Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Exploring Financially Constrained SMEs based on Multi-Relation Translational Graph Attention Network

Qianqiao Liang[1], Hua Wei[2], Yaxi Wu[2], Feng Wei[2], Deng Zhao[3], Jianshan He[3], Xiaolin Zheng[†1], Guofang Ma[4], Bing Han[2]

[1]*College of Computer Science, Zhejiang University, Hangzhou, China*
[2]*MYbank, Ant Group, Hangzhou, China*
[3]*Ant Group, Hangzhou, China*
[4]*School of Computer and Information Engineering, Zhejiang Gongshang University*
[†]E-mail: xlzheng@zju.edu.cn

Abstract: Financing needs exploration (FNE), which explores financially constrained small and medium-sized enterprises (SMEs), becomes increasingly important in industry for financial institutions to facilitate SMEs' developments. In this paper, we first perform an insightful exploratory analysis on an industrial dataset to exploit the financing needs transfer phenomenon among SMEs, which motivates us to fully exploit the multi-relational enterprise social network for boosting the effectiveness of FNE. The main challenges lies in modeling two kinds of heterogeneity, i.e., transfer heterogeneity and SME's behavior heterogeneity, under different relation types simultaneously. To address these challenges, we propose a graph neural network named M-RIGHT, which not only models the transfer heterogeneity of financing needs along different relation types based on a novel entity-relation composition operator, but also enables heterogeneous SMEs' representations based on a translation mechanism on relational hyperplanes to distinguish SMEs' heterogeneous behaviors under different relation types. Extensive experiments on two large-scale real-world datasets demonstrate M-RIGHT's superiority over the state-of-the-art methods in the FNE task.

Key words: Financing needs exploration, Graph representation learning, Transfer heterogeneity, Behavior heterogeneity

https://doi.org/ **CLC number:** TP391.4

1 Introduction

As the financial markets becomes increasingly volatile with the outbreak of crises such as the COVID-19 pandemic and geopolitical wars, more and more small and medium-sized enterprises (SMEs) are facing financial stress and are in need of financing, which motivates us to study the problem of financial needs exploration (FNE). The FNE problem aims to exploit those financially constrained SMEs among millions of SMEs, which is significant for financial institutions to provide financial support to those struggling SMEs and facilitate their developments. However, the transfer of financing needs among SMEs within the enterprise social network causes the complicated external factors of SMEs' financing needs (Ceptureanu et al., 2019), which makes it very difficult to address the FNE problem effectively. Therefore, it is in high demand to devise an effective graph-based method that can model the financing needs transfer phenomenon for accurate financial needs exploration.

Motivating Example. An SME's financing need will transfer to its related SMEs, which will have influences on their degree of financing needs. Fig. 1 depicts SMEs' financing needs transfer via a toy example, in which a color bar denotes an SME's financing need degree and the colors of edges denote

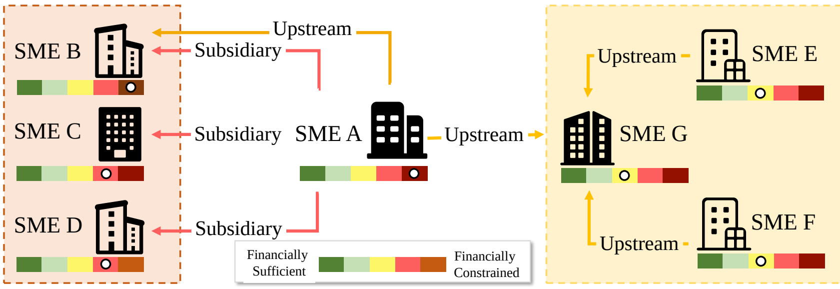


Figure 1 A motivating example. The financing need transferred from SME A to different SMEs under different relation types.

the transfer intensities. In this example, SME A is extremely financially constrained, which will transfer to its subsidiary SMEs and its upstream SMEs, resulting in the increases of the financing needs degrees of its related SMEs. This phenomenon is common in real practice that a parent SME is likely to reduce financial support for its subsidiary, while a downstream company tends to owe the upstream company money if it is financially constrained. An effective FNE method is supposed to consider such significant transfer factor to improve its performance. More details about the financing needs transfer phenomenon among SMEs will be elaborated in Section 3 based on an industrial data collected from MYbank [1].

The above example motivates us to construct an SMEs graph, in which the nodes are SMEs and the edges are the relations among SMEs, and apply graph representation learning methods to model the transfer phenomenon of financing needs among SMEs for facilitating the FNE task.

However, in real scenario, SMEs graph contains multiple relation types, and the above idea faces two kinds of heterogeneity, i.e., transfer heterogeneity and behavior heterogeneity, which make it very challenging to model the financing needs transfer among SMEs. On the one hand, *transfer heterogeneity* **(CH1)** refers that the transferred financing needs are different under different relation types. As shown in Fig. 1, in general, the financing needs transferred from SME A to its subsidiaries are stronger in intensity than those transferred to its upstream SMEs. This is because a "parent" SME usually has substantial capital control and management decision-making power over its subsidiary, which means that the parent SME generally has a great influence on its subsidiaries. In comparison, SME A is only one of the many downstream enterprises of SME G, which means that SME A's constrained financing condition will only partially affect SME G's overall operating conditions, and the transfer intensity of SME A's financing needs is unlikely to be too large. Such transfer heterogeneity increases the complexity of an effective FNE method. On the other hand, *behavior heterogeneity* **(CH2)** refers that the each SME has different behaviors under different relation types. As shown in the Fig. 1, SMEs B, C, and D are all subsidiaries of SME A, whose financing needs are also affected by SME A. Intuitively, SMEs B, C, and D should have similar financing needs conditions, which however, does not hold. Even though SME B has similar behavior to SMEs C and D under the relation type subsidiary, it may behave differently under other relation types, which leads to their differences in financing needs conditions. The behavior heterogeneity is likely to originates from the fact that SMEs have different roles under different relationships. An FNE method failing to recognize the behavior heterogeneity will have a decrease in its performance.

In this paper, to address the two kinds of heterogeneity for facilitating the FNE task, we propose a novel **M**ulti-relation t**R**anslat**I**onal **G**rap**H** a**T**tention network, named **M-RIGHT**, which includes two key modules, i.e., the transfer heterogeneity learning module and the behavior heterogeneity learning module. Specifically, the *transfer heterogeneity learning module* attentively transfers the message of financing needs among connected SMEs under different relation types based on our devised entity-relation composition operator, where the operator is able to distinguish heterogeneous transferred messages by utilizing the heterogeneous representations of different relation types

[1] MYbank refers to Zhejiang E-commerce Bank, which is a Chinese company that offers banking services for SMEs.

(for addressing CH1). The *behavior heterogeneity learning module* first learns the heterogeneous representations of SMEs under different relation types by performing SMEs representation translations on each relational hyperplane, which helps distinguish each SME's heterogeneous behaviors (for addressing CH2), then leverages SMEs' heterogeneous representations to predict the graph triplets and compute the corresponding loss for the model's update in a self-supervised learning manner. Finally, M-RIGHT leverages a tree-based method to predict financially constrained SMEs based on the learned SMEs' representations.

To sum up, our main contributions are:

- We have provided an in-depth exploratory analyses to exploit the financing needs transfer phenomenon in the enterprise social network, i.e., SME graph, based on an industrial dataset, which indicates the transfer heterogeneity and behavior heterogeneity in real practice.
- We propose a novel graph representation learning based method, named M-RIGHT, which effectively models the two kinds of heterogeneity in the financing needs transfer phenomenon. To the best of our knowledge, this is the first method to model the financing needs transfer phenomenon in SME graphs, which is beneficial to the FNE task.
- We have conducted comprehensive experiments on two real-world datasets, which demonstrate M-RIGHT's superiority over the state-of-the-art methods in the FNE task.

2 Related Works

In this section, we survey two lines of studies highly relevant to FNE.

2.1 Financing Needs Exploration

Existing methods for the FNE task can be categorized into two kinds, i.e., rule based methods and machine learning based methods. Rule-based method designs various objective functions manually based on empirical rules and utilizes rating models (Angilella and Mazzù, 2015; Luo et al., 2021) to predict the financially constrained probability of SMEs. Such methods depend heavily on expert advice, which fail to capture SMEs' financial patterns automatically. To facilitate automatic FNE, some financial institutes have been utilizing machine learning based methods, which models SMEs' financial patterns (Graesch et al., 2021; Jeon, 2021) and explores the financially constrained SMEs (Kshetri, 2016; Tian et al., 2018) based on machine learning algorithms. Although few studies have reported their methods' details, they claimed that some methods for exploring potential customers could be reference solutions. For example, Zhang et al. (2021) designed a clustering algorithm to target customers in the e-commerce platform. Xu et al. (2021) leveraged contextual bandit to model the funnel structure in the email marketing campaigns. Rogic and Kascelan (2019) and Rogić et al. (2022) utilized support vector machines to predict the values of customers. Graesch et al. (2021) devised a method to direct the marketing campaigns in retail banking based on a deep belief network. Chen et al. (2020) and Duan and Ma (2018) leveraged extreme gradient boosting algorithm, i.e., XGBoost, to mine the potential customers that require their products. Lessmann et al. (2021) proposed an ensemble learning framework including XGboost to target the potential customers for profit. In addition to the studies of customer targeting, deep learning based models (Guo et al., 2017; Cheng et al., 2016) and tree based models (Chen et al., 2020) are common solutions for the FNE task in real industrial practice.

Despite the success of these methods, they can hardly achieve satisfying performance in the FNE task because these methods only focus on SMEs' financing needs induced by their own operational conditions and ignore the financing needs transfer phenomenon within the enterprise social network. In real scenarios, the financing needs of SMEs are affected by not only the independent internal factors but also the external factors conveyed by the relations between SMEs (Ceptureanu et al., 2019).

2.2 Graph Representation Learning

Graph representation learning methods are popular in embedding graphical structural data, which can be applied in exploring SMEs graphs for boosting FNE. Existing graph representation learning methods can be categorized into two kinds.

2.2.1 Homogeneous graph representation learning

Homogeneous graph representation learning methods are designed to model graph data with homogeneous node types or homogeneous relation types. Early studies are relatively shallow, which first perform random walks (Grover and Leskovec, 2016; Perozzi et al., 2014) on graph data to generate node sequences, then input these node sequences into word2vec (Mikolov et al., 2013) to obtain node representations. With the development of deep learning, graph neural networks (Wu et al., 2020) attract great attention from researchers, in which the graph convolution network (Welling and Kipf, 2017) has achieved great success. Based on it, Hamilton et al. (2017) designed an inductive graph convolution network, which can efficiently perform message passing between connected nodes. To further improve the effectiveness, Velickovic et al. (2018) incorporated attention mechanisms and Xu et al. (2019) incorporated the Weisfeiler-Lehman test into the message passing process. The attention based graph neural network have been proven to be effective in various tasks (Wang et al., 2022; Liao et al., 2022). Despite the success of homogeneous graph representation learning methods, they are not referenced solutions for the FNE problem due to their lack of considerations in the graph's multiple relation types, meaning that they are unable to model the two kinds of heterogeneity in FNE.

2.2.2 Heterogeneous graph representation learning

Heterogeneous graph representation learning methods are designed to model graph data with different node types or different relation types, which can be categorized into three kinds (Yang et al., 2022). The first kind of methods is the proximity-preserving method, which obtains node representations by preserving the similarity of the node to its heterogeneous neighbors, such as the similarities in random walks (Dong et al., 2017; Shi et al., 2018b; Wang et al., 2019), under different relation types (Tang et al., 2015), from different perspectives (Shi et al., 2018a), or in different meta-paths (Zhang et al., 2020b). However, these methods are too shallow to be applicable to the increasingly complex high-order data. The second kind of method is the message passing method, which aggregates node representations from their heterogeneous neighbors in the deep graph neural network. Schlichtkrull et al. (2018) and Shang et al. (2019) proposed extensions of graph convolution networks on heterogeneous graphs, which first models the message passing under different relation types separately then aggregates the node representations from various message passing paths. Furthermore, Ye et al. (2019) and Vashishth et al. (2020a) proposed to learn the node and relation representations simultaneously for improving the modeling of heterogeneous graphs. In addition, Zhang et al. (2020b), Fu et al. (2020) and Zhao et al. (2021) proposed to learn node representations under different meta-paths, where the design of the meta-paths is task-specific and highly depends on expert knowledge. To eliminate the dependency of the meta-paths, Yu et al. (2022) automatically captured the meta-paths in heterogeneous graph neural network. The disadvantage of these kind of methods is its lack of knowledge modeling in heterogeneous graphs, which is significant in the SME graphs, such as the knowledge of subsidiary and upstream. The third kind of methods is relation-learning methods (Ji et al., 2021), which focuses on preserving knowledge structure, i.e., knowledge triplets, based on different triplet scoring functions and is common for knowledge graph embedding. Traditional relation-learning methods includes triple translation methods (Bordes et al., 2013; Wang et al., 2014; Sadeghian et al., 2021) and semantic-based triplet matching methods (Yang et al., 2015; Nickel et al., 2016; Trouillon et al., 2016), which are too shallow to capture the complex knowledge in the triplets. Instead, deep neural network-based methods utilize deep node embeddings for calculating the triplets (Dettmers et al., 2018; Shang et al., 2019; Vashishth et al., 2020b; Li et al., 2022). However, existing relation-learning methods do not take into consideration the heterogeneous representations of nodes under different relation, which means that it is difficult to guarantee their performances on recognizing the behavior heterogeneity in the FNE task.

In conclusion, existing graph representation learning methods are ineffective in modeling SMEs graphs because of their inabilities in addressing the challenging transfer heterogeneity and behavior heterogeneity simultaneously in the FNE task.

3 Exploratory Analyses

In this section, we conduct in-depth exploratory analyses based on an industrial bank dataset collected from one of the largest fintech banks in China, and discover the financing needs transfer patterns in the enterprise social network, i.e., SME graph.

The dataset contains 42.5 million SMEs and 1.3 billion relations among SMEs. Specifically, we refer SMEs that brought a loan product as financially constrained SMEs, otherwise it is financially sufficient. The details of the analyses are elaborated as follows.

3.1 Correlation between SMEs' Relations and Financing Needs

To explore the correlation between SMEs' relations and their financing needs, we conduct two statistical analysis based on six randomly chosen relation types, i.e., affiliated, parent, supplier, stock owner, stock guarantor, and joint owner .

In the first statistical analysis, we first separate SMEs into a financially constrained group and a financially sufficient group, then calculate the number of SMEs that have previously financially constrained neighbors to the size of each group. The results are shown in Fig. 2. Obviously, under the six types of relations, financially constrained SMEs are more likely to have previously financially constrained neighbors, which indicates the importance of SMEs' relations on their financing needs.

In the second statistical analysis, we investigate the correlation between the financing needs of SMEs and the number of their financially constrained neighbors. Specifically, under each relation, we group SMEs by the number of financially constrained neighbors that they have and report the ratio of financially constrained SMEs in each group. The results are shown in Fig. 3. It is worth noticing that the more financially constrained neighbors an SME has, the more likely it is also financially constrained, which indicates that relations among SMEs have a great impact on SMEs' financing needs.

The results of the above statistical analyses demonstrate that connected SMEs tends to have similar financing conditions, which verify the necessity of modeling SMEs' relations in the FNE task.

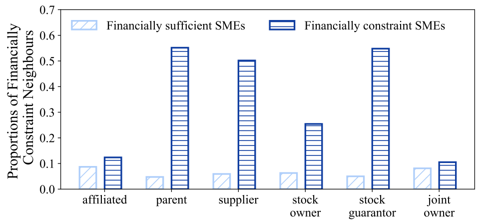


Figure 2 Comparisons of financially constrained neighbors under two different SME groups.

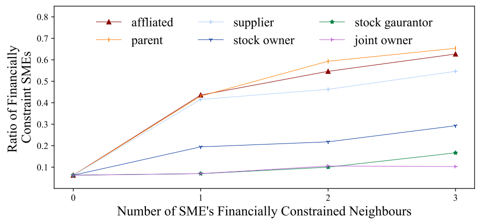


Figure 3 Ratio comparisons of financially constraint SMEs under different SME groups. SMEs in each group have a specific number of financially constrained neighbors.

3.2 Financing Needs Transfer Under Heterogeneous Relations

We further investigate the results under different relation types. Interestingly, the proportions in Fig. 2 are different under various relation types. The differences in the results under six relation types are even more significant in Fig. 3, which demonstrates that the transferred messages are heterogeneous under different types of relations. This observation motivates us to pay attention to transfer heterogeneity when modeling the SMEs graphs.

In addition, we discover that the relations in the SMEs graph have multi-structure properties. For example, the relation type 'affiliated', 'parent', and 'stock guarantor' have a one-to-many mapping structure, the 'stock owner' and 'supplier' are structured to be many-to-many mapping, whereas the relation type 'stock guarantor' and 'joint owner' are structured to be many-to-one mapping. In addition, each SME has more than two types of relations on average. These observations indicate that each SME may have different roles and behaves differently under different relation types. Such behavior heterogeneity is common in real enterprises' social network, which motivates us to allow heterogeneous representations of each SME under different relation types.

Table 1 Main notations of our proposed M-RIGHT.

Notation	Description
$\mathbf{h}_i^l \in \mathbb{R}^{d_l}$	Node i's input feature in the l-th layer
$\mathbf{r}_i^l \in \mathbb{R}^{d_l}$	Relation i's input feature in the l-th layer
L	Total number of layers
K	Total number of attention head
$\mathcal{N}(h)$	Neighbor set of a node h
$\mathcal{T}(i,j)$	Relation type between node i and j
$\mathbf{W}_{\mathbf{h},k}^l \in \mathbb{R}^{d_l \times \frac{d_{l+1}}{k}}$	Node's transformation metric in the k-th attention heads of the l-th layer
$\mathbf{W}_{\mathbf{r}}^l \in \mathbb{R}^{(d_l)\times(d_{l+1})}$	Relation transformation metric in the l-th layer
$\mathbf{W}_{\mathbf{w}}^l \in \mathbb{R}^{(d_l)\times(d_{l+1})}$	Relational hyperplane transformation metric in the l-th layer
$\mathbf{w}_r^l \in \mathbb{R}^{d_l}$	Hyperplane's projection vector for relation type r in the l-th hidden layer

4 Methodology

In this section, we first provide formal definitions related to the FNE tasks. Then, we present the overview of our proposed M-RIGHT, followed by the details of its two key modules. Finally, we analyze the properties of M-RIGHT generally.

4.1 Problem Settings and Preliminaries

Definition 1 (SME Graph) An SME graph is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \triangle, \mathbf{H}, \mathbf{R})$. Here, $\mathcal{V}$ is the set of nodes that represent SMEs, $\mathcal{E}$ is the set of heterogeneous relation types among SMEs, $\triangle$ is the set of triplets, each of which $(s, r, o) \in \triangle$ contains the head node s, relation type r, and tail node o, correspondingly, $\mathbf{H} \in \mathbb{R}^{|\mathcal{V}|\times d_0^h}$ and $\mathbf{R} \in \mathbb{R}^{|\mathcal{E}|\times d_0^e}$ denote the original nodes feature matrix of all SME nodes and the relation feature matrix of all relation types respectively.

Definition 2 (Financial Needs Exploration (FNE)) Given the SME graph $\mathcal{G}$ and the labeled financially constrained SMEs as the training set, in which $y = 1$ denotes that the SME is financially constrained and $y = 0$ denotes that the SME is financially sufficient, the goal of the FNE task is to predict the financially constrained SMEs in the future.

Definition 3 (Graph Representation Learning) Graph representation learning methods aim to learn the representations of nodes and their relations that encode structural information of the graph given an SMEs graph $\mathcal{G}$. In the FNE task, the learned representations can be used as inputs of the downstream model to predict financially constrained SMEs.

The main notations in this paper are shown in Table 1, which we will elaborate in following sections.

4.2 Overview of M-RIGHT

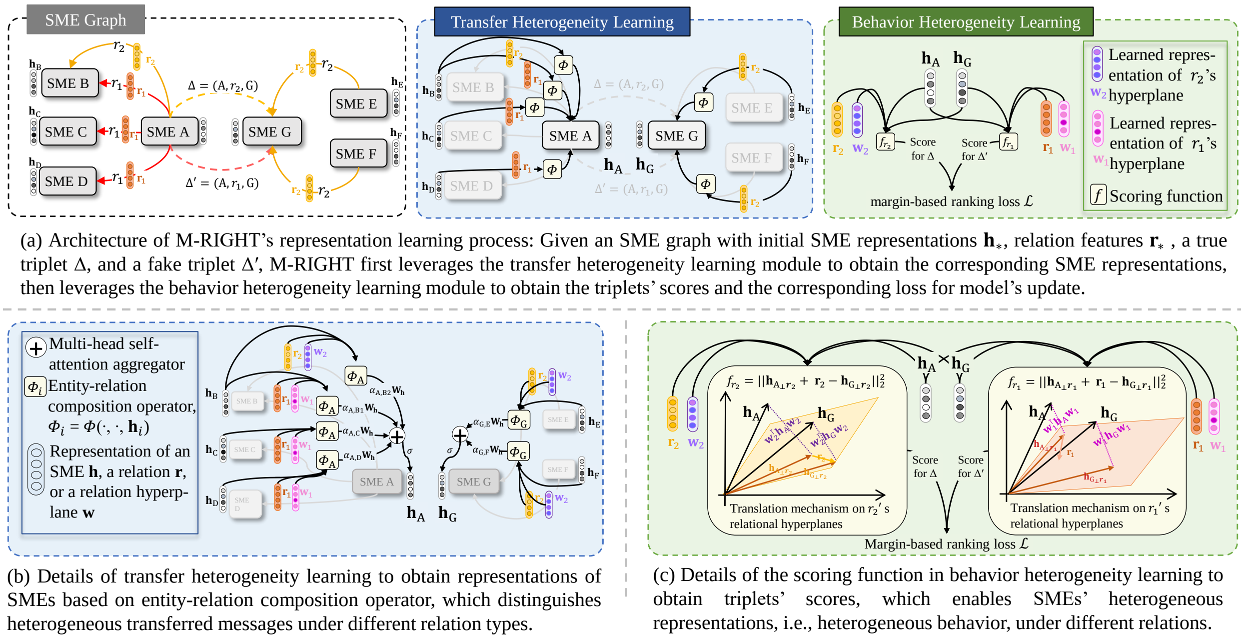
Overall, M-RIGHT first learns the representations of SMEs and their relation types, then leverages the learned representations to facilitate the downstream FNE task. The architecture of M-RIGHT's representation learning process is presented in Fig. 4, which includes two key modules, i.e., the transfer heterogeneity learning module and the behavior heterogeneity learning module. The *transfer heterogeneity learning module* attentively transfers representations among connected SMEs based on our devised entity-relation composition operator, which is able to distinguish heterogeneous transferred messages under different relation types, and obtains a representation of each SME after the messages transfer. The *behavior heterogeneity learning module* first enables heterogeneous representations of each SME under different relation types by performing SMEs representation translations on each relational hyperplane, which helps distinguish each SME's heterogeneous behavior, then leverages SMEs' heterogeneous representations to predict the scores of graph triplets for the model's update in a self-supervised learning manner. Finally, with the learned SMEs' representations, M-RIGHT leverages XGBoost to predict financially constrained SMEs. The following sections elaborate on M-RIGHT's transfer heterogeneity learning module and the behavior heterogeneity learning module.

4.3 Transfer Heterogeneity Learning Module

In this module, M-RIGHT performs heterogeneous message passing via L graph convolution layers and obtains the representations of SMEs. Specifically, M-RIGHT leverages our devised entity-relation composition operator in the message passing process to distinguish heterogeneous transferred messages under different relation types. To further improve the effectiveness of the message passing, M-RIGHT utilize a self-attention mechanism to differentiate the importance of neighbors. In this section, we introduce the message passing process and the self-attention mechanism, respectively.

4.3.1 Message passing with entity relation composition operator

In the l-th layer of the graph convolution network, M-RIGHT obtains a set of SMEs representa-



(a) Architecture of M-RIGHT's representation learning process: Given an SME graph with initial SME representations $\mathbf{h}_*$, relation features $\mathbf{r}_*$, a true triplet Δ, and a fake triplet Δ', M-RIGHT first leverages the transfer heterogeneity learning module to obtain the corresponding SME representations, then leverages the behavior heterogeneity learning module to obtain the triplets' scores and the corresponding loss for model's update.

(b) Details of transfer heterogeneity learning to obtain representations of SMEs based on entity-relation composition operator, which distinguishes heterogeneous transferred messages under different relation types.

(c) Details of the scoring function in behavior heterogeneity learning to obtain triplets' scores, which enables SMEs' heterogeneous representations, i.e., heterogeneous behavior, under different relations.

Figure 4 The architecture of M-RIGHT's representation learning process, which includes the transfer heterogeneity learning module and behavior heterogeneity learning module.

tions, i.e., node embeddings $\mathbf{h} = \{\mathbf{h}_1^l, \mathbf{h}_2^l, \ldots, \mathbf{h}_{|\mathcal{V}|}^l\}$, where $\mathbf{h}_i^l \in \mathbb{R}^{d_l}$, and a set of relation embeddings $\mathbf{r} = \left\{\mathbf{r}_1^l, \mathbf{r}_2^l, \ldots, \mathbf{r}_{|\mathcal{E}|}^l\right\}$, where $\mathbf{r}_i^l \in \mathbb{R}^{d_l}$. Then, M-RIGHT calculates the representation of each SME by message passing from their one-hop neighbors. To distinguish the information transferred from different relation types, we devise an entity-relation composition operator $\Phi(\cdot) : \mathbb{R}^{d_l} \times \mathbb{R}^{d_l} \to \mathbb{R}^{d_l}$, which aggregates the representations from both the neighbors and the corresponding relation types during the message passing process. Then, each SME's output representation in this layer is obtained by a shared linear transformation parametrized by a matrix, $\mathbf{W}_{\mathbf{h}}^l \in \mathbb{R}^{d_l \times d_{l+1}}$ along with a nonlinear transformation parametrized by a function $\sigma(\cdot)$. The representations of each SME u in this layer is calculated as:

$$\mathbf{h}_u^{l+1} = f\left(\sum_{v \in \mathcal{N}(u)} \mathbf{W}_{\mathbf{h}}^l \Phi\left(\mathbf{h}_v^l, \mathbf{r}_{\mathcal{T}(u,v)}^l, \mathbf{h}_u^l\right)\right), \quad (1)$$

where $\mathcal{N}(v)$ denotes the one-hop neighbor set of SME u, $\mathcal{T}(u,v)$ denotes the relation type between SMEs u and v, and Φ is the entity-relation composition operator that distinguish heterogeneous transferred messages from different relation types. Note that the number of relations among two SMEs is allowed to be more than one. The instantiation of the entity-relation composition operation will be elaborated in Section 4.4.

In addition, because the update of SMEs' representations in Eq. (1) will transform the original vector space, the representations of each relation type i in layer l should be transformed similarly with transformation matrix $\mathbf{W}_r^l \in \mathbb{R}^{d_l \times d_{l+1}}$ to obtain the output relation type representation as follows:

$$\mathbf{r}_i^{l+1} = \sigma\left(\mathbf{W}_{\mathbf{r}}^l \mathbf{r}_i^l\right). \quad (2)$$

Overall, Eq. (1) and Eq.(2) allow our proposed M-RIGHT to model the transfer heterogeneity while keeping the space complexity of relation type modeling to be ($\mathcal{O}(|\mathcal{E}|d_l)$), i.e., linear in the number of feature dimensions.

4.3.2 Convolution with multi-head self-attention

One drawback of the SME representation update introduced in Eq. (1) is that it can neither deal with variable-sized neighbors as input nor focuses on the most relevant neighbors for message passing. To address these problems, we utilized a self-attention mechanism (Velickovic et al., 2018) that updates the each SME's representation by attending over its neighbors. Similar to the graph attention network (GAT) (Velickovic et al., 2018), to stabilize the learning process of self-attention, we extend the mechanism to employ multi-head attention, in which K independent self-attention processes are performed independently and concatenated jointly to replace Eq. (1).

Specifically, in the k-th head, for each SME u, the attention coefficient of its neighbor v is computed with a shared attentional single-layer feed-forward neural network $a : \mathbb{R}^{d_{l+1}} \times \mathbb{R}^{d_{l+1}} \to \mathbb{R}$ as follows:

$$e_{u,v}^{k} = a(\mathbf{W}_{\mathbf{h},k}^{l}\mathbf{h}_{u}^{l}, \mathbf{W}_{\mathbf{h},k}^{l}\Phi(\mathbf{h}_{v}^{l}, \mathbf{r}_{\mathcal{T}(i,j)}^{l})), \tag{3}$$

where $\mathbf{W}_{h,k}^{l} \in \mathbb{R}^{d_l \times \frac{d_{l+1}}{K}}$ is the linear transformation metric shared among all SMEs in the attention head k of layer l. We choose the attentional single-layer feed-forward neural network a to be parameterized by a weight vector $\mathbf{a} \in \mathbb{R}^{\frac{2d_{l+1}}{K}}$, and apply the LeakyReLU nonlinearity (Maas et al., 2013). Fully expanded out, Eq. (3) may then be expressed as:

$$e_{u,v}^{k} = \frac{\exp\left(\text{ LeakyReLU }\left(\mathbf{a}^{T}\left[\mathbf{W}_{\mathbf{h},k}^{l}\mathbf{h}_{u}^{l} \| \mathbf{W}_{\mathbf{h},k}^{l}\Phi(\mathbf{h}_{v}^{l}, \mathbf{r}_{\mathcal{T}(u,v)}^{l})\right]\right)\right)}{\sum_{m\in\mathcal{N}(u)}\exp\left(\text{LeakyReLU}\left(\mathbf{a}^{T}\left[\mathbf{W}_{\mathbf{h},k}^{l}\vec{h}_{u}^{l} \| \mathbf{W}_{\mathbf{h},k}^{l}\Phi(\mathbf{h}_{m}^{l}, \mathbf{r}_{\mathcal{T}(i,m)}^{l})\right]\right)\right)}, \tag{4}$$

where $\|$ is the concatenation operation.

The attention coefficient computed by Eq.(3) indicates the importance of SME u's representation to that of SME v. To make coefficients easily comparable across different neighbors, the attention coefficients of SME u is normalized across all its neighbors j using the softmax function:

$$\alpha_{u,v}^{k} = \text{softmax}_{v}\left(e_{u,v}^{k}\right) = \frac{\exp\left(e_{u,v}^{k}\right)}{\sum_{m\in\mathcal{N}(u)}\exp\left(e_{u,m}^{k}\right)}. \tag{5}$$

With these normalized attention coefficients, M-RIGHT first calculates the representation for every SME u under the k-th head attention, then concatenates the representations from K independent attention mechanisms to obtains the output representation of each SME. In other words, the output representation for every SME u is calculated in Eq.(6), which is used to substitute Eq.(1) under the multi-head attention setting as follows:

$$\mathbf{h}_{u}^{l+1} = \|_{k=1}^{K} f\left(\sum_{v\in\mathcal{N}(u)}\alpha_{u,v}^{k}\mathbf{W}_{\mathbf{h},k}^{l}\Phi(\mathbf{h}_{v}^{l}, \mathbf{r}_{\mathcal{T}(u,v)}^{l}, \mathbf{h}_{u}^{l}),\right), \tag{6}$$

where $\|$ is the concatenation operation.

In the last layer, we substitute the concatenation operation by an averaging operation as follows:

$$\mathbf{h}_{u}^{l+1} = \frac{1}{K}\sum_{k=1}^{K} f\left(\sum_{v\in\mathcal{N}(u)}\alpha_{u,v}^{k}\mathbf{W}_{\mathbf{h},k}^{l}\Phi(\mathbf{h}_{v}^{l}, \mathbf{r}_{\mathcal{T}(u,v)}^{l})\right). \tag{7}$$

Such an attention mechanism is efficient because it is parallelizable across node neighbor pairs. Moreover, the model is directly applicable to inductive learning problems, including tasks where the model has to generalize to completely unseen nodes.

Note that our transfer heterogeneity learning module follows the message passing framework in most graph neural networks (Hamilton et al., 2017; Velickovic et al., 2018; Vashishth et al., 2020a), in which our terms SME and relations correspond to the terms node and edge, respectively, in traditional graph neural networks. The main novelty of our proposed transfer heterogeneity learning process over traditional process, e.g., the attention based message passing process in GAT, is that we introduce an effective entity-relation composition operator to consider transfer heterogeneity along different relation types in calculating both the neighbours' attentions and the transferred messages.

4.4 Behaviour Heterogeneity Learning Module

In this module, M-RIGHT first obtains heterogeneous representations of each SME under different relations, which correspond to SME's heterogeneous behaviours in different environments, then reads out the graph's structure, i.e., calculates margin-based ranking loss of each triplet pair, for model's update.

To enable SME's heterogeneous representations, we introduce a translation mechanism on relational hyperplanes. Specifically, under each relation type r, this mechanism utilizes a vector $\mathbf{w}_r$ to project the representation $\mathbf{h}_s$ of each SME s into a hyperplane and obtain the SME's representation as

$$\mathbf{h}_{s\perp r} = \mathbf{h}_s - \mathbf{w}_r^{\top}\mathbf{h}_s\mathbf{w}_r \tag{8}$$

Then, the score of each triples (s, r, o) can be calculated in Eq. (9) as follows:

$$f_r(\mathbf{h}_s, \mathbf{h}_o) = \left\|(\mathbf{h}_s - \mathbf{w}_r^{\top}\mathbf{h}_s\mathbf{w}_r) + \mathbf{r}_r - (\mathbf{h}_o - \mathbf{w}_r^{\top}\mathbf{h}_o\mathbf{w}_r)\right\|_2^2. \tag{9}$$

In other words, when calculating the triplets, the translation mechanism on relational hyperplanes enables each SME to have distinguishable representations under different relation types, which avoids collapsing SMEs' representations to be the same.

With this scoring function and treating in-relation and out-relation separately, Φ in each layer l in the transfer heterogeneity learning process can be instantiated as follows:

$$\Phi(\mathbf{h}_s, \mathbf{r}_r, \mathbf{h}_o) = \begin{cases} \mathbf{h}_{s\perp r} + \mathbf{r}_r + \mathbf{h}_o - \mathbf{h}_{o\perp r}, (s, r, o) \in \triangle_{\text{in}} \\ \mathbf{h}_{s\perp r} - \mathbf{r}_r + \mathbf{h}_o - \mathbf{h}_{o\perp r}, (s, r, o) \in \triangle_{\text{out}} \end{cases} \tag{10}$$

Algorithm 1 M-RIGHT's representation learning process

Input: SME graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \triangle, \mathbf{H}, \mathbf{R})$; depth L; number of attention head K; neighborhood function $\mathcal{N} : h \rightarrow 2^{\mathcal{V}}$.

Output: Final representations of SMEs $\{\mathbf{z}_i, \forall i \in \mathcal{V}\}$.

while *not converged* **do**

 # Embedding of each layer

 for $l = 0, 2, ..., L$ **do**

  # Attention under each head

  for $k = 1, 2, ..., K$ **do**

   # Representation of each SME

   for $u \in \mathcal{V}$ **do**

    # Attention of each neighbor

    for $v \in \mathcal{N}(u)$ **do**

     $e^k_{u,v} = a(\mathbf{W}^l_{\mathbf{h},k}\mathbf{h}^l_u, \mathbf{W}^l_{\mathbf{h},k}\Phi(\mathbf{h}^l_v, \mathbf{r}^l_{\mathcal{T}(i,j)}))$.

    for $v \in \mathcal{N}(u)$ **do**

     $\alpha^k_{u,v} = \mathrm{softmax}_v(e^k_{u,v})$.

  $\mathbf{h}^{l+1}_u = \|^K_{k=1} f\left(\sum_{v\in\mathcal{N}(u)} \alpha^k_{u,v}\mathbf{W}^l_{\mathbf{h},k}\Phi(\mathbf{h}^l_v, \mathbf{r}^l_{\mathcal{T}(u,v)}), \mathbf{h}^l_u\right)$, $\mathbf{r}^{l+1}_i = f\left(\mathbf{W}^l_{\mathbf{r}}\mathbf{r}^l_i\right)$, $\mathbf{r}^{l+1}_i = f\left(\mathbf{W}^l_{\mathbf{r}}\mathbf{r}^l_i\right)$.

 $\forall u \in \mathcal{V} : \mathbf{z}_u = \mathbf{h}^{L+1}_u$.

 # Backpropagation with loss

 $\mathcal{L} = \sum_{\substack{(s,r,o)\in\triangle \\ (s',r',o')\in\triangle'}} [f_r(\mathbf{z}_s, \mathbf{z}_o), +\gamma - f_{r'}(\mathbf{z}'_s, \mathbf{z}'_o)]_+ + C\left\{\sum_{v\in\mathcal{V}}\left[\|\mathbf{z}_v\|^2_2 - 1\right]_+ + \sum_{r\in\mathcal{R}}\left[\frac{\left(\mathbf{w}^\top_r\mathbf{r}^{L+1}_r\right)^2}{\left\|\mathbf{r}^{L+1}_r\right\|^2_2} - \epsilon^2\right]_+\right\}$.

where $\triangle_{\text{in}}$ and $\triangle_{\text{out}}$ denote the in-relational triplet and out-relational triplets, respectively. In this entity-relation composition operator, the first two terms is equivalent to passing heterogeneous messages on the relational hyperplane, and the last two terms is equivalent to projecting the aggregated messages back to the original space.

In addition, in each of the l-layer in the transfer heterogeneity learning module, each relational hyperplanes projection vectors $\mathbf{w}_r$ will be transformed by the matrix $\mathbf{W}^l_{\mathbf{w}} \in \mathbb{R}^{d_l\times d_{l+1}}$ as follows:

$$\mathbf{w}^{l+1}_r = f\left(\mathbf{W}^l_{\mathbf{w}}\mathbf{w}^l_r\right). \tag{11}$$

The score is expected to be higher for a ground-truth triplet $\triangle$ and lower for an generated fake triplet $\triangle'$. To maximize the discriminations between ground-truth triplets and the generated fake triplets, we use the following margin-based ranking loss:

$$\mathcal{L} = \sum_{\substack{(s,r,o)\in\triangle \\ (s',r',o')\in\triangle'}} \left[f_r(\mathbf{z}_s, \mathbf{z}_o) + \gamma - f_{r'}(\mathbf{z}'_s, \mathbf{z}'_o)\right]_+, \tag{12}$$

where $\mathbf{z}_i$ denotes the SME i's representation after L graph convolution layers, and γ is the margin separating positive and negative triplets. To guarantee that the output representation of each relation type r from the graph convolution network, i.e., $\mathbf{r}^{L+1}_r$, is in the relational hyperplane and is regularized, the following constraints are considered when we minimize the loss $\mathcal{L}$:

$$\begin{aligned}
&\forall r \in \mathcal{E}, \left|\mathbf{w}^\top_r\mathbf{r}^{L+1}_r\right| / \left\|\mathbf{r}^{L+1}_r\right\|_2 \le \epsilon,\\
&\forall r \in \mathcal{E}, \|\mathbf{w}_r\|_2 = 1,\\
&\forall v \in \mathcal{V}, \|\mathbf{z}_v\|_2 \le 1,
\end{aligned} \tag{13}$$

where ϵ is an error vector to ensure orthogonality. With these constraints, Eq.(12) can be re-written as:

$$\begin{aligned}
\mathcal{L} = &\sum_{\substack{(s,r,o)\in\triangle \\ (s',r',o')\in\triangle'}} \left[f_r(\mathbf{z}_s, \mathbf{z}_o), +\gamma - f_{r'}\left(\mathbf{z}'_s, \mathbf{z}'_o\right)\right]_+\\
&+ C\left\{\sum_{v\in\mathcal{V}}\left[\|\mathbf{z}_v\|^2_2 - 1\right]_+ + \sum_{r\in\mathcal{R}}\left[\frac{\left(\mathbf{w}^\top_r\mathbf{r}^{L+1}_r\right)^2}{\left\|\mathbf{r}^{L+1}_r\right\|^2_2} - \epsilon^2\right]_+\right\}.
\end{aligned} \tag{14}$$

We adopt stochastic gradient descent (SGD) to minimize the above loss function, with which M-RIGHT's parameters, including the shared linear transformation metric $\mathbf{W}^l_{\mathbf{h},k}$, $\mathbf{W}^l_{\mathbf{r}}$ and $\mathbf{W}^l_{\mathbf{w}}$ in each l-th layer, the weight vector $\mathbf{a}$, and the hyperplane's norm vector $\mathbf{w}_r$ of each relation r, can be updated. The set of ground-truth triplets are randomly traversed multiple times. When a ground-truth triplet is visited, a fake triplet with the same nodes and a randomly selected fake relation is constructed based on the self-adversarial negative sampling mechanism (Sun et al., 2019).

Note that compared with most graph neural networks, e.g., GAT, that directly utilize the node representations after message passing for the node classification or link prediction task, our proposed behaviour heterogeneity learning module is equivalent to an additional step for further optimizing node representations before performing downstream tasks. Such an additional step is not only beneficial for modeling the heterogeneity of node representations, but also increases the flexibility of M-RIGHT in performing various downstream tasks because it is trained in a self-supervised manner that is not limited to a specific task.

Overall, M-RIGHT's representation learning process is shown in Algorithm 1.

4.5 Analysis on M-RIGHT

Here we give the analysis of our proposed M-RIGHT as follows:

- The proposed M-RIGHT is able to effectively address the transfer heterogeneity by calculating the neighbours' attentions and the transferred messages based on our devised entity-relation composition operator, which distinguishes heterogeneously transferred messages from different relation types.

- The proposed M-RIGHT is able to address the behaviour heterogeneity based on the translation mechanism on relational hyperplanes, which enables each SME to have distinguishable representations under different relation types.

- The modeling of SMEs and various relation types is based on shared parameters, which alleviates the inefficient problem of over-parameterization while applying graph convolution networks on relational graphs and allows any available features as initial representations.

5 Experiments

In this section, we verify the effectiveness of our proposed M-RIGHT on the FNE task in MYbank. We start by introducing the task settings in MYbank and the experimental settings in this paper. Then, we report experimental results with comparisons to some related and state-of-the-art methods. The comprehensive experiments are conducted on two large-scale real-world datasets to answer the following questions:

- Q1: How does M-RIGHT perform on the FNE task compared with state-of-the-art methods?

- Q2: How do the SMEs' relations contribute to M-RIGHT's performance on the FNE task?

- Q3: How do the two key modules, transfer heterogeneity learning module and behavior heterogeneity learning module, contribute to M-RIGHT's performance?

5.1 Experiment Settings

5.1.1 FNE scenario in MYbank

MYbank is an online financial platform that provides financial loan services to SMEs. Because SMEs' credits have been validated as a separate task, the main focus of the FNE task in MYbank is to target those financially constrained SMEs among the credit-validated SMEs. Then, MYbank launches some marketing campaigns on those targeted SMEs, such as presenting a loan product on the owner's homepage of the application program (APP) or sending short messages (SMS) to those SMEs' owners. Then, the feedback of whether the targeted SMEs have bought a loan product afterward, are collected as offline datasets, in which the SMEs that bought a loan product are regarded as financially constraint SMEs. In a word, two datasets, i.e., the APP dataset and the SMS dataset, are available to verify the effectiveness of our proposed method.

5.1.2 Datasets descriptions

In this section, we provide detailed descriptions of the two datasets, i.e., APP and SMS, each of which contains an SME graph and the financing need labels of SMEs. Each SME graph contains 27 relation types. The initial node features are attributes and

Table 2 Datasets descriptions.

dataset	# SMEs (nodes)	# relations (edges)	# relation types
APP	42.45 million	1.26 billion	27
SMS	62.92 million	1.54 billion	27

dataset	# node features	# positive SMEs
APP	1,164	3.89 million
SMS	885	5.68 million

historical behaviors of SMEs, whereas the initial relation features are randomly generated, which are used as input of the graph representation learning model. The detailed information of the two datasets is presented in Table 2. Note that the data used in the experiments is sampled from the server's logs, and we have performed data desensitization to ensure that any private and sensitive SMEs' information cannot be extracted from this data.

Each dataset is organized chronologically, in which the earlier 85% data is used for training and validation while the later 15% data is used for testing. In this way, we can guarantee that the data from the training set and the validation set is ahead of the test set, which ensures predictions are on the future.

5.1.3 Evaluation metrics

We evaluate the performance of different FNE methods using three metrics, i.e., classification accuracy (CA), micro-averaged F1 scores (micro-F1), and ROC-AUC (AUC), which are widely used in graph representation learning studies (Velickovic et al., 2018; Xu et al., 2019) and financial studies (Yang et al., 2020). Classification accuracy summarizes the performance of a classification model as the number of correct predictions divided by the total number of predictions. The micro-averaged F1 score is used to assess the quality of our binary classification problems, which measures the F1-score of the aggregated contributions of the two classes. ROC-AUC is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. In summary, the higher values of the three metrics indicate better performance.

5.1.4 Comparison methods

The comparison methods can be categorized into two groups:

(1) Graph-free methods:

- **ANOVA-XGBoost** (Chen et al., 2020) utilizes the extreme gradient boosted tree method to perform the prediction task.
- **SVM-RE** (Rogic and Kascelan, 2019) utilizes a hybrid support vector machine rule extraction method to predict the targeted SMEs.
- **DeepFM** (Guo et al., 2017) combines the power of factorization machines and deep learning to predict financially constrained SMEs, which is a common method in real industrial practice.

(2) Graph-based methods:

- **GIN** (Xu et al., 2019) is a homogeneous graph representation learning method that incorporates the Weisfeiler-Lehman test into the neighborhood aggregation process.
- **GraphSAGE** (Hamilton et al., 2017) is a homogeneous graph representation learning method that generates node embeddings in an inductive manner without a pre-defined graph Laplacian.
- **GAT** (Velickovic et al., 2018) leverages masked self-attentional layers to specify different weights to different nodes in a neighborhood of a homogeneous graph convolution network.
- **RGCN** (Schlichtkrull et al., 2018) first models the message passing under different relation types separately then aggregates the node representations from various message passing paths in a heterogeneous graph.
- **CompGCN** (Vashishth et al., 2020a) leverages a composition operators to jointly learn the representations of nodes and relations in a heterogeneous graph.
- **MHGCN** (Yu et al., 2022) extracts useful heterogeneous meta-paths of different lengths in heterogeneous networks through multi-layer convolution aggregation to enhance the effectiveness of message passing.

Table 3 The performance of all methods on CA, micro-F1, and AUC values (mean ± range, computed across 10 runs).

Categories	Methods	APP			SMS		
		CA	micro-F1	AUC	CA	micro-F1	AUC
Graph-free Methods	ANOVA-XGBoost	0.6688 ± 0.006	0.2303 ± 0.035	0.8585 ± 0.012	0.9780 ± 0.002	0.4094 ± 0.003	0.9306 ± 0.000
	SVM-RE	0.5072 ± 0.005	0.1794 ± 0.026	0.5567 ± 0.022	0.9803 ± 0.002	0.1078 ± 0.001	0.7324 ± 0.000
	DeepFM	0.5821 ± 0.005	0.2041 ± 0.029	0.8344 ± 0.033	0.9769 ± 0.002	0.3668 ± 0.002	0.9234 ± 0.001
Graph-based Methods	GIN	0.6708 ± 0.000	0.2344 ± 0.005	0.8550 ± 0.001	0.9783 ± 0.001	0.4110 ± 0.001	0.9289 ± 0.000
	GraphSage	0.6645 ± 0.001	0.2302 ± 0.004	0.8601 ± 0.000	0.9724 ± 0.001	0.4141 ± 0.002	0.9285 ± 0.001
	GAT	0.6752 ± 0.000	0.2377 ± 0.005	0.8649 ± 0.000	0.9833 ± 0.001	0.4094 ± 0.001	0.9275 ± 0.000
	RGCN	0.6697 ± 0.001	0.2294 ± 0.005	0.8648 ± 0.001	0.9752 ± 0.003	0.4152 ± 0.003 *	0.9336 ± 0.001
	CompGCN	0.6695 ± 0.001	0.2295 ± 0.007	0.8660 ± 0.001	0.9828 ± 0.003	0.4120 ± 0.004	0.9298 ± 0.001
	MHGCN	0.6690 ± 0.007	0.2305 ± 0.011	0.8659 ± 0.002	0.9835 ± 0.001 *	0.4133 ± 0.001	0.9312 ± 0.000
	HRAN	0.6761 ± 0.001 *	0.2395 ± 0.006 *	0.8664 ± 0.004 *	0.9831 ± 0.004	0.4150 ± 0.007	0.9338 ± 0.001 *
Our Proposed Methods	M-RIGHT	0.6906 ± 0.001	0.2456 ± 0.007	0.9006 ± 0.001	0.9841 ± 0	0.4287 ± 0.003	0.9469 ± 0.001
	M-RIGHT-w/o-rt	0.6760 ± 0.001	0.2382 ± 0.009	0.8788 ± 0.001	0.9790 ± 0.002	0.4158 ± 0.003	0.9339 ± 0
	M-RIGHT-w/o-rs	0.6743 ± 0.001	0.2334 ± 0.007	0.8737 ± 0.001	0.9830 ± 0.003	0.4181 ± 0.002	0.9368 ± 0
Improvement (%) [1]		2.1447	2.5470	3.9474	0.0610	3.2514	1.4029
p-value		0.000	0.004	0.000	0.001	0.000	0.000

[1] Improvement of M-RIGHT over the best-performing comparison methods.
[2] Statistically not different from the best-performing comparison methods if p-value < 0.05 (p-value with paired t-test).
* Results of the best-performing comparison methods.

- **HRAN** (Li et al., 2022) is a deep neural network based relation-learning methods that fuses and attends each type of semantic-specific information through the relation-path in the heterogeneous graphs.

5.1.5 Implementation details

We implement all graph representation learning methods in TensorFlow with the Adam optimizer (Kingma and Ba, 2015), and set the hyperparameters according to the best results in the validation set. All the graph convolution networks in the comparison models involve two convolution layers, each of which computes 128 features ($d_1 = d_2 = 128$) as output, followed by an exponential linear unit nonlinearity. For GAT and M-RIGHT, the attention head is set to be 2, each of which computes 64 features. All methods are trained on a cluster of 15 Dual-CPU servers with AGL framework (Zhang et al., 2020a). We use early stopping with a patience of 100, i.e. we stop training if the loss does not decrease for 100 consecutive epochs. Results on all experiments are averaged over 10 runs.

5.2 Experiments Results and Analysis

5.2.1 Evaluation on FNE (for Q1 and Q2)

To answer **Q1**, we compare M-RIGHT with all comparison methods. Among the comparison methods, the graph-free methods utilize the initial graph-free SMEs' features as inputs to predict the financially constrained SMEs. M-RIGHT utilizes the initial graph-free features as input to train its graph model and output the 128-dimensional SMEs' representations, which are then concatenated with the initial graph-free SMEs' features as the inputs of the ANOVA-XGBoost to predict the SMEs' financing needs. To better investigate the usefulness of the graph-based methods in exploiting relations and ensure a fair comparison in answering **Q2**, we process all the comparison graph-based methods similarly to M-RIGHT. In other words, we obtain the SMEs' representations from each graph-based method, which are then concatenated with the initial graph-free SMEs' features as the inputs to the ANOVA-XGBoost for calculating the final classification results. The results are reported in Table 3.

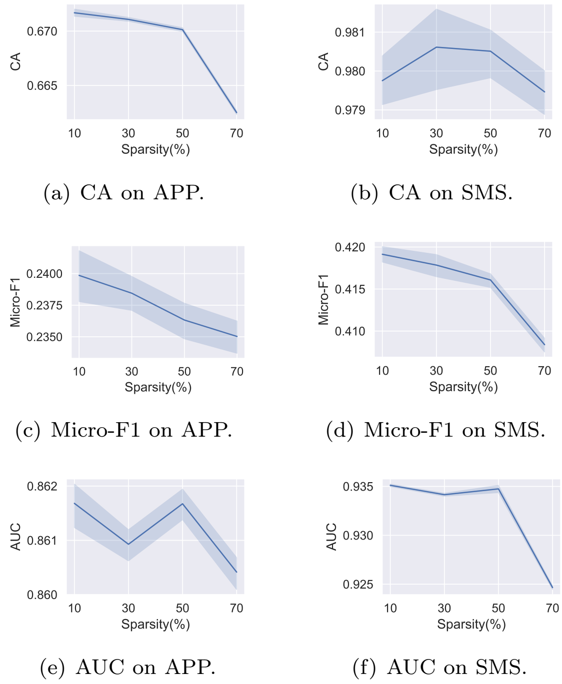
Results & Analysis: In Table 3, M-RIGHT outperforms all the comparison methods in terms of CA, micro-F1, and AUC values, with an average improvement of 1.10%, 2.90% and 2.69%, respectively, over the best-performing comparison methods, which demonstrates the effectiveness of M-RIGHT in the FNE task. On the one hand, M-RIGHT outperforms the comparison graph-free methods because M-RIGHT is able to consider the SMEs' relations among SMEs. Note that all the graph-based methods outperform the graph-free methods with respect to all three metrics on APP and SMS, respectively. Such improvements demonstrate the importance of modeling SMEs' relations among SMEs in the FNE task.

On the other hand, M-RIGHT outperforms the comparison graph-based methods. M-RIGHT outperforms the homogeneous graph representation learning methods, i.e., GIN, GraphSage, and GAT, because M-RIGHT can address the transfer heterogeneity with our novel entity-relation composition operator. M-RIGHT outperforms the heterogeneous graph representation learning methods, i.e., RGCN, CompGCN, MHGCN, and HRAN, because M-RIGHT can address the hebavior heterogeneity based on the translation mechanism on relational hyperplanes. These results not only demonstrate that the relation modeling is significant but also indicate that M-RIGHT's relation modeling mechanism is superior to the state-of-the-art graph-based methods in the FNE task.

5.2.2 Evaluation under different relation sparsity (for Q2)

To investigate how the relation information among SMEs affects M-RIGHT performance, we evaluate M-RIGHT's performance on CA, micro-F1, and AUC values under different relation sparsities. Specifically, we randomly filter each training dataset into four sub-datasets based on the different degrees of relation sparsity. For example, a sparsity of 10% suggests that 10% of the relations are filtered out in the training dataset. The performance of M-RIGHT under different data sparsities is presented in Fig. 5. Furthermore, to investigate the properties of modeling relations, we calculate the following values in the two datasets: (1) relation densities, i.e., the number of actual relations among SMEs divided by the maximal number of relations in a pseudo fully connected graph. This is calculated according to statistics in Table 2; (2) improvements of modeling relations, i.e., the improvements of graph-based methods to the graph-free methods; (3) M-RIGHT's degradation of missing relations, i.e., the degradation of M-RIGHT's performance under the 10% sparsity to the 70% sparsity. We present the results in Table 4.

Results & Analysis: In Fig. 5, we can make three conclusions. Firstly, M-RIGHT has a degradation in performance when some existing relations are unavailable. Specifically, M-RIGHT's CA, micro-F1, and AUC values decrease with the increase of relation sparsity, which indicates that its performance is affected by the availability of the relations. Without



(a) CA on APP. (b) CA on SMS.
(c) Micro-F1 on APP. (d) Micro-F1 on SMS.
(e) AUC on APP. (f) AUC on SMS.

Figure 5 The performance of M-RIGHT under different dataset sparsities (mean ± range, computed across 10 runs).

sufficient information on the relations among SMEs, M-RIGHT is unable to capture the external source of SMEs' financing needs, i.e., the financing needs transferred from their neighbors, which results in its degradation in performance. Secondly, the performance degradation under more relation-dependent scenarios is even more severe. Specifically, Table 4 shows that SMS contains denser relations, and modeling relations leads to more improvement on SMS than on APP, which indicates that SMEs' financing conditions depend more on relations under the SMS scenario. From the last column in Table 4, we can conclude that lacking relations on more relation-dependent scenarios may bring more severe degradation in performance. Third, M-RIGHT outperforms the graph-free methods even under the most sparse training dataset, demonstrating the importance of relations in transferring neighborhood information.

5.2.3 Ablation experiments (for Q3)

To investigate how the two important mechanisms in M-RIGHT contributes to its performance, we evaluate the performance of two simplified versions of M-RIGHT: (1) **M-RIGHT-w/o-rt** refers to M-RIGHT without considering transfer heterogeneity, namely, the entity-relation composition operation is discarded and the representations of a node's neighbors in the message passing process does not consider the relation types; (2) **M-RIGHT-w/o-rf** refers to M-RIGHT without considering behavior heterogeneity, namely, Eq. (9) is replaced with $f_r(\mathbf{h}_s, \mathbf{h}_o) = \|\mathbf{h}_s + \mathbf{r}_r - \mathbf{h}_o\|_2^2$ and Eq. (10) is replaced similarly. The results are shown in Table 3.

Results & Analysis: In Table 3, M-RIGHT outperforms M-RIGHT-w/o-rt on all three metrics with average improvements of 2.58% and 1.67% on the APP and SMS datasets respectively. Without considering the transfer heterogeneity when aggregating the nodes' embedding in the graph convolution network, M-RIGHT is unable to distinguish the differences in financing needs transferred under different relation types, which leads to a decrease in performance. Comparing M-RIGHT with M-RIGHT-w/o-rs, M-RIGHT achieves the higher values on all three metrics with average improvements of 3.57% and 1.24% on the APP and SMS datasets, respectively. Without considering the behavior heterogeneity, M-RIGHT is unable to distinguish SMEs' roles under different relation types, which leads to a decrease in performance.

5.2.4 Case Study

To intuitively demonstrate the capabilities of M-RIGHT, we conduct a case study of its predicted results on six randomly selected SMEs in the APP dataset. In Table 5, given a specified pair of a head SME s and a tail SME o, we present the top 4 relations with M-RIGHT's highest predicted triplet scores $f_r(\mathbf{h}_s, \mathbf{h}_o)$ out of a total of 27 relations and the predicted FNE label of each SME, in which the bold results mean that the predicted results hit the ground-truth results. From Table 5, we can draw two conclusions. Firstly, M-RIGHT is able to assign higher scores to ground-truth triplets and exploit relations between SMEs correctly. Such effectiveness in predicting triplet scores relies highly on M-RIGHT's accuracy in learning not only the heterogeneous relation representations but also SME's heterogeneous representations on different relational hyperplanes. Secondly, M-RIGHT is accurate in predicting SMEs' financing needs labels, which indicates that the learned SMEs' representations according to Algorithm 1 can well facilitate the FNE task.

Table 4 The relation densities, the improvements of modeling relations, and degradation of missing relations on two datasets.

Dataset	Relation Density	Improvement of Modeling Relations	Degradation of Missing Relations
APP	30.07 %	15.11 %	0.69 %
SMS	61.10 %	16.37 %	1.21 %

Table 5 Case study on six SMEs from two industries.

SME pairs	Top 4 relations	Predicted labels
s: fabric purchase SME 1 o: fiber preprocess SME 2	(1) money transfer (2) **upstream** (3) purchase (4) parent	SME 1: $\mathbf{y} = \mathbf{1}$ SME 2: $\mathbf{y} = \mathbf{1}$ SME 3: $\mathbf{y} = \mathbf{0}$
s: closthing design SME 3 o: fiber preprocess SME 2	(1) **subsidiary** (2) **share holder** (3) client (4) upstream	
s: technology SME 4 o: electronic SME 5	(1) **subsidiary** (2) invoice (3) **share holder** (4) stock gaurantor	SME 4: $\mathbf{y} = \mathbf{1}$ SME 5: $\mathbf{y} = \mathbf{1}$ SME 6: $\mathbf{y} = \mathbf{0}$
s: server manufactor SME 6 o: electronic SME 5	(1) invoice (2) **upstream** (3) joint venture (4) contrast	

Interestingly, we observe that even though SME 2 are connected with SME 1 and SME 3, SME 2' predicted label is the same as that of SME1 whereas different from that of SME 3. By analyzing SME 2's projection representations on relations 'subsidiary' (denoted as r_1), 'upstream' (denoted as r_2), and 'share holder' (denoted as r_3), we find that $||\mathbf{h}_{\text{SME2}\perp r_2}||_2 > ||\mathbf{h}_{\text{SME2}\perp r_1}||_2 > ||\mathbf{h}_{\text{SME2}\perp r_3}||_2$, which means that SME 2 has larger sub-component on r_2's hyperplane, thereby more affected by the message from r_2, i.e., the message from SME 1, compared to messages from SME3. This may explain why the prediction on SME 2 is the same as that of SME1. Similar phenomenon can be observed on SME 4, SME 5, and SME 6, and similar conclusion can be drawn that SME5 is more affected by the message from relation type 'subsidiary' than those from 'upstream' and 'share holder'. This interesting observation indicates that each SME may have different 'sensibility' under different relations. Therefore, allowing SME's heterogeneous representations under different relation types can facilitate effective representation learning of SMEs.

6 Conclusions and Future Work

In this paper, we have conducted exploratory analysis on an industrial dataset, which indicates the importance of modeling SMEs' relations with two kinds of heterogeneity in the FNE task. Then, we have proposed a novel method named M-RIGHT, whose main novelty is that it simultaneously addresses two challenging heterogeneity, i.e., transfer heterogeneity and behavior heterogeneity, in modeling SME graph with multiple relations. Specifically, to address the transfer heterogeneity, M-RIGHT leverages a novel entity-relation composition operator in the neighborhood message passing process, which distinguishes heterogeneous transferred messages under different relation types. To address the behavior heterogeneity, M-RIGHT enables heterogeneous representations of each SME under different relation types, which correspond to SME's heterogeneous behaviours in different environments, by performing SMEs representation translations on each relational hyperplane. Comprehensive experiments on two real-world datasets have demonstrated the superiority of M-RIGHT compared with state-of-the-art methods in exploring financially constrained SMEs.

In the future, we intend to extend our work in two potential directions. Firstly, given the fact that the relations among SMEs are changing dynamically, we plan to incorporate temporal factors and mechanisms into M-RIGHT such that it can catch up with the changes. Secondly, considering that SMEs can be clustered into several groups and SMEs in different groups may have different behavior patterns, we plan to introduce such cluster information into the relational learning process in the future.

7 Acknowledgments

This work was supported by the National Key R&D Program of China (No. 2018YFB1403001).

References

Angilella S, Mazzù S, 2015. The financing of innovative smes: A multicriteria credit rating model. *European Journal of Operational Research*, 244(2):540-554. https://doi.org/10.1016/j.ejor.2015.01.033

Bordes A, Usunier N, García-Durán A, et al., 2013. Translating embeddings for modeling multi-relational data. 27th Annual Conference on Neural Information Processing Systems, p.2787-2795.

Ceptureanu EG, Ceptureanu S, Herteliu C, 2019. Evidence regarding external financing in manufacturing mses using partial least squares regression. *Annals of Operations Research*, 299(1):1-14. https://doi.org/10.1007/s10479-019-03291-2

Chen X, Zhao Y, Wei Z, et al., 2020. Intelligent identification of potential customers for electricity substitution. Proceedings of Fuzzy Systems and Data Mining, p.168-179. https://doi.org/10.3233/FAIA200697

Cheng HT, Koc L, Harmsen J, et al., 2016. Wide & deep learning for recommender systems. Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, p.7-10. https://doi.org/10.1145/2988450.2988454

Dettmers T, Minervini P, Stenetorp P, et al., 2018. Convolutional 2d knowledge graph embeddings. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1):1811-1818.

Dong Y, Chawla NV, Swami A, 2017. metapath2vec: Scalable representation learning for heterogeneous networks. Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, p.135-144. https://doi.org/10.1145/3097983.3098036

Duan G, Ma X, 2018. A coupon usage prediction algorithm based on xgboost. 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, p.178-183. https://doi.org/10.1109/FSKD.2018.8686924

Fu X, Zhang J, Meng Z, et al., 2020. Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding. Proceedings of The Web Conference 2020, p.2331-2341. https://doi.org/10.1145/3366423.3380297

Graesch JP, Hensel-Börner S, Henseler J, 2021. Information technology and marketing: an important partnership for decades. *Industrial Management and Data Systems*, 121(1):123-157. https://doi.org/10.1108/IMDS-08-2020-0510

Grover A, Leskovec J, 2016. node2vec: Scalable feature learning for networks. Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, p.855-864. https://doi.org/10.1145/2939672.2939754

Guo H, Tang R, Ye Y, et al., 2017. Deepfm: A factorization-machine based neural network for CTR prediction. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, p.1725-1731. https://doi.org/10.24963/ijcai.2017/239

Hamilton WL, Ying Z, Leskovec J, 2017. Inductive representation learning on large graphs. 31th Annual Conference on Neural Information Processing Systems, p.1024-1034.

Jeon H, 2021. Investment and financing decisions in the presence of time-to-build. *European Journal of Operational Research*, 288(3):1068-1084. https://doi.org/10.1016/j.ejor.2020.06.034

Ji S, Pan S, Cambria E, et al., 2021. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):494-514. https://doi.org/10.1109/TNNLS.2021.3070843

Kingma DP, Ba J, 2015. Adam: A method for stochastic optimization. 3rd International Conference on Learning Representations, p.1-10.

Kshetri N, 2016. Big data's role in expanding access to financial services in china. *International Journal of Information Management*, 36(3):297-308. https://doi.org/10.1016/j.ijinfomgt.2015.11.014

Lessmann S, Haupt J, Coussement K, et al., 2021. Targeting customers for profit: An ensemble learning framework to support marketing decision-making. *Information Sciences*, 557:286-301.

Li Z, Liu H, Zhang Z, et al., 2022. Learning knowledge graph embedding with heterogeneous relation attention networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33(8):3961-3973.

Liao H, Hu J, Li T, et al., 2022. Deep linear graph attention model for attributed graph clustering. *Knowledge-Based Systems*, 246:108665.

Luo Y, Deng T, Wei Q, et al., 2021. Optimal financing decision in a contract food supply chain with capital constraint. *Complexity*, 2021(1):8925102:1-8925102:17.

Maas AL, Hannun AY, Ng AY, et al., 2013. Rectifier nonlinearities improve neural network acoustic models. Proceedings of International Conference on Machine Learning, 30(1):1-6.

Mikolov T, Chen K, Corrado G, et al., 2013. Efficient estimation of word representations in vector space. 1st International Conference on Learning Representations, p.1-10.

Nickel M, Rosasco L, Poggio T, 2016. Holographic embeddings of knowledge graphs. Proceedings of the AAAI Conference on Artificial Intelligence, 30(1).

Perozzi B, Al-Rfou R, Skiena S, 2014. Deepwalk: Online learning of social representations. Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, p.701-710.

Rogic S, Kascelan L, 2019. Customer value prediction in direct marketing using hybrid support vector machine rule extraction method. European Conference on Advances in Databases and Information Systems, p.283-294. https://doi.org/10.1007/978-3-030-30278-8_30

Rogić S, Kašćelan L, Pejić Bach M, 2022. Customer response model in direct marketing: Solving the problem of unbalanced dataset with a balanced support vector machine. *Journal of Theoretical and Applied Electronic Commerce Research*, 17(3):1003-1018.

Sadeghian A, Armandpour M, Colas A, et al., 2021. Chronor: rotation based temporal knowledge graph embedding. Proceedings of the AAAI Conference on Artificial Intelligence, 35(7):6471-6479.

Schlichtkrull MS, Kipf TN, Bloem P, et al., 2018. Modeling relational data with graph convolutional networks. 15th International Extended Semantic Web Conference, p.593-607. https://doi.org/10.1007/978-3-319-93417-4_38

Shang C, Tang Y, Huang J, et al., 2019. End-to-end structure-aware convolutional networks for knowledge base completion. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01):3060-3067.

Shi Y, Gui H, Zhu Q, et al., 2018a. Aspem: Embedding learning by aspects in heterogeneous information networks. Proceedings of the 2018 SIAM international conference on data mining, p.144-152. https://doi.org/10.1137/1.9781611975321.16

Shi Y, Zhu Q, Guo F, et al., 2018b. Easing embedding learning by comprehensive transcription of heterogeneous information networks. Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, p.2190-2199. https://doi.org/10.1145/3219819.3220006

Sun Z, Deng Z, Nie J, et al., 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019, p.1-10.

Tang J, Qu M, Mei Q, 2015. Pte: Predictive text embedding through large-scale heterogeneous text networks. Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining, p.1165-1174. https://doi.org/10.1145/2783258.2783307

Tian Z, Hassan A, Razak N, 2018. Big data and sme financing in china. Journal of Physics: Conference Series, 1018(1):012002.

Trouillon T, Welbl J, Riedel S, et al., 2016. Complex embeddings for simple link prediction. International conference on machine learning, p.2071-2080.

Vashishth S, Sanyal S, Nitin V, et al., 2020a. Composition-based multi-relational graph convolutional networks. 8th International Conference on Learning Representations, p.1-10.

Vashishth S, Sanyal S, Nitin V, et al., 2020b. Interacte: Improving convolution-based knowledge graph embeddings by increasing feature interactions. Proceedings of the AAAI conference on artificial intelligence, 34(03):3009-3016.

Velickovic P, Cucurull G, Casanova A, et al., 2018. Graph attention networks. 6th International Conference on Learning Representations, p.1-10.

Wang X, Zhang Y, Shi C, 2019. Hyperbolic heterogeneous information network embedding. Proceedings of the AAAI conference on artificial intelligence, 33(01):5337-5344.

Wang Y, Jing C, Xu S, et al., 2022. Attention based spatiotemporal graph attention networks for traffic flow forecasting. *Information Sciences*, 607(1):869-883. https://doi.org/10.1016/j.ins.2022.05.127

Wang Z, Zhang J, Feng J, et al., 2014. Knowledge graph embedding by translating on hyperplanes. Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, p.1112-1119.

Welling M, Kipf TN, 2017. Semi-supervised classification with graph convolutional networks. J International Conference on Learning Representations, p.1-10.

Wu Z, Pan S, Chen F, et al., 2020. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4-24. https://doi.org/10.1109/TNNLS.2020.2978386

Xu K, Hu W, Leskovec J, et al., 2019. How powerful are graph neural networks? 7th International Conference on Learning Representations, p.1-10.

Xu Z, Meisami A, Tewari A, 2021. Decision making problems with funnel structure: A multi-task learning approach with application to email marketing campaigns. International Conference on Artificial Intelligence and Statistics, p.127-135.

Yang B, Yih SWt, He X, et al., 2015. Embedding entities and relations for learning and inference in knowledge bases. Proceedings of the International Conference on Learning Representations (ICLR) 2015, p.1-10.

Yang C, Xiao Y, Zhang Y, et al., 2022. Heterogeneous network representation learning: A unified framework with survey and benchmark. *IEEE Transactions on Knowledge and Data Engineering*, 34(10):4854-4873. https://doi.org/10.1109/TKDE.2020.3045924

Yang S, Zhang Z, Zhou J, et al., 2020. Financial risk analysis for smes with graph-based supply chain mining. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, p.4661-4667. https://doi.org/10.24963/ijcai.2020/643

Ye R, Li X, Fang Y, et al., 2019. A vectorized relational graph convolutional network for multi-relational network alignment. Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, p.4135-4141. https://doi.org/10.24963/ijcai.2019/574

Yu P, Fu C, Yu Y, et al., 2022. Multiplex heterogeneous graph convolutional network. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, p.2377-2387. https://doi.org/10.1145/3534678.3539482

Zhang B, Wang L, Li Y, 2021. Precision marketing method of e-commerce platform based on clustering algorithm. *Complexity*, 2021:1-10.

Zhang D, Huang X, Liu Z, et al., 2020a. AGL: A scalable system for industrial-purpose graph machine learning. *Proceedings of the VLDB Endowment*, 13(12):3125-3137. https://doi.org/10.14778/3415478.3415539

Zhang W, Fang Y, Liu Z, et al., 2020b. mg2vec: Learning relationship-preserving heterogeneous graph representations via metagraph embedding. *IEEE Transactions on Knowledge and Data Engineering*, 34(03):1-14. https://doi.org/10.1109/TKDE.2020.2992500

Zhao J, Wang X, Shi C, et al., 2021. Heterogeneous graph structure learning for graph neural networks. Proceedings of the 35th AAAI Conference on Artificial Intelligence, 35(5):4697-4705.