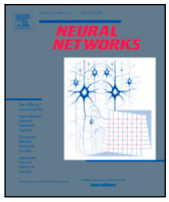
Neural Networks 181 (2025) 106757

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet



Full Length Article

Multi-relational graph contrastive learning with learnable graph augmentation

Xian Mo [a,d,*], Jun Pang [b], Binyuan Wan [a,**], Rui Tang [c], Hao Liu [a,d], Shuyu Jiang [c]

[a] School of Information Engineering, Ningxia University, Yinchuan 750021, China
[b] Faculty of Science, Technology and Medicine, University of Luxembourg, Esch-sur-Alzette L-4364, Luxembourg
[c] School of Cyber Science and Engineering, Sichuan University, Chengdu 610065, Sichuan, China
[d] Ningxia Key Laboratory of Artificial Intelligence and Information Security for Channeling Computing Resources from the East to the West, Ningxia University, Yinchuan 750021, China

ARTICLE INFO

Dataset link: https://github.com/Legendary-L/MRGCL

Keywords:
Contrastive learning
Learnable graph augmentation
Multi-relational graph learning

ABSTRACT

Multi-relational graph learning aims to embed entities and relations in knowledge graphs into low-dimensional representations, which has been successfully applied to various multi-relationship prediction tasks, such as information retrieval, question answering, and etc. Recently, contrastive learning has shown remarkable performance in multi-relational graph learning by data augmentation mechanisms to deal with highly sparse data. In this paper, we present a Multi-Relational Graph Contrastive Learning architecture (MRGCL) for multi-relational graph learning. More specifically, our MRGCL first proposes a Multi-relational Graph Hierarchical Attention Networks (MGHAN) to identify the importance between entities, which can learn the importance at different levels between entities for extracting the local graph dependency. Then, two graph augmented views with adaptive topology are automatically learned by the variant MGHAN, which can automatically adapt for different multi-relational graph datasets from diverse domains. Moreover, a subgraph contrastive loss is designed, which generates positives per anchor by calculating strongly connected subgraph embeddings of the anchor as the supervised signals. Comprehensive experiments on multi-relational datasets from three application domains indicate the superiority of our MRGCL over various state-of-the-art methods. Our datasets and source code are published at https://github.com/Legendary-L/MRGCL.

1. Introduction

Multi-relational graphs, also known as knowledge graphs (KGs) (Chen, Fang, Meng, Zhang and Liang, 2022), are composed of different types of entities as nodes and relations as edges, which can be used to store abundance of factual knowledge. For instance, KGs are usually stored as a triad (s, r, o), where s and o respectively define the different types of source and object entities, and r defines the different types of relations. Multi-relational graph learning, also known as knowledge graph embedding (KGE) (Chen, Fang et al., 2022), aims to embed both entities and relations into low-dimensional vector representations, which can preserve the inherent structure of KGs. It has been successfully applied to various downstream multi-relationship prediction tasks utilizing the representation vectors, such as relation extraction (Qu, Gao, Xhonneux, & Tang, 2020), information retrieval (Zhou, Chen, He, Ye, & Sun, 2022), personalized recommendation (Chen, Yang et al., 2022), question answering (Liu, Chen, Das, Yang, & Tong, 2023), and drug–drug interaction prediction (Xiong et al., 2023), etc.

Recent researches on KGE have focused on employing graph convolutional networks (GCN) (Kipf & Welling, 2017) to encode inherent semantic and structural information of entities and relations into a low-dimensional vector representations. For example, RGCN (Schlichtkrull et al., 2018) presents a relational GCN to address the highly multi-relational data feature of KGs, and GTN (Yun, Jeong, Kim, Kang, & Kim, 2019) introduces a self-adaption weighting message passing to encode entities and relations in KGs. GGPN (Chen, Fang et al., 2022) later proposes a novel multi-relational graph Gaussian Process network with attention for multi-relational graph representation learning. Nevertheless, different entities under diverse relations in KGs contribute differently to target entity embeddings. The above approaches ignore the different importance of different entities under diverse relations for target entities. Some heterogeneous network embedding methods

* Corresponding author at: School of Information Engineering, Ningxia University, Yinchuan 750021, China.
** Corresponding author.
E-mail addresses: mxian168@nxu.edu.cn (X. Mo), jun.pang@uni.lu (J. Pang), binyuanw@outlook.com (B. Wan), tangrscu@scu.edu.cn (R. Tang), liuhao@nxu.edu.cn (H. Liu), jiang.shuyu07@gmail.com (S. Jiang).

https://doi.org/10.1016/j.neunet.2024.106757
Received 22 February 2024; Received in revised form 9 August 2024; Accepted 21 September 2024
Available online 26 September 2024

considering multiple meta-path relations based on hierarchical attention have also emerged recently, which can identify the importance of different levels between entities. For example, HAN (Wang et al., 2019) presents a heterogeneous GCN based on the hierarchical attention containing node-level and semantic-level attentions for heterogeneous graph embeddings. ie-HGCN (Yang et al., 2023) designs a object-level attention and type-level attention to learn the vector representations of entities. However, these methods focus on heterogeneous graph rather than KGs. Thus, such methods cannot be used directly for multi-relationship prediction tasks. More recently, DHAN (Zhao, Wei et al., 2023) designs a hierarchical attention architecture including intra-type attention and inter-type attention to learn the same type of nodes and different types of neighbor in bi-typed multi-relational graphs. However, only two types of entities are considered in this approach. Multi-relational graphs in the real world are often composed of multiple types of entities and relations, resulting in insufficient research on KGs based on hierarchical attention.

Currently, most of the existing GCN-based KGM models are trained in a supervised manner, but collecting a large number of labeled data often requires a lot of resources and time. In recent years, graph contrastive learning (GCL) (Xiong et al., 2023) has made great progress in the multi-relational graph learning, which aims to learn vector representations from unlabeled graphs. Its core idea is to enhance node representations by agreement between the created contrastive views by contrasting the defined positive pairs with negative instance counterparts. For instance, MRCGNN (Xiong et al., 2023) generates two contrastive views by randomly shuffling edge relations and node features for multi-relational drug–drug interaction event prediction, while CMGNN (Fang, Zhang, Hu, Wu and Xu, 2023) encodes two views containing connected neighbors and a knowledge graph diffusion for multi-modal knowledge graph learning. Nevertheless, these approaches mainly utilize human-designed graph augmentations for domain-specific datasets. A good enhanced view should be rather diverse for different domains while keeping task-related information intact (Tian et al., 2020). However, the existing handcraft graph augmentation methods based on random perturbation would fail to keep the task-relevant information intact in different domains. More recently, some adaptive graph contrastive learning models have emerged, such as AdaGCL (Jiang, Huang, & Huang, 2023) and ADGCL (Zhuang, Wang, Zhao, & Sun, 2023). However, these methods focus on graphs learn with limited relation types and cannot be directly applied to multi-relational graphs. Therefore, *how to design a learnable graph augmentation strategy with an adaptive topology that can automatically adapt to diverse application domain multi-relational graph datasets remains as a challenge.*

Furthermore, existing contrastive loss functions in GCL mainly adopt contrastive losses originally developed in computer vision (Shen, Sun, Pan, Zhou, & Yang, 2023), which pull positive pairs together and push negative pairs far apart for guiding node learning. For example, the classic graph contrastive loss method InfoNCE (van den Oord, Li, & Vinyals, 2018) generates positive pairs of each anchor by creating different augmented views, which treat the same nodes in the different views as the positive pair of the anchor. It treats the other different nodes as negative pairs from different view, even if they are edge-connected with the same nodes of the anchor in the different views. NT-Xent (Zhu et al., 2020) treats all the other different nodes within same view and different views as negative pairs. On the other hand, graph homophily assumption (McPherson, Smith-Lovin, & Cook, 2001) indicates that connected neighboring nodes usually should be close to each other rather than far apart from each other. InfoNCE and NT-Xent treat neighboring nodes as negative pairs and pushed apart from the anchor, which contradicts the graph homophily assumption. Recently, NCLA (Shen et al., 2023) introduces a neighbor contrastive loss to generate multiple positive pairs per anchor by considering the 1-hop neighbors of the anchor within same view and different views. However, graphs reveals high-order properties by topologies, considering only the 1-hop neighbors of the anchor results in limited performance on downstream tasks. MRCGNN (Xiong et al., 2023) generates positives per anchor by calculating global graph representations of the anchor within same view. However, anchors tends to have a closer relationship with neighboring nodes with smaller hops. Calculating global graph representations for anchors as its positives may lead to inaccurate supervised signals. From the above discussion, it is clear that *how to design a contrastive loss strategies that utilize local neighbor relationships to generate positive pairs per anchor for a high-quality node learning remains as another challenge.*

In light of the two challenges identified above, our work presents a Multi-Relational Graph Contrastive Learning architecture (MRGCL) for multi-relational graph learning as shown in Fig. 1. Specifically, our MRGCL first proposes a Multi-relational Graph Hierarchical Attention Networks (MGHAN) to identify the importance between entities, which consists of entity-level, relation-level, and layer-level attentions. In more detail, the entity-level attention can identify the importance between the entities and their neighbor for a specific relation type, the relation-level attention can identify the importance of different relation-type for a specific entity, and the layer-level attention can identify the importance of different propagation layers of the MGHAN for a specific entity. In this way, MGHAN can learn the importance at different levels between entities for extracting the local graph dependency. We then utilize the variant MGHAN to automatically learn two graph augmented views with adaptive topology. In more detail, we remove the relation-level attention from MGHAN to learn contrast view 1 and remove both entity-level and relation-level attentions from MGHAN to learn contrast view 2. In this way, two graph augmented views are automatically learned by the variant MGHAN, which would keep exactly the same nodes and edges as the original graph but with different adaptive edge weights by removing different levels of attentions. Thus, it can keep task-related information intact and automatically adapt for diverse domain multi-relational graph datasets. Finally, we design a subgraph contrastive loss to generate positive pairs per anchor. Specifically, we first construct all k-hop neighbors of anchor in same view as their strongly connected subgraph. Then, we calculate the subgraph embeddings as positive pairs for each anchor, which can extract local higher-order relations of anchors for a high-quality node learning.

In summary, our work makes the following contributions:

- We propose an effective method MRGCL for multi-relational graph learning, which introduces the MGHAN architecture to learn the importance of different levels between entities for extracting the local graph dependency.
- We employ the proposed variant MGHAN for contrastive augmentation, which can keep task-related information intact and automatically adapt diverse domain-specific multi-relational graph datasets.
- We design a subgraph contrastive loss to generate positive pairs per anchor, which can extract local higher-order relations of anchors for a high-quality node learning.
- Comprehensive experiments on five multi-relational datasets from three application domains demonstrate the superiority of our MRGCL over various SOTA approaches.

2. Related work

GCN-based multi-relational graph learning. In recent years, multi-relational graph learning based on GCN has shown superior performance in learning entity and relation representations on KGs, which extends GCN to deal with multi-relational graphs and learns embeddings of KGs based on the message-passing mechanism. For example, RGCN (Schlichtkrull et al., 2018) presents a relational GCN to address the highly multi-relational data feature of KGs. SACN (Shang et al., 2019) adopts an encoder of a weighted graph convolutional

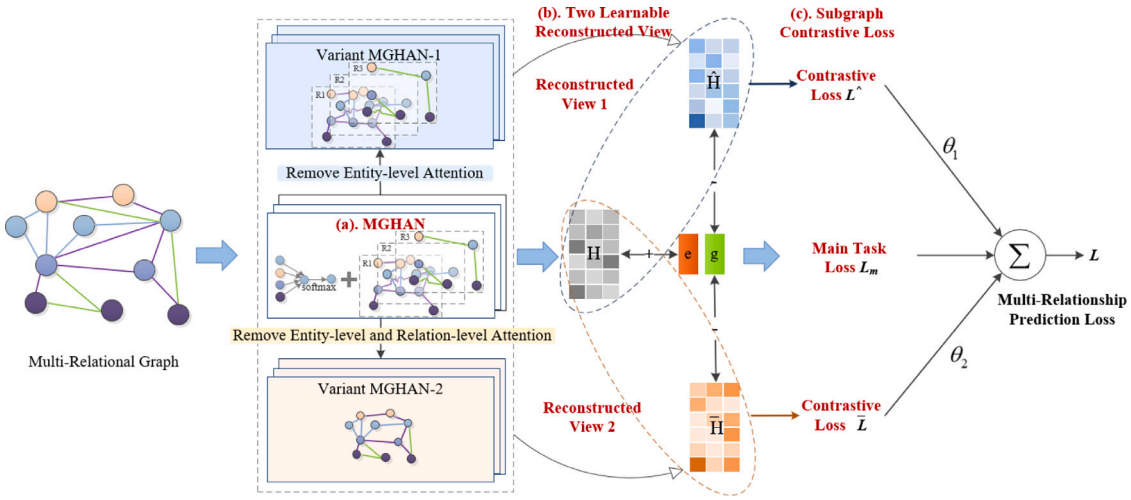


Fig. 1. Overall architecture of MRGCL. We adopt different colored circles represent different types of entities and different colored lines represent different types of relations to describe the multi-relational graph. (a): a multi-relational graph hierarchical attention networks consisting of entity-level, relation-level, and layer-level attentions. (b): two graph augmented views with adaptive topology are automatically learned by the variant MGHAN-1/2. (c): a subgraph contrastive loss is designed to generate positives per anchor by calculating strongly connected subgraph embeddings e of the anchor as the supervised signals.

network utilizing node structure, node attributes and edge relation types, and a decoder of a convolutional network for knowledge base completions. VR-GCN (Ye, Li, Fang, Zang, & Wang, 2019) presents a vectorized relational GCN to embed entities and relations simultaneously for multi-relational graphs. CompGCN (Vashishth, Sanyal, Nitin, & Talukdar, 2020) adopts a variety of entity-relation composition operations based on GCN architectures to learn entities and relations in KGs. MBGMN (Xia, Xu, Huang, Dai and Bo, 2021) merges the multi-behavior pattern to construct a graph meta-network for multi-behavior recommendations. To effectively predict drug–drug interactions, TrimNet-DDI (Li et al., 2021) utilizes a TrimNet (Li et al., 2021) to learn drug embeddings for multi-relational drug–drug interaction event prediction (DDI event prediction). MUFFIN (Chen et al., 2021) fuses multi-scale drug feature using deep learning model to learn drug embeddings for DDI event prediction. More recently, NMuR (Khatir, Choudhary, Choudhury, Agarwal, & Reddy, 2023) utilizes the proposed nonlinear hyperbolic normalization to the multi-relational graph MuR (Balazevic, Allen, & Hospedales, 2019) for multi-relational reasoning. ERGCN (Fang, Li et al., 2023) introduces a relation-aware GCN including entity convolution and relation convolution to embed entities and relations for multi-relational network alignment. Nevertheless, these methods ignore the attention mechanism on the multi-relational graphs.

Some multi-relational graph learning models that consider attention mechanisms appear. A2N (Bansal, Juan, Ravi, & McCallum, 2019) presents a novel attention-based strategy to embed query-dependent entity representations in KGs. GTN (Yun et al., 2019) introduces a self-adaption weighting message passing to encode entities and relations in KGs. KBAT (Nathani, Chauhan, Sharma, & Kaul, 2019) employs entity and relation features to construct a attention-based feature embedding framework for relation prediction. Recently, GGPN (Chen, Fang et al., 2022) proposes a novel multi-relational graph Gaussian Process network with attention for multi-relational graph representation learnings. SSI-DDI (Nyamabo, Yu, Liu, & Shi, 2022) employs multiple GAT (Chen, Fang et al., 2022) layers with a co-attention to learn embeddings for DDI event prediction. MRGAT (Dai, Wang, Zou, Liu, & Cen, 2022) introduces a multi-relational graph attention network to learn the importance of different neighborings for knowledge graph completion. More recently, DanSmp (Zhao, Du et al., 2023) introduces a bi-typed hybrid-relational market knowledge graph via dual attention Networks for stock movement prediction. HyperFormer (Hu, Gutiérrez-Basulto, Xiang, Li, & Pan, 2023) utilizes local-level sequential information, which encodes the content of the entities, relations and qualifiers of a triple via multi-head attention for knowledge graph completion. NYLON (Yu, Yang, & Yang, 2024) introduce element-wise confidence measuring the fine-grained confidence of each entity or relation of a hyper-relational fact using self-attention networks for link prediction. However, these above approaches do not consider the hierarchical attention, which cannot learn the different importance of different entities under diverse relations for target entities. KHGT (Xia, Huang et al., 2021) introduces a knowledge-enhanced hierarchical graph transformer network for multi-behavior recommendations. DHAN (Zhao, Wei et al., 2023) designs a hierarchical attention architecture including intra-type attention and inter-type attention to learn the same type of nodes and different types of neighbor in bi-typed multi-relational graphs. HAHE (Luo et al., 2023) introduces global-level and local-level attentions for hyper-relational KGs learning. However, these GCN-based KGM models are trained in a supervised manner, which has limited ability for unlabeled graphs.

GCL-based multi-relational graph learning. GCL-based models aim to generate graph augmented views with discrepancies by different graph augmentations. Some node-level augmentation methods randomly perturb graph topology to generate augmented views, such as SGL (Wu et al., 2021), SimGCL (Yu et al., 2022), NCL (Lin, Tian, Hou, & Zhao, 2022), etc. However, node-level augmentation methods may drop an important edge and can heavily damage the graph topology resulting in sub-optimal performance of downstream tasks. Some feature-level augmentation methods appear. Fox example, DGI (Velickovic et al., 2019) shuffles row-wise node attributes to augment the original graph, while GRACE (Zhu et al., 2020) removes edges and masking attributes to corrupt graphs for data augmentation. To improve GRACE, GCA (Zhu et al., 2021) assigns different probabilities to remove edges and mask attributes adaptively for data augmentation. Recent studies have introduced contrastive learning (CL) into multi-relational graph learning to handle the label sparsity problem with self-supervised signals. Fox example, NC-KGE (Fan, Yang, Xu, & Chen, 2023) introduces a node-based CL approach for knowledge graph learning. CMGNN (Fang, Zhang et al., 2023) encodes two views containing connected neighbors and a knowledge graph diffusion for multi-modal knowledge graph learning. VMCL (Li et al., 2023) designs two CL objectives that work across entities and meta-KGs to simulate the transfer mode for inductive knowledge graph embedding. For effectively multi-relational recommendations, CKGC (Cao et al., 2022) develops a cross-modal KGs contrastive learning method from descriptive attributes and structural connections for recommendations.

KGCL (Yang, Huang, Xia, & Li, 2022) proposes a contrastive learning model that eases the information noise for knowledge graph-enhanced using recommendations. CML (Wei et al., 2022) presents a multi-behavior contrastive learning model for recommendations. RCL (Wei, Xia, & Huang, 2023) performs behavior-level augmentation for multi-relational recommendation. In addition, some other application domains of multi-relational graph learning models appear. Fox example, ConvQA (Kacupaj, Singh, Maleshkova, & Lehmann, 2022) presents a contrastive learning method to rank KG paths for conversational question answering. MRCGNN (Xiong et al., 2023) generates two contrastive views by randomly shuffling edge relations and node features for multi-relational drug–drug interaction event prediction. KRACL1 (Tan et al., 2023) utilizes contrastive loss with cross entropy loss to ease the sparsity in KGs for sparse knowledge graph completion. However, most of these existing models mainly utilize human-designed graph augmentations for domain-specific datasets. Their applicability to multi-relational graph datasets from diverse domains is rather limited.

3. Problem statement

A multi-relational graph can be represented as $G = (V, R, \xi)$, where V defines the set of entities (nodes), R defines the set of relations (types of edges), ξ defines the set of facts with the form of triples: $\{(s, r, o) \in V \times R \times V\}$, respectively. The s and o represent the source and object entities, respectively, and r represents the types of relations.

As discussed in Section 1, research on KGs based on hierarchical attention are insufficient. Furthermore, designing a learnable graph augmentation strategy for diverse application domain multi-relational graph datasets and designing a contrastive lose strategies for a high-quality node learning are the current research challenges. Therefore, multi-relational graph learning in this paper is formally defined as follows:

Given a multi-relational graph $G = (V, R, \xi)$, we first adopt the proposed MGHAN architecture to learn the importance of different levels between entities V for extracting the local graph dependency. Then, variant MGHAN-guided augmentation generates two learnable contrast views and local higher-order relations-guided generates a subgraph contrastive loss for unsupervised high-quality node learning. The local higher-order relations of anchors are obtained by calculating all k-hop neighbors of anchor in same view Finally, the goal of our multi-relational graph learning is to learn representation vectors H of entities utilizing the strategies described above for downstream multi-relationship prediction tasks.

4. Methods

In this section, we elaborately enumerate all components of the proposed model MRGCL that is displayed in Fig. 1.

4.1. Multi-relational graph learning

Most existing multi-relational graph learning models focus on GCN (Kipf & Welling, 2017) to learn entity and relation representations, which has shown superior performance for downstream tasks. In this section, we propose a Multi-relational Graph Hierarchical Attention Networks (MGHAN) based on GCN to learn entity representations. which consists of entity-level, relation-level, and layer-level attentions and can learn the importance of different levels between entities. Each propagation layer i is defined as Formula (1).

$$\mathrm{h}_{\mathrm{v}}^{(\mathrm{i}+1)} = \sigma\Big(\sum_{r\in R} b_v^r \sum_{u\in N_v^r} a_{vu}^r \mathrm{W}_{\mathrm{r}}^{(\mathrm{i})} \mathrm{h}_{\mathrm{u}}^{(\mathrm{i})} + \mathrm{W}_0^{(\mathrm{i})} \mathrm{h}_{\mathrm{v}}^{(\mathrm{i})}\Big) \tag{1}$$

where $\mathrm{h}_{\mathrm{v}}^{(\mathrm{i})}$ and $\mathrm{h}_{\mathrm{u}}^{(\mathrm{i})}$ represent the vector representations of node v and node u in layer $i \in I$, respectively. σ represents an activation function and N_v^r represents the set of neighbor nodes of v under a specific relation $r \in R$. $\mathrm{W}_0^{(\mathrm{i})}$ and $\mathrm{W}_{\mathrm{r}}^{(\mathrm{i})}$ represent trainable weight matrices in layer i. We adopt a_{vu}^r to define the entity-level attention describing the importance of node (v, u) in relation type r, which can be constructed by Formula (2):

$$a_{vu}^r = \frac{exp(\sigma(\mathrm{a}_{\mathrm{r}}^{\mathrm{T}} \left[\mathrm{h}_{\mathrm{v}}^{\mathrm{r}} \parallel \mathrm{h}_{\mathrm{u}}^{\mathrm{r}}\right]))}{\sum_{w\in N_v^r} exp(\sigma(\mathrm{a}_{\mathrm{r}}^{\mathrm{T}} \left[\mathrm{h}_{\mathrm{v}}^{\mathrm{r}} \parallel \mathrm{h}_{\mathrm{w}}^{\mathrm{r}}\right]))} \tag{2}$$

where $\mathrm{h}_{\mathrm{v}}^{\mathrm{r}}$ and $\mathrm{h}_{\mathrm{u}}^{\mathrm{r}}$ define the input representation vectors of nodes v and u in a specific relation type r, respectively. $\parallel$ defines concatenation operators, a_{r} defines weight vector parameters of the attention function for relation type r. In this way, a_{vu}^r represents the importance coefficient that based on softmax function over the neighborhoods in a specific relation type r to reveal the importance of node u to node v. We adopt b_v^r to the define relation-level attentions describing the importance of different relation-type for a specific entity v, which can be constructed by Formula (3):

$$b_v^r = \frac{exp(\mathrm{q}^{\mathrm{T}} \cdot \sigma(\left[\mathrm{W}_{\mathrm{r}} \mathrm{h}_{\mathrm{v}}^{\mathrm{r}} + \mathrm{b}_{\mathrm{r}}\right]))}{\sum_{r=1}^{R} exp(\mathrm{q}^{\mathrm{T}} \cdot \sigma(\left[\mathrm{W}_{\mathrm{r}} \mathrm{h}_{\mathrm{v}}^{\mathrm{r}} + \mathrm{b}_{\mathrm{r}}\right]))} \tag{3}$$

where q^{T} defines the entity-level attention vector. R defines the set of relation types, W_{r} and b_{r} are trainable parameters. In this way, b_v^r can identify the importance of different relation-type for a specific entity v. Considering that the representation vectors at different propagation layers of MGHAN include interactive information of different levels. Hence, we design a layer-level attention c_i to combine these embeddings and get the final embeddings of each entity v, which can be constructed by Formula (4):

$$\mathrm{h}_{\mathrm{v}} = \sum_{i=1}^{I} c_i \mathrm{h}_{\mathrm{v}}^{\mathrm{i}} \tag{4}$$

where c_i defines the layer-level attention for adaptive contribution of the ith layer embedding to the final entity representation, which can be constructed by Formula (5):

$$c_i = \frac{exp(\mathrm{p}^{\mathrm{T}} \cdot \sigma(\left[\mathrm{W}_{\mathrm{i}} \mathrm{h}_{\mathrm{v}}^{\mathrm{i}} + \mathrm{o}_{\mathrm{i}}\right]))}{\sum_{i=1}^{I} exp(\mathrm{p}^{\mathrm{T}} \cdot \sigma(\left[\mathrm{W}_{\mathrm{i}} \mathrm{h}_{\mathrm{v}}^{\mathrm{i}} + \mathrm{o}_{\mathrm{i}}\right]))} \tag{5}$$

where p^{T} defines the layer-level attention vector. I defines the number of propagation layers, W_{i} and o_{i} are trainable parameters. In this way, c_i can identify the importance of different propagation layers of the MGHAN for a specific entity v.

4.2. Learnable multi-relational graph augmentation

As discussed in Section 1, most existing handcraft graph augmentation approaches based on random perturbation would fail to keep the task-relevant information intact in different domains. Designing a learnable graph augmentation strategy for multi-relational graph datasets from diverse domains remains a challenge. In this section, we utilize the variant MGHAN to automatically learn two graph augmented views with adaptive topology. We first remove the entity-level attention a_{vu}^r from MGHAN to learn contrast view 1 and each propagation layer i is defined as Formula (6).

$$\hat{\mathrm{h}}_{\mathrm{v}}^{(\mathrm{i}+1)} = \sigma\Big(\sum_{r\in R} b_v^r \sum_{u\in N_v^r} \hat{\mathrm{W}}_{\mathrm{r}}^{(\mathrm{i}+1)} \hat{\mathrm{h}}_{\mathrm{u}}^{(\mathrm{i})} + \hat{\mathrm{W}}_0^{(\mathrm{i})} \hat{\mathrm{h}}_{\mathrm{v}}^{(\mathrm{i})}\Big) \tag{6}$$

where $\hat{\mathrm{h}}_{\mathrm{v}}^{(\mathrm{i})}$ and $\hat{\mathrm{h}}_{\mathrm{u}}^{(\mathrm{i})}$ define the input representation vectors of node v and u for a specific relation type r in the contrast view 1 separately and the initial value is the same as the main view. We employ the similar MGHAN parameters to obtain the reconstructed view $\hat{G} = (\hat{V}, \hat{R}, \hat{\xi})$. $\hat{\mathrm{W}}_0^{(\mathrm{i})}$ and $\hat{\mathrm{W}}_{\mathrm{r}}^{(\mathrm{i}+1)}$ represent trainable weight matrices in layer i. Then, we remove both entity-level attention a_{vu}^r and relation-level attention b_v^r from MGHAN to learn contrast view 2 and each propagation layer i is defined as Formula (7).

$$\bar{\mathrm{h}}_{\mathrm{v}}^{(\mathrm{i}+1)} = \sigma\Big(\sum_{r\in R} \sum_{u\in N_v^r} \bar{\mathrm{W}}_{\mathrm{r}}^{(\mathrm{i})} \bar{\mathrm{h}}_{\mathrm{u}}^{(\mathrm{i})} + \bar{\mathrm{W}}_0^{(\mathrm{i})} \bar{\mathrm{h}}_{\mathrm{v}}^{(\mathrm{i})}\Big) \tag{7}$$

where $\bar{h}_v^{(i)}$ and $\bar{h}_u^{(i)}$ define the input representation vectors of node v and u for a specific relation type r in the contrast view 2 separately and the initial value is the same as the main view. We also employ the similar MGHAN parameters to obtain the reconstructed view $\bar{G} = (\bar{V}, \bar{R}, \bar{\xi})$. $\bar{W}_0^{(i)}$ and $\bar{W}_r^{(i)}$ represent trainable weight matrices in layer i.

MRGCL utilize the variant MGHAN to yield two learnable graph augmented views with adaptive topology. Such learnable augmentation can be automatically compatible with diverse graph datasets without prior domain knowledge. Moreover, in contrast to inappropriate handcraft graph augmentations which might heavily harm the original topology, the learnable augmented views would keep the same nodes and edges as the original graph but with different adaptive edge weights. In addition, the two reconstructed views do not share weights with the main view MGHAN, and the weights are obtained by learning.

4.3. Subgraph contrastive loss

To train the model parameters, we designed a subgraph contrastive loss to contrast two augmented view embeddings with main-view embeddings. We define the contrastive loss for view 1 as Formula (8).

$$\hat{L} = -\frac{1}{|V| + |\hat{V}|}\left(\sum_{v \in V}[log\, D(\mathrm{h_v}, \mathrm{e})] + \sum_{u \in \hat{V}}[log(1 - D(\hat{\mathrm{h}}_u, \mathrm{g}))]\right) \tag{8}$$

where g defines global graph representations of the main-view $G = (V, R, \xi)$, which can be obtained by a readout function $\mathrm{g} = \Gamma(\mathrm{H})$ (Binkowski, Sawczyn, Janiak, Bielak, & Kajdanowicz, 2023). e defines subgraph representations of anchors in the main-view $G = (V, R, \xi)$, which can be obtained by calculating all k-hop neighbor embeddings of anchors in same view. For anchor v, the subgraph representations e can be obtained from formula (9).

$$e = \frac{\sum_{i \in N_v^k} h_i}{k} \tag{9}$$

where N_v^k represents the set of all k-hop neighbors of node v and h_i represents embeddings of node i. $D(\mathrm{h_v}, \mathrm{e}) = \sigma(\mathrm{h_v^T W e})$, where W is a trainable parameter matrix. We define the contrastive loss for view 2 in a similar way as shown in Formula (10).

$$\bar{L} = -\frac{1}{|V| + |\bar{V}|}\left(\sum_{v \in V}[log\, D(\mathrm{h_v}, \mathrm{e})] + \sum_{u \in \bar{V}}[log(1 - D(\bar{\mathrm{h}}_u, \mathrm{g}))]\right) \tag{10}$$

The training goal of contrastive learning is to maximize the consistency between H and e, and the difference between $\hat{\mathrm{H}}/\bar{\mathrm{H}}$ and g, where $\hat{\mathrm{H}}/\bar{\mathrm{H}}$ represents the embedding vectors of the contrast view $\hat{G} = (\hat{V}, \hat{R}, \hat{\xi})$ and $\bar{G} = (\bar{V}, \bar{R}, \bar{\xi})$, respectively. In this way, our contrastive loss can pull subgraph representation of anchors together for guiding a high-quality node learning. For a specific multi-relationship prediction task, we jointly optimize the contrastive losses with main objective function, which can be constructed by Formula (11):

$$L = L_m + \theta_1 \hat{L} + \theta_2 \bar{L} \tag{11}$$

where θ_1 and θ_2 are hyper-parameter for control contribution of the contrastive losses. L_m is the main objective function, which can be constructed by Formula (12).

$$L_m = -\sum_{(v,u) \in \Omega} \sum_{r \in R} y^r_{(v,u)} log\ \hat{y}^r_{(v,u)} \tag{12}$$

where Ω defines training sets, $y^r_{(v,u)}$ represents the probability that the entity pair (v, u) belongs to relation type r and $\hat{y}^r_{(v,u)}$ represents corresponding label.

5. Evaluation

In this section, we conduct comprehensive experiments to verify the performance of our MRGCL framework in three application domains, namely recommender systems, drug–drug interaction event prediction (DDI event prediction) for drug discovery and multi-relational reasoning over KGs.

Table 1
Datasets.

Datasets		Entities	Relations	Interactions
Domain 1	IJCAI	58,011	4	199,654
Domain 2	Deng's	570	65	37,264
	Ryu's	1700	86	191,570
Domain 3	WN18RR	40,943	11	93,003
	FB15k-237	14,541	237	310,116

5.1. Experimental settings

5.1.1. Datasets

We first select the IJCAI dataset (Wei et al., 2023) for recommendations. Then, we select Deng's and Ryu's datasets (Xiong et al., 2023) for DDI event prediction. Finally, we select WN18RR (Dettmers, Minervini, Stenetorp, & Riedel, 2018) and FB15k-237 (Bordes, Usunier, García-Durán, Weston, & Yakhnenko, 2013) datasets for multi-relational reasoning. The statistical attributes of three domain datasets are shown in Table 1.

- **IJCAI:** It is an online retailing dataset containing 22,438 user entities and 35,573 item entities with four types of user online activities, and 199,654 interactions.
- **Deng's:** It is a drug–drug interactions datasets containing 570 drugs, 65 types relations, and 37,264 interactions.
- **Ryu's:** It is also a drug–drug interactions datasets containing 1700 drugs, 86 types relations, and 191,570 interactions.
- **WN18RR:** It is a hierarchical WordNet relational graph containing 40,943 words, 11 semantic relations, and 93,003 interactions.
- **FB15k-237:** It is a knowledge graph extracted from Wikipedia, etc., containing 14,541 entities, 237 relations, and 310,116 interactions.

5.1.2. Baselines

We choose different baselines in three application domains to demonstrate the performance of the MRGCL model, and baselines for DDI event prediction are shown below.

- **RGCN** (Schlichtkrull et al., 2018)**:** It presents a relational GCN to address the highly multi-relational data feature of KGs.
- **TrimNet-DDI** (Li et al., 2021)**:** It utilizes a TrimNet (Li et al., 2021) to learn drug embeddings for DDI event prediction.
- **MUFFIN** (Chen et al., 2021)**:** It fuses multi-scale drug feature to learn drug embeddings for DDI event prediction.
- **SSI-DDI** (Nyamabo et al., 2022)**:** It employs multiple GAT layers with a co-attention to learn embeddings of drug pairs.
- **MRCGNN** (Xiong et al., 2023)**:** It generates two contrastive views by randomly shuffling edge relations and node features for DDI event prediction.

We select the following SOTA methods for multi-relational reasoning over KGs as baselines.

- **A2N** (Bansal et al., 2019)**:** It presents a novel attention-based strategy to embed query-dependent entity representations in KGs.
- **SACN** (Shang et al., 2019)**:** It adopts an encoder of a weighted graph convolutional network utilizing node structure, node attributes and edge relation types, and a decoder of a convolutional network for knowledge base completions.
- **KBAT** (Nathani et al., 2019)**:** It employs entity and relation features to construct a attention-based feature embedding framework for relation prediction.
- **CompGCN** (Vashishth et al., 2020)**:** It adopts a variety of entity-relation composition operations based on GCN architectures to learn entities and relations in KGs.

Table 2
Parameter explanation.

Parameter	Explanation
θ_1/θ_2	Controlling contribution of the contrastive losses
k	Defining k-hop neighbors of an anchor
R	Defining the set of relation types
a^r_{vu}	Defining the entity-level attention
b^r_v	Defining relation-level attention
c_i	Defining the layer-level attention

- **GGPN** (Chen, Fang et al., 2022)**:** It proposes a novel multi-relational graph Gaussian Process network with attention for multi-relational graph representation learnings.
- **NMuR** (Khatir et al., 2023)**:** It utilizes the proposed nonlinear hyperbolic normalization to the multi-relational graph MuR (Balazevic et al., 2019) for multi-relational reasoning.

We select the following baselines for recommendations.

- **KHGT** (Xia, Huang et al., 2021)**:** It introduces a knowledge-enhanced hierarchical graph transformer network for multi-behavior recommendations.
- **MBGMN** (Xia, Xu et al., 2021)**:** It merges the multi-behavior pattern to construct a graph meta-network for multi-behavior recommendations.
- **EHCF** (Chen et al., 2020)**:** It presents a novel non-sampling transfer learning as additional supervision signals for recommendations.
- **CML** (Wei et al., 2022)**:** It presents a multi-behavior contrastive learning model for recommendations.
- **RCL** (Wei et al., 2023)**:** It learns behavior-level augmentation by a proposed dynamic cross-relational memory network for multi-relational recommendations.

5.1.3. Parameter settings

We set the hyper-parameter for control contribution of the contrastive losses $\theta_1 = 0.2$ and $\theta_2 = 0.15$ for the IJCAI dataset; $\theta_1 = 0.1$ and $\theta_2 = 0.05$ for Deng's and Ryu's datasets; $\theta_1 = 0.2$ and $\theta_2 = 0.1/0.15$ for FB15K-237 and WN18RR datasets. The learning rate was set 0.001. The k was set 2 for the IJCAI dataset and 3 for others. Some important parameters and explanations are shown in Table 2. The baseline parameters are set to their default values. We employ the Accuracy, Macro-F1, Macro-Rec., and Macro-Pre. Xiong et al. (2023) to estimate the MRGCL model for DDI event prediction, MRR, Hits@3, and Hits@10 (Khatir et al., 2023) for multi-relational reasoning, and NDCG@N and HR@N (Wei et al., 2023) with the default $N = 10$ for multi-relational recommendations. We run 5 times experiments reporting average metrics. The experiments are run on the Ubuntu 20.04.6 operating system with a Intel(R) Xeon(R) Gold 5317 CPU @ 3.00 GHz machine, 512 GB memory, Tesla A100 80G, and Python 3.9. Our datasets and source code are published.[1]

5.2. DDI event prediction

Table 3 displays MRGCL prediction performance on two datasets and also for the five baselines, with the best results depicted in bold. Our model achieves the best predictive performance as shown in Table 3. RGCN, TrimNet-DDI, and MUFFIN consider drug structural features and adopt deep learning model to learn drug. However, these methods ignore the attention mechanism, resulting in unsatisfactory performance. Even though SSI-DDI utilizes multiple GAT layers with a co-attention to learn embeddings of drug pairs, our model outperforms it because our model can learn the importance of different levels between entities. MRCGNN generates two contrastive views by randomly shuffling edge relations and node features. In contrast, MRCGNN adopts handcraft graph augmentation strategies, which has limited ability to keep the task-relevant information intact. In summary, we can see that MRGCL has the following main advantages over baseline methods. First, MRGCL introduces the MGHAN architecture containing entity-level, relation-level, and layer-level attentions to learn the importance of different levels between entities. And more importantly, MRGCL adopts a learnable contrastive augmentation to keep task-related information intact. Furthermore, MRGCL introduces a subgraph contrastive loss to generate positive pairs per anchor, which can extract local higher-order relations of anchors for a high-quality node learning. Thus, our MRGCL model beats all the baselines models in the DDI event prediction tasks.

5.3. Multi-relational reasoning over KGs

Table 4 shows the multi-relational reasoning results of MRGCL and baselines on WN18RR and FB15K-237 datasets. The experiment results indicate that the MRGCL outperforms all the baselines in both datasets. Specifically, MRGCL improves the performance by 1.5%, 0.6% and 0.7% on MRR, Hits@3 and Hits@10, respectively, compared with the second-best baseline GGPN on the FB15K-237 dataset. In the WN18RR dataset, the average Hits@10 value of MRGCL is slightly lower than NMuR, but the other two metrics of the MRGCL are significantly superior to the NMuR. We can also obtain similar results compared with the second-best baseline GGPN on the WN18RR dataset. However, the overall performance of GGPN is significantly better than NMuR and lower than our MRGCL. The experiment results shown that MRGCL is effective for multi-relational reasoning tasks. The reason may be that SACN, CompGCN, and NMuR ignore the attention mechanism, and even though A2N, KBAT, and GGPN consider the attention mechanism, these models do not construct hierarchical attention to learn the different levels of importance between entities. And more importantly, our MRGCL adopts a learnable contrastive augmentation to keep task-related information intact, which can automatically adapt for diverse application domains. Hence, experimental results in multi-relational reasoning show that our MRGCL can beat various SOTA methods.

5.4. Multi-relational recommendation

Table 5 shows the multi-relational recommendation results of MRGCL and baselines on the IJCAI dataset. The experiment results also demonstrate that the MRGCL outperforms all the baselines in the IJCAI dataset. Specifically, KHGT and MBGMN mainly utilize a GCN-based architecture to learn embedded representations of multi-relational graphs, and the performance is significantly lower than that of models based on self-supervised paradigms, such as EHCF and CML. Moreover, RCL adaptively learns behavior-level augmentation and beats both EHCF and CML. However, it is still defeated by our MRGCL model, probably because our model designs a subgraph contrastive loss to generate the model's supervisory signal, which can yield high-quality node vectors. Thus, experimental results in multi-relational recommendations demonstrate that our MRGCL can beat diverse SOTA approaches.

5.5. Ablation study

In this section, we design an ablation study to verify the effectiveness of components including contrastive learning, subgraph contrastive loss and hierarchical attention in predicting DDI events and multi-relational recommendations. We execute 5 experiments to show the average Accuracy, Macro-F1, Macro-Rec., and Macro-Pre. values on the Deng's and Ryu's datasets, and average HR@10 and NDCG@10 on the IJCAI dataset.

[1] https://github.com/Legendary-L/MRGCL.

Table 3
Results of MRGCL for DDI event prediction.

Methods	Deng's dataset				Ryu's dataset			
	Accuracy	Macro-F1	Macro-Rec.	Macro-Pre.	Accuracy	Macro-F1	Macro-Rec.	Macro-Pre.
RGCN	0.870	0.703	0.688	0.750	0.928	0.849	0.829	0.888
TrimNet-DDI	0.857	0.655	0.636	0.705	0.935	0.829	0.813	0.863
MUFFIN	0.827	0.525	0.484	0.620	0.951	0.857	0.834	0.898
SSI-DDI	0.787	0.422	0.390	0.514	0.901	0.666	0.629	0.751
MRCGNN	0.898	0.779	0.769	0.810	0.957	0.889	0.873	0.922
MRGCL	**0.902**	**0.806**	**0.780**	**0.859**	**0.961**	**0.917**	**0.905**	**0.939**

Table 4
Results of MRGCL for multi-relational reasoning.

Methods	WN18RR dataset			FB15k-237 dataset		
	MRR	Hits@3	Hits@10	MRR	Hits@3	Hits@10
A2N	0.450	0.460	0.510	0.317	0.348	0.486
SACN	0.470	0.480	0.540	0.350	0.390	0.540
KBAT	0.410	0.451	0.501	0.318	0.362	0.499
CompGCN	0.479	0.494	0.546	0.355	0.390	0.535
NMuR	0.447	0.481	**0.579**	0.322	0.355	0.506
GGPN	0.481	0.499	0.548	0.361	0.396	0.540
MRGCL	**0.493**	**0.519**	0.570	**0.376**	**0.402**	**0.547**

Table 5
Results of MRGCL for multi-relational recommendation on IJCAI datasets.

	NDCG@10	HR@10
KHGT	0.145	0.278
MBGMN	0.176	0.329
EHCF	0.207	0.362
CML	0.235	0.410
RCL	0.312	0.510
MRGCL	**0.319**	**0.541**

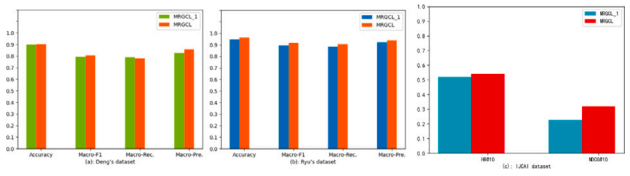


Fig. 2. Effect of contrastive learning.

5.5.1. *Effect of contrastive learning*

To verify the effectiveness of contrastive learning for DDI event predict and multi-relational recommendations, we remove the $\hat{L}$ and $\bar{L}$ losses from L (denoted as MRGCL _1) to verify its contribution to the MRGCL model. As shown in Fig. 2, the average Accuracy of MRGCL is 0.3% higher than MRGCL_1, the average Macro-F1 value is 1.2% higher than MRGCL_1, and the average Macro-Pre. value is 3.4% higher than MRGCL_1 on the Deng's dataset; the average Accuracy of MRGCL is 1.5% higher than MRGCL_1, the average Macro-F1 value is 2.5% higher than MRGCL_1, the average Macro-Rec. value is 2.0% higher than MRGCL_1, and the average Macro-Pre. value is 1.6% higher than MRGCL_1 on the Ryu's dataset; the average HR@10 of MRGCL is 2.2% higher than MRGCL_1, and the average NDCG@10 value is 9.1% higher than MRGCL_1 on the IJCAI dataset. In the Deng's dataset, the average Macro-Rec. value of MRGCL is lower 0.9% than MRGCL_1, but the other three metrics of the MRGCL are superior to the MRGCL_1. In addition, all metrics of MRGCL are superior to the MRGCL_1 in the Ryu's dataset. Therefore, the overall performance on both datasets of our MRGCL model is superior to the MRGCL_1. The experimental results demonstrate the effectiveness of the contrastive learning. The reason may be that MRGCL utilizes the proposed variant MGHAN for learnable contrastive augmentation.

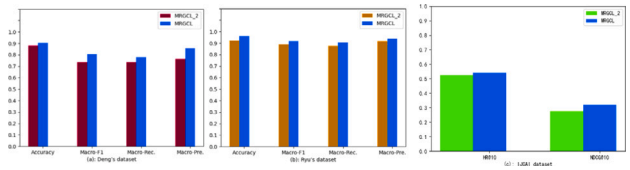


Fig. 3. Effect of subgraph contrastive loss.

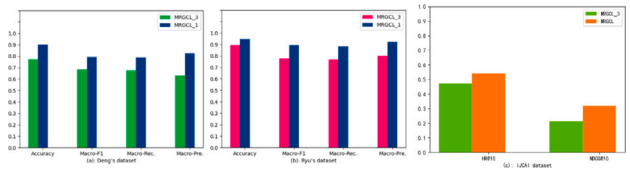


Fig. 4. Effect of hierarchical attention.

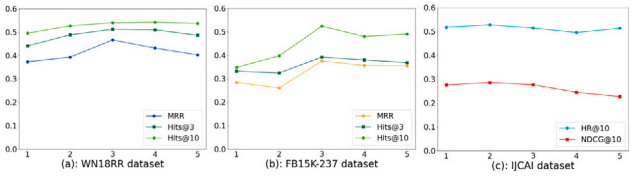


Fig. 5. Hyperparameter analysis for k.

5.5.2. *Effect of subgraph contrastive loss*

To verify the effectiveness of the subgraph contrastive loss strategy, we replace all subgraph representation of anchors e with the global representation of anchors g (denoted as MRGCL _2) in subgraph contrastive loss to verify its contribution to the MRGCL model. As shown in Fig. 3, the average Accuracy of MRGCL is 2% higher than MRGCL_2, the average Macro-F1 value is 7.2% higher than MRGCL_2, the average Macro-Rec. value is 4.7% higher than MRGCL_2, and the average Macro-Pre. value is 9.5% higher than MRGCL_2 on the Deng's dataset; the average Accuracy of MRGCL is 3.9% higher than MRGCL_2, the average Macro-F1 value is 2.6% higher than MRGCL_2, the average Macro-Rec. value is 2.9% higher than MRGCL_2, and the average Macro-Pre. value is 2.3% higher than MRGCL_2 on the Ryu's dataset; the average HR@10 of MRGCL is 1.6% higher than MRGCL_2, and the average NDCG@10 value is 4.2% higher than MRGCL_2 on the IJCAI dataset. The experimental results demonstrate the remarkable effectiveness of the subgraph contrastive loss. The reason may be that MRGCL calculates the subgraph embeddings as positive pairs for anchors, which can effectively extract local valuable higher-order relations of anchors compared with the global representations.

5.5.3. *Effect of hierarchical attention*

To verify the effectiveness of the hierarchical attention mechanisms, we remove all hierarchical attentions from MRGCL _1(denoted as MRGCL _3) to verify its contribution to the MRGCL model. As shown in Fig. 4, the average Accuracy of MRGCL_1 is 12.8% higher than MRGCL_3, the average Macro-F1 value is 10.7% higher than MRGCL_3,

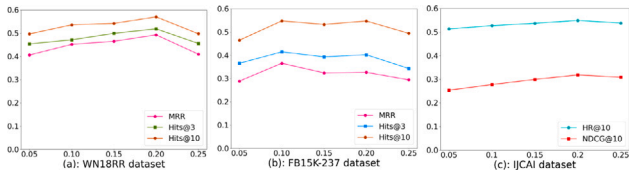


Fig. 6. Hyperparameter analysis for θ_1.

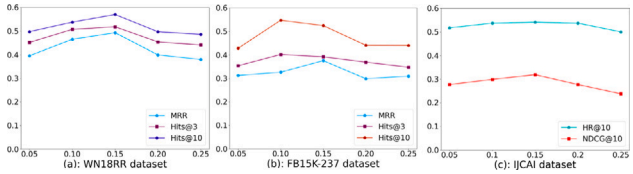


Fig. 7. Hyperparameter analysis for θ_2.

the average Macro-Rec. value is 11.2% higher than MRGCL_3, and the average Macro-Pre. value is 19.7% higher than MRGCL_3 on the Deng's dataset; the average Accuracy of _1 is 5.2% higher than MRGCL_3, the average Macro-F1 value is 11.6% higher than MRGCL_3, the average Macro-Rec. value is 11.5% higher than MRGCL_3, and the average Macro-Pre. value is 11.9% higher than MRGCL_3 on the Ryu's dataset; the average HR@10 of MRGCL_1 is 7% higher than MRGCL_3, and the average NDCG@10 value is 10.5% higher than MRGCL_3 on the IJCAI dataset. The experimental results demonstrate the effectiveness of the hierarchical attention mechanic, which can significantly improve predictive performance on all metrics in the three datasets. The reason may be that MRGCL_1 can identify the importance of different levels between entities, which can learn the different level's contribution to target entity embeddings.

5.6. Hyperparameter analysis

In this section, we conduct hyperparameter analysis, and the performances are shown in Figs. 5, 6, and 7. Specifically, we estimate how different k-hop neighbors and the parameter θ_1 and θ_2 can impact the multi-relational reasoning and recommendations performance. We execute 5 experiments to show the average MRR, Hits@3, and Hits@10 values on the WN18RR and FB15k-237 datasets, and average HR@10 and NDCG@10 on the IJCAI dataset.

Parameter k: We set the parameter k to 1, 2, 3, 4, and 5 to evaluate the multi-relational reasoning and recommendations performance of our MRGCL. Our model employs k to decide the hop of the subgraph. In the experiment, we discovered that a small value of k can get the appropriate results. The experimental results indicate in Fig. 5 that when $k = 3$ on multi-relational reasoning and $k = 2$ on recommendations, our model can achieve the best performance.

Parameter θ_1 **and** θ_2: We change the parameter θ_1 and θ_2 to explore the contribution of different contrastive learning tasks. The θ_1 and θ_2 controls the contribution of contrastive learning. In the experiment, the parameter θ_1 and θ_2 are searched in the range of $\{0.05, 0.10, 0.15, 0.20, 0.25\}$. The experimental results are shown in Figs. 6 and 7 that the best performance is achieved when $\theta_1 = 0.2$ and $\theta_2 = 0.1/0.15$ on both tasks. In the experiment, we found that the larger the values of both parameters, the larger the contribution to contrastive learning, but the worse the performance. The reason may be that a larger value make the model pay too much attention to the contrastive learning task and reduce the focus on the main task, resulting in decreased performance.

6. Conclusion

In this paper, we have proposed an effective learnable augmentation method MRGCL to the graph contrastive learning architecture for multi-relational graph learning. Especially, our MRGCL first presents a MGHAN framework to learn the importance between entities, which consists of entity-level, relation-level, and layer-level attentions and can learn the importance of different levels between entities. We then remove the relation-level attention from MGHAN to learn contrast view 1 and remove both entity-level and relation-level attentions from MGHAN to learn contrast view 2, which can automatically learn two graph augmented views to keep task-related information intact and adapt for diverse application domains. Finally, we design a subgraph contrastive loss to generate positive pairs per anchor, which can extract local higher-order relations of anchors for a high-quality node learning. Comprehensive experiments in three application domain multi-relational datasets demonstrate the superiority of our MRGCL over various SOTA approaches. Our future work will explore semantic relationships (Ding, Zhang, Ma, Zhang, & Zhong, 2024; Zhu et al., 2024) of multi-relational graph to construct new contrastive losses for a high-quality multi-relational graph learning.

CRediT authorship contribution statement

Xian Mo: Writing – original draft, Supervision, Methodology, Funding acquisition, Conceptualization. **Jun Pang:** Writing – review & editing, Methodology, Conceptualization. **Binyuan Wan:** Validation, Methodology, Formal analysis, Data curation. **Rui Tang:** Writing – review & editing, Methodology, Conceptualization. **Hao Liu:** Writing – review & editing, Methodology, Conceptualization. **Shuyu Jiang:** Writing – review & editing, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Our datasets and source code are published at https://github.com/Legendary-L/MRGCL.

Acknowledgments

The work is financed by the National Natural Science Foundation of China (62306157, 62202320, 62102201), the Natural Science Foundation of Ningxia Province, China (2024A1315, 2021A0966, 2022AAC05010), the Natural Science Foundation of Sichuan Province, China (2024NSFSC1449), the Fundamental Research Funds for the Central Universities, China (2023SCU12126), and the Science and Technology Leading Talent Training Projects of Ningxia, China (2022GKLRLX03).

References

Balazevic, I., Allen, C., & Hospedales, T. M. (2019). Multi-relational poincaré graph embeddings. In *Proceedings of the 32th annual conference on neural information processing systems* (pp. 4465–4475).

Bansal, T., Juan, D., Ravi, S., & McCallum, A. (2019). A2N: Attending to neighbors for knowledge graph inference. In *Proceedings of the 57th conference of the association for computational linguistics* (pp. 4387–4392).

Binkowski, J., Sawczyn, A., Janiak, D., Bielak, P., & Kajdanowicz, T. (2023). Graph-level representations using ensemble-based readout functions. In *Proceedings of the 23th international conference on computational science* (pp. 393–405).

Bordes, A., Usunier, N., García-Durán, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In *Proceedings of the 27th annual conference on neural information processing systems 2013* (pp. 2787–2795).

Cao, X., Shi, Y., Wang, J., Yu, H., Wang, X., & Yan, Z. (2022). Cross-modal knowledge graph contrastive learning for machine learning method recommendation. In *Proceedings of the 30th ACM international conference on multimedia* (pp. 3694–3702).

Chen, G., Fang, J., Meng, Z., Zhang, Q., & Liang, S. (2022). Multi-relational graph representation learning with Bayesian Gaussian process network. In *Proceedings of the 36th AAAI conference on artificial intelligence* (pp. 5530–5538).

Chen, Y., Ma, T., Yang, X., Wang, J., Song, B., & Zeng, X. (2021). MUFFIN: multi-scale feature fusion for drug-drug interaction prediction. *Bioinformatics*, *37*(17), 2651–2658.

Chen, Y., Yang, Y., Wang, Y., Bai, J., Song, X., & King, I. (2022). Attentive knowledge-aware graph convolutional networks with collaborative guidance for personalized recommendation. In *Proceedings of the 38th IEEE international conference on data engineering* (pp. 299–311).

Chen, C., Zhang, M., Zhang, Y., Ma, W., Liu, Y., & Ma, S. (2020). Efficient heterogeneous collaborative filtering without negative sampling for recommendation. In *Proceedings of the 34th AAAI conference on artificial intelligence* (pp. 19–26).

Dai, G., Wang, X., Zou, X., Liu, C., & Cen, S. (2022). MRGAT: Multi-relational graph attention network for knowledge graph completion. *Neural Networks*, *154*, 234–245.

Dettmers, T., Minervini, P., Stenetorp, P., & Riedel, S. (2018). Convolutional 2D knowledge graph embeddings. In *Proceedings of the 32th AAAI conference on artificial intelligence* (pp. 1811–1818).

Ding, X., Zhang, H., Ma, C., Zhang, X., & Zhong, K. (2024). User identification across multiple social networks based on naive Bayes model. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(3), 4274–4285.

Fan, Z., Yang, Y., Xu, M., & Chen, H. (2023). Node-based knowledge graph contrastive learning for medical relationship prediction. CoRR abs/2310.10138.

Fang, Y., Li, X., Ye, R., Tan, X., Zhao, P., & Wang, M. (2023). Relation-aware graph convolutional networks for multi-relational network alignment. *ACM Transactions on Intelligent Systems and Technology*, *14*(2), 37:1–37:23.

Fang, Q., Zhang, X., Hu, J., Wu, X., & Xu, C. (2023). Contrastive multi-modal knowledge graph representation learning. *IEEE Transactions on Knowledge and Data Engineering*, *35*(9), 8983–8996.

Hu, Z., Gutiérrez-Basulto, V., Xiang, Z., Li, R., & Pan, J. Z. (2023). HyperFormer: Enhancing entity and relation interaction for hyper-relational knowledge graph completion. In *Proceedings of the 32nd ACM international conference on information and knowledge management* (pp. 803–812).

Jiang, Y., Huang, C., & Huang, L. (2023). Adaptive graph contrastive learning for recommendation. In *Proceedings of the 29th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 4252–4261).

Kacupaj, E., Singh, K., Maleshkova, M., & Lehmann, J. (2022). Contrastive representation learning for conversational question answering over knowledge graphs. In M. A. Hasan, & L. Xiong (Eds.), *Proceedings of the 31st ACM international conference on information & knowledge management* (pp. 925–934).

Khatir, M., Choudhary, N., Choudhury, S., Agarwal, K., & Reddy, C. K. (2023). A unification framework for euclidean and hyperbolic graph neural networks. In *Proceedings of the 32th international joint conference on artificial intelligence* (pp. 3875–3883).

Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th international conference on learning representations* (pp. 1–14).

Li, Q., Joty, S., Wang, D., Feng, S., Zhang, Y., & Qin, C. (2023). Contrastive learning with generated representations for inductive knowledge graph embedding. In *Proceedings of the 61tt annual meeting of the association for computational linguistics* (pp. 14273–14287).

Li, P., Li, Y., Hsieh, C., Zhang, S., Liu, X., Liu, H., et al. (2021). TrimNet: learning molecular representation from triplet messages for biomedicine. *Briefings in Bioinformatics*, *22*(4).

Lin, Z., Tian, C., Hou, Y., & Zhao, W. X. (2022). Improving graph collaborative filtering with neighborhood-enriched contrastive learning. In *Proceedings of the 7th ACM web conference 2022* (pp. 2320–2329).

Liu, L., Chen, Y., Das, M., Yang, H., & Tong, H. (2023). Knowledge graph question answering with ambiguous query. In *Proceedings of the ACM web conference* (pp. 2477–2486).

Luo, H., E, H., Yang, Y., Guo, Y., Sun, M., Yao, T., et al. (2023). HAHE: Hierarchical attention for hyper-relational knowledge graphs in global and local level. In *Proceedings of the 61tt annual meeting of the association for computational linguistics* (pp. 8095–8107).

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444.

Nathani, D., Chauhan, J., Sharma, C., & Kaul, M. (2019). Learning attention-based embeddings for relation prediction in knowledge graphs. In *Proceedings of the 57th conference of the association for computational linguistics* (pp. 4710–4723).

Nyamabo, A. K., Yu, H., Liu, Z., & Shi, J. (2022). Drug-drug interaction prediction with learnable size-adaptive molecular substructures. *Briefings in Bioinformatics*, *23*(1).

Qu, M., Gao, T., Xhonneux, L. A. C., & Tang, J. (2020). Few-shot relation extraction via Bayesian meta-learning on relation graphs. In *Proceedings of the 37th international conference on machine learning* (pp. 7867–7876).

Schlichtkrull, M. S., Kipf, T. N., Bloem, P., van den Berg, R., Titov, I., & Welling, M. (2018). Modeling relational data with graph convolutional networks. In *Proceedings of the 15th international conference on semantic web* (pp. 593–607).

Shang, C., Tang, Y., Huang, J., Bi, J., He, X., & Zhou, B. (2019). End-to-end structure-aware convolutional networks for knowledge base completion. In *Proceedings of the 33th AAAI conference on artificial intelligence* (pp. 3060–3067).

Shen, X., Sun, D., Pan, S., Zhou, X., & Yang, L. T. (2023). Neighbor contrastive learning on learnable graph augmentation. In *Proceedings of the 37th AAAI conference on artificial intelligence* (pp. 9782–9791).

Tan, Z., Chen, Z., Feng, S., Zhang, Q., Zheng, Q., Li, J., et al. (2023). KRACL: Contrastive learning with graph context modeling for sparse knowledge graph completion. In *Proceedings of the ACM web conference 2023* (pp. 2548–2559).

Tian, Y., Sun, C., Poole, B., Krishnan, D., Schmid, C., & Isola, P. (2020). What makes for good views for contrastive learning? In *Proceedings of the 33th advances in neural information processing systems* (pp. 6827–6839).

van den Oord, A., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. CoRR abs/1807.03748.

Vashishth, S., Sanyal, S., Nitin, V., & Talukdar, P. P. (2020). Composition-based multi-relational graph convolutional networks. In *Proceedings of the 8th international conference on learning representations* (pp. 1–16).

Velickovic, P., Fedus, W., Hamilton, W. L., Liò, P., Bengio, Y., & Hjelm, R. D. (2019). Deep graph infomax. In *Proceedings of the 7th international conference on learning representations* (pp. 1–17).

Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., et al. (2019). Heterogeneous graph attention network. In *Proceedings of the world wide web conference* (pp. 2022–2032).

Wei, W., Huang, C., Xia, L., Xu, Y., Zhao, J., & Yin, D. (2022). Contrastive meta learning with behavior multiplicity for recommendation. In *Proceedings of the 15th ACM international conference on web search and data mining* (pp. 1120–1128).

Wei, W., Xia, L., & Huang, C. (2023). Multi-relational contrastive learning for recommendation. In *Proceedings of the 17th ACM conference on recommender systems* (pp. 338–349).

Wu, J., Wang, X., Feng, F., He, X., Chen, L., Lian, J., et al. (2021). Self-supervised graph learning for recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval* (pp. 726–735).

Xia, L., Huang, C., Xu, Y., Dai, P., Zhang, X., Yang, H., et al. (2021). Knowledge-enhanced hierarchical graph transformer network for multi-behavior recommendation. In *Proceedings of the 35th AAAI conference on artificial intelligence* (pp. 4486–4493).

Xia, L., Xu, Y., Huang, C., Dai, P., & Bo, L. (2021). Graph meta network for multi-behavior recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval* (pp. 757–766).

Xiong, Z., Liu, S., Huang, F., Wang, Z., Liu, X., Zhang, Z., et al. (2023). Multi-relational contrastive learning graph neural network for drug-drug interaction event prediction. In *Proceedings of the 37th AAAI conference on artificial intelligence* (pp. 5339–5347).

Yang, Y., Guan, Z., Li, J., Zhao, W., Cui, J., & Wang, Q. (2023). Interpretable and efficient heterogeneous graph convolutional network. *IEEE Transactions on Knowledge and Data Engineering*, *35*(2), 1637–1650.

Yang, Y., Huang, C., Xia, L., & Li, C. (2022). Knowledge graph contrastive learning for recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval* (pp. 1434–1443).

Ye, R., Li, X., Fang, Y., Zang, H., & Wang, M. (2019). A vectorized relational graph convolutional network for multi-relational network alignment. In *Proceedings of the 28th international joint conference on artificial intelligence* (pp. 4135–4141).

Yu, W., Yang, J., & Yang, D. (2024). Robust link prediction over noisy hyper-relational knowledge graphs via active learning. In *Proceedings of the ACM on web conference* (pp. 2282–2293).

Yu, J., Yin, H., Xia, X., Chen, T., Cui, L., & Nguyen, Q. V. H. (2022). Are graph augmentations necessary?: Simple graph contrastive learning for recommendation. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval* (pp. 1294–1303).

Yun, S., Jeong, M., Kim, R., Kang, J., & Kim, H. J. (2019). Graph transformer networks. In *Proceedings of the advances in neural information processing systems* (pp. 11960–11970).

Zhao, Y., Du, H., Liu, Y., Wei, S., Chen, X., Zhuang, F., et al. (2023). Stock movement prediction based on bi-typed hybrid-relational market knowledge graph via dual attention networks. *IEEE Transactions on Knowledge and Data Engineering*, *35*(8), 8559–8571.

Zhao, Y., Wei, S., Du, H., Chen, X., Li, Q., Zhuang, F., et al. (2023). Learning bi-typed multi-relational heterogeneous graph via dual hierarchical attention networks. *IEEE Transactions on Knowledge and Data Engineering*, *35*(9), 9054–9066.

Zhou, Y., Chen, X., He, B., Ye, Z., & Sun, L. (2022). Re-thinking knowledge graph completion evaluation from an information retrieval perspective. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval* (pp. 916–926).

Zhu, P., Wang, B., Tang, K., Zhang, H., Cui, X., & Wang, Z. (2024). A knowledge-guided graph attention network for emotion-cause pair extraction. *Knowledge-Based Systems*, *286*, Article 111342.

Zhu, Y., Xu, Y., Yu, F., Liu, Q., Wu, S., & Wang, L. (2020). Deep graph contrastive representation learning. CoRR abs/2006.04131.

Zhu, Y., Xu, Y., Yu, F., Liu, Q., Wu, S., & Wang, L. (2021). Graph contrastive learning with adaptive augmentation. In *Proceedings of the web conference* (pp. 2069–2080).

Zhuang, L., Wang, H., Zhao, J., & Sun, Y. (2023). Adaptive dual graph contrastive learning based on heterogeneous signed network for predicting adverse drug reaction. *Information Sciences*, *642*, Article 119139.