# Mining Denoising Complementarity and Consistent Consensus for Unsupervised Multiplex Graph Representation Learning

Chang Liu[1], Zengyi Wo[1], Minglai Shao[2(✉)], Wenjun Wang[1,3,4], and Qin Tian[1]

[1] College of Intelligence and Computing, Tianjin University, China
{changliu0425,wozengyi1999,wjwang,tianqin123}@tju.edu.cn
[2] School of New Media and Communication, Tianjin University, China
shaoml@tju.edu.cn
[3] Yazhou Bay Innovation Institute, Hainan Tropical Ocean University, China
[4] State Key Laboratory of Synthetic Biology, Tianjin University, China

**Abstract.** Unsupervised multiplex graph representation learning (UMGRL) has attracted significant attention due to its ability to capture complex interaction information effectively without relying on labeled data. However, few studies have adequately addressed the prominent noise issue while simultaneously capturing the complementarity and consistency within the multiplex graph. To alleviate the above limitation, we propose a UMGRL method, named MD3C. It designs denoising encoders by integrating multi-perspective information from original node features, multi-hop neighborhoods, and different iterations. Subsequently, it mines intra-graph denoising complementarity through reconstruction and dimensional decorrelation constraints. Additionally, it achieves inter-graph consistent consensus by the cross-graph guided contrast and correlation maximization constraints. MD3C collaboratively counteracts noise and comprehensively extracts valuable information in the multiplex graph, thereby generating high-quality representations. Extensive thorough experiments validate the effectiveness and superiority of the proposed method across various downstream tasks.

**Keywords:** Graph representation learning · Multiplex graph · Unsupervised learning.

## 1 Introduction

Most graph representation learning methods focus on single-structure graphs [1, 2], where a single type of relationship connects nodes. However, the multiplex graph (MultiG), consisting of multiple graphs sharing the node features but with distinct structures indicating multiple relationships among nodes, is extensively employed in practical applications, such as social networks [3], biological networks [4] and citation networks [5]. MultiG can model complex multi-relational interactions, such as interactions between papers sharing the same author or subject, enriching the relational information and enabling more comprehensive learning compared to single-structure graphs.

The key challenge in multiplex graph representation learning lies in effectively capturing complementarity and consistency in MultiG. Complementarity refers to the unique and valuable information in specific relationships that facilitates the identification of potential patterns in data from multiple perspectives. For instance, the same author and common subject relationships offer two aspects to distinguish paper classes. Consistency refers to similar characteristics and patterns across graphs. It reveals deep associations between different relational perspectives and has been proven to be a key component in sample identification [6].

Some studies [7] attempt to alleviate the high cost and scarcity of labels through semi-supervised multiplex graph representation learning but still fail to eliminate reliance on labels. Due to its independence from labels and its ability to capture complex interaction information, unsupervised multiplex graph representation learning (UMGRL) has gained considerable attention. Traditional UMGRL methods, represented by the random walk strategy [8], excessively focus on the structural proximity while neglecting the important discriminative information hidden in node features. Recently, self-supervised multiplex graph representation learning (SMGRL) methods have demonstrated remarkable effectiveness [9–11]. Most SMGRL methods are based on intra-graph or inter-graph processing to mine complementarity within each graph or consistency across graphs. Intra-graph contrastive learning (intra-GCL) and inter-graph contrastive learning (inter-GCL) are widely used intra-graph and inter-graph processing approaches [12–14]. For example, CKD [6] leverages both intra-GCL and inter-GCL to collaboratively extract regional and global knowledge. Additionally, several methods design specific constraints for both intra-graph and inter-graph processing. For instance, MGDCR [15] employs intra-graph and inter-graph decorrelation losses to improve representation learning. Moreover, some studies [5] introduce disentangled representation learning to capture clean common information and complementary private information.

Despite the achievements of previous UMGRL methods, there are still several limitations. On the one hand, the graph structure may contain noisy edges that connect nodes from different classes, for example, action and comedy movies may be linked by the same lead actor. We show the noisy edge ratio across all graphs in MultiG datasets ACM, IMDB, and DBLP in Fig. 1, highlighting the prominent noise issue in MultiG. Without a targeted design, aggregation mechanisms commonly adopted by existing methods [12, 14] smooth adjacent node representations. This process incorporates noisy structural patterns into node embeddings, leading to suboptimal performance [16]. Although some studies [3, 15, 17, 18] have made preliminary attempts to mitigate the noise issue by strategies such as exploring common information or introducing the Multilayer Perceptron encoders, these approaches remain relatively limited and have not adequately addressed the noise issue. On the other hand, some methods independently learn representations within each graph and then combine them through certain fusion mechanisms, such as attention mechanisms [13, 14]. However, these approaches ignore the intrinsic consistent relationships between different graphs.

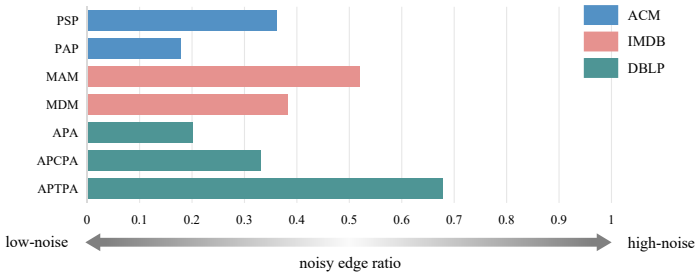


**Fig. 1.** Noisy edge ratio of multiplex graph datasets.

Motivated by the above limitations, we propose a UMGRL method, **MD3C**, for **M**ining **D**enoising **C**omplementarity and **C**onsistent **C**onsensus. Specifically, we first design specialized denoising encoders to integrate multi-perspective discriminative information from original node features, multi-hop neighborhoods, and different iterations. Then, the proposed method collaboratively mitigates the prominent noise issue and comprehensively identifies complex patterns in MultiG through the following two components, thereby generating high-quality representations. Intra-Graph Denoising Complementarity Mining captures unique and valuable information from multiple specific relationships by reconstruction and dimensional decorrelation constraints. Inter-Graph Consistent Consensus Mining extracts the latent associations among multiple relationships via constructing the cross-graph guided contrastive learning and inter-graph correlation maximization constraint. The contributions of our method are summarized as follows:

- To the best of our knowledge, we make the first attempt to design unsupervised denoising encoders for MultiG. This encoding mechanism comprehensively combines multi-perspective discriminative information from original node features, multi-hop neighborhoods, and different iterations. Simultaneously, we explore consistency to capture stable relationships across multiple graphs. Together, these efforts collaboratively counteract the severe and detrimental noise within MultiG.
- The proposed UMGRL method, MD3C, models complex interactions in MultiG by exploring the intra-graph denoising complementarity and inter-graph consistent consensus, generating informative representations.
- For various downstream tasks, including node classification and node clustering, extensive experiments on numerous public datasets demonstrate the effectiveness of the proposed MD3C compared to several baseline approaches.

## 2 Methodology

In this section, we propose a novel UMGRL method, MD3C. We first introduce the problem statement (Section 2.1), followed by the two components of the proposed method (Section 2.2 and 2.3). The overall framework is shown in Fig. 2.

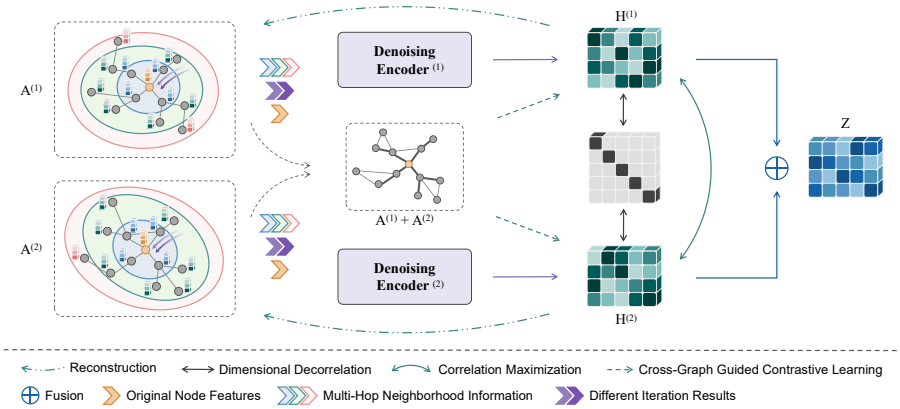


**Fig. 2.** The overall framework of MD3C. It leverages denoising encoders to combine node features, multi-hop neighborhood information, and different iteration results, mining intra-graph complementarity through reconstruction and dimensional decorrelation constraints, while achieving inter-graph consistency via cross-graph guided contrast and correlation maximization constraints, ultimately obtaining fused representations.

### 2.1 Problem Statement

Let $\mathcal{G} = \{\mathcal{G}^{(1)}, \ldots, \mathcal{G}^{(M)}\}$ denote MultiG, where $\mathcal{G}^{(m)} = \{\mathcal{V}, \mathcal{E}^{(m)}\}$ is $m$-th graph comprised of a node set $\mathcal{V}$ and an edge set $\mathcal{E}^{(m)}$, $M$ denotes the number of graphs. $\mathrm{X} \in \mathbb{R}^{N \times D}$ denotes node features shared by all graphs, $N$ is the number of nodes and $D$ is the dimension of node features. $\mathrm{A}^{(m)} \in \mathbb{R}^{N \times N}$ is the adjacency matrix of $\mathcal{G}^{(m)}$ where $\mathrm{A}_{ij}^{(m)} = 1$ if edge $e_{ij}^{(m)} \in \mathcal{E}^{(m)}$ between node $v_i$ and $v_j$ exists, otherwise $\mathrm{A}_{ij}^{(m)} = 0$.

Given MultiG $\mathcal{G}$, the goal is to learn low-dimensional fused representations $\mathrm{Z} \in \mathbb{R}^{N \times d}$ by leveraging node features and multiple graph structures, without using any labels.

### 2.2 Intra-Graph Denoising Complementarity Mining

Most existing methods utilize message-passing mechanisms, represented by Graph Convolutional Networks (GCNs) [1] smoothing adjacent node representations, to learn node embeddings [13, 14]. When these encoders operate on noisy edges connecting nodes of different classes, interference from other classes can be encoded into the node embeddings, thereby reducing distinguishability [19, 20].

To tackle the above issue, we specifically design denoising encoders $f^{(m)}$. First, original node features inherently contain unique information with distinguishability from other nodes. The encoding mechanism should ensure that original node features develop independently to avoid excessive smoothing during aggregation. Based on the above considerations, we adopt a linear encoding approach $\mathrm{H}_{ego}^{(m)} = \sigma\left(\mathrm{X}\mathrm{W}_{f}^{(m)}\right)$ instead of GCNs to obtain the ego node

embeddings $\mathrm{H}_{ego}^{(m)}$ of $\mathcal{G}^{(m)}$, where $\sigma$ is the nonlinear activation function and $\mathrm{W}_{f}^{(m)} \in \mathbb{R}^{D\times e}$ is a learnable weight matrix. Second, the different distance neighborhoods in a graph often contain more diverse information, especially in the presence of noisy edges. The central node may have different classes from its directly adjacent nodes but share the same class as nodes in higher-order neighborhoods, indicating that multi-hop neighborhoods can provide more relevant context. Therefore, our method incorporates multi-hop neighborhood structural information, with the $k$-hop neighborhood structure of $\mathcal{G}^{(m)}$ represented as $\mathrm{S}_{k}^{(m)} = \left(\mathrm{A}^{(m)}\right)^{k} - \left(\mathrm{A}^{(m)}\right)^{k-1} - \ldots - \mathrm{A}^{(m)} - \mathrm{I}$, which captures the higher-order interactions between the central node and its $k$-hop neighborhood that strip away the relative influence of previous $k-1$ hop neighborhoods and self-loops in the message-passing space. Then, we obtain the node embedding matrix $\mathrm{H}_{k,l}^{(m)}$ of $\mathcal{G}^{(m)}$ corresponding to the $k$-hop neighborhood through graph convolution, as follows:

$$\mathrm{H}_{k,l}^{(m)} = \left(\mathrm{D}_{k}^{(m)}\right)^{-\frac{1}{2}} \tilde{\mathrm{S}}_{k}^{(m)} \left(\mathrm{D}_{k}^{(m)}\right)^{-\frac{1}{2}} \mathrm{H}_{l-1}^{(m)} \tag{1}$$

where $\tilde{\mathrm{S}}_{k}^{(m)} = \mathbb{1}(\mathrm{S}_{k}^{(m)} > 0)$, $\mathbb{1}$ is the indicator function that sets elements greater than 0 to 1 and others to 0 in the matrix, and $\mathrm{D}_{k}^{(m)}$ is the degree matrix of $\tilde{\mathrm{S}}_{k}^{(m)}$, with the process being repeated for $L$ iterations and $\mathrm{H}_{0}^{(m)} = \mathrm{H}_{ego}^{(m)}$. To avoid mixing information from each hop neighborhood and allow them to develop independently, we concatenate them together: $\mathrm{H}_{l}^{(m)} = \sigma\left(\mathrm{H}_{1,l}^{(m)} \parallel \ldots \parallel \mathrm{H}_{K,l}^{(m)}\right)$, where $\parallel$ denotes the concatenation operation. Third, the information extracted from different iterations varies. Earlier iterations focus on local information, while later ones tend to capture global information. We concatenate results from different iterations to preserve the diverse information at each stage: $\mathrm{H}_{topo}^{(m)} = \mathrm{H}_{1}^{(m)} \parallel \ldots \parallel \mathrm{H}_{L}^{(m)}$. Finally, we combine $\mathrm{H}_{ego}^{(m)}$ and $\mathrm{H}_{topo}^{(m)}$ to generate the final representations of $\mathcal{G}^{(m)}$, which captures the semantic and structural information: $\mathrm{H}^{(m)} = \left(\mathrm{H}_{ego}^{(m)} \parallel \mathrm{H}_{topo}^{(m)}\right) \mathrm{W}_{o}^{(m)}$, where $\mathrm{W}_{o}^{(m)} \in \mathbb{R}^{\frac{e\left(k^{l+1}-1\right)}{k-1}\times d}$ is a learnable weight matrix.

We derive supervisory signals from the graph data itself to construct a loss $\mathcal{L}_{intra}$ for mining intra-graph denoising complementarity, which involves reconstructing the actual structure from node representations and applying dimensional decorrelation constraint to reduce redundancy [21], as follows:

$$\begin{aligned}\mathcal{L}_{intra} =& \alpha_1 \sum_{m=1}^{M} \left( -\frac{1}{N^2} \sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{A}_{ij}^{(m)} log \tilde{\mathrm{A}}_{ij}^{(m)} + \left(1 - \mathrm{A}_{ij}^{(m)}\right) log\left(1 - \tilde{\mathrm{A}}_{ij}^{(m)}\right) \right) \\ &+ \alpha_2 \sum_{m=1}^{M} \left\| \left(\mathrm{H}^{(m)}\right)^{\top} \mathrm{H}^{(m)} - \mathrm{I} \right\|_{F}^{2}\end{aligned} \tag{2}$$

where $\tilde{\mathrm{A}}^{(m)} = \sigma\left(\mathrm{H}^{(m)}\left(\mathrm{H}^{(m)}\right)^{\top}\right)$ is the graph structure reconstructed by the inner product decoder, and $\sigma$ is a non-linear activation function.

### 2.3 Inter-Graph Consistent Consensus Mining

The different graphs in MultiG are not completely independent, as they reflect the connections of the same nodes from different relational perspectives, implying that there must be deep correlations among them. Meanwhile, although different graphs provide unique complementary information, they may contain noise. Mining consistency across graphs helps to discover stable relationships to mitigate noise impact. Therefore, we design a cross-graph guided contrastive learning mechanism. The correlation between nodes is mainly reflected in structure. Intuitively, if two nodes are associated across multiple views, they are likely to belong to the same class. Specifically, we first define a relevant function $\mathbb{F}_i$ to count the number of connections between node $v_i$ and other nodes: $\mathbb{F}_i(j) = \sum_{m=1}^{M} \mathbb{1}\left(v_j \in N_i^{(m)}\right)$, where $\mathbb{1}$ is the indicator function, $N_i^{(m)}$ is the neighboring node set of node $v_i$ in $\mathcal{G}^{(m)}$. We sort the nodes in set $P_i = \{v_j | v_j \in \mathcal{V} \text{ and } \mathbb{F}_i(j) \neq 0\}$ in descending order according to $\mathbb{F}_i(j)$ and construct positive samples according to the following rules:

$$\begin{cases} N_i^{(pos)} = P_i[1:t_p] \text{ if } |P_i| > t_p \\ N_i^{(pos)} = P_i \qquad\quad \text{if } |P_i| \leq t_p \end{cases} \tag{3}$$

where $N_i^{(pos)}$ is the positive sample set of node $v_i$, the remaining nodes are assigned to the negative sample set $N_i^{(neg)}$, and $t_p$ is threshold, $P_i[1:t_p]$ denotes top $t_p$ nodes in the set $P_i$. The cross-graph guided contrastive loss is given by:

$$\mathcal{L}_{con}^{(m)}(i) = -log\frac{\sum_{v_j \in N_i^{(pos)}} exp\left(sim\left(h_i^{(m)}, h_j^{(m)}\right)/\tau\right)}{\sum_{v_k \in \left\{N_i^{(pos)} \cup N_i^{(neg)}\right\}} exp\left(sim\left(h_i^{(m)}, h_k^{(m)}\right)/\tau\right)} \tag{4}$$

where $sim$ is the cosine similarity, and $\tau$ denotes a temperature parameter. Moreover, to achieve inter-graph consistent consensus, we introduce a correlation maximization constraint between different graph representations. The final loss function $\mathcal{L}_{inter}$ for mining inter-graph consistent consensus is as follows:

$$\mathcal{L}_{inter} = \beta_1 \frac{1}{N}\sum_{m=1}^{M}\sum_{i=1}^{N}\mathcal{L}_{con}^{(m)}(i) - \beta_2 \frac{1}{N}\sum_{m_1=1}^{M}\sum_{m_2>m_1}^{M}\sum_{i=1}^{N}\frac{h_i^{(m_1)}\left(h_i^{(m_2)}\right)^{\top}}{\left\|h_i^{(m_1)}\right\|\left\|h_i^{(m_2)}\right\|} \tag{5}$$

With $\mathcal{L}_{intra}$ and $\mathcal{L}_{inter}$, the overall objective function of the proposed MD3C is given by:

$$\mathcal{J} = \mathcal{L}_{intra} + \mathcal{L}_{inter} \tag{6}$$

The proposed MD3C is continuously optimized through back-propagation to extract complementary information and consistent information. Finally, we use $\mathrm{Z} = \frac{1}{M}\sum_{m=1}^{M}\mathrm{H}^{(m)}$ for downstream tasks.

# 3 Experiments

## 3.1 Experimental Setup

**Datasets.** The public benchmark datasets used include two citation multiplex graph datasets, ACM [7] and DBLP [7], and a movie multiplex graph dataset, IMDB [7]. The dataset splitting follows the public setting [7].

**Comparison Methods.** The comparison methods include four single-view and thirteen MultiG methods. Single-view methods include two semi-supervised methods (GCN [1], GAT [2]) and two unsupervised methods (Deepwalk [22], DGI [23]). MultiG methods include one semi-supervised method (HAN [7]) and twelve unsupervised methods: MNE [8], DMGI [14], DMGIattn [14], HDMI [13], HeCo [11], MCGC [10], CKD [6], RGRL [9], DMG [5], MGDCR [15], CoCoMG [18], UDW [17]. For fairness, single-view methods are trained independently on each graph in MultiG, then all representations are concatenated for downstream tasks.

**Evaluation Metrics.** As in previous works [13, 14], we evaluate MD3C through node classification and node clustering tasks. The evaluation metrics are Macro-F1 and Micro-F1 for node classification, and Accuracy and Normalized Mutual Information (NMI) for node clustering. The fused representations obtained by training MD3C are fed into the logistic regression classifier and K-means algorithm to perform node classification and node clustering tasks, respectively. Results are averaged over five repetitions of the experiment.

**Table 1.** Node classification (Macro-F1 and Micro-F1) and clustering performance (Accuracy and NMI).

| Method | ACM | | | | IMDB | | | | DBLP | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MaF1 | MiF1 | Acc | NMI | MaF1 | MiF1 | Acc | NMI | MaF1 | MiF1 | Acc | NMI |
| Deep Walk | 73.9 | 74.8 | 64.5 | 41.6 | 42.5 | 43.3 | 42.1 | 1.5 | 88.1 | 89.5 | 89.5 | 69.0 |
| GCN | 86.9 | 87.0 | - | - | 45.7 | 49.8 | - | - | 90.2 | 90.9 | - | - |
| GAT | 85.0 | 84.9 | - | - | 49.4 | 53.6 | - | - | 91.0 | 92.1 | - | - |
| DGI | 89.1 | 88.2 | 81.1 | 64.0 | 45.1 | 46.7 | 48.9 | 8.3 | 90.3 | 91.1 | 85.4 | 65.6 |
| MNE | 79.2 | 79.7 | 69.1 | 54.5 | 44.7 | 45.6 | 46.5 | 4.6 | 89.3 | 90.6 | 86.3 | 68.4 |
| HAN | 89.4 | 89.2 | - | - | 49.8 | 54.2 | - | - | 91.2 | 92.0 | - | - |
| DMGI | 89.8 | 89.8 | 88.4 | 68.7 | 52.2 | 53.7 | 52.5 | 13.1 | 92.1 | 92.9 | 91.8 | 76.4 |
| DMGIattn | 88.7 | 88.7 | 90.9 | 70.2 | 52.6 | 53.6 | 52.6 | 9.2 | 90.9 | 91.8 | 91.3 | 75.2 |
| HDMI | 90.1 | 90.1 | 90.8 | 69.5 | 55.6 | 57.3 | 57.6 | 14.5 | 91.3 | 92.2 | 90.1 | 73.1 |
| HeCo | 88.2 | 88.3 | 88.4 | 67.8 | 50.8 | 51.7 | 50.9 | 10.1 | 91.0 | 91.6 | 89.2 | 71.0 |
| MCGC | 90.2 | 90.0 | 90.4 | 69.0 | 56.3 | 57.5 | 56.5 | 14.9 | 91.9 | 92.1 | 91.9 | 76.5 |
| CKD | 90.4 | 90.5 | 90.6 | 69.3 | 54.8 | 57.7 | 53.9 | 13.8 | 92.0 | 92.3 | 91.4 | 75.9 |
| RGRL | 90.3 | 90.2 | 90.7 | 69.4 | 52.1 | 55.5 | 52.5 | 13.3 | 91.7 | 92.0 | 91.0 | 74.5 |
| DMG | 90.9 | 90.9 | **92.9** | 74.5 | 57.6 | 58.9 | **60.3** | 17.0 | 93.3 | 93.9 | **94.1** | 80.0 |
| MGDCR | 91.6 | 91.5 | 91.9 | 72.1 | 54.2 | 55.5 | 57.2 | 13.2 | 93.4 | 94.1 | 93.8 | 80.2 |
| CoCoMG | 92.8 | 92.7 | 92.3 | 73.6 | 58.1 | 59.2 | 59.0 | 17.4 | 91.9 | 92.7 | 91.3 | 75.6 |
| UDW | 91.6 | 91.7 | 91.5 | 70.5 | 57.2 | 58.8 | 56.8 | 14.5 | 93.5 | 94.2 | 92.6 | 76.9 |
| **MD3C** | **93.4** | **93.1** | **92.9** | **74.9** | **59.1** | **59.9** | 58.5 | **18.3** | **93.9** | **94.5** | 94.0 | **81.0** |

### 3.2 Effectiveness Analysis

We report the evaluation performance of MD3C on node classification and clustering tasks in Table 1. The best results are in bold, and the second-best are underlined. MD3C outperforms all baseline methods in node classification and matches or exceeds the best results in node clustering. Our method significantly improves node classification and node clustering task performance across all datasets compared to single-view methods. Additionally, MultiG methods generally outperform single-view methods, proving that MultiG methods can explore complex associations across graphs. MD3C extracts more comprehensive information than single-view methods by mining inter-graph consistent consensus.

MD3C achieves the best overall performance in different tasks across all datasets compared to MultiG baseline methods. Because MD3C designs specialized denoising encoders and explores intra-graph denoising complementarity as well as inter-graph consistent consensus, collaboratively counteracting the severe and adverse noise, and effectively modeling the complex interactions within MultiG, thereby generating discriminative, high-quality node representations.

**Parameter Analysis.** We investigate the node classification performance under different neighborhood ranges $K$ on the IMDB dataset in Fig. 3. The performance is optimal with $K = 2$, as 1-hop neighborhood provides limited information, while 3-hop neighborhoods introduce too much irrelevant information.

### 3.3 Ablation Study

**Effectiveness of Each Loss.** MD3C primarily includes $\mathcal{L}_{intra}$ and $\mathcal{L}_{inter}$ to extract denoising complementarity and consistent consensus. To validate the effectiveness of each loss, we investigate their performance on various downstream tasks and report the results in Table 2. Evidently, MD3C with a complete objective function outperforms the variants with only one type of loss, indicating that both designed loss functions capture valuable information.

**Effectiveness of Each Graph.** To validate the uniqueness and complementarity of each graph in MultiG, we visualize the node classification performance across all datasets with different graph combinations in Fig. 4. Each graph provides useful information, but their importance varies due to differing amounts of complementary information. Using all graphs yields the best results, showing that MD3C effectively captures complementary information from different graphs to generate more informative representations.

**Table 2.** Node classification (Macro-F1 and Micro-F1) and clustering performance (Accuracy and NMI) for ablation study.

| | | **ACM** | | | | **IMDB** | | | | **DBLP** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\mathcal{L}_{intra}$ | $\mathcal{L}_{inter}$ | MaF1 | MiF1 | Acc | NMI | MaF1 | MiF1 | Acc | NMI | MaF1 | MiF1 | Acc | NMI |
| ✓ | - | 89.2 | 89.1 | 66.7 | 44.7 | 48.7 | 49.3 | 42.1 | 2.2 | 93.1 | 93.8 | 87.7 | 66.9 |
| - | ✓ | 84.0 | 84.1 | 54.6 | 54.3 | 56.5 | 58.6 | 51.3 | 12.5 | 92.9 | 93.4 | 92.4 | 76.1 |
| ✓ | ✓ | **93.4** | **93.1** | **92.9** | **74.9** | **59.1** | **59.9** | **58.5** | **18.3** | **93.9** | **94.5** | **94.0** | **81.0** |

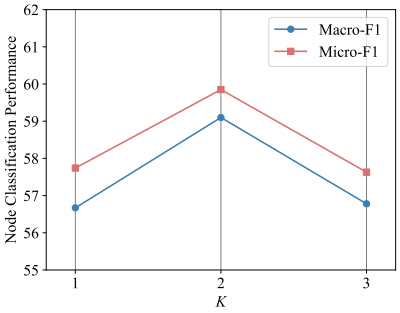


**Fig. 3.** Node classification performance with different $K$.

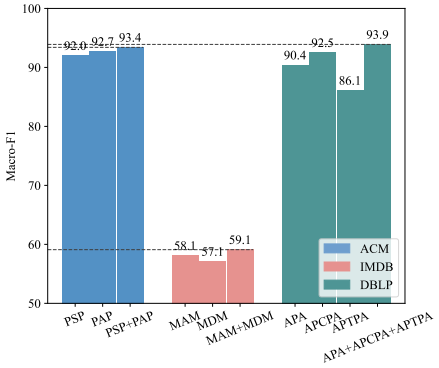


**Fig. 4.** Macro-F1 for different graph combinations.

## 4 Conclusion

In this paper, to tackle the prominent noise issue in MultiG and extract valuable information from multiple relationships, we propose MD3C. This method collaboratively counters noise by exploring intra-graph denoising complementarity and inter-graph consistent consensus, while modeling complex interactions within MultiG to generate informative representations. Extensive comprehensive experiments validate the effectiveness and superiority of MD3C in terms of various downstream tasks.

**Acknowledgments.** This work is supported by the National Natural Science Foundation of China (No. 62472305, No. 62272338), and the Shenzhen Sustainable Development Project (KCXFZ20211020172544004).

## References

1. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
2. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., et al.: Graph attention networks. stat **1050**(20), 10–48550 (2017)
3. Wang, Q., Jiang, H., Jiang, Y., Yi, S., Nie, Q., Zhang, G.: Multiplex network infomax: Multiplex network embedding via information fusion. Digital Communications and Networks **9**(5), 1157–1168 (2023)
4. Li, Z., Wang, H., Zhang, J.: A multiplex graph learning framework for predicting drug-disease associations. IEEE/ACM Transactions on Computational Biology and Bioinformatics **18**(3), 807–818 (2021)
5. Mo, Y., Lei, Y., Shen, J., Shi, X., Shen, H.T., Zhu, X.: Disentangled multiplex graph representation learning. In: International Conference on Machine Learning. pp. 24983–25005. PMLR (2023)

6. Wang, C., Zhou, S., Yu, K., Chen, D., Li, B., Feng, Y., Chen, C.: Collaborative knowledge distillation for heterogeneous information network embedding. In: Proceedings of the ACM Web Conference 2022. pp. 1631–1639 (2022)
7. Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., Yu, P.S.: Heterogeneous graph attention network. In: The world wide web conference. pp. 2022–2032 (2019)
8. Zhang, H., Qiu, L., Yi, L., Song, Y.: Scalable multiplex network embedding. In: IJCAI. vol. 18, pp. 3082–3088 (2018)
9. Lee, N., Hyun, D., Lee, J., Park, C.: Relational self-supervised learning on graphs. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. pp. 1054–1063 (2022)
10. Pan, E., Kang, Z.: Multi-view contrastive graph clustering. Advances in neural information processing systems **34**, 2148–2159 (2021)
11. Wang, X., Liu, N., Han, H., Shi, C.: Self-supervised heterogeneous graph neural network with co-contrastive learning. In: Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining. pp. 1726–1736 (2021)
12. Jing, B., Feng, S., Xiang, Y., Chen, X., Chen, Y., Tong, H.: X-goal: Multiplex heterogeneous graph prototypical contrastive learning. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. pp. 894–904 (2022)
13. Jing, B., Park, C., Tong, H.: Hdmi: High-order deep multiplex infomax. In: Proceedings of the Web Conference 2021. pp. 2414–2424 (2021)
14. Park, C., Kim, D., Han, J., Yu, H.: Unsupervised attributed multiplex network embedding. In: Proceedings of the AAAI conference on artificial intelligence. vol. 34, pp. 5371–5378 (2020)
15. Mo, Y., Chen, Y., Lei, Y., Peng, L., Shi, X., Yuan, C., Zhu, X.: Multiplex graph representation learning via dual correlation reduction. IEEE Transactions on Knowledge and Data Engineering **35**(12), 12814–12827 (2023)
16. Zhao, J., Wang, X., Shi, C., Hu, B., Song, G., Ye, Y.: Heterogeneous graph structure learning for graph neural networks. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 4697–4705 (2021)
17. Mo, Y., Shen, H.T., Zhu, X.: Unsupervised multi-view graph representation learning with dual weight-net. Information Fusion **114**, 102669 (2025), available online 10 September 2024
18. Peng, L., Wang, X., Zhu, X.: Unsupervised multiplex graph learning with complementary and consistent information. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 454–462 (2023)
19. Wang, J., Guo, Y., Yang, L., Wang, Y.: Understanding heterophily for graph neural networks. arXiv preprint arXiv:2401.09125 (2024)
20. Park, M., Heo, J., Kim, D.: Mitigating oversmoothing through reverse process of GNNs for heterophilic graphs. In: Forty-first International Conference on Machine Learning (2024)
21. Jing, L., Vincent, P., LeCun, Y., Tian, Y.: Understanding dimensional collapse in contrastive self-supervised learning. arXiv preprint arXiv:2110.09348 (2021)
22. Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 701–710 (2014)
23. Velickovic, P., Fedus, W., Hamilton, W.L., Liò, P., Bengio, Y., Hjelm, R.D.: Deep graph infomax. ICLR (Poster) **2**(3), 4 (2019)