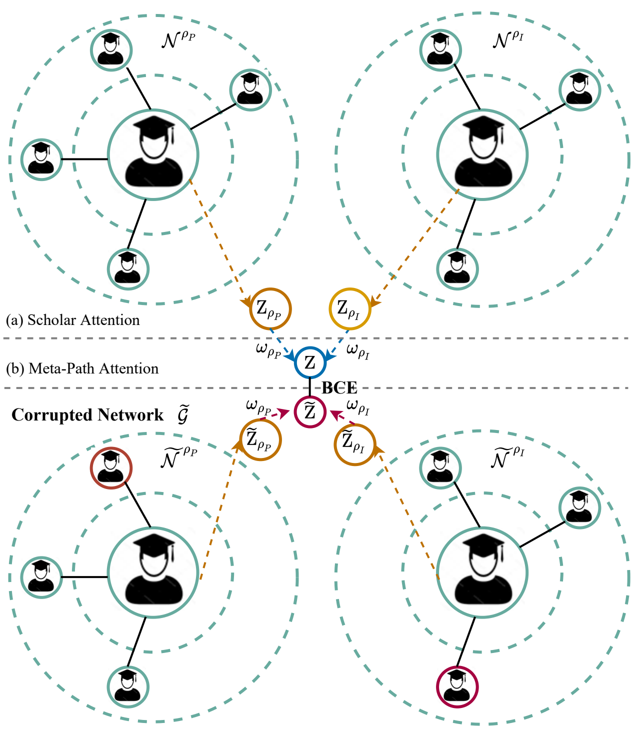


Figure 2: The overall framework of HAI.

4.2 Scholar Attention

Supervisors and collaborators of early-career researchers may influence their choice on which target institutions to apply for. Hence, we present the scholar attention to learn the influence by calculating the attention score through meta-path neighbors and represent the embeddings for each scholar in the HSN.

We leverage Doc2vec [10] to represent the abstracts of each scholar into feature spaces. The projection process can be formulated as follows:

$$H_i = \sum_{k=1}^{N_i} Doc2vec(ab_k)/N_i, \tag{1}$$

where H_i denotes the represented feature of scholar i, N_i and ab_k is the number of papers and the kth abstract of the paper that scholar i had published, respectively.

Next, we leverage the masked self-attention mechanism [20, 22] to learn the information between collaborators and colleagues. Given a scholar pair (i, j) that are connected with meta-path ρ, the attention score $\alpha_{i,j}$ can learn how relevant is the scholar j to the scholar i on the network. Notice that $\alpha_{i,j} \neq \alpha_{j,i}$ because they have different orders when calculating attention scores with masked attention. The attention score of scholar pair (i, j) on meta-path ρ can be formulated as follows:

$$e_{i,j}^{\rho} = att_{scholar}(\mathbf{W}H_i, \mathbf{W}H_j, \rho), \tag{2}$$

where $att_{scholar}$ denotes the shared attentional mechanism on the deep neural network which performs the scholar attention through meta-path ρ, and the weight matrix $\mathbf{W} \in \mathbb{R}^{D\times D}$ is a shared linear transformation which is applied to every node on meta-path ρ.

We only calculate the attention score on the meta-path neighbors, so there comes the masked attention. The masked attention will mask other nodes and only operate on each scholar $j \in \mathcal{N}_i^{\rho}$, where $\mathcal{N}_i^{\rho}$ denotes the meta-path neighbors of scholar i. In this study, j will be the second-order neighbor of i on meta-path *APA* or *AIA*. To make weight coefficients more clear to compare different scholars, we normalize them by using an activate function softmax:

$$\alpha_{i,j}^{\rho} = softmax_j(e_{i,j}^{\rho}), \tag{3}$$

which can be expanded as:

$$\alpha_{i,j}^{\rho} = \frac{\exp(\sigma(\mathrm{a}^{\mathrm{T}}[\mathbf{W}H_i \| \mathbf{W}H_j]))}{\sum_{k\in\mathcal{N}_i^{\rho}} \exp(\sigma(\mathrm{a}^{\mathrm{T}}[\mathbf{W}H_i \| \mathbf{W}H_j]))}, \tag{4}$$

where $\alpha_{i,j}^{\rho}$ denotes the scholar attention weight coefficient, σ is the activate function such as LeakyReLU, $\cdot^{\mathrm{T}}$ and $\|$ is the transpose operation and the concatenation operation on matrix. The Figure 2 (a) shows the process of scholar attention.

Corresponding the scholar attention weight coefficients, we can obtain the aggregated scholar features of a_i by the operation formulated as follows:

$$z_i^{\rho} = \sigma\left(\sum_{j\in\mathcal{N}_i^{\rho}} \alpha_{i,j}^{\rho} \mathbf{W}H_j\right), \tag{5}$$

where z_i^{ρ} is the embedding of scholar a_i on meta-path ρ. To learn more stable embeddings on the heterogeneous network, we have found expanding Equation 5 to multi-head will benefit attention process a lot. Embedding features of each head that are concatenated results in the final scholar attention feature representation formulated as follows:

$$z_i^{\rho} = \Big\Vert_{k=1}^{K} \sigma\left(\sum_{j\in\mathcal{N}_i^{\rho}} \alpha_{i,j}^{\rho,k} \mathbf{W}^k H_j\right). \tag{6}$$

Here, $\alpha_{i,j}^{\rho,k}$ represents the kth-head scholar attention weight coefficient on meta-path ρ, and $\mathbf{W}^k$ is the kth weight matrix. We denote $\mathbf{Z}_{\rho_i}$ as the feature representation on the ith meta-path.

4.3 Meta-Path Attention

For a specific scholar, we aggregate scholar attention weight coefficients from different meta-paths to one representation space. The meta-path attention will learn the weights of different types of meta-paths, which does not have the same influence on scholar representations. In the constructed heterogeneous network, we denote the meta-path *APA* and *AIA* as ρ_P and ρ_I, respectively. Every scholar in the heterogeneous network has two types of semantic information, which significantly differs from that of the homogeneous network. Making full use of the heterogeneity of the network, the meta-path attention automatically learns the weights of different meta-path and aggregates the weights to representation space for the institution recommendation. To learn the different influence of meta-path *APA* and *AIA* on each scholar, we deploy the final nonlinear transformation layer to measure the influence which can be formulated as follows:

$$\omega_{\rho_i} = \sigma\left(\frac{1}{|\mathcal{V}^a|}\sum_{j\in\mathcal{V}^a} \mathrm{q}^{\mathrm{T}} \tanh(\mathbf{W}z_j^{\rho_i} + \mathrm{b})\right), \tag{7}$$

where b is the bias vector, q is a scholar attention vector, and $z_j^{\rho_i}$ represents the feature representation of scholar j on the ith meta-path ρ_i. The higher ω_{ρ_i} is, the more important meta-path ρ_i is for scholars on the institution recommendation task. We finally get the embeddings of each scholar by aggregate ω_{ρ_i} and scholar attention feature $\mathbf{Z}_{\rho_i}$:

$$\mathbf{Z} = \sum_{i=1}^{|\mathcal{P}|} \omega_{\rho_i}\mathbf{Z}_{\rho_i}. \tag{8}$$

4.4 Scholar Mutual Information

In the institution recommendation task, we do not have any labels about where the scholar will move to. Therefore, we propose a novel unsupervised framework to tackle this problem by importing Deep InfoMax [6, 21] into heterogeneous networks. The original Deep InfoMax algorithm learns the representation for the downstream task by maximizing the mutual information between the input and the output of encoder, while our approach learns the representation for the recommendation by maximizing the mutual information between local features and global features. We set the scholar embedding $\mathbf{Z}$ to be the local features and obtain the global features s by a readout function, which can be formulated as follows:

$$\mathcal{R}(\mathbf{Z}) = \frac{1}{|\mathcal{V}|}\sum_{i=1}^{|\mathcal{V}|} \mathrm{z}_i. \tag{9}$$

Next, some negative samples are provided by constructing a fake heterogeneous network from the original one. Shuffling the nodes in each type of node, a corrupted function C is constructed. Subsequently, we get a fake heterogeneous network $\widetilde{\mathcal{G}} = C(\mathcal{G}) = (\widetilde{\mathcal{V}}, \widetilde{\mathcal{E}}, \widetilde{\mathcal{A}}, \widetilde{\mathcal{X}})$. A discriminator is deployed to classify the negative samples, which is a linear binary classification formulated as follows:

$$\mathcal{D}(z_i, s) = \sigma(z_i^{\mathrm{T}}\mathbf{W}s). \tag{10}$$

Following the intuitions from Deep infomax, we use a noise-contrastive type objective with a standard binary cross-entropy (BCE) between positive samples and negative samples as the objective function:

$$\begin{aligned}\mathcal{L} = \frac{1}{N+M}\Bigg(&\sum_{i=1}^{N}\mathbb{E}_{(\mathcal{X},\mathcal{A})}\left[\log\mathcal{D}(z_i,s)\right]\\ &+\sum_{i=1}^{M}\mathbb{E}_{(\tilde{\mathcal{X}},\tilde{\mathcal{A}})}\left[\log\left(1-\mathcal{D}\left(\tilde{z_i},\tilde{s}\right)\right)\right]\Bigg).\end{aligned} \tag{11}$$

This approach effectively maximizes the mutual information between local features and global features by the Adam optimizer.

4.5 Model Analysis

The analysis of the model HAI are described as follows:

- HAI handles various types of nodes and edges, which can be trained on the heterogeneous network without labels.
- The overall algorithmic complexity of HAI is $O(VC_{att})$, where V is the number of nodes, and C_{att} represents the cost of attention operation on the node pairs of meta-path neighbors. $C_{att} = O(KNF)$ where K is the number of attention heads, N is the number of meta-path neighbors, and F is the number of input features. The low cost of computation makes HAI, which is linear to the number of nodes, efficiently capture the information from meta-path by maximizing mutual information and make decisions for the downstream institution recommendation task.
- A general challenge of heterogeneous network embedding is the low interpretability ("black box problem"), while the model we proposed has good explainability using the attention mechanism. For a certain scholar in the HSN, HAI can automatically discover the relationships between collaborators and colleagues by calculating the attention score on the list of meta-path neighbors. Meanwhile, which scholar or meta-path is more related to the task can be discovered to explain the results of our model.

5 EXPERIMENTS

In this section, we analyze the patterns of scholar career moves and present the results of the experiments to evaluate the efficiency of our model on the institution recommendation task and show the explainability.

5.1 Preprocessing

The purpose of preprocessing is to construct an HSN and generate testing set to validate our proposed model. The model is designed for those junior scholars whose academic year is between five and ten years. In this study, we construct an HSN with three types of nodes, containing authors, papers, and institutions. Considering nodes in different types have different features, we set the abstract of papers as the features of papers, the abstract of papers that authors have published during 2010 to 2015 as the features of authors, and the features of institutions are encoded to one-hot embeddings especially. If we represent the institution features in the same way of authors and papers, there may be over-smoothing in the representation space because of too many overlapping papers in different institutions. We construct the HSN on the ASN [19] which is extracted by the AMiner from DBLP, ACM, MAG (Microsoft Academic Graph), and other sources. We collect the top 100 institutions in terms of the number of articles published between 2010 and 2015. Further, we collect the postgraduate students who satisfy the following three conditions:

Table 1: Heterogeneous Scholarly Network Statistics

# Scholars	# Papers	# Institutions	# Works-With	# Writes
2563	3992	100	2563	4276

(1) academic age ranging from 5 to 10,
(2) have published papers when worked with the top 100 institutions during 2010 to 2015,
(3) staying in the same institution during these 5 years.

Table 1 shows the analysis of HSN, which has 6,655 nodes totally. The nodes consist of 2563 scholars, 3992 papers, and 100 institutions. Additionally, it contains 2563 edges connected with scholars and institutions and 4276 edges connected with scholars and papers. To evaluate the effectiveness of our model, we extract the institution from where scholars firstly employed during 2015 to 2020 as the testing set. We utilize the meta-path *APA* and *AIA* to perform the experiment.

5.2 Experimental Settings

5.2.1 Baselines. We evaluate HAI on the institution recommendation task against three state-of-the-art unsupervised heterogeneous network learning models HeGAN [8], GTN [37], HDGI-C [14]; two classical unsupervised heterogeneous network learning models MetaPath2Vec [2], HIN2Vec [4]; one classical recommendation algorithm Collaborative Filtering [15]; and one representation learning model Doc2Vec [10].

- *Collaborative Filtering* [15] is a classical widely deployed recommendation algorithm in the industry.
- *Doc2Vec* [10] is a model that represents arbitrary documents to a specific feature space.
- *MetaPath2Vec* [2] is a model for heterogeneous graph embedding, which generates meta-path based on random walks and embeds nodes through the skip-gram algorithm.
- *HIN2Vec* [4] is a neural network based model for heterogeneous network representation learning, which utilizes random walk and negative sampling to generate meta-path and represent nodes and meta-paths through neural networks.
- *HeGAN* [8] is a deep model into adversarial learning on heterogeneous information networks, which is inspired by generative adversarial networks.
- *GTN* [37] is a neural network based model for heterogeneous network, which automatically select the best meta-path.
- *HDGI-C* [14] is a model that aggregates mutual information and Graph Convolutional Networks (GCN).

5.2.2 Evaluation Metric. To quantitatively evaluate the performance of our model on the institution recommendation, we consider two widely used performance metrics in the recommendation system: AUC and Hit Ratio (HR). The metric AUC is one of the popular metrics used in the industry, which stands for "Area Under The ROC Curve". The ROC curve [3] is plotted with TPR (true positive rate) against the FPR (true positive rate). ROC is a probability curve and AUC represents the degree or measure of separability. It indicates how well the model is able to distinguish between classes: the higher the AUC is, the better the model performs at recommendation. The metric HR [17] is defined in Equation 12:

$$HR = \frac{\#hits}{\#scholars}, \tag{12}$$

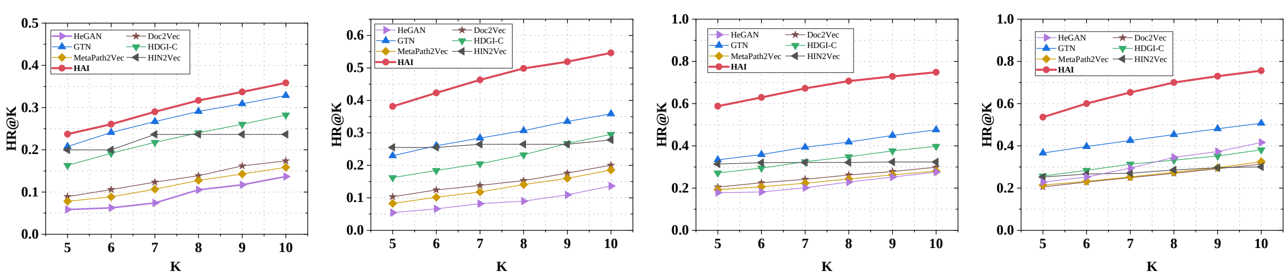
where $\#hits$ is the number of scholars whose ground-truth institution appears in the top K institutions of preference list Pre we recommend. In our experiment, we truncate the ranked list at $K \in [5\%, 6\%, 7\%, 8\%, 9\%, 10\%]$. We utilize the percentage of the rank list due to the difference in the length of preference list in the different models.

5.3 Results and Analysis

Table 2: AUC Comparisons of Different Methods under Different Dimensions and Execution Time

Models	$D64$	$D128$	$D256$	$D512$	Time
Doc2Vec	0.6209	0.6309	0.6378	0.6412	**3.16m**
MetaPath2Vec	0.6915	0.7466	0.7601	0.7701	19.26m
HIN2Vec	0.6440	0.7377	0.7513	0.6906	23.34 m
HeGAN	0.6674	0.6677	0.6769	0.7062	40.2m
GTN	**0.7353**	0.7581	0.7672	0.7880	53.46m
HDGI-C	0.7186	0.7212	0.7283	0.7122	20.30m
HAI	0.7352	**0.8442**	**0.8798**	**0.8973**	13.2m

Table 2 shows the AUC of different methods on the different feature space dimension D and the average running time of training. The collaborative filtering with AUC 0.7887 outperforms other baselines even though it is not a representation learning method, which shows that collaborative information on the ASN deeply influences the behavior of career moves. Experiment shows that the change in dimensionality has little effect on the performance of Doc2Vec, and Doc2Vec shows weak performance on the institution recommendation task, which indicates that a single abstract feature



(a) HR@K of 64 embedding dimensions (b) HR@K of 128 embedding dimensions (c) HR@K of 256 embedding dimensions (d) HR@K of 512 embedding dimensions

Figure 3: The Hit Ratio at top K of different embedding dimensions.

of scholars without structural information has no positive effect on the results. MetaPath2Vec and HIN2Vec are random walk based models for embedding heterogeneous graph, which show good performances on recommendation but consuming too much time. These two random walk based models randomly generate millions of meta-paths on our heterogeneous scholarly graph, costing too much time for walking on each node. The generative adversarial based model HeGAN shows a good performance, but shows instability while training. GTN with mutual information performs second only to our model, while performance improves slowly as the dimension rises. Our model outperforms other baselines on most conditions, and the AUC score close to 0.9 indicates that the ground-truth institution is ranked on the top.

Figure 3 shows the HR of different dimensions. Only semantic information based model Doc2Vec and only structure information based model MetaPath2Vec both get a poor performance at HR, while our model that aggregates semantic information and structure information outperforms other baselines. Models that utilize the semantic information of the scholar's abstract have significantly higher Hit Ratio values than models that only utilize structural information. The MetaPath2Vec performs good at AUC but average at HR indicates that the model ranks most ground-truth institutions between 20 and 30 percent of the preference list but not ranks the ground-truth institutions high enough to hit top K. Although the GTN with mutual information performs similar to our model at 64 feature dimensions, the transformer of GTN consumes lots of time to transform the network while our model only needs about ten iterations to produce good results in a few minutes. Overall, we recommend using our model to represent scholars into 128-dimensional feature space for institution recommendation to achieve a balance between performance and time-consuming.

5.4 Case Study

5.4.1 Attention Behavior. With the case study, we find that even without collaboration information, HAI still can recommend for early-career researchers institutions by research topic. Figure 4 illustrates the true case of our model recommending institution to the early-career researcher Sam Staton. Sam Staton and other scholars are on the list of *APA* meta-path neighbors, where attention is carried on. We can see that the scholar attention score from Sam Staton to Dimitri Kartsaklis is significantly higher than others, at the same time Dimitri Kartsaklis is ranked on the top of preference list by our model. Thus, our model recommends Sam Staton to move to the institution Dimitri Kartsaklis works with. Sam Station moved from University of Cambridge to University of Oxford in March 2015, which matched our recommendation result. Even Sam Station had not collaborated with Dimitri Kartsaklis, the meta-path in scholarly network reveal the similarity between them. And the papers they had published embedded to the node features reveal that they have the same research interest "Quantum Computing Languages (QCL)".

5.4.2 Recommendation Influence. To quantitatively demonstrate the effectiveness of our recommendations, we analyze the average publication numbers and average citation numbers of Hit scholars and Non-Hit scholars between 2015 and 2020. The scholars are from 100 famous institutions over the world and have similar academic

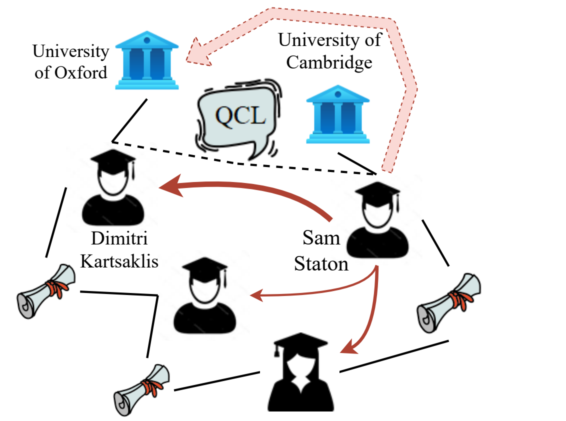


Figure 4: The figure illustrates a case study on how the attention mechanism carries on the meta-path neighbors. The black solid lines represent the edge in the network. The dotted lines indicate the hidden research topic relationships in the network. The red arrows indicate the attention scores and the thickness of the arrow indicates the value of the score. The red arrow with a dashed boundary indicates the career move.

abilities. We consider those scholars that joined the institution in the Top K of preference list our model recommended as Hit scholars. Similarly, we consider those who did not join the institution in the Top K of preference list as Non-Hit scholars. Figure 5 shows that

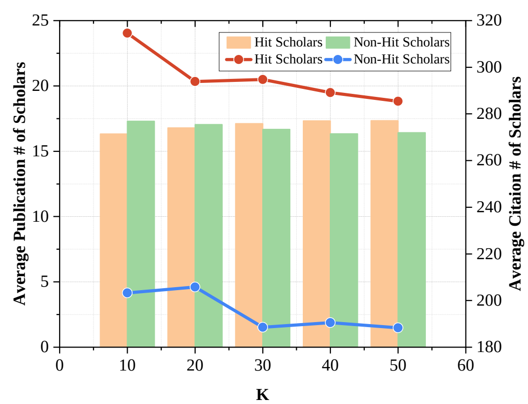


Figure 5: The bar graph illustrates the average publication numbers of Hit scholars and Non-Hit scholars between 2015 and 2020. The line graph illustrates the average citation numbers of Hit scholars and Non-Hit scholars between 2015 and 2020.

the average numbers of citations are significantly greater for Hit scholars than Non-Hit scholars with a small difference in the average numbers of publications. As the institutions that Hit scholars joined getting ranked further down the preference list, the average numbers of citations get lower and lower. The above results show that Hit scholars have higher average citations than Non-Hit scholars, which means our recommendations are meaningful for early-career scholars.

6 CONCLUSION

In this work, we found that nearly half of the junior scholars had changed academic institutions between 2010 and 2015, with 74% moved to their collaborator's institution. Based on these findings, we have designed a novel unsupervised learning algorithm HAI for institution recommendation. We applied an attention mechanism to calculate the scholar attention scores on the meta-path neighbors, and we further integrated the abstract representation embedded by Doc2Vec and semantic information of meta-path to represent local features and global features of scholars. Finally, we proposed a mutual information based approach to aggregate these features. Experimental results on the ASN demonstrated the effectiveness and efficiency of our approach. Based on the findings presented in this paper, it is possible to explore the motivation of senior scholars' career moves, as well as learning scholar information on the dynamic scholarly network in the future.

REFERENCES

[1] Monya Baker. 2015. Social media: A network boost. *Nature* 518, 7538 (2015), 263–265.

[2] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 135–144.

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

[4] Tao-yang Fu, Wang-Chien Lee, and Zhen Lei. 2017. Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. 1797–1806.

[5] Teng Guo, Feng Xia, Shihao Zhen, Xiaomei Bai, Dongyu Zhang, Zitao Liu, and Jiliang Tang. 2020. Graduate Employment Prediction with Bias. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 670–677.

[6] R. Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Philip Bachman, Adam Trischler, and Yoshua Bengio. 2018. Learning deep representations by mutual information estimation and maximization. In *International Conference on Learning Representations*.

[7] Mingliang Hou, Jing Ren, Da Zhang, Xiangjie Kong, Dongyu Zhang, and Feng Xia. 2020. Network embedding: Taxonomies, frameworks and applications. *Computer Science Review* 38 (2020), 100296.

[8] Binbin Hu, Yuan Fang, and Chuan Shi. 2019. Adversarial learning on heterogeneous information networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 120–129.

[9] 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 *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1531–1540.

[10] Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International conference on machine learning*. PMLR, 1188–1196.

[11] Yuanfu Lu, Chuan Shi, Linmei Hu, and Zhiyuan Liu. 2019. Relation Structure-Aware Heterogeneous Information Network Embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 4456–4463.

[12] Meng Qu, Jian Tang, and Jiawei Han. 2018. Curriculum Learning for Heterogeneous Star Network Embedding via Deep Reinforcement Learning. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 468–476.

[13] Jing Ren, Lei Wang, Kailai Wang, Shuo Yu, Mingliang Hou, Ivan Lee, Xiangjie Kong, and Feng Xia. 2019. API: An Index for Quantifying a Scholar's Academic Potential. *IEEE Access* 7 (2019), 178675–178684.

[14] Yuxiang Ren, Bo Liu, Chao Huang, Peng Dai, Liefeng Bo, and Jiawei Zhang. 2019. Heterogeneous deep graph infomax. *arXiv preprint arXiv:1911.08538* (2019).

[15] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*. 285–295.

[16] J Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. 2007. Collaborative filtering recommender systems. In *The adaptive web*. Springer, 291–324.

[17] Chuan Shi, Xiaotian Han, Li Song, Xiao Wang, Senzhang Wang, Junping Du, and Philip S. Yu. 2021. Deep Collaborative Filtering with Multi-Aspect Information in Heterogeneous Networks. *IEEE Transactions on Knowledge and Data Engineering* 33, 4 (2021), 1413–1425.

[18] Chuan Shi, Binbin Hu, Wayne Xin Zhao, and Philip S. Yu. 2019. Heterogeneous Information Network Embedding for Recommendation. *IEEE Transactions on Knowledge and Data Engineering* 31, 2 (2019), 357–370.

[19] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. 990–998.

[20] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 6000–6010.

[21] Petar Velickovic, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, and R. Devon Hjelm. 2018. Deep Graph Infomax. In *International Conference on Learning Representations*.

[22] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *International Conference on Learning Representations*.

[23] Liangtian Wan, Feng Xia, Xiangjie Kong, Ching-Hsien Hsu, Runhe Huang, and Jianhua Ma. 2020. Deep Matrix Factorization for Trust-Aware Recommendation in Social Networks. *IEEE Transactions on Network Science and Engineering* 8, 1 (2020), 511–528.

[24] Wei Wang, Jiaying Liu, Tao Tang, Suppawong Tuarob, Feng Xia, Zhiguo Gong, and Irwin King. 2020. Attributed collaboration network embedding for academic relationship mining. *ACM Transactions on the Web (TWEB)* 15, 1 (2020), 1–20.

[25] Wei Wang, Tao Tang, Feng Xia, Zhiguo Gong, Zhikui Chen, and Huan Liu. 2020. Collaborative Filtering with Network Representation Learning for Citation Recommendation. *IEEE Transactions on Big Data* (2020), 1–1.

[26] Wei Wang, Feng Xia, Jian Wu, Zhiguo Gong, Hanghang Tong, and Brian D. Davison. 2021. Scholar2vec: Vector Representation of Scholars for Lifetime Collaborator Prediction. *ACM Transactions on Knowledge Discovery From Data* 15, 3 (2021), 40.

[27] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous graph attention network. In *The World Wide Web Conference*. 2022–2032.

[28] Yakun Wang and Xiaodong Han. 2021. Attractive community detection in academic social network. *Journal of Computational Science* 51 (2021), 101331.

[29] Zekai Wang, Hongzhi Liu, Yingpeng Du, Zhonghai Wu, and Xing Zhang. 2019. Unified Embedding Model over Heterogeneous Information Network for Personalized Recommendation.. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. 3813–3819.

[30] Lukasz Wiechetek, Kongkiti Phusavat, and Zbigniew Pastuszak. 2020. An analytical system for evaluating academia units based on metrics provided by academic social network. *Expert Systems with Applications* 159 (2020), 113608.

[31] Chris Woolston. 2017. Graduate survey: A love–hurt relationship. *Nature* 550, 7677 (2017), 549–552.

[32] Feng Xia, Jiaying Liu, Hansong Nie, Yonghao Fu, Liangtian Wan, and Xiangjie Kong. 2020. Random Walks: A Review of Algorithms and Applications. *IEEE Transactions on Emerging Topics in Computational Intelligence* 4, 2 (2020), 95–107.

[33] Feng Xia, Ke Sun, Shuo Yu, Abdul Aziz, Liangtian Wan, Shirui Pan, and Huan Liu. 2021. Graph Learning: A Survey. *IEEE Transactions on Artificial Intelligence* 2, 2 (2021), 109–127.

[34] Feng Xia, Wei Wang, Teshome Megersa Bekele, and Huan Liu. 2017. Big scholarly data: A survey. *IEEE Transactions on Big Data* 3, 1 (2017), 18–35.

[35] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L. Hamilton, and Jure Leskovec. 2018. Graph Convolutional Neural Networks for Web-Scale Recommender Systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 974–983.

[36] Shuo Yu, Jiaying Liu, Feng Xia, Haoran Wei, and Hanghang Tong. 2021. How to optimize an academic team when the outlier member is leaving. *IEEE Intelligent Systems* (2021), 1–1.

[37] Seongjun Yun, Minbyul Jeong, Raehyun Kim, Jaewoo Kang, and Hyunwoo J. Kim. 2019. Graph Transformer Networks. In *33rd Annual Conference on Neural Information Processing Systems, NeurIPS 2019*, Vol. 32. 11960–11970.

[38] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. 2019. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 793–803.

[39] Huan Zhao, Quanming Yao, Jianda Li, Yangqiu Song, and Dik Lun Lee. 2017. Meta-graph based recommendation fusion over heterogeneous information networks. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 635–644.

[40] Zhongying Zhao, Xuejian Zhang, Hui Zhou, Chao Li, Maoguo Gong, and Yongqing Wang. 2020. HetNERec: Heterogeneous network embedding based recommendation. *Knowledge Based Systems* 204 (2020), 106218.