Multiplex Graph Representation Learning via Common and Private Information Mining

Yujie Mo[1*], Zongqian Wu[4*], Yuhuan Chen[1], Xiaoshuang Shi[1], Heng Tao Shen[1, 3], Xiaofeng Zhu[1, 2†]

[1]School of Computer Science and Engineering,
University of Electronic Science and Technology of China, Chengdu 611731, China
[2]Shenzhen Institute for Advanced Study,
University of Electronic Science and Technology of China, Shenzhen 518000, China
[3]Peng Cheng Laboratory, Shenzhen 518000, China
[4]Guangxi Key Lab of Multi-Source Information Mining and Security,
Guangxi Normal University, Guilin 541004, China
moyujie2017@gmail.com

Abstract

Self-supervised multiplex graph representation learning (SMGRL) has attracted increasing interest, but previous SMGRL methods still suffer from the following issues: (i) they focus on the common information only (but ignore the private information in graph structures) to lose some essential characteristics related to downstream tasks, and (ii) they ignore the redundant information in node representations of each graph. To solve these issues, this paper proposes a new SMGRL method by jointly mining the common information and the private information in the multiplex graph while minimizing the redundant information within node representations. Specifically, the proposed method investigates the decorrelation losses to extract the common information and minimize the redundant information, while investigating the reconstruction losses to maintain the private information. Comprehensive experimental results verify the superiority of the proposed method, on four public benchmark datasets.

Introduction

Multiplex graph representation learning (MGRL) is a powerful approach to extracting multiple relationships among nodes in the graph data, and has recently attracted much attention in real applications (Chu et al. 2019; Zhang and Kou 2022; Peng et al. 2022). The multiplex graph can be regarded as a combination of multiple graphs. Different graphs consist of shared node features and different graph structures, and each graph structure focuses on one type of relationship. Therefore, MGRL is available to explore more relationships than traditional graph representation learning (GRL) (Kipf and Welling 2017; Fan et al. 2019).

Since node labels are usually difficult to be obtained, self-supervised multiplex graph representation learning (SMGRL) has emerged as a hot research topic with great success (Hwang et al. 2020; Pan and Kang 2021; Lin et al. 2021). SMGRL aims at extracting the hidden information of the multiplex graph and generating discriminative representations without the help of labeled nodes. Existing SMGRL methods can be broadly classified into two categories, *i.e.,* the contrastive learning within the same graph and the contrastive learning across graphs. The contrastive learning within the same graph aims to improve the quality of node representations by capturing the global properties (Park et al. 2020; Ren et al. 2020; Jing, Park, and Tong 2021). However, it treats every graph independently and ignores the intrinsic correlation (*i.e.,* common information) among different graphs. To alleviate this, the contrastive learning across graphs is designed to extract the common information among different graphs by contrasting their representations (Zhu et al. 2022; Li, Jing, and Tong 2022; Zhou et al. 2022). For instance, to extract the common information, STENCIL (Zhu et al. 2022) conducts multi-view contrastive learning between each graph and the aggregation graph, while CKD (Zhou et al. 2022) conducts the contrastive learning between different graphs.

Despite its effectiveness, the contrastive learning across graphs still has limitations to be addressed. On the one hand, these methods only focus on the common information and ignore the private information in each graph structure. However, it has been demonstrated that private information contains discriminative characteristics, which is important for downstream tasks in real applications such as cross-modal retrieval and image recognition (Xie et al. 2020; Wang et al. 2022). Similarly, private information in graph structures may convey significant connection characteristics between nodes in the multiplex graph. Therefore, ignoring the private information in graph structures may lead to decreased effectiveness in downstream tasks. On the other hand, these methods ignore the redundant information in node representations of each graph. That is, different dimensionalities of node representations may share the same content. As a result, the redundancy may weaken the discriminative information of representations and even lead to collapse (Zbontar et al. 2021).

Based on the above observations, jointly capturing the common information and the private information while minimizing the redundant information in node representations is a possible solution to achieve an effective SMGRL. However, few methods address these issues in a unified framework due

[*]These authors contributed equally.
[†]Corresponding author.

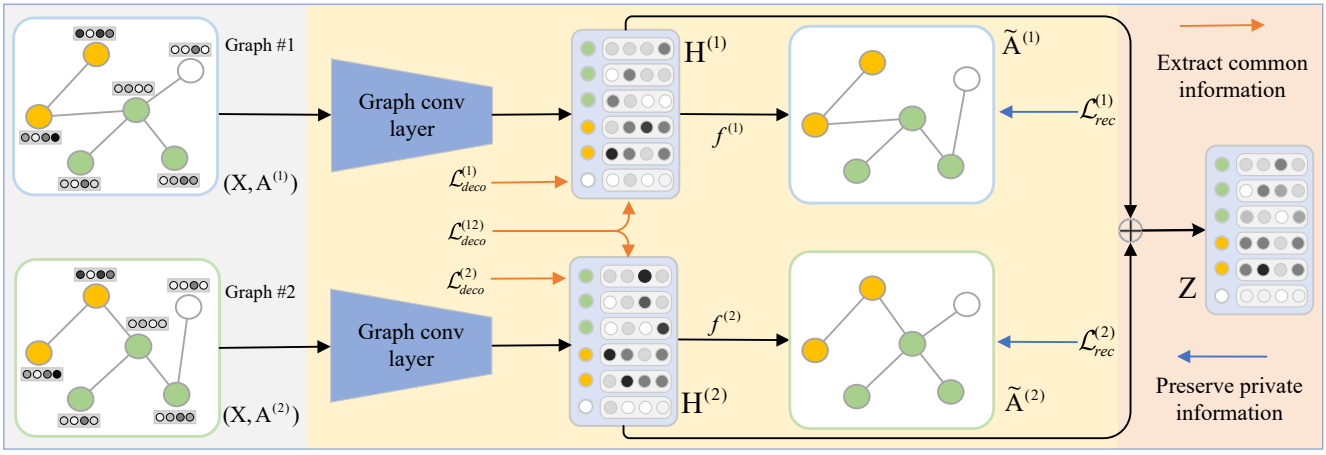

Figure 1: The flowchart of the proposed CPIM on the dataset with two graphs. Specifically, given the node features and graph structures, CPIM first employs the graph convolutional layer to generate the node representations, and then investigates the decorrelation losses (*i.e.,* $\mathcal{L}_{deco}^{(1)}$, $\mathcal{L}_{deco}^{(2)}$, and $\mathcal{L}_{deco}^{(12)}$) to mine the common information among different graphs and minimize the redundant information within node representations, while investigates the reconstruction losses (*i.e.,* $\mathcal{L}_{rec}^{(1)}$ and $\mathcal{L}_{rec}^{(2)}$) to maintain the private information within each graph. Finally, the node representations of different graphs are fused by the attention mechanism to obtain the final representations.

to the following challenges: (i) it is difficult to simultaneously mine the common information and private information in the multiplex graph due to their conflicting nature, and (ii) a new framework needs to be designed to minimize the redundant information within node representations to maintain the effectiveness of the SMGRL method.

In this paper, to address the above issues, we propose a self-supervised multiplex graph representation learning framework via Common and Private Information Mining (CPIM) to conduct effective SMGRL. To do this, CPIM explores both the common information and the private information in the multiplex graph and minimizes the redundant information within node representations in a unified framework, as shown in Figure 1. Specifically, we first employ the graph convolutional layer (Kipf and Welling 2017) to generate node representations for each graph. After that, we investigate the decorrelation losses on the node representations to extract the common information among them and reduce the redundant information within node representations. Meanwhile, we investigate the reconstruction losses on the reconstructed graph structures to preserve the private information within each graph. Finally, we employ the attention mechanism (You et al. 2016) to fuse the node representations from different graphs for downstream tasks.

Compared to previous SMGRL methods, the main contributions of our method are summarized as follows:

- To the best of our knowledge, we make the first attempt to simultaneously explore both the common information and the private information in the multiplex graph in a unified framework for SMGRL.
- The proposed method extracts the common information among different graphs and reduces the redundant information within node representations by the decorrelation losses, and preserves the private information within each graph by the reconstruction losses.
- Extensive experiments on four datasets demonstrate the superiority of the proposed method, compared to twelve comparison methods, on the node classification task.

Related Work

Contrastive Learning

Recently, the contrastive learning has drawn much attention due to its great success in many research areas (Liu et al. 2021a), such as computer vision (Xu et al. 2022a; Zhang et al. 2022; Xu et al. 2022b), natural language processing (Devlin et al. 2019; Gao, Yao, and Chen 2021; Lan et al. 2020) and graph representation learning (Cen et al. 2019; Zhou et al. 2020; Mo et al. 2022). Contrastive learning aims to learn discriminative representations by contrasting the positive pairs and negative pairs without labeled data. As a result, contrastive learning avoids the issue of relying on labels and is able to mine the hidden information in the original data.

Previous contrastive learning methods typically encourage the encoder to learn representations by maximizing the mutual information between the local representations and their related representations. Depending on pretext tasks, the related representations can be either the global representations of the input or the local representations of another view. The contrastive learning between the local and global representations aims to maximize the mutual information between them by contrasting each other. For example, Deep InfoMax (Hjelm et al. 2019) conducts contrastive learning by maximizing the mutual information between a local patch and its global context. CPC (Oord, Li, and Vinyals 2018) achieves great results on the speech recognition task by maximizing

the mutual information between the local audio representations and its global audio representations.

Different from local-global contrastive learning, contrastive learning between the local and local representations aims to learn the invariance between the original local representations and the augmented local representations. For example, CMC (Tian, Krishnan, and Isola 2020) conducts contrastive learning between local representations from multiple views of an image, and learns the invariant information among them. MoCo (He et al. 2020) further enhances the effectiveness of local-local contrastive learning by introducing the momentum mechanism, which substantially increases the number of negative samples. SimCLR (Chen et al. 2020) conducts the contrastive learning between the original view and the augmented view, and argues that data augmentation plays an important role in contrastive learning.

Multiplex Graph Representation Learning

MGRL aims to reveal the relationships between nodes in the graph from multiple aspects (Tang et al. 2015; Park et al. 2020). Thus, MGRL is more relevant to practical applications than GRL since entities in the real world are generally connected by multiple relationships.

The pioneering MGRL methods are mainly based on the proximity among nodes to generate representations. For example, the very first work PMNE (Liu et al. 2017) extends the random walk based strategies (Perozzi, Al-Rfou, and Skiena 2014) to the multiplex graph by maximizing the probabilities of the sampled neighbors. Moreover, Metapath2vec (Dong, Chawla, and Swami 2017) designs a meta-path guided random walk and employs the skip-gram to generate node representations. Despite effectiveness, these methods ignore the node features, which may contain the critical contents (*e.g.,* user profiles in social networks). To alleviate this, HAN (Wang et al. 2019) introduces self-attention mechanisms (Vaswani et al. 2017) to aggregate representations at both the node-level and semantic-level. Moreover, MAGNN (Fu et al. 2020) further improves HAN by utilizing intermediate node features along each meta-path.

Regarding the cost of obtaining labels, self-supervised multiplex graph representation learning (SMGRL) has drawn the attention of researchers. Previous SMGRL methods generally conduct the contrastive learning within the same graph or across graphs. For example, DMGI (Park et al. 2020), HDMI (Jing, Park, and Tong 2021), and HDGI (Ren et al. 2020) perform the contrastive learning by maximizing the mutual information between node representations and the graph summary generated from the same graph. To extract the common information among different graphs, STENCIL (Zhu et al. 2022) conducts the contrastive learning across graphs by contrasting node representations from each graph and an aggregation graph. GCool (Li, Jing, and Tong 2022) conducts contrastive learning by contrasting node representations and community representations from different graphs.

Despite their success, existing methods generally ignore the non-common information (*i.e.,* private information) within each graph. This may lead to the loss of some critical information related to downstream tasks and weakens the effectiveness of previous SMGRL methods.

Method

Notations. Let $\mathcal{G} = \{\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \ldots, \mathcal{G}^{(\mathcal{R})}\}$ to denote the multiplex graph, where $\mathcal{G}^{(r)} = \{\mathbf{X}, \mathbf{A}^{(r)}\}$ is the r-th graph in the multiplex graph, $\mathcal{R}$ is the number of graphs. $\mathbf{X} \in \mathbb{R}^{N \times F}$ represents the node features and $\mathbf{A}^{(r)} \in \mathbb{R}^{N \times N}$ represents the graph structure of each graph $\mathcal{G}^{(r)}$, where N and F denote the number of nodes and the dimension of node features, respectively. The goal of the proposed CPIM is to learn the encoders to map the original inputs into low-dimensional representations $\mathbf{Z} \in \mathbb{R}^{N \times d}$, where $d \ll F$ is the dimension of the representation space.

Motivation

The SMGRL methods aim at generating discriminative node representations by extracting the hidden information (*e.g.,* the common and private information) in the multiplex graph. We argue that both the common information among different graphs and the private information within each graph are important for node representation learning. On the one hand, node representations should contain the common information among different graphs since they share the same node features and there are correlations among different graph structures (*i.e.,* different types of relationships among nodes). For example, in citation networks, if papers are connected by the same author, they tend to be connected by the same research topic as well. On the other hand, node representations should contain the private information within each graph. The reason is that each graph structure focuses on one type of relationship among nodes, and some significant connections may exist in only one graph structure, *i.e.,* the private information within one graph. Considering the same example of citation networks, if papers are only connected by the same author rather than by the same research topic, these private connection characteristics should be preserved. The reason is that the relationship connected by authors is important for the citation multiplex graph (Sun et al. 2022).

However, recent works (Zhou et al. 2022; Li, Jing, and Tong 2022; Zhu et al. 2022) are mainly proposed to focus on the common information while ignoring the private information within each graph, which leads to the loss of some essential connection characteristics in the multiplex graph and weakens the effectiveness of previous SMGRL methods. In addition, recent works generally ignore the redundant information within node representations, which may lead to the homogenization of different dimensionalities of node representations and weaken the discriminative information among them (Zhu et al. 2022). To solve these issues, the proposed CPIM investigates the decorrelation losses to mine the common information among different graphs and minimize the redundant information within node representations. Meanwhile, the proposed method also investigates the graph structure reconstruction losses to mine the private information within each graph. We show the framework in Figure 1 and introduce the details as follows.

Common Information Mining

Given the node features $\mathbf{X}$ and multiple graph structures $\{\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(\mathcal{R})}\}$, our method first employs the encoder

(*i.e.,* the graph convolutional layer) $g^{(r)} : \mathbb{R}^{N\times F} \times \mathbb{R}^{N\times N} \rightarrow \mathbb{R}^{N\times d}$ to generate node representations $\mathbf{H}^{(r)}$ for each graph, formulated as:

$$\mathbf{H}^{(r)} = \sigma(\widehat{\mathbf{A}}^{(r)}\mathbf{X}\Theta^{(r)}), \tag{1}$$

where σ is the activation function, and $\Theta^{(r)}$ is the weight matrix of the encoder $g^{(r)}$. $\widehat{\mathbf{A}}^{(r)} = \hat{\mathbf{D}}^{(r)^{-1/2}}(\mathbf{A}^{(r)} + w\mathbf{I}_N)\hat{\mathbf{D}}^{(r)^{-1/2}} \in \mathbb{R}^{N\times N}$ is the symmetrically normalized graph structure of each graph, $\hat{\mathbf{D}}^{(r)}$ is the degree matrix of $\mathbf{A}^{(r)} + w\mathbf{I}_N$, and w is the weight of identity matrix $\mathbf{I}_N$.

Based on the node representations $\mathbf{H}^{(r)}$ of each graph, previous methods generally conduct the contrastive learning within the same graph or across graphs. The contrastive learning within the same graph aims at improving the quality of node representations by capturing global properties (*i.e.,* graph summary) (Wu et al. 2021). Despite its success, it is counterintuitive to treat every graph independently and ignore the common information among different graphs (Zhou et al. 2022). To solve this issue, recent works (Zhou et al. 2022; Zhu et al. 2022; Li, Jing, and Tong 2022) conduct the contrastive learning across graphs by maximizing the mutual information between representations from different graphs, to extract the common information among them. However, these efforts generally ignore the redundant information within node representations, which weakens the effectiveness of previous SMGRL methods (Zhang et al. 2021; Ermolov et al. 2021; Bardes, Ponce, and Lecun 2022).

In this paper, inspired by the canonical correlation analysis (Andrew et al. 2013; Zbontar et al. 2021), we introduce the decorrelation losses to extract the common information among different graphs, while minimizing the redundant information within node representations. Specifically, we first conduct the decorrelation loss on node representations from different graphs, aiming to maintain the correlation among them. After that, we further conduct the decorrelation loss on node representations in each graph, aiming to filter the redundancy among different dimensionalities of the node representations. To conduct the decorrelation loss on node representations from different graphs, we first generate the correlation matrix $\mathbf{C}^{(rr')} \in \mathbb{R}^{d\times d}$ between $\mathbf{H}^{(r)}$ and $\mathbf{H}^{(r')}$ ($r \neq r'$), by defining its element as:

$$c_{ij}^{(rr')} = \frac{\sum_n \mathbf{h}_{n,i}^{(r)}\mathbf{h}_{n,j}^{(r')}}{\|\sum_n \mathbf{h}_{n,i}^{(r)}\|\|\sum_n \mathbf{h}_{n,j}^{(r')}\|}, \tag{2}$$

where $i, j \in [1, d]$ indicate the i-th and j-th dimensionalities of the n-th node representations from different graphs. We then enforce the correlation matrix $\mathbf{C}^{(rr')}$ converge to the identity matrix $\mathbf{I} \in \mathbb{R}^{d\times d}$, formulated as:

$$\begin{aligned}\mathcal{L}_{deco}^{(rr')} &= \sum(\mathbf{C}^{(rr')} - \mathbf{I})^2 \\ &= \sum_{i=1}^{d}(c_{ii}^{(rr')} - 1)^2 + \sum_{i=1}^{d}\sum_{j\neq i}(c_{ij}^{(rr')})^2.\end{aligned} \tag{3}$$

We further introduce a non-negative parameter $\lambda^{(rr')}$ to balance the first term and the second term of Eq. (3). Thus, the decorrelation loss between different graphs becomes:

$$\mathcal{L}_{deco}^{(rr')} = \sum_{i=1}^{d}(c_{ii}^{(rr')} - 1)^2 + \lambda^{(rr')}\sum_{i=1}^{d}\sum_{j\neq i}(c_{ij}^{(rr')})^2. \tag{4}$$

Intuitively, Eq. (4) mainly focuses on enforcing the corresponding node representations from different graphs to agree with each other, and maintaining the correlation among them. To further filter the redundancy among different dimensionalities of the node representations, we also conduct the decorrelation loss on node representations in each graph. Similarly, we first generate the correlation matrix $\mathbf{C}^{(r)} \in \mathbb{R}^{d\times d}$ for $\mathbf{H}^{(r)}$ and $\mathbf{H}^{(r)^{\mathrm{T}}}$ by defining its element as:

$$c_{ij}^{(r)} = \frac{\sum_n \mathbf{h}_{n,i}^{(r)}\mathbf{h}_{n,j}^{(r)}}{\|\sum_n \mathbf{h}_{n,i}^{(r)}\|\|\sum_n \mathbf{h}_{n,j}^{(r)}\|}. \tag{5}$$

We then enforce the correlation matrix $\mathbf{C}^{(r)}$ to converge to the identity matrix $\mathbf{I} \in \mathbb{R}^{d\times d}$. As a result, we obtain the decorrelation loss within each graph as:

$$\mathcal{L}_{deco}^{(r)} = \sum_{i=1}^{d}(c_{ii}^{(r)} - 1)^2 + \lambda^{(r)}\sum_{i=1}^{d}\sum_{j\neq i}(c_{ij}^{(r)})^2, \tag{6}$$

where $\lambda^{(r)}$ is a non-negative parameter to balance the first term and the second term of Eq. (6). Eq. (6) mainly focuses on minimizing the agreement among different dimensionalities of node representations in each graph, and reducing the redundancy among them.

Finally, the decorrelation losses in Eq. (4) and Eq. (6) maintain the correlation among node representations from different graphs, and filter the redundancy among different dimensionalities of node representations in each graph. In other words, $\mathcal{L}_{deco}^{(rr')}$ and $\mathcal{L}_{deco}^{(r)}$ extract the common information among node representations from different graphs, and minimize the redundant information within node representations of each graph. Moreover, as a byproduct, compared to recent SMGRL methods, our method avoids negative samples to achieve efficiency.

Private Information Mining

Apart from extracting the common information among different graphs, it is important to preserve the private information within each graph as well. However, recent works (Li, Jing, and Tong 2022; Zhu et al. 2022; Zhou et al. 2022) tend to focus only on the common information and fail to preserve the private information within each graph. As a result, some essential connection characteristics in graph structures may be ignored, resulting in the degradation of downstream task performance.

To solve this issue, we investigate the graph structure reconstruction loss (Kipf and Welling 2016; Wu et al. 2020) to preserve the private information within each graph. Specifically, we first reconstruct the graph structure based on the node representations and then conduct the reconstruction loss on it, aiming to encourage the reconstructed graph structure $\widetilde{\mathbf{A}}^{(r)}$ to preserve the connection characteristics in the original graph structure $\widehat{\mathbf{A}}^{(r)}$. To do this, we first generate the

Datasets	Nodes	Meta-paths	Edges	Features	Labeled Nodes	Classes
ACM	3,025	Paper-Subject-Paper (PSP) Paper-Author-Paper (PAP)	2,210,761 29,281	1,830 (Paper Abstract)	600	3
IMDB	4,780	Movie-Actor-Movie (MAM) Movie-Director-Movie (MDM)	98,010 21,018	1,232 (Movie Plot)	300	3
DBLP	4,057	Author-Paper-Author (APA) Author-Paper-Conference-Paper-Author (APCPA) Author-Paper-Term-Paper-Author (APTPA)	11,113 5,000,495 6,776,335	334 (Paper Abstract)	800	4
Amazon	11,944	User-Product-User (UPU) User-Star-User (USU) User-Review similarity-User (URU)	363,160 7,144,902 2,085,418	25 (Handcrafted Features)	600	2

Table 1: Statistics of the datasets.

reconstructed graph structure $\widetilde{\mathbf{A}}^{(r)}$ with the reconstruction function $f^{(r)}: \mathbb{R}^{N\times d} \times \mathbb{R}^{d\times N} \rightarrow \mathbb{R}^{N\times N}$, *i.e.,*

$$\widetilde{\mathbf{a}}_{nm}^{(r)} = \text{sigmoid}(\mathbf{h}_n^{(r)} \mathbf{h}_m^{(r)^{\mathrm{T}}}). \tag{7}$$

We then conduct the reconstruction loss on the reconstructed graph structure $\widetilde{\mathbf{A}}^{(r)}$, which thus converges to the original graph structure $\widehat{\mathbf{A}}^{(r)}$ to have:

$$\mathcal{L}_{rec}^{(r)} = \frac{1}{N^2} \sum_{n=1}^{N} \sum_{m=1}^{N} (\widetilde{\mathbf{a}}_{nm}^{(r)} - \widehat{\mathbf{a}}_{nm}^{(r)})^2. \tag{8}$$

In Eq. (8), the graph structure reconstruction loss enforces every element in the reconstructed graph structure $\widetilde{\mathbf{A}}^{(r)}$ equivalent to the corresponding element in original graph structure $\widehat{\mathbf{A}}^{(r)}$. Such consistency between $\widetilde{\mathbf{A}}^{(r)}$ and $\widehat{\mathbf{A}}^{(r)}$ enables $\mathcal{L}_{rec}^{(r)}$ to preserve the private information in each original graph structure in node representations. As a result, the node representations learned by our method are expected to contain more discriminative information than those learned by previous SMGRL methods.

Finally, integrating the decorrelation losses (*i.e.,* Eq. (4) and Eq. (6)) with the graph structure reconstruction loss in Eq. (8), the objective function of the proposed CPIM is:

$$\mathcal{J} = \sum_{r=1, r'\neq r}^{\mathcal{R}} \mathcal{L}_{deco}^{(rr')} + \sum_{r=1}^{\mathcal{R}} \omega^{(r)} \mathcal{L}_{deco}^{(r)} + \sum_{r=1}^{\mathcal{R}} \gamma^{(r)} \mathcal{L}_{rec}^{(r)}, \tag{9}$$

where $\omega^{(r)}$ and $\gamma^{(r)}$ are the non-negative parameters.

Representation Fusion

After obtaining the node representations of every graph containing both the common information and the private information, we fuse them for downstream tasks. To do this, the simplest way is the average pooling (LeCun et al. 1989). However, the average pooling is generally counterintuitive since the importance of different graphs is different. In this work, we employ the semantic attention mechanism (You et al. 2016) to assign different weights to node representations from different graphs. Specifically, we first generate the value $b_n^{(r)}$ of node representations in each graph with the semantic attention:

$$b_n^{(r)} = \tanh(\mathbf{q}^{(r)^{\mathrm{T}}} \mathbf{W}^{(r)} \mathbf{h}_n^{(r)}), \tag{10}$$

where $\mathbf{W}^{(r)}$ is the weight matrix and $\mathbf{q}^{(r)}$ is the representations of $\mathbf{h}_n^{(r)}$ after the Multi-Layer Perceptron. We then obtain the weights of node representations from different graphs, *i.e.,*

$$\alpha_n^{(r)} = \frac{\exp(b_n^{(r)})}{\sum_{r'=1}^{\mathcal{R}} \exp(b_n^{(r')})}, \tag{11}$$

where $\alpha_n^{(r)}$ represents the importance of the r-th graph in the multiplex graph. With the weights $\boldsymbol{\alpha}^{(r)}$ of each graph, the fused node representations $\mathbf{Z}$ can be obtained by:

$$\mathbf{z}_n = \sum_{r=1}^{\mathcal{R}} \alpha_n^{(r)} \mathbf{h}_n^{(r)}. \tag{12}$$

The fused node representations $\mathbf{Z}$ can be used for downstream tasks such as node classification.

Experiments

In this section, we conduct extensive experiments on four public benchmark datasets to evaluate the effectiveness and efficiency of the proposed method, compared to twelve comparison methods, on the node classification task.

Experimental Setup

Datasets. We use four public benchmark datasets from various domains, including two citation multiplex graph networks (*i.e.,* ACM (Wang et al. 2019) and DBLP (Wang et al. 2019)), one movie multiplex graph network (*i.e.,* IMDB (Wang et al. 2019)), and one review multiplex graph network (*i.e.,* Amazon (Liu et al. 2021b)). Table 1 summarizes the data statistics.

Comparison methods. The comparison methods include four methods designed for the single graph and eight for the multiplex graph. The former includes three traditional methods (*i.e.,* DeepWalk (Perozzi, Al-Rfou, and Skiena 2014), GCN (Kipf and Welling 2017), and GAT (Velickovic et al. 2018)), and one self-supervised method, *i.e.,* DGI (Velickovic et al. 2019). Multiplex graph methods include two traditional methods (*i.e.,* MNE (Zhang et al. 2018) and HAN (Wang et al. 2019)), and six self-supervised methods (*i.e.,* DMGI (Park et al. 2020), DMGIattn (Park et al. 2020), HDMI (Jing, Park, and Tong 2021), HeCo (Wang et al. 2021), CKD (Zhou et al. 2022), and STENCIL (Zhu et al. 2022)). To conduct a fair comparison, in graph network methods for single graph,

Method	ACM		IMDB		DBLP		Amazon	
	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
Deep Walk	73.9 ± 0.3	74.1 ± 0.1	42.5 ± 0.2	43.3 ± 0.4	88.1 ± 0.2	89.5 ± 0.3	80.2 ± 0.2	92.9 ± 0.4
GCN	86.9 ± 0.2	87.0 ± 0.3	45.7 ± 0.4	49.8 ± 0.2	90.2 ± 0.2	90.9 ± 0.5	82.3 ± 0.2	93.3 ± 0.4
GAT	85.0 ± 0.4	84.9 ± 0.3	49.4 ± 0.2	53.6 ± 0.4	91.0 ± 0.4	92.1 ± 0.2	78.6 ± 0.3	92.7 ± 0.1
DGI	89.1 ± 0.4	88.2 ± 0.4	45.1 ± 0.2	46.7 ± 0.2	90.3 ± 0.1	91.1 ± 0.4	83.6 ± 0.4	93.4 ± 0.2
MNE	79.2 ± 0.4	79.7 ± 0.3	44.7 ± 0.5	45.6 ± 0.3	89.3 ± 0.2	90.6 ± 0.4	81.2 ± 0.2	93.0 ± 0.4
HAN	89.4 ± 0.2	89.2 ± 0.2	49.8 ± 0.5	54.2 ± 0.3	91.2 ± 0.4	92.0 ± 0.5	83.9 ± 0.3	93.6 ± 0.2
DMGI	89.8 ± 0.1	89.8 ± 0.1	52.2 ± 0.2	53.7 ± 0.3	92.1 ± 0.2	92.9 ± 0.3	83.2 ± 0.2	92.4 ± 0.5
DMGIattn	88.7 ± 0.3	88.7 ± 0.5	52.6 ± 0.2	53.6 ± 0.4	90.9 ± 0.2	91.8 ± 0.3	83.0 ± 0.3	92.3 ± 0.3
HDMI	90.1 ± 0.3	90.1 ± 0.3	55.6 ± 0.3	57.3 ± 0.3	91.3 ± 0.2	92.2 ± 0.5	86.1 ± 0.3	94.9 ± 0.3
HeCo	88.3 ± 0.3	88.2 ± 0.2	50.8 ± 0.3	51.7 ± 0.3	91.0 ± 0.3	91.6 ± 0.2	84.3 ± 0.4	93.8 ± 0.4
CKD	90.4 ± 0.3	90.5 ± 0.2	54.8 ± 0.2	57.7 ± 0.3	92.0 ± 0.2	92.3 ± 0.5	86.3 ± 0.1	95.1 ± 0.3
STENCIL	90.7 ± 0.2	90.7 ± 0.2	54.1 ± 0.2	**58.2 ± 0.1**	92.3 ± 0.2	92.7 ± 0.4	86.9 ± 0.1	95.2 ± 0.2
CPIM	**91.4 ± 0.3**	**91.3 ± 0.2**	**55.9 ± 0.4**	57.8 ± 0.2	**93.2 ± 0.3**	**93.8 ± 0.2**	**90.1 ± 0.1**	**97.2 ± 0.1**

Table 2: Classification performance (*i.e.,* Macro-F1 and Micro-F1) of all methods on all datasets.

we separately learn the representations for every graph and further concatenate them for downstream tasks.

Setting-up. All experiments are implemented in PyTorch and conducted on a server with eight NVIDIA GeForce 3090. In all experiments, we repeat the experiments five times for all methods and report the average results. To evaluate the effectiveness, we use Macro-F1 and Micro-F1 for the node classification task. To evaluate the efficiency, we collect the training time of all SMGRL methods.

Effectiveness Analysis

We first evaluate the effectiveness of the proposed method on the semi-supervised node classification task by reporting the results (*i.e.,* Macro-F1, Micro-F1) of all methods on four datasets in Table 2. Obviously, our method achieves the best effectiveness on the node classification task.

First, compared to methods designed for the single graph (*i.e.,* Deep Walk, GCN, GAT, and DGI), our CPIM always outperforms them by a large margin. For example, our method on average improves by 8.9%, compared to the best single graph method (*i.e.,* DGI), in terms of Macro-F1 and Micro-F1, on all datasets. This demonstrates the superiority of multiplex graph methods, which can learn high-quality node representations for the multiplex graph by exploiting the correlation among different graphs.

Second, compared to traditional methods designed for the multiplex graph (*i.e.,* MNE and HAN), our CPIM obtains promising improvements as well. For example, our method on average improves by 4.8%, compared to the semi-supervised method HAN, in terms of Macro-F1 and Micro-F1, on all datasets. This indicates that self-supervised methods are sufficient to fully exploit the hidden information of the data and generate discriminative representations.

Third, compared to previous SMGRL methods, our CPIM achieves the best results, followed by STENCIL, CKD, HDMI, DMGI, DMGIattn, and HeCo. For example, our method on average improves by 1.5%, compared to the best comparison method STENCIL, in terms of Macro-F1 and Micro-F1, on all datasets. This can be attributed to the fact that our CPIM jointly mines the common information by the decorrelation losses and the private information by the reconstruction losses. As a result, this introduces more task-relevant information in node representations, leading to better downstream task performance.

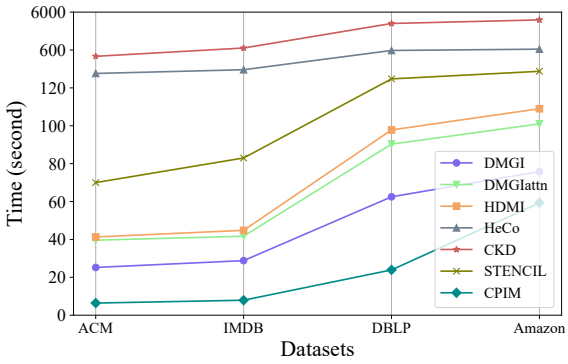

Figure 2: Training time of all SMGRL methods on all datasets.

Efficiency Analysis

We evaluate the efficiency of the proposed method by reporting the training time of all self-supervised methods on all datasets in Figure 2. Obviously, compared to other self-supervised methods, the proposed method CPIM consistently achieves the best efficiency on all datasets. For example, CPIM is on average 2.9× faster than the fastest SMGRL method DMGI, on all datasets. The reason is that CPIM extracts the common information among different graphs and minimizes the redundant information by the decorrelation losses, while preserving the private information within each graph by the reconstruction loss instead of contrastive learning. Thus, the proposed method avoids the drawbacks (*e.g.,* negative sample encoding) associated with contrastive learning, which weakens the efficiency of

$\mathcal{L}_{deco}^{(rr')}$	$\mathcal{L}_{deco}^{(r)}$	$\mathcal{L}_{rec}^{(r)}$	ACM		IMDB		DBLP		Amazon	
			Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
–	–	√	63.3 ± 0.3	63.1 ± 0.4	39.7 ± 0.2	40.8 ± 0.3	77.9 ± 0.2	78.2 ± 0.3	80.6 ± 0.3	95.5 ± 0.1
–	√	–	68.2 ± 0.3	68.6 ± 0.4	43.6 ± 0.2	44.7 ± 0.3	65.0 ± 0.2	69.1 ± 0.4	74.1 ± 0.3	94.8 ± 0.2
√	–	–	83.1 ± 0.2	83.2 ± 0.3	52.8 ± 0.2	54.5 ± 0.4	92.6 ± 0.3	93.4 ± 0.2	81.8 ± 0.2	96.2 ± 0.4
–	√	√	69.4 ± 0.2	69.8 ± 0.4	45.8 ± 0.2	46.3 ± 0.3	78.1 ± 0.4	78.7 ± 0.4	80.8 ± 0.1	95.6 ± 0.4
√	–	√	84.2 ± 0.4	83.8 ± 0.2	52.3 ± 0.4	54.4 ± 0.2	81.8 ± 0.3	82.5 ± 0.1	81.2 ± 0.4	95.9 ± 0.4
√	√	–	89.7 ± 0.1	89.5 ± 0.1	54.4 ± 0.2	55.8 ± 0.3	92.3 ± 0.1	93.2 ± 0.3	87.0 ± 0.2	96.1 ± 0.3
√	√	√	**91.4 ± 0.3**	**91.3 ± 0.2**	**55.9 ± 0.4**	**57.8 ± 0.2**	**93.2 ± 0.3**	**93.8 ± 0.2**	**90.1 ± 0.1**	**97.2 ± 0.1**

Table 3: Classification performance (*i.e.,* Macro-F1 and Micro-F1) of each component in our proposed method on all datasets.

previous contrastive-based SMGRL methods.

Ablation Study

The proposed CPIM employs decorrelation losses (*i.e.,* $\mathcal{L}_{deco}^{(rr')}$ and $\mathcal{L}_{deco}^{(r)}$) to extract the common information among different graphs, and the reconstruction loss (*i.e.,* $\mathcal{L}_{rec}^{(r)}$) to preserve the private information within each graph.

To verify the effectiveness of each loss in the proposed framework, we investigate the performance of each component on the node classification task by reporting the results in Table 3. First, our method with complete losses on average improves by 15.1%, compared to the method with the decorrelation losses or reconstruction loss only, indicating that both the decorrelation losses and the reconstruction loss are necessary for our method. This is consistent with our argument above, *i.e.,* both the common information and private information in the multiplex graph are important for node representations. Second, the method with decorrelation losses on average improves by 25.4%, compared to the method with reconstruction loss, indicating that the decorrelation losses may play a more important role than the reconstruction loss. That is, the common information is more important for the multiplex graph, while the private information provides some complementary contents. Thus, the effectiveness of each component is verified.

Parameter Analysis

In the objective function of the proposed method CPIM, we employ a non-negative parameter $\gamma^{(r)}$ to achieve a trade-off between the decorrelation losses and the reconstruction loss. To investigate the impact of $\gamma^{(r)}$ in Eq. (9) with different settings, we conduct the node classification on all datasets by varying the value of $\gamma^{(r)}$ in the range of $[10^{-3},10^{3}]$ and reporting the results in Figure 3. Obviously, the proposed method CPIM consistently achieves significant performance while the values of $\gamma^{(r)}$ are set appropriately (*e.g.,* $[10^{-1},10^{1}]$). Moreover, if the values of $\gamma^{(r)}$ are too large (*e.g.,* $> 10^{2}$) or too small (*e.g.,* $< 10^{-2}$), the proposed CPIM obtains inferior results due to the failure to extract the common information among different graphs or the failure to preserve the private information within each graph. This again validates the importance of the common information and the private information in the multiplex graph, and validates the effectiveness of our method.

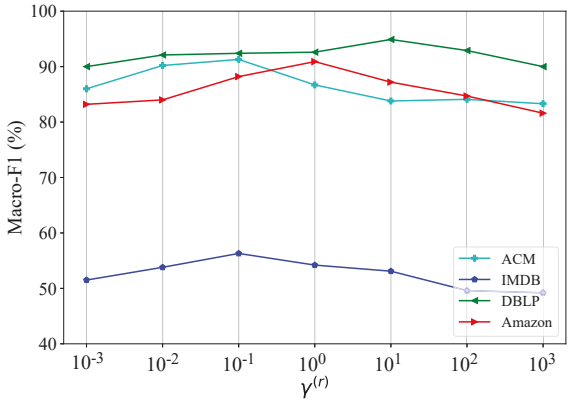

Figure 3: Classification results of our method at different parameter settings (*i.e.,* $\gamma^{(r)}$) on all datasets.

Conclusion

In this paper, we proposed a self-supervised multiplex graph representation learning framework, by jointly mining the common information and the private information in the multiplex graph while minimizing the redundant information within node representations. Specifically, we conducted the decorrelation losses on node representations, aiming to extract the common information among different graphs and minimize the redundant information within node representations of each graph. Moreover, we conducted the reconstruction loss on the reconstructed graph structure, aiming to preserve the private information within each graph. Extensive experimental results demonstrate that the proposed method consistently achieves state-of-the-art performance for both effectiveness and efficiency on the node classification task.

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China under Grant No. 2022YFA1004100, the National Natural Science Foundation of China under Grant No. 61876046, the Medico-Engineering Cooperation Funds from the University of Electronic Science and Technology of China under Grant No. ZYGX2022YGRH009 and Grant No. ZYGX2022YGRH014, and the Guangxi "Bagui" Teams for Innovation and Research, China.

References

Andrew, G.; Arora, R.; Bilmes, J.; and Livescu, K. 2013. Deep canonical correlation analysis. In *ICML*, 1247–1255.

Bardes, A.; Ponce, J.; and Lecun, Y. 2022. VICReg: Variance-Invariance-Covariance Regularization for Self-Supervised Learning. In *ICLR*.

Cen, Y.; Zou, X.; Zhang, J.; Yang, H.; Zhou, J.; and Tang, J. 2019. Representation learning for attributed multiplex heterogeneous network. In *KDD*, 1358–1368.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In *ICML*, 1597–1607.

Chu, X.; Fan, X.; Yao, D.; Zhu, Z.; Huang, J.; and Bi, J. 2019. Cross-Network Embedding for Multi-Network Alignment. In *WWW*, 273–284.

Devlin, J.; Chang, M.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT*, 4171–4186.

Dong, Y.; Chawla, N. V.; and Swami, A. 2017. metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In *KDD*, 135–144.

Ermolov, A.; Siarohin, A.; Sangineto, E.; and Sebe, N. 2021. Whitening for self-supervised representation learning. In *ICML*, 3015–3024.

Fan, S.; Zhu, J.; Han, X.; Shi, C.; Hu, L.; Ma, B.; and Li, Y. 2019. Metapath-guided Heterogeneous Graph Neural Network for Intent Recommendation. In *KDD*, 2478–2486.

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

Gao, T.; Yao, X.; and Chen, D. 2021. SimCSE: Simple Contrastive Learning of Sentence Embeddings. In *EMNLP*, 6894–6910.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *CVPR*, 9729–9738.

Hjelm, R. D.; Fedorov, A.; Lavoie-Marchildon, S.; Grewal, K.; Bachman, P.; Trischler, A.; and Bengio, Y. 2019. Learning deep representations by mutual information estimation and maximization. In *ICLR*, 1–24.

Hwang, D.; Park, J.; Kwon, S.; Kim, K.; Ha, J.; and Kim, H. J. 2020. Self-supervised Auxiliary Learning with Meta-paths for Heterogeneous Graphs. In *NeurIPS*, 10294–10305.

Jing, B.; Park, C.; and Tong, H. 2021. Hdmi: High-order deep multiplex infomax. In *WWW*, 2414–2424.

Kipf, N. T.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*, 1–14.

Kipf, T. N.; and Welling, M. 2016. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*.

Lan, Z.; Chen, M.; Goodman, S.; Gimpel, K.; Sharma, P.; and Soricut, R. 2020. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. In *ICLR*, 1–14.

LeCun, Y.; Boser, B.; Denker, J.; Henderson, D.; Howard, R.; Hubbard, W.; and Jackel, L. 1989. Handwritten digit recognition with a back-propagation network. In *NeurIPS*, volume 2.

Li, B.; Jing, B.; and Tong, H. 2022. Graph Communal Contrastive Learning. In *WWW*, 1203–1213.

Lin, Z.; Kang, Z.; Zhang, L.; and Tian, L. 2021. Multi-view attributed graph clustering. *IEEE Transactions on Knowledge and Data Engineering*, 1–9.

Liu, W.; Chen, P.-Y.; Yeung, S.; Suzumura, T.; and Chen, L. 2017. Principled multilayer network embedding. In *ICDM Workshop*, 134–141.

Liu, X.; Zhang, F.; Hou, Z.; Mian, L.; Wang, Z.; Zhang, J.; and Tang, J. 2021a. Self-supervised learning: Generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering*, 1–20.

Liu, Y.; Ao, X.; Qin, Z.; Chi, J.; Feng, J.; Yang, H.; and He, Q. 2021b. Pick and Choose: A GNN-based Imbalanced Learning Approach for Fraud Detection. In *WWW*, 3168–3177.

Mo, Y.; Peng, L.; Xu, J.; Shi, X.; and Zhu, X. 2022. Simple Unsupervised Graph Representation Learning. In *AAAI*, 7797–7805.

Oord, A. v. d.; Li, Y.; and Vinyals, O. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Pan, E.; and Kang, Z. 2021. Multi-view Contrastive Graph Clustering. In *NeurIPS*, volume 34, 2148–2159.

Park, C.; Kim, D.; Han, J.; and Yu, H. 2020. Unsupervised Attributed Multiplex Network Embedding. In *AAAI*, 5371–5378.

Peng, L.; Wang, N.; Dvornek, N.; Zhu, X.; and Li, X. 2022. FedNI: Federated Graph Learning with Network Inpainting for Population-Based Disease Prediction. *IEEE Transactions on Medical Imaging*, 1–12.

Perozzi, B.; Al-Rfou, R.; and Skiena, S. 2014. Deepwalk: Online learning of social representations. In *KDD*, 701–710.

Ren, Y.; Liu, B.; Huang, C.; Dai, P.; Bo, L.; and Zhang, J. 2020. HDGI: An Unsupervised Graph Neural Network for Representation Learning in Heterogeneous Graph. In *AAAI Workshop*.

Sun, Y.; Zhu, D.; Du, H.; and Tian, Z. 2022. MHNF: Multi-Hop Heterogeneous Neighborhood Information Fusion Graph Representation Learning. *IEEE Transactions on Knowledge and Data Engineering*, 1–14.

Tang, J.; Qu, M.; Wang, M.; Zhang, M.; Yan, J.; and Mei, Q. 2015. Line: Large-scale information network embedding. In *WWW*, 1067–1077.

Tian, Y.; Krishnan, D.; and Isola, P. 2020. Contrastive Multiview Coding. In *ECCV*, volume 12356, 776–794.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is All you Need. In *NeurIPS*, 5998–6008.

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2018. Graph Attention Networks. In *ICLR*, 1–12.

Velickovic, P.; Fedus, W.; Hamilton, W. L.; Liò, P.; Bengio, Y.; and Hjelm, R. D. 2019. Deep Graph Infomax. In *ICLR*, 1–17.

Wang, H.; Guo, X.; Deng, Z.-H.; and Lu, Y. 2022. Rethinking minimal sufficient representation in contrastive learning. In *CVPR*, 16041–16050.

Wang, X.; Ji, H.; Shi, C.; Wang, B.; Ye, Y.; Cui, P.; and Yu, P. S. 2019. Heterogeneous Graph Attention Network. In *WWW*, 2022–2032.

Wang, X.; Liu, N.; Han, H.; and Shi, C. 2021. Self-supervised Heterogeneous Graph Neural Network with Co-contrastive Learning. In *KDD*, 1726–1736.

Wu, L.; Lin, H.; Tan, C.; Gao, Z.; and Li, S. Z. 2021. Self-supervised learning on graphs: Contrastive, generative, or predictive. *IEEE Transactions on Knowledge and Data Engineering*, 1–20.

Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; and Philip, S. Y. 2020. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1): 4–24.

Xie, D.; Deng, C.; Li, C.; Liu, X.; and Tao, D. 2020. Multi-task consistency-preserving adversarial hashing for cross-modal retrieval. *IEEE Transactions on Image Processing*, 29: 3626–3637.

Xu, J.; Ren, Y.; Tang, H.; Yang, Z.; Pan, L.; Yang, Y.; Pu, X.; Yu, P. S.; and He, L. 2022a. Self-supervised discriminative feature learning for deep multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*.

Xu, J.; Tang, H.; Ren, Y.; Peng, L.; Zhu, X.; and He, L. 2022b. Multi-Level Feature Learning for Contrastive Multi-View Clustering. In *CVPR*, 16051–16060.

You, Q.; Jin, H.; Wang, Z.; Fang, C.; and Luo, J. 2016. Image Captioning with Semantic Attention. In *CVPR*, 4651–4659.

Zbontar, J.; Jing, L.; Misra, I.; LeCun, Y.; and Deny, S. 2021. Barlow Twins: Self-Supervised Learning via Redundancy Reduction. In *ICML*, volume 139, 12310–12320.

Zhang, H.; and Kou, G. 2022. Role-based Multiplex Network Embedding. In *ICML*, 26265–26280.

Zhang, H.; Qiu, L.; Yi, L.; and Song, Y. 2018. Scalable Multiplex Network Embedding. In *IJCAI*, 3082–3088.

Zhang, H.; Wu, Q.; Yan, J.; Wipf, D.; and Yu, P. S. 2021. From canonical correlation analysis to self-supervised graph neural networks. In *NeurIPS*, volume 34, 76–89.

Zhang, J.; Song, J.; Gao, L.; and Shen, H. 2022. Free-Lunch for Cross-Domain Few-Shot Learning: Style-Aware Episodic Training with Robust Contrastive Learning. In *ACM MM*, 2586–2594.

Zhou, S.; Yu, K.; Chen, D.; Li, B.; Feng, Y.; and Chen, C. 2022. Collaborative Knowledge Distillation for Heterogeneous Information Network Embedding. In *WWW*, 1631–1639.

Zhou, X.; Shen, F.; Liu, L.; Liu, W.; Nie, L.; Yang, Y.; and Shen, H. T. 2020. Graph Convolutional Network Hashing. *IEEE Transactions on Cybernetics*, 50(4): 1460–1472.

Zhu, Y.; Xu, Y.; Cui, H.; Yang, C.; Liu, Q.; and Wu, S. 2022. Structure-enhanced heterogeneous graph contrastive learning. In *ICDM*, 82–90.