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 small- and medium-sized enterprises based on a 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 and Technology, Zhejiang University, Hangzhou 310000, China*
[2]*MYbank, Ant Group, Hangzhou 310000, China*
[3]*Department of Ant Intelligent Engine Technology, Ant Group, Hangzhou 310000, China*
[4]*School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou 310000, China*
†E-mail: xlzheng@zju.edu.cn

Received Apr. 14, 2022; Revision accepted Dec. 1, 2022; Crosschecked Feb. 24, 2023

Abstract: Financing needs exploration (FNE), which explores financially constrained small- and medium-sized enterprises (SMEs), has become increasingly important in industry for financial institutions to facilitate SMEs' development. In this paper, we first perform an insightful exploratory analysis to exploit the transfer phenomenon of financing needs among SMEs, which motivates us to fully exploit the multi-relation enterprise social network for boosting the effectiveness of FNE. The main challenge lies in modeling two kinds of heterogeneity, i.e., transfer heterogeneity and SMEs' behavior heterogeneity, under different relation types simultaneously. To address these challenges, we propose a graph neural network named Multi-relation tRanslatIonal GrapH aTtention network (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/10.1631/FITEE.2200151 **CLC number:** TP391

1 Introduction

As financial markets become increasingly volatile with the outbreak of crises such as the coronavirus disease-2019 (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 financing needs exploration (FNE). The FNE problem aims to exploit those financially constrained SMEs, which is

‡ Corresponding author
* Project supported in part by the National Natural Science Foundation of China (No. 72192823), the "Ten Thousand Talents Program" of Zhejiang Province for Leading Experts (No. 2021R52001), and the Cooperation Project of MYbank, Ant Group
# Electronic supplementary materials: The online version of this article (https://doi.org/10.1631/FITEE.2200151) contains supplementary materials, which are available to authorized users
ORCID: Qianqiao LIANG, https://orcid.org/0000-0002-6205-988X; Xiaolin ZHENG, https://orcid.org/0000-0001-5483-0366
© Zhejiang University Press 2023

significant for financial institutions to facilitate the development of those struggling SMEs. 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., 2021), which makes it very difficult to address the FNE problem effectively. Therefore, it is of utmost necessity to devise an effective graph-based method that can model the transfer phenomenon of financing needs for accurate FNE.

A motivating example of FNE is given below. An SME's financing need will transfer to its related SMEs, which will have influences on their degree of financing needs. Fig. 1 depicts this phenomenon via a toy example, in which a color bar denotes the degree of an SME's financing need and the colors of edges denote the transfer intensities. In this example, SME A is extremely financially constrained, which will transfer this constraint to its subsidiary SMEs and its upstream SMEs, resulting in increases in the degree of financing needs 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 money to the upstream company if it is financially constrained. An effective FNE method is supposed to consider such a significant transfer factor to improve its performance. More details about the transfer phenomenon of financing needs among SMEs are elaborated in the supplementary materials based on an industrial dataset from MYbank (refers to Zhejiang E-commerce Bank, which is a Chinese company that offers banking services for SMEs).

The above example motivates us to construct an SME 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 a real scenario, the SME graph contains multiple relation types; the above idea faces two kinds of heterogeneity, i.e., transfer heterogeneity and behavior heterogeneity, which makes it very challenging to model the transfer of financing needs among SMEs. On one hand, transfer heterogeneity (CH1) indicates 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 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 financing condition will only partially affect that of SME G, 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) indicates that each SME has different behaviors under different relation types. As shown in Fig. 1, SMEs B, C, and D are all subsidiaries of SME A, and their financing needs are also affected by those of SME A. Intuitively, SMEs B, C, and D should have similar financing need 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 need conditions. Such behavior heterogeneity is likely to originate from the fact that SMEs have different roles under different relations, which complicates an FNE method.

In this paper, to address the two kinds of heterogeneity for facilitating the FNE task, we propose a novel Multi-relation tRanslatIonal GrapH aTtention 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 using the heterogeneous representations of different relation types (for addressing CH1). The behavior heterogeneity learning module first learns the heterogeneous representations of SMEs under different relation types by performing SME representation translations on each relational hyperplane, which helps distinguish each SME's heterogeneous behaviors (for addressing CH2), and then leverages SMEs' representations to predict the graph triplets and compute the corresponding loss for the model's update in a self-supervised learning manner. Finally,

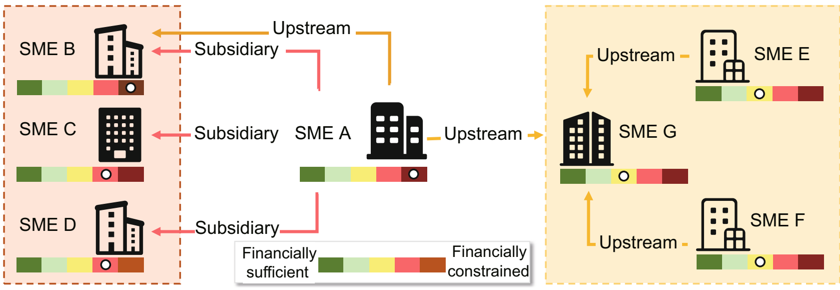

Fig. 1 A motivating example: financing needs transferred from SME A to different SMEs under different relation types (SME: small- and medium-sized enterprise). References to color refer to the online version of this figure

M-RIGHT leverages a tree-based method to predict financially constrained SMEs based on the learned SMEs' representations.

Our main contributions are as follows:

1. We provide in-depth exploratory analyses to exploit the transfer phenomenon of financing needs in the enterprise social network, i.e., an SME graph, which indicates the transfer heterogeneity and behavior heterogeneity in real practice.

2. We propose a novel graph representation learning based method, named M-RIGHT, which effectively models the two kinds of heterogeneity in the transfer phenomenon of financing needs. To the best of our knowledge, this is the first method to model the transfer phenomenon of financing needs in SME graphs, which is beneficial to the FNE task.

3. We conduct comprehensive experiments on two real-world datasets, demonstrating 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. A rule-based method designs various objective functions manually based on empirical rules and uses rating models (Angilella and Mazzù, 2015; Luo et al., 2021) to predict the probability of financial constraint of SMEs. Such methods depend heavily on expert advice, which fail to capture SMEs' financial patterns automatically. To facilitate automatic FNE, some financial institutions have been using machine learning based methods, which model SMEs' financial patterns (Graesch et al., 2021; Jeon, 2021) and explore the financially constrained SMEs (Kshetri, 2016; Tian et al., 2018) based on machine learning algorithms. Although few studies have reported their methods in detail, they claimed that methods for exploring potential customers could be reference solutions. For example, Zhang B et al. (2021) designed a clustering algorithm to target customers in an e-commerce platform. Xu ZP et al. (2021) leveraged the contextual bandit to model the funnel structure in email marketing campaigns. Rogic and Kascelan (2019) and Rogić et al. (2022) used 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. Duan and Ma (2018) and Chen et al. (2020) leveraged an extreme gradient boosting algorithm, XGBoost, to mine the potential customers that require their products. Lessmann et al. (2021) proposed an ensemble learning framework including XGBoost to target potential customers for profit. In addition to the studies on customer targeting, deep learning based models (Cheng et al., 2016; Guo et al., 2017) 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 focus only on SMEs' financing needs induced by their own operational conditions and ignore the transfer phenomenon of financing needs within the enterprise social network. In real scenarios, the financing needs of SMEs are affected by not only independent internal factors, but also the external factors conveyed by the relations between SMEs (Ceptureanu et al., 2021).

2.2 Graph representation learning

Graph representation learning methods are popular for embedding graphical structural data, which can be applied for exploring SME 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 relation types. Early studies are relatively shallow, which first perform random walks (Perozzi et al., 2014; Grover and Leskovec, 2016) on graph data to generate node sequences and 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., 2021) have attracted great attention from researchers, among which the graph convolution network (Kipf and Welling, 2017) has achieved great success. Based on the graph convolution network, Hamilton et al. (2017) designed an inductive graph convolution network, which can efficiently perform message passing between connected nodes. To further improve the effectiveness, Veličković et al. (2018) incorporated attention mechanisms and Xu KYL et al. (2019) incorporated the Weisfeiler–Lehman test into the message-passing process. Attention-based graph neural networks have been proven to be effective in various tasks (Liao et al., 2022; Wang Y 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 of the graph's multiple relation types.

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 C et al., 2022). The first kind of method 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 X et al., 2019), under different relation types (Tang et al., 2015), from different perspectives (Shi et al., 2018a), or in different meta-paths (Zhang WT et al., 2022). 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 model the message passing under different relation types separately and then aggregate the node representations from various message-passing paths. Furthermore, Ye et al. (2019) and Vashishth et al. (2020a) proposed learning node and relation representations simultaneously to improve the modeling of heterogeneous graphs. In addition, Fu et al. (2020), Zhao et al. (2021), and Zhang WT et al. (2022) proposed learning node representations under different meta-paths, where the design of meta-paths is task-specific and highly dependent on expert knowledge. To eliminate the dependency on the meta-paths, Yu et al. (2022) automatically captured the meta-paths in heterogeneous graph neural networks. The disadvantage of these kinds of methods is their lack of knowledge modeling in heterogeneous graphs, such as the knowledge of subsidiaries and upstream SMEs in SME graphs, which is of great importance. The third kind of method is relation-learning method (Ji et al., 2022), which focuses on preserving the knowledge structure, i.e., knowledge triplets, based on different triplet scoring functions and is common for knowledge graph embedding. Traditional relation-learning methods include triple translation methods (Bordes et al., 2013; Wang Z et al., 2014; Sadeghian et al., 2021) and semantics-based triplet-matching methods (Yang BS 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 use 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 relations, which means that it is difficult to guarantee their performances in recognizing the behavior heterogeneity in the FNE task.

In conclusion, existing graph representation learning methods are ineffective in modeling SME graphs because of their inability to address the two kinds of heterogeneity in FNE.

3 Methodology

3.1 Problem settings and preliminaries

Definition 1 (SME graph) An SME graph is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \triangle, \boldsymbol{H}, \boldsymbol{R})$. Here, $\mathcal{V}$ is the set of nodes that represent the SMEs; $\mathcal{E}$ is the set of heterogeneous relation types among the 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; $\boldsymbol{H} \in \mathbb{R}^{|\mathcal{V}| \times d_0^{\mathrm{h}}}$ and $\boldsymbol{R} \in \mathbb{R}^{|\mathcal{E}| \times d_0^{\mathrm{e}}}$ denote the original node feature matrix of all SME nodes and the relation feature matrix of all relation types, respectively.

Definition 2 (Financing needs exploration) 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 the structural information of the graph given an SME 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 study are shown in the supplementary materials.

3.2 Overview of M-RIGHT

Overall, M-RIGHT first learns the representations of SMEs and their relation types and then leverages the learned representations to facilitate the downstream FNE task. The architecture of M-RIGHT's representation learning process is presented in Fig. 2, 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 then obtains a representation of each SME after the message transfer. The behavior heterogeneity learning module first enables heterogeneous representations of each SME under different relation types by performing SME representation translations on each relational hyperplane (which helps distinguish each SME's heterogeneous behavior) and then leverages the 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 subsections elaborate on M-RIGHT's transfer heterogeneity learning module and behavior heterogeneity learning module.

3.3 Transfer heterogeneity learning

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 uses a self-attention mechanism to differentiate the importance of neighbors. Here, we introduce the message-passing process and the self-attention mechanism.

3.3.1 Message passing with entity–relation composition operator

In the l^{th} layer of the graph convolution network, M-RIGHT obtains a set of SME representations, i.e., node embeddings $\boldsymbol{h} = \{\boldsymbol{h}_1^l, \boldsymbol{h}_2^l, \ldots, \boldsymbol{h}_{|\mathcal{V}|}^l\}$, where $\boldsymbol{h}_i^l \in \mathbb{R}^{d_l}$, and a set of relation embeddings $\boldsymbol{r} = \left\{\boldsymbol{r}_1^l, \boldsymbol{r}_2^l, \ldots, \boldsymbol{r}_{|\mathcal{E}|}^l\right\}$, where $\boldsymbol{r}_i^l \in \mathbb{R}^{d_l}$. 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 $\Phi(\cdot) : \mathbb{R}^{d_l} \times \mathbb{R}^{d_l} \rightarrow \mathbb{R}^{d_l}$, which aggregates their representations from both the neighbors and the corresponding relation types during the message-passing process. Then, each SME's

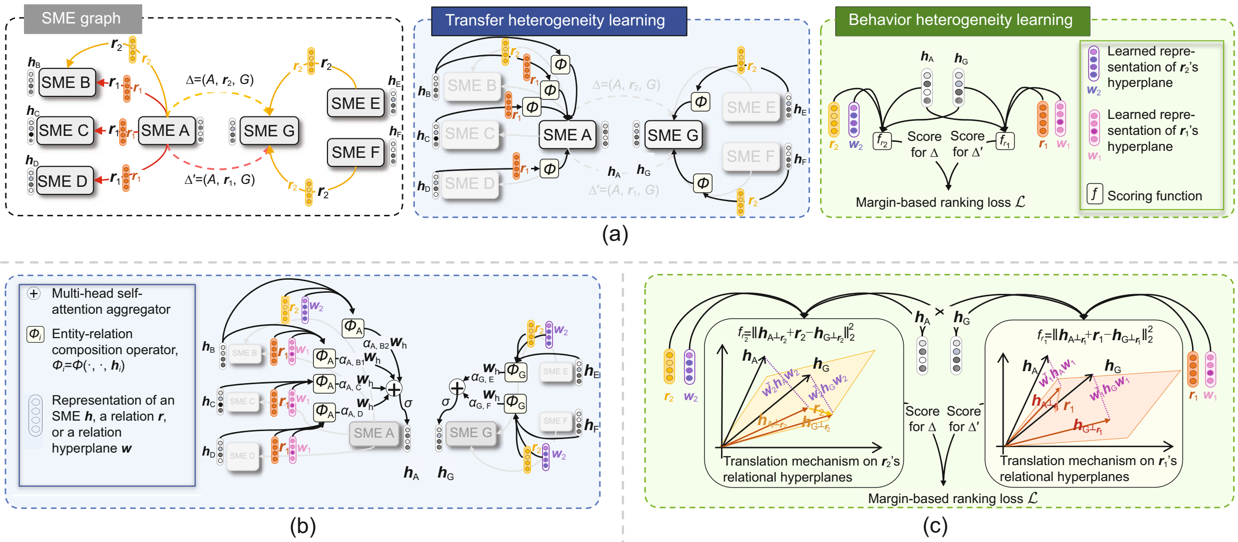

Fig. 2 Architecture of M-RIGHT's representation learning process, which includes the transfer heterogeneity learning module and behavior heterogeneity learning module: (a) given an SME graph with initial SME representations $\boldsymbol{h}_*$, relation features $\boldsymbol{r}_*$, a true triplet $\triangle$, and a fake triplet $\triangle'$, M-RIGHT first leverages the transfer heterogeneity learning module to obtain the corresponding SME representations, and then leverages the behavior heterogeneity learning module to obtain the triplets' scores and the corresponding loss for the model's update; (b) details of transfer heterogeneity learning to obtain representations of SMEs based on the 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 enable SMEs' heterogeneous representations, i.e., heterogeneous behavior, under different relations (M-RIGHT: Multi-relation tRanslatIonal GrapH aTtention network; SME: small- and medium-sized enterprise)

output representation in this layer is obtained by a shared linear transformation parameterized by a matrix, $\boldsymbol{W}_h^l \in \mathbb{R}^{d_l \times d_{l+1}}$, along with a nonlinear transformation parameterized by a non-linear function $f(\cdot)$. The representations of each SME u in this layer are calculated as follows:

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

where $\mathcal{N}(u)$ denotes the one-hop neighbor set of SME u, and $\mathcal{T}(u,v)$ denotes the relation type between SMEs u and v. Note that the number of relations between two SMEs is allowed to be more than one. The instantiation of the entity–relation composition operation will be elaborated in Section 3.4.

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 $\boldsymbol{W}_r^l \in \mathbb{R}^{d_l \times d_{l+1}}$ to obtain the output relation type representation as follows:

$$\boldsymbol{r}_i^{l+1} = f\left(\boldsymbol{W}_r^l \boldsymbol{r}_i^l\right). \quad (2)$$

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

3.3.2 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 focus on the most relevant neighbors for message passing. To address these problems, we use a self-attention mechanism (Veličković et al., 2018) that updates each SME's representation by attending over its neighbors. Similar to the graph attention network (GAT) (Veličković et al., 2018), to stabilize the learning process of self-attention, we extend the mechanism to apply 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 $a : \mathbb{R}^{d_{l+1}} \times \mathbb{R}^{d_{l+1}} \rightarrow \mathbb{R}$ as follows:

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

where $\boldsymbol{W}_{h,k}^{l} \in \mathbb{R}^{d_l \times \frac{d_{l+1}}{K}}$ 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 $\boldsymbol{a} \in \mathbb{R}^{\frac{2d_{l+1}}{K}}$, and apply the leaky rectified linear unit (LeakyReLU) (Maas et al., 2013). Fully expanded out, Eq. (3) may then be expressed as follows:

$$\begin{aligned} e_{u,v}^{k} =& \left\{ \exp\left(\text{LeakyReLU}\left(\boldsymbol{a}^{\mathrm{T}}\left[\boldsymbol{W}_{h,k}^{l}\boldsymbol{h}_{u}^{l} \| \boldsymbol{W}_{h,k}^{l}\right.\right.\right.\right. \\ & \left.\left.\left.\left. \cdot \Phi(\boldsymbol{h}_{v}^{l}, \boldsymbol{r}_{\mathcal{T}(u,v)}^{l}, \boldsymbol{h}_{u}^{l})\right]\right)\right)\right\} \\ & \cdot \left\{ \sum_{m\in\mathcal{N}(u)} \exp\left(\text{LeakyReLU}\left(\boldsymbol{a}^{\mathrm{T}}\right.\right.\right. \\ & \left.\left.\left. \cdot \left[\boldsymbol{W}_{h,k}^{l}\boldsymbol{h}_{u}^{l} \| \boldsymbol{W}_{h,k}^{l}\Phi(\boldsymbol{h}_{m}^{l}, \boldsymbol{r}_{\mathcal{T}(i,m)}^{l}, \boldsymbol{h}_{i}^{l})\right]\right)\right)\right\}^{-1}, \end{aligned} \quad (4)$$

where "$\|$" is the concatenation operation.

The attention coefficient computed using 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 coefficient of SME u is normalized across all its neighbors using the softmax function, as follows:

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

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

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

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

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

Such an attention mechanism is efficient because it is parallelizable across pairs of neighbor nodes. Moreover, the model is directly applicable to inductive learning problems, including tasks in which 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; Veličković 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 the 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 neighbors' attentions and the transferred messages.

3.4 Behavior heterogeneity learning

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

To enable an SME's heterogeneous representations, we introduce a translation mechanism on relational hyperplanes. Specifically, under each relation type r, this mechanism uses a vector $\boldsymbol{w}_r$ to project the representation $\boldsymbol{h}_s$ of each SME s into a hyperplane, and the representation of SME can be obtained as follows:

$$\boldsymbol{h}_{s\perp r} = \boldsymbol{h}_s - \boldsymbol{w}_r^{\mathrm{T}}\boldsymbol{h}_s\boldsymbol{w}_r. \quad (8)$$

Then, the score of each triple (s, r, o) can be calculated as follows:

$$\begin{aligned} f_r(\boldsymbol{h}_s, \boldsymbol{h}_o) = &\left\|\left(\boldsymbol{h}_s - \boldsymbol{w}_r^{\mathrm{T}}\boldsymbol{h}_s\boldsymbol{w}_r\right)\right. \\ &\left.+\boldsymbol{r}_r - \left(\boldsymbol{h}_o - \boldsymbol{w}_r^{\mathrm{T}}\boldsymbol{h}_o\boldsymbol{w}_r\right)\right\|_2^2. \end{aligned} \quad (9)$$

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 the SMEs' representations to be the same.

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

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

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

In addition, in each layer l in the transfer heterogeneity learning module, each relational hyperplane projection vector $\boldsymbol{w}_r$ will be transformed by matrix $\boldsymbol{W}_w^l \in \mathbb{R}^{d_l \times d_{l+1}}$ as follows:

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

The score is expected to be higher for a ground-truth triplet $\triangle$ and lower for a 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:

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

where $\boldsymbol{z}_i$ denotes the SME i's representation after L graph convolution layers, γ is the margin separating positive and negative triplets, and "$[\ ]_+$" is an operator that converts a negative value to zero. To guarantee that the output representation of each r from the graph convolution network, i.e., $\boldsymbol{r}_r^{L+1}$, is in the relational hyperplane and is regularized, the following constraints are considered when we minimize $\mathcal{L}$:

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

where ϵ is an error term to ensure orthogonality. Then, Eq. (12) can be rewritten as follows:

$$\begin{aligned} \mathcal{L} = & \sum_{\substack{(s,r,o) \in \triangle \\ (s',r',o') \in \triangle'}} \left[f_r(\boldsymbol{z}_s, \boldsymbol{z}_o) + \gamma - f_{r'}\left(\boldsymbol{z}'_s, \boldsymbol{z}'_o\right)\right]_+ \\ & + C\left\{\sum_{v \in \mathcal{V}} \left[\|\boldsymbol{z}_v\|_2^2 - 1\right]_+ \right. \\ & \left. + \sum_{r \in \mathcal{E}} \left[\frac{\left(\boldsymbol{w}_r^{\mathrm{T}} \boldsymbol{r}_r^{L+1}\right)^2}{\left\|\boldsymbol{r}_r^{L+1}\right\|_2^2} - \epsilon^2\right]_+ \right\}. \end{aligned} \tag{14}$$

We adopt stochastic gradient descent to minimize the above loss function, with which M-RIGHT's parameters, including the shared metrics $\boldsymbol{W}_{h,k}^l$, $\boldsymbol{W}_r^l$, and $\boldsymbol{W}_w^l$ in each l^{th} layer, the weight vector $\boldsymbol{\alpha}$, and the hyperplane's norm vector $\boldsymbol{w}_r$ of each relation r, can be updated. The sets 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, which directly use the node representations after message passing for the node classification or link prediction task, our proposed behavior 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 it 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.

Detailed analysis of our proposed M-RIGHT is presented in the supplementary materials.

4 Experiments

In this section, we verify the effectiveness of our proposed M-RIGHT on the FNE task in MYbank. 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

Algorithm 1 M-RIGHT's representation learning process

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

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

while not converged **do**

  # Embedding of each layer

  for $l=0,1,...,L$ **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**

          $e_{u,v}^k = a\big(\boldsymbol{W}_{h,k}^l \boldsymbol{h}_u^l, \boldsymbol{W}_{h,k}^l \varPhi(\boldsymbol{h}_v^l, \boldsymbol{r}_{\mathcal{T}(i,j)}^l, \boldsymbol{h}_u^l)\big)$

          $\alpha_{u,v}^k = \mathrm{softmax}_v(e_{u,v}^k)$

        end for

      end for

    end for

    $\boldsymbol{h}_u^{l+1} = \|_{k=1}^K f\big(\sum_{v\in\mathcal{N}(u)} \alpha_{u,v}^k \boldsymbol{W}_{h,k}^l \cdot \varPhi(\boldsymbol{h}_v^l, \boldsymbol{r}_{\mathcal{T}(u,v)}^l, \boldsymbol{h}_u^l)\big)$

    $\forall i \in \mathcal{E},\ \boldsymbol{r}_i^{l+1} = f\left(\boldsymbol{W}_r^l \boldsymbol{r}_i^l\right)$

    $\forall i \in \mathcal{E},\ \boldsymbol{w}_i^{l+1} = f\left(\boldsymbol{W}_w^l \boldsymbol{w}_i^l\right)$

  end for

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

  # Backpropagation with loss

$$\mathcal{L} = \sum_{\substack{(s,r,o)\in\triangle \\ (s',r',o')\in\triangle'}} \left[f_r(\boldsymbol{z}_s,\boldsymbol{z}_o)+\gamma-f_{r'}(\boldsymbol{z}'_s,\boldsymbol{z}'_o)\right]_+ + C\left\{\sum_{v\in\mathcal{V}}\left[\|\boldsymbol{z}_v\|_2^2-1\right]_+ + \sum_{r\in\mathcal{E}}\left[\frac{(\boldsymbol{w}_r^{\mathrm{T}}\boldsymbol{r}_r^{L+1})^2}{\|\boldsymbol{r}_r^{L+1}\|_2^2}-\epsilon^2\right]_+\right\}$$

end while

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?

4.1 Experimental settings

4.1.1 Dataset description

In this subsection, we use two datasets, i.e., application program (APP) and short messaging service (SMS), collected from MYbank, each of which contains an SME graph and the financing need labels of SMEs. The detailed FNE scenario in MYbank is presented in the supplementary materials. The details of the two datasets are presented in Table 1.

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

4.1.2 Evaluation metrics

We evaluate the performance of different FNE methods using three metrics, i.e., classification accuracy (CA), micro-averaged F1 score (micro-F1), and area under the receiver operating characteristic (ROC) curve (AUC), which are widely used in graph representation learning studies (Veličković et al., 2018; Xu KYL et al., 2019) and financial studies (Yang S et al., 2020). More details of the metrics are presented in the supplementary materials.

4.1.3 Comparison methods

The comparison methods can be categorized into two groups, as follows:

1. Graph-free methods

ANOVA-XGBoost (Chen et al., 2020) uses the extreme gradient boosted tree method for prediction. SVM-RE (Rogic and Kascelan, 2019) uses a hybrid support vector machine rule extraction method for prediction. DeepFM (Guo et al., 2017) combines factorization machines and deep learning for prediction, which is a common method in real practice.

2. Graph-based methods

GIN (Xu KYL et al., 2019) is a homogeneous graph representation learning method that leverages the Weisfeiler–Lehman test for message passing. GraphSAGE (Hamilton et al., 2017) is an inductive homogeneous graph representation learning method. GAT (Veličković et al., 2018) is an attentive homogeneous graph convolution network. RGCN (Schlichtkrull et al., 2018) is a heterogeneous graph convolution network that models the message passing under different relation types. CompGCN (Vashishth et al., 2020a) is a heterogeneous graph convolution network that jointly learns the representations of nodes and relations. MHGCN (Yu et al., 2022) is a meta-path-based heterogeneous graph neural network. HRAN (Li et al., 2022) is a heterogeneous graph neural network that fuses and attends

Table 1 Dataset description

Dataset	Number of SMEs (nodes)	Number of relations (edges)	Number of relation types	Number of node features	Number of positive SMEs
APP	42.45 million	1.26 billion	27	1164	3.89 million
SMS	62.92 million	1.54 billion	27	885	5.68 million

APP: application program; SMS: short messaging service; SME: small- and medium-sized enterprise

semantic-specific information through the relation path.

More details on the experimental settings are presented in the supplementary materials.

4.2 Experimental results and analysis

4.2.1 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 use the initial graph-free SMEs' features as inputs to predict the financially constrained SMEs. M-RIGHT uses the initial graph-free features as input to train its graph model and outputs 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 2.

In Table 2, 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 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. Such improvements demonstrate the importance of modeling SMEs' relations 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 behavior 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.

4.2.2 Evaluation with respect to relation sparsity (for Q2)

To investigate how the relation information among SMEs affects the performance of M-RIGHT, we evaluate M-RIGHT's performance under different relation sparsities. Specifically, we randomly filter each training dataset into four sub-datasets based on different sparsities. 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 sparsities is presented in Fig. 3. Furthermore, we calculate the following values in the datasets: (1) relation densities, i.e., the number of actual relations among the SMEs divided by the maximum number of relations in a fully connected graph; (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 from 10% to 70% sparsity. The results are shown in Table 3.

From Fig. 3, we can make three conclusions. First, M-RIGHT shows degradation in performance when some existing relations are unavailable. Specifically, M-RIGHT's CA, micro-F1, and AUC values

Table 2 Performance of all methods on CA, micro-F1, and AUC values (mean±range, computed across 10 runs)

Category	Method	APP		
		CA	Micro-F1	AUC
Graph-free methods	ANOVA-XGBoost	0.6688±0.006	0.2303±0.035	0.8585±0.012
	SVM-RE	0.5072±0.005	0.1794±0.026	0.5567±0.022
	DeepFM	0.5821±0.005	0.2041±0.029	0.8344±0.033
Graph-based methods	GIN	0.6708±0.000	0.2344±0.005	0.8550±0.001
	GraphSAGE	0.6645±0.001	0.2302±0.004	0.8601±0.000
	GAT	0.6752±0.000	0.2377±0.005	0.8649±0.000
	RGCN	0.6697±0.001	0.2294±0.005	0.8648±0.001
	CompGCN	0.6695±0.001	0.2295±0.007	0.8660±0.001
	MHGCN	0.6690±0.007	0.2305±0.011	0.8659±0.002
	HRAN	0.6761±0.001*	0.2395±0.006*	0.8664±0.004*
Our proposed methods	M-RIGHT	0.6906±0.001	0.2456±0.007	0.9006±0.001
	M-RIGHT-w/o-rt	0.6760±0.001	0.2382±0.009	0.8788±0.001
	M-RIGHT-w/o-rs	0.6743±0.001	0.2334±0.007	0.8737±0.001
Improvement (%)[1]		2.1447	2.5470	3.9474
p-value[2]		0.000	0.004	0.000

Category	Method	SMS		
		CA	Micro-F1	AUC
Graph-free methods	ANOVA-XGBoost	0.9780±0.002	0.4094±0.003	0.9306±0.000
	SVM-RE	0.9803±0.002	0.1078±0.001	0.7324±0.000
	DeepFM	0.9769±0.002	0.3668±0.002	0.9234±0.001
Graph-based methods	GIN	0.9783±0.001	0.4110±0.001	0.9289±0.000
	GraphSAGE	0.9724±0.001	0.4141±0.002	0.9285±0.001
	GAT	0.9833±0.001	0.4094±0.001	0.9275±0.000
	RGCN	0.9752±0.003	0.4152±0.003*	0.9336±0.001
	CompGCN	0.9828±0.003	0.4120±0.004	0.9298±0.001
	MHGCN	0.9835±0.001*	0.4133±0.001	0.9312±0.000
	HRAN	0.9831±0.004	0.4150±0.007	0.9338±0.001*
Our proposed methods	M-RIGHT	0.9841±0.000	0.4287±0.003	0.9469±0.001
	M-RIGHT-w/o-rt	0.9790±0.002	0.4158±0.003	0.9339±0.000
	M-RIGHT-w/o-rs	0.9830±0.003	0.4181±0.002	0.9368±0.000
Improvement (%)[1]		0.0610	3.2514	1.4029
p-value[2]		0.001	0.000	0.000

APP: application program; SMS: short messaging service; CA: classification accuracy; Micro-F1: micro-averaged F1 score; AUC: area under the receiver operating characteristic (ROC) curve. [1] Improvement of M-RIGHT over the best-performing comparison methods. [2] Statistical improvement over the best-performing comparison methods if p-value<0.05 (p-value with paired t-test). * Results of the best-performing comparison methods

decrease with the increase of relation sparsity, which indicates that its performance is affected by the availability of the relations. Without 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. Second, the performance degradation under more relation-dependent scenarios is even more severe. Specifically, Table 3 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 3, 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 sparsest training dataset, demonstrating the importance of relations in transferring neighborhood information.

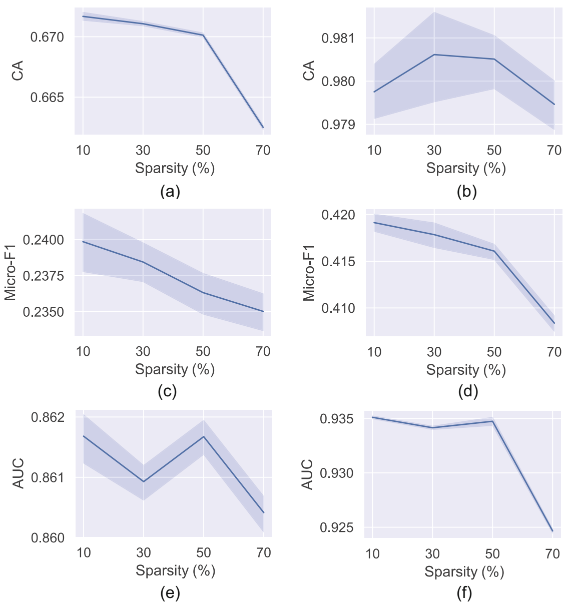

Fig. 3 Performance of M-RIGHT under different dataset sparsities (mean±range, computed across 10 runs): (a) CA on APP; (b) CA on SMS; (c) Micro-F1 on APP; (d) Micro-F1 on SMS; (e) AUC on APP; (f) AUC on SMS (CA: classification accuracy; Micro-F1: micro-averaged F1 score; AUC: area under the receiver operating characteristic (ROC) curve; APP: application program; SMS: short messaging service)

Table 3 Phenomenon with respect to sparisity

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%

APP: application program; SMS: short messaging service

4.2.3 Ablation experiments (for Q3)

To investigate how the two important mechanisms in M-RIGHT contribute 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 do not consider the relation types); (2) M-RIGHT-w/o-rs refers to M-RIGHT without considering behavior heterogeneity (namely, Eq. (9) is replaced with $f_r(\boldsymbol{h}_s, \boldsymbol{h}_o) = \|\boldsymbol{h}_s + \boldsymbol{r}_r - \boldsymbol{h}_o\|_2^2$ and Eq. (10) is replaced similarly). The results are shown in Table 2.

In Table 2, 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, leading to a decrease in performance. Comparing M-RIGHT with M-RIGHT-w/o-rs, M-RIGHT achieves 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, leading to a decrease in performance.

4.2.4 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. Due to the space limit, we present the results and analysis in the supplementary materials.

5 Conclusions and future work

In this paper, we have conducted exploratory analysis on the financing needs exploration task, which indicates the importance of modeling SMEs' relations. Then, we have proposed a novel method named M-RIGHT, whose main novelty is that it simultaneously addresses two kinds of challenging heterogeneity, i.e., transfer heterogeneity and behavior heterogeneity, in modeling SME graphs 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 SMEs' heterogeneous behaviors in different environments, by performing SME representation translations on each relational hyperplane. Comprehensive experiments on two real-world datasets have demonstrated the superiority of M-RIGHT to the state-of-the-art methods in exploring financially constrained SMEs.

In the future, we intend to extend our work in two potential directions. First, 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. Second, considering that SMEs can be clustered into several groups and that SMEs in different groups may have different behavior patterns, we plan to introduce such cluster information into the relational learning process in the future.

Contributors

Qianqiao LIANG designed the research. Hua WEI, Yaxi WU, and Feng WEI conducted the experiments. Qianqiao LIANG drafted the paper. Deng ZHAO, Jianshan HE, and Guofang MA helped organize the paper. Xiaolin ZHENG and Bing HAN revised and finalized the paper.

Compliance with ethics guidelines

Qianqiao LIANG, Hua WEI, Yaxi WU, Feng WEI, Deng ZHAO, Jianshan HE, Xiaolin ZHENG, Guofang MA, and Bing HAN declare that they have no conflict of interest.

Data availability

Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data are not available.

References

Angilella S, Mazzù S, 2015. The financing of innovative SMEs: a multicriteria credit rating model. *Eur J Oper Res*, 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. Proc 26th Int Conf on Neural Information Processing Systems, p.2787-2795. https://doi.org/10.5555/2999792.2999923

Ceptureanu EG, Ceptureanu S, Herteliu C, 2021. Evidence regarding external financing in manufacturing MSEs using partial least squares regression. *Ann Oper Res*, 299(1-2):1189-1202. https://doi.org/10.1007/s10479-019-03291-2

Chen XF, Zhao YD, Wei ZF, et al., 2020. Intelligent identification of potential customers for electricity substitution. In: Tallón-Ballesteros AJ (Ed.), Fuzzy Systems and Data Mining, VI. IOS Press, p.168-179. https://doi.org/10.3233/FAIA200697

Cheng HT, Koc L, Harmsen J, et al., 2016. Wide & deep learning for recommender systems. Proc 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. Proc 32nd AAAI Conf on Artificial Intelligence, p.1811-1818. https://doi.org/10.1609/aaai.v32i1.11573

Dong YX, Chawla NV, Swami A, 2017. metapath2vec: scalable representation learning for heterogeneous networks. Proc 23rd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining, p.135-144. https://doi.org/10.1145/3097983.3098036

Duan GL, Ma X, 2018. A coupon usage prediction algorithm based on XGBoost. Proc 14th Int Conf on Natural Computation, Fuzzy Systems and Knowledge Discovery, p.178-183. https://doi.org/10.1109/FSKD.2018.8686924

Fu XY, Zhang JN, Meng ZQ, et al., 2020. MAGNN: metapath aggregated graph neural network for heterogeneous graph embedding. Proc Web Conf, 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. *Ind Manag Data Syst*, 121(1):123-157. https://doi.org/10.1108/IMDS-08-2020-0510

Grover A, Leskovec J, 2016. node2vec: scalable feature learning for networks. Proc 22nd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining, p.855-864. https://doi.org/10.1145/2939672.2939754

Guo HF, Tang RM, Ye YM, et al., 2017. DeepFM: a factorization-machine based neural network for CTR prediction. Proc 26th Int Joint Conf 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. Proc 31st Int Conf on Neural Information Processing Systems, p.1025-1035. https://doi.org/10.5555/3294771.3294869

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

Ji SX, Pan SR, Cambria E, et al., 2022. A survey on knowledge graphs: representation, acquisition, and applications. *IEEE Trans Neur Netw Learn Syst*, 33(2):494-514. https://doi.org/10.1109/TNNLS.2021.3070843

Kipf TN, Welling M, 2017. Semi-supervised classification with graph convolutional networks. Proc 5th Int Conf on Learning Representations, p.1-10.

Kshetri N, 2016. Big data's role in expanding access to financial services in China. *Int J Inform Manag*, 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. *Inform Sci*, 557:286-301. https://doi.org/10.1016/j.ins.2019.05.027

Li ZF, Liu H, Zhang ZL, et al., 2022. Learning knowledge graph embedding with heterogeneous relation attention networks. *IEEE Trans Neur Netw Learn Syst*, 33(8):3961-3973. https://doi.org/10.1109/TNNLS.2021.3055147

Liao HF, Hu J, Li TR, et al., 2022. Deep linear graph attention model for attributed graph clustering. *Knowl-Based Syst*, 246:108665. https://doi.org/10.1016/j.knosys.2022.108665

Luo Y, Deng TY, Wei Q, et al., 2021. Optimal financing decision in a contract food supply chain with capital constraint. *Complexity*, 2021:8925102. https://doi.org/10.1155/2021/8925102

Maas AL, Hannun AY, Ng AY, 2013. Rectifier nonlinearities improve neural network acoustic models. Proc 30th Int Conf on Machine Learning, p.1-6.

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

Nickel M, Rosasco L, Poggio T, 2016. Holographic embeddings of knowledge graphs. Proc 30th AAAI Conf on Artificial Intelligence, p.1955-1961. https://doi.org/10.5555/3016100.3016172

Perozzi B, Al-Rfou R, Skiena S, 2014. DeepWalk: online learning of social representations. Proc 20th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining, p.701-710. https://doi.org/10.1145/2623330.2623732

Rogic S, Kascelan L, 2019. Customer value prediction in direct marketing using hybrid support vector machine rule extraction method. Proc 23rd European Conf 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. *J Theor Appl Electron Commer Res*, 17(3):1003-1018. https://doi.org/10.3390/jtaer17030051

Sadeghian A, Armandpour M, Colas A, et al., 2021. ChronoR: rotation based temporal knowledge graph embedding. Proc AAAI Conf on Artificial Intelligence, p.6471-6479. https://doi.org/10.1609/aaai.v35i7.16802

Schlichtkrull M, Kipf TN, Bloem P, et al., 2018. Modeling relational data with graph convolutional networks. Proc 15th Int Conf on Semantic Web Conf, 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. Proc AAAI Conf on Artificial Intelligence, p.3060-3067. https://doi.org/10.1609/aaai.v33i01.33013060

Shi Y, Gui H, Zhu Q, et al., 2018a. ASPEM: embedding learning by aspects in heterogeneous information networks. Proc SIAM Int Conf 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. Proc 24th ACM SIGKDD Int Conf on Knowledge Discovery & Data Mining, p.2190-2199. https://doi.org/10.1145/3219819.3220006

Sun ZQ, Deng ZH, Nie JY, et al., 2019. Rotate: knowledge graph embedding by relational rotation in complex space. Proc 7th Int Conf on Learning Representations, p.1-10.

Tang J, Qu M, Mei QZ, 2015. PTE: predictive text embedding through large-scale heterogeneous text networks. Proc 21st ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining, p.1165-1174. https://doi.org/10.1145/2783258.2783307

Tian Z, Hassan AFS, Razak NHA, 2018. Big data and SME financing in China. *J Phys Conf Ser*, 1018:012002. https://doi.org/10.1088/1742-6596/1018/1/012002

Trouillon T, Welbl J, Riedel S, et al., 2016. Complex embeddings for simple link prediction. Proc 33rd Int Conf on Machine Learning, p.2071-2080. https://doi.org/10.5555/3045390.3045609

Vashishth S, Sanyal S, Nitin V, et al., 2020a. Composition-based multi-relational graph convolutional networks. Proc 8th Int Conf 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. Proc AAAI Conf on Artificial Intelligence, p.3009-3016. https://doi.org/10.1609/aaai.v34i03.5694

Veličković P, Cucurull G, Casanova A, et al., 2018. Graph attention networks. Proc 6th Int Conf on Learning Representations, p.1-10.

Wang X, Zhang YD, Shi C, 2019. Hyperbolic heterogeneous information network embedding. Proc AAAI Conf on Artificial Intelligence, p.5337-5344. https://doi.org/10.1609/aaai.v33i01.33015337

Wang Y, Jing CF, Xu SS, et al., 2022. Attention based spatiotemporal graph attention networks for traffic flow forecasting. *Inform Sci*, 607:869-883. https://doi.org/10.1016/j.ins.2022.05.127

Wang Z, Zhang JW, Feng JL, et al., 2014. Knowledge graph embedding by translating on hyperplanes. Proc 28th AAAI Conf on Artificial Intelligence, p.1112-1119. https://doi.org/10.5555/2893873.2894046

Wu ZH, Pan SR, Chen FW, et al., 2021. A comprehensive survey on graph neural networks. *IEEE Trans Neur Netw Learn Syst*, 32(1):4-24. https://doi.org/10.1109/TNNLS.2020.2978386

Xu KYL, Hu WH, Leskovec J, et al., 2019. How powerful are graph neural networks? Proc 7th Int Conf on Learning Representations, p.1-10.

Xu ZP, Meisami A, Tewari A, 2021. Decision making problems with funnel structure: a multi-task learning approach with application to email marketing campaigns. Proc 24th Int Conf on Artificial Intelligence and Statistics, p.127-135.

Yang BS, Yih SWT, He XD, et al., 2015. Embedding entities and relations for learning and inference in knowledge bases. Proc 3rd Int Conf on Learning Representations, p.1-10.

Yang C, Xiao YX, Zhang Y, et al., 2022. Heterogeneous network representation learning: a unified framework with survey and benchmark. *IEEE Trans Knowl Data Eng*, 34(10):4854-4873. https://doi.org/10.1109/TKDE.2020.3045924

Yang S, Zhang ZQ, Zhou J, et al., 2020. Financial risk analysis for SMEs with graph-based supply chain mining. Proc 29th Int Joint Conf on Artificial Intelligence, p.4661-4667. https://doi.org/10.24963/ijcai.2020/643

Ye R, Li X, Fang YJ, et al., 2019. A vectorized relational graph convolutional network for multi-relational network alignment. Proc 28th Int Joint Conf on Artificial Intelligence, p.4135-4141. https://doi.org/10.24963/ijcai.2019/574

Yu PY, Fu CF, Yu YW, et al., 2022. Multiplex heterogeneous graph convolutional network. Proc 28th ACM SIGKDD Conf on Knowledge Discovery and Data Mining, p.2377-2387. https://doi.org/10.1145/3534678.3539482

Zhang B, Wang LQ, Li YY, 2021. Precision marketing method of e-commerce platform based on clustering algorithm. *Complexity*, 2021:5538677. https://doi.org/10.1155/2021/5538677

Zhang WT, Fang Y, Liu ZM, et al., 2022. mg2vec: learning relationship-preserving heterogeneous graph representations via metagraph embedding. *IEEE Trans Knowl Data Eng*, 34(3):1317-1329. 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. Proc AAAI Conf on Artificial Intelligence, p.4697-4705. https://doi.org/10.1609/aaai.v35i5.16600

List of supplementary materials