Strategy-Architecture Synergy: A Multi-View Graph Contrastive Paradigm for Consistent Representations

Shuman Zhuang[1] , **Zhihao Wu**[2] , **Yuhong Chen**[3] , **Zihan Fang**[1] , **Jiali Yin**[1*] , **Ximeng Liu**[1]
[1]College of Computer and Data Science, Fuzhou University, Fuzhou, China
[2]College of Computer Science and Technology, Zhejiang University, Hangzhou, China
[3]Institute of Artificial Intelligence, Xiamen University, Xiamen, China
shumanzhuang@163.com, zhihaowu1999@gmail.com, {yhchen2320, fzihan11}@163.com,
jlyin@fzu.edu.cn, snbnix@gmail.com

Abstract

Facing the growing diversity of multi-view data, multi-view graph-based models have made encouraging progress in handling multi-view data modeled as graphs. Graph Contrastive Learning (GCL) naturally fits multi-view graph data by treating their inherent views as augmentations. However, the development of GCL on multi-view graph data is still in the infant stage. Challenges remain in designing strategies that coordinate preprocessing and contrastive learning, and in developing model architectures that automatically meet the needs of diverse views. To tackle these, we propose a framework named CAMEL, which refines consistency learning by introducing a tailored contrastive paradigm for multi-view graphs. Initially, we theoretically analyze the positive effect of edge-dropping preprocessing on the consistency and quantify the factors that influence it. Paired with a learnable model architecture, the proposed adaptive edge-dropping preprocessing strategy is guided by dynamic topology, making the heterogeneity of views more controllable and better aligned with contrastive learning. Finally, we design a neighborhood consistency multi-view contrastive objective that enhances consistency information interaction by extending positive samples. Extensive experiments on downstream tasks, including node classification and clustering, validate the superiority of our proposed model.

1 Introduction

Advancements in multimedia technologies have enabled diverse data collection ways, capturing rich and multi-faceted information. For instance, in multimedia analysis, complementary information can be captured via images, audio, and textual descriptions of the same content. Such data deriving from different sources can be modeled as multi-view data [Wan *et al.*, 2024b; Xu *et al.*, 2025a; Zhuang *et al.*, 2024]. Numerous traditional multi-view learning methods have been proposed, aiming to integrate these complementary views into cohesive representations [Zhang *et al.*, 2025; Fu *et al.*, 2023]. However, they struggle to model the complex intrinsic relationships among samples, particularly when such relationships extend beyond the feature space. In real-world scenarios, samples often exhibit interdependencies and intricate semantic relationships, making graph data a universal way to model these complex connections [Wu *et al.*, 2023b; Tu *et al.*, 2025]. By structuring multi-view data into graphs, latent relationships can be effectively captured and utilized.

Graph Contrastive Learning (GCL), as an unsupervised paradigm for graph-structured data analysis, leverages unlabeled samples to provide rich information [Li *et al.*, 2023; Zhuang *et al.*, 2025]. Notably, multi-view graph data, with its inherent multiple views as augmentations, is naturally well-suited for GCL. However, the development of GCL for multi-view data is still in its early stages [Liu *et al.*, 2023a; Wang *et al.*, 2023]. Contrastive learning follows the "augmentation-contrast" paradigm, where augmentations and contrastive strategies are interdependent. Typically, augmentations are handcrafted to align with the contrastive objective. For instance, the contrastive objective in GRACE [Zhu *et al.*, 2020] aims to maximize the mutual information between identical nodes across two augmented views. To achieve this, it employs perturbation strategies like edge dropout and feature masking to generate two lightly corrupted views for contrastive learning. Similarly, the contrastive objective in DGI [Velickovic *et al.*, 2019] is to maximize the mutual information between patch-summary pairs, using stochastic corruption techniques, such as feature shuffling, to introduce negative samples.

However, directly applying the GCL paradigm to multi-view graphs is suboptimal, as traditional GCL methods fail to account for the intricate heterogeneity inherent in multi-view scenarios. Specifically, in contrastive learning, multiple views are typically generated through manually defined perturbations, often utilizing simple, controllable transformations where the relationships between views are explicit and relatively straightforward to model. In contrast, the views inherent to multi-view data exhibit complex consistency and complementarity relationships that are difficult to capture [Sun *et al.*, 2024; Wang *et al.*, 2021]. These relationships encompass view-shared, view-specific task-relevant, and view-specific task-irrelevant information, which defy traditional

*Corresponding Author.

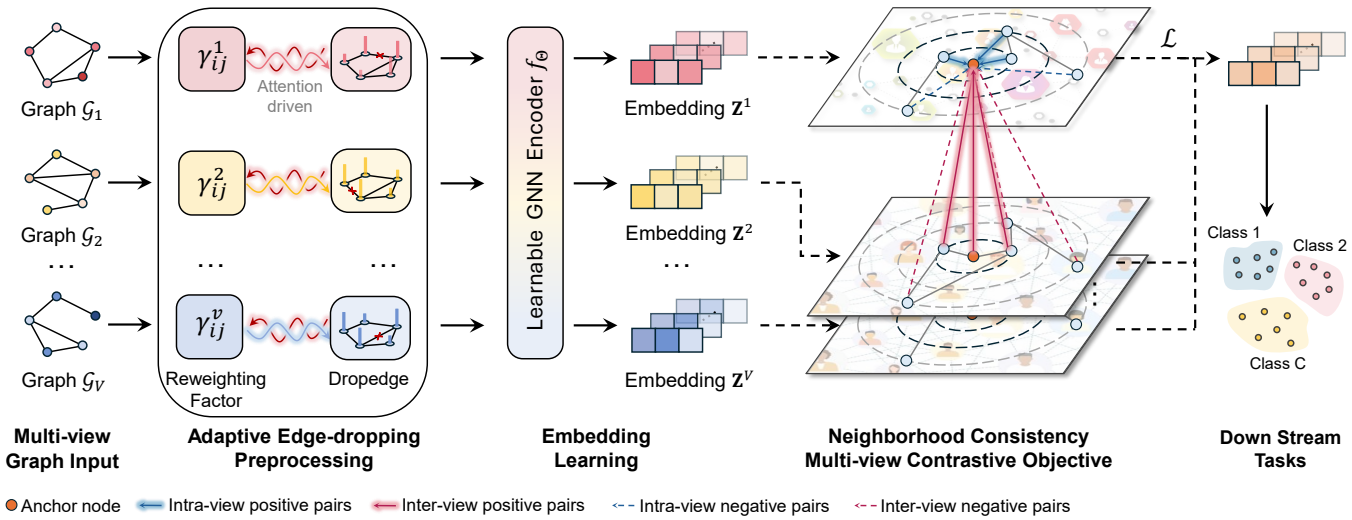


Figure 1: Overview of CAMEL. Within the learnable GNN architecture, edge-dropping preprocessing is applied to each view, guided by the reweighting factor. Then a neighborhood consistency multi-view contrastive objective is employed to obtain consistent representations.

quantification and modeling paradigms. Directly treating these multiple views as augmentations neglects potential distributional discrepancies and alignment issues, which may result in the shared GNN encoder failing to learn the desired representations across heterogeneous views. Thus, preprocessing strategies tailored for multi-view data are essential to make the discrepancies and interdependencies between views more manageable. Moreover, the contrastive objective and model architecture in GCL also lack consideration of the complexities of multi-view data and graph-specific properties, resulting in incomplete or potentially erroneous information interactions.

Based on the above analysis, we aim to design a novel contrastive learning paradigm tailored for multi-view graph data. Two key challenges remain for contrastive learning in the multi-view context: (1) *Strategically, how to design a preprocessing strategy that better aligns multiple views with GCL and a contrastive strategy that captures richer information?* (2) *Structurally, how to design a model architecture that is well-suited for multi-view graphs?*

To tackle these challenges, we propose a novel multi-view learning framework, termed Consistency-Aware Multi-view graph contrastivE Learning (CAMEL), which integrates meticulously designed strategies and architecture specifically tailored for multi-view graph data. Through rigorous mathematical analysis, we first reveal that the edge-dropping preprocessing promotes the consistency of multi-view representations. Moreover, we quantify the relationship between the impact of edge-dropping operations and the reweighting factor. To better learn consistency information and adapt to the heterogeneity across various views, we develop a learnable model architecture that enables each view to autonomously assimilate information and select the most suitable structure. Naturally, an adaptive edge-dropping preprocessing strategy is designed, where edge-dropping operations are steered by adaptive topology to ensure compatibility across different views. Finally, we introduce a neighborhood consistency multi-view contrastive strategy that redefines the construction of positive and negative pairs by considering graph properties. Specifically, the first-order neighbors of the anchor node, both within the same view and across views, are expanded as positive samples, facilitating the flow of consistency information. The flowchart of CAMEL is shown in Figure 1. Our contributions are summarized as

- We conduct a theoretical analysis that confirms edge dropping preprocessing promotes consistency learning and quantifies the factors influencing consistency.
- We propose a novel contrastive paradigm focused on multi-view graph data, termed CAMEL, which integrates a learnable model architecture with an adaptive edge-dropping preprocessing strategy and a neighborhood consistency multi-view contrastive objective.
- Extensive experiments on three types of datasets demonstrate the superiority of the proposed framework compared to state-of-the-art multi-view GNNs.

2 Related Work

2.1 Graph Contrastive Learning

Graph Contrastive Learning (GCL) has emerged as a pivotal research direction in graph data learning, focusing on maximizing mutual information between positive pairs while minimizing it for negative pairs. The core research focus of GCL lies in designing graph augmentation strategies, with edge deletion and attribute masking as the most prevalent techniques [Zhu *et al.*, 2020; Liu *et al.*, 2024b; Bu *et al.*, 2024]. Besides, some methods employ contrastive losses commonly used in computer vision, such as InfoNCE and Triplet loss [Wu *et al.*, 2022; Koromilas *et al.*, 2024], to extract the shared core information across different augmented views, following the InfoMax principle. Despite the significant advancements in GCL, the standard paradigm still faces

some limitations, particularly with regard to graph augmentations and contrastive objectives. Thus, other novel graph enhancement paradigms have also been proposed to expand the range of augmentation techniques. For instance, [Shen *et al.*, 2023] adopts a multi-head graph attention mechanism to learn graph augmentations. Other notable approaches include motif centrality-based adaptive augmentations [Li *et al.*, 2023] and similarity-preserving adversarial augmentations [In *et al.*, 2023]. Nonetheless, the utilization of GCL for multi-view data is notably limited, primarily due to a critical shortage of augmentation and contrastive strategies that are tailored specifically for such datasets.

2.2 Multi-view Learning

Multi-view learning aims to integrate and encode information from multiple sources to derive a low-dimensional representation that captures both the consistency and complementarity across different views. Graph-based multi-view learning [Wan *et al.*, 2024a; Lu *et al.*, 2024; Wu *et al.*, 2024], with its superior capability to model complex relationships, has become a widely adopted learning paradigm. For instance, [Liang *et al.*, 2022] defines a min-max formulation for robust learning that addresses issues of local optima. [Huang *et al.*, 2023] enhances the graph-based multi-view model by utilizing an attention allocation approach and a sample-weighting strategy. Recent methodologies have also introduced contrastive techniques into multi-view learning to harness and contrast the unique features from each view. For instance, [Shen *et al.*, 2023] introduces a contrastive multi-view kernel that implicitly embeds the views into a joint semantic space, while [Su *et al.*, 2024] builds on this to address the issue of false negatives. Despite the notable strides of these GCL-incorporated approaches in capturing graph-specific characteristics, their reliance on empirically derived heuristics and lack of theoretical rigor may limit generalization across diverse tasks and datasets.

3 Proposed Method

3.1 Preliminaries and Notations

Let $\mathcal{G} = \{\mathcal{G}^v\}_{v=1}^V$ denote a multi-view graph, where V is the number of views and $\mathcal{G}^v$ is the v-th view data. The feature matrix $\mathbf{X} \in \mathbb{R}^{n\times m}$ contains the node information, with each node having an m-dimensional feature vector. For the v-th view $\mathcal{G}^v = (\mathcal{V}, \mathcal{E}^v)$, where $\mathcal{V}$ is the node set with $|\mathcal{V}| = n$ and $\mathcal{E}^v \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set. The adjacency matrix of v-th view is defined as $\mathbf{A}^v \in \{0,1\}^{n\times n}$ with the degree matrix $\mathbf{D}^v$, where $d_i^v = \sum_j \mathbf{A}_{ij}^v$ and $\mathbf{A}_{ij}^v = 1$ iff $(v_i, v_j) \in \mathcal{E}^v$.

GCL aims to maximize the agreement of representations across views, which naturally fits the processing of multi-view data. Unlike conventional CL, Graph contrastive multi-view learning treats multi-view data as a natural augmentation of samples. Specifically, pair construction is generated by categorizing the instances from identical samples as positive pairs and the others as negative [Liu *et al.*, 2023b]. Typically, each view graph is first encoded by a shared GNN encoder to obtain normalized data representations, denoted as $\mathbf{Z}^v = f_\Theta(\mathbf{A}^v, \mathbf{X})$. Following InfoNCE [Oord *et al.*, 2018], the loss for the i-th data sample between the v-th and v'-th view can be defined as

$$\ell_{f_\Theta}(\boldsymbol{z}_i^v) = -\log \frac{e^{\theta(\boldsymbol{z}_i^v, \boldsymbol{z}_i^{v'})/\tau}}{\underbrace{e^{\theta(\boldsymbol{z}_i^v, \boldsymbol{z}_i^{v'})/\tau}}_{\text{positive pairs}} + \underbrace{\sum_{j\neq i} e^{\theta(\boldsymbol{z}_i^v, \boldsymbol{z}_j^v)/\tau} + \sum_{j\neq i} e^{\theta(\boldsymbol{z}_i^v, \boldsymbol{z}_j^{v'})\tau}}_{\text{negtive pairs}}}, \quad (1)$$

where θ is a similarity measure defined over node representations and τ is a temperature parameter. The encoder f_Θ can be substituted into any GNN model. Typically, the i-th node representation in the v-th view can be generalized using a basic form of message passing in single-layer GNN encoders:

$$\boldsymbol{z}_i^v = \sum_{j\in\{\mathcal{N}_i \cup i\}} \gamma_{i,j}^v \boldsymbol{z}_j^v, \quad (2)$$

where γ_{ij}^v is the weight for message passing trough the edge (v_i, v_j), termed reweighting factor, and is exemplified in vanilla GCN as $\gamma_{i,j}^v = \frac{1}{\sqrt{d_i d_j}}$.

3.2 Consistency Analyses on Preprocessing

Based on the previous analysis, directly treating the views inherent in multi-view data as augmentations is unreasonable. Here, inspired by [Xu *et al.*, 2025b], a theoretical analysis is conducted to explore the impact of view preprocessing on the consistency information within multi-view graphs. Initially, the Disruptive Index in each view can be evaluated by tracking the propagation of information through non-consensual edges, as defined by the following:

Definition 1 (Disruptive Index)**.** *Given a multi-view graph $\mathcal{G}^v = (\mathcal{V}, \mathcal{E}^v)$, the Disruptive Index (DI) quantifies the extent to which each view hinders consistent decision-making across views, defined as*

$$\mathfrak{I}^v = \frac{\sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{N}_i^{\text{non}}} \gamma_{ij}^v}{\sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{N}_i} \gamma_{ij}^v}, \quad (3)$$

where $\mathcal{N}_i$ denotes the neighbor set of node i, $\mathcal{N}_i^{\text{non}} \subseteq \mathcal{N}_i$ is the subset of neighbors connected by edges that hinder consensus learning, i.e., non-consensual edges. Conversely, the consensual edge set are defined as the edges in $\mathcal{N}_i \setminus \mathcal{N}_i^{non}$, which positively contribute to multi-view learning.

After DI is defined for each view based on the ratio of non-consensual information, the effective preprocessing strategy seeks a smaller DI. Here, we focus on commonly used edge perturbation techniques, analyzing their impact on enhancing representation consistency across views and mitigating the negative influence of non-consensual edges. Specifically, we examine the effect of dropping or adding an arbitrary edge $(a, b) \in \mathcal{E}^v$ on DI. Let $\tilde{\mathfrak{I}}^v$ denote the DI of the modified view with the corresponding reweighting factor $\tilde{\gamma}_{ij}^v$, the change of DI can be further represented as

$$\Delta = \tilde{\mathfrak{I}}^v - \mathfrak{I}^v. \quad (4)$$

Assume $s = \sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{N}_i} \gamma_{ij}^v$ and $\tilde{s} = \sum_{i\in\mathcal{V}} \sum_{j\in\mathcal{N}_i} \tilde{\gamma}_{ij}^v$, we deduce the following quantification operator ω as

$$\omega = s \cdot \tilde{s} \cdot \Delta, \quad (5)$$

which satisfies $\omega \propto \Delta$, effectively reflecting the change in non-consensual information within the view. To further explore the changes of DI, we precisely quantify ω:

Lemma 1. *Assuming each node v_i in the view receives non-consensual information at a ratio m relative to total information, it can be derived that:*

$$\begin{cases} \omega = 2(m - \mathbb{1}_{\{a,b\}\in\mathscr{E}^v}) \cdot \tilde{s} \cdot \gamma_{ij}^v, & \textit{if an edge is dropped}, \\ \omega = 2(\mathbb{1}_{\{a,b\}\in\mathscr{E}^v} - m) \cdot \tilde{s} \cdot \gamma_{ij}^v, & \textit{if an edge is added}, \end{cases} \tag{6}$$

where $\mathscr{E}^v \in \mathcal{E}^v$ is the set of non-consensual edges.

The full derivation of Eq. (5) and Lemma 1 can be found in Appendix B. Overall, combining (5) and (6), the following theorem can be deduced:

Theorem 1. *When any edge $(a, b) \in \mathcal{E}^v$ is dropped or added from a graph $\mathcal{G}^v$, it results in:*

$$\begin{cases} \Delta = \frac{2(m-\mathbb{1}_{\{a,b\}\in\mathscr{E}^v)}}{s} \cdot \gamma_{ij}^v, & \textit{if an edge is dropped} \\ \Delta = \frac{2(\mathbb{1}_{\{a,b\}\in\mathscr{E}^v}-m)}{s} \cdot \gamma_{ij}^v, & \textit{if an edge is added} \end{cases} \tag{7}$$

For edge-dropping, the following holds:

$$\begin{cases} \Delta = \frac{2(m-1)}{s} \cdot \gamma_{ij}^v, & \textit{if } (a,b) \in \mathscr{E}^v \\ \Delta = \frac{2m}{s} \cdot \gamma_{ij}^v, & \textit{if } (a,b) \notin \mathscr{E}^v \end{cases} \tag{8}$$

For edge-adding, the following holds:

$$\begin{cases} \Delta = \frac{2(1-m)}{s} \cdot \gamma_{ij}^v, & \textit{if } (a,b) \in \mathscr{E}^v \\ \Delta = \frac{-2m}{s} \cdot \gamma_{ij}^v, & \textit{if } (a,b) \notin \mathscr{E}^v \end{cases} \tag{9}$$

Assuming that the DI for the views in multi-view data is below 0.5, i.e., $m < 0.5$, which is typical in most cases, we observe the following effects of edge perturbation: For edge-dropping, If the dropped edge is non-consensual, $\Delta < 0$, indicating an improvement in consistency with a greater consistency gain; If the dropped edge is consensual, $\Delta > 0$, resulting in disruption in consistency, but with less significant damage. Edge dropping results in a larger consistency gain than loss, while the trend is reversed for the edge adding operation. Thus, Theorem 1 confirms that *edge dropping promotes consistency learning more effectively than edge adding.*

3.3 Adaptive Edge-dropping GNN Architecture

Following the theoretical analysis indicating that edge-dropping promotes multi-view consistency learning, thus we adopt edge-dropping as a preprocessing step for each view. As per Theorem 1, the change in DI satisfies $\Delta \propto \gamma_{ij}^v$, motivating the use of γ_{ij}^v to control edge-dropping. Previous GCL models employing GNN encoders indiscriminately smooth all node representations within a neighborhood, often struggling to accommodate the heterogeneity of multi-view data. Therefore, we propose a more general and flexible model architecture suitable for multi-view contexts, adopting an adaptive reweighting factor $\gamma_{ij}^v = \eta_{ij}^v / \sqrt{d_i d_j}$, where η_{ij}^v ranges between 0 and 1. It enables nodes to selectively assimilate information during message passing across diverse views. Specifically, we employ a gating mechanism to learn η_{ij}^v:

$$\eta_{i,j}^v = \mathrm{softmax}(\sigma(\mathbf{W}^\top [\boldsymbol{z}_i^v || \boldsymbol{z}_j^v])), \tag{10}$$

where $\sigma(\cdot)$ is the activation function, $\mathbf{W} \in \mathbb{R}^{2F'}$ is a trainable weight matrix with the hidden uints F'. Combining reweighting factors with Theorem 1, the probability of an edge (v_i, v_j) being dropped is guided by the reweighting factor as follows:

$$p_{ij}^v = \min\left(\frac{\gamma_{ij}^v - \min(\gamma^v)}{\mathrm{mean}(\gamma^v) - \min(\gamma^v)} \cdot \alpha, p_\tau\right), \tag{11}$$

where α is a hyper-parameter used to control the drop probability, and p_τ is a cutoff probability that does not exceed 1. Based on the edge-dropping probability p_{ij}^v, a masking variable s_{ij}^v can be sampled, whose value is drawn from a Bernoulli distribution: $s_{ij}^v \sim \mathcal{B}(p_{ij}^v)$. The reweighting factor can be computed as $\gamma_{ij}^v = (1 - s_{ij}^v) \cdot \gamma_{ij}^v$ if $s_{ij}^v = 1$ and $\gamma_{ij}^v = 0$ otherwise, with the probability of $s_{ij}^v = 1$ determined by p_{ij}^v.

For each v-th view, a single-layer learnable GNN encoder $f_\Theta(\cdot)$ learns the embedding of each node by aggregating the embeddings of its neighbors with adaptive structure summarized as:

$$\boldsymbol{z}_i^v = \sigma(\mathbf{W}_{\mathrm{in}}^\top \boldsymbol{x}_i), \tag{12}$$

$$\gamma_{i,j}^v = \mathrm{softmax}(\sigma(\mathbf{W}^\top [\boldsymbol{z}_i^v || \boldsymbol{z}_j^v]))/\sqrt{d_i d_j}, \tag{13}$$

$$\gamma_{ij}^v = (1 - s_{ij}^v) \cdot \gamma_{ij}^v, \quad \text{s.t. } s_{ij}^v \in \{0, 1\} \tag{14}$$

$$\boldsymbol{z}_i^v = (1 - \beta) \sum_{j \in \mathcal{N}_i} \gamma_{i,j}^v \boldsymbol{z}_j^v + \beta \boldsymbol{z}_i^v, \tag{15}$$

where $(1 - \beta)$ is added for the trade-off between self feature and message from neighbors. The advantages of the proposed adaptive edge-dropping architecture method can be summarized as follows: (1) The edge-dropping preprocessing promotes multi-view consistency, ensuring that the discrepancies between views are controllable, in accordance with the theoretical analysis. (2) Unlike fixed structures, the shared learnable model architecture enables each view to autonomously adapt and discover an optimal structure tailored to its specific needs, without relying on prior knowledge. (3) The edge-dropping operation is highly related to the dynamically learned topology, enabling compatibility with each individual view.

3.4 Neighborhood Consistency Multi-View Contrastive Objective

Multi-view data modeled as graphs inherently exhibit graph properties like homogeneity, where connected nodes often share the same label. However, the pairing strategy in Eq. (1), which designates instances from the same sample as positive pairs while treating all others as negatives, risks separating nodes linked by consensual edges. To address this issue, we propose a neighborhood consistency contrastive objective tailored for multi-view graph data, redefining the construction of positive and negative pairs. Specifically, we expand positive samples by incorporating the first-order neighbors of anchors within the same view and across different views. The loss of the i-th data sample between the v-th and v'-th view as defined in Eq. (1) can also be represented as:

$$\ell_{f_\Theta}(\boldsymbol{z}_i^v) = -\log \frac{\mathrm{pos}_{f_\Theta(\boldsymbol{z}_i^v)}}{\mathrm{pos}_{f_\Theta(\boldsymbol{z}_i^v)} + \mathrm{neg}_{f_\Theta(\boldsymbol{z}_i^v)}}. \tag{16}$$

Datasets	ACM		DBLP		IMDB		YELP	
# Nodes	3,025		4,057		4,780		2,614	
Metrics	Macro F1	Micro F1	Macro F1	Micro F1	Macro F1	Micro F1	Macro F1	Micro F1
GCN	76.7 (5.2)	78.0 (4.5)	90.7 (0.6)	91.4 (0.5)	23.6 (0.1)	54.6 (0.0)	55.5 (0.9)	74.0 (0.4)
SGC	67.5 (1.1)	70.2 (0.8)	66.3 (0.5)	81.5 (0.6)	27.0 (1.4)	54.8 (0.7)	51.9 (0.7)	67.4 (1.5)
HAN	87.9 (0.4)	88.0 (0.4)	89.3 (0.2)	90.4 (0.2)	24.0 (1.2)	55.9 (1.2)	55.3 (4.5)	68.0 (5.0)
DMGI	66.7 (1.8)	70.4 (1.1)	75.4 (1.3)	81.2 (0.7)	38.3 (3.1)	57.0 (0.4)	52.7 (2.3)	69.5 (0.7)
IGNN	82.9 (0.0)	82.7 (0.0)	86.8 (0.0)	87.5 (0.0)	45.3 (0.0)	54.8 (0.0)	71.4 (0.0)	75.0 (0.0)
MRGCN	87.6 (0.2)	87.5 (0.2)	89.5 (1.6)	90.5 (1.2)	45.2 (2.3)	47.7 (2.3)	54.4 (0.4)	73.7 (0.5)
SSDCM	84.3 (3.5)	85.2 (2.9)	87.7 (2.7)	88.8 (3.9)	35.3 (3.5)	50.8 (2.7)	55.9 (3.0)	68.9 (5.6)
MHGCN	88.8 (1.6)	89.1 (0.7)	**92.5 (0.3)**	**93.0 (0.2)**	**51.2 (1.3)**	**61.2 (1.3)**	60.9 (1.0)	73.3 (0.2)
AMOGCN	**91.4 (0.5)**	**90.0 (0.6)**	92.3 (0.4)	92.8 (0.4)	49.1 (0.8)	61.0 (1.3)	**71.8 (2.1)**	**77.4 (0.4)**
CAMEL	**93.1 (0.1)**	**92.2 (0.3)**	**93.1 (0.1)**	**93.6 (0.1)**	**50.2 (0.4)**	**61.3 (0.0)**	**91.3 (0.5)**	**90.1 (0.6)**

Table 1: Macro F1 and Micro F1 scores of all methods on multi-relational datasets, where the best results are highlighted in **orange** and the second-best results are highlighted in **blue**.

In the proposed contrastive objective, the positive pairs and the negative pairs are reconstructed based on the first-order neighborhood:

$$\mathrm{pos}_{f_\Theta}(\boldsymbol{z}_i^v) = \underbrace{\sum_{j\in\mathcal{N}_i^v} e^{\theta\left(\boldsymbol{z}_i^v,\boldsymbol{z}_j^v\right)/\tau}}_{\text{intra-view positive pairs}} + \underbrace{\sum_{j\in\{\mathcal{N}_i^{v'}\cup i\}} e^{\theta\left(\boldsymbol{z}_i^v,\boldsymbol{z}_j^{v'}\right)/\tau}}_{\text{inter-view positive pairs}}, \tag{17}$$

$$\mathrm{neg}_{f_\Theta}(\boldsymbol{z}_i^v) = \underbrace{\sum_{j\notin\{\mathcal{N}_i^v\cup i\}} e^{\theta\left(\boldsymbol{z}_i^v,\boldsymbol{z}_j^v\right)/\tau}}_{\text{intra-view negative pairs}} + \underbrace{\sum_{j\notin\{\mathcal{N}_i^{v'}\cup i\}} e^{\theta\left(\boldsymbol{z}_i^v,\boldsymbol{z}_j^{v'}\right)/\tau}}_{\text{inter-view negative pairs}}, \tag{18}$$

where $\mathcal{N}_i^v$, $\mathcal{N}_i^{v'}$ are the neighbor sets of node i in view v and v'. Specifically, with $\boldsymbol{z}_i^v$ as the anchor, positive pairs are not only derived from the instance $\boldsymbol{z}_i^{v'}$ in a different view v' but also incorporate neighbors from both the same and across views, i.e., $\{\boldsymbol{z}_j^v | j \in \mathcal{N}_i^v\}$ and $\{\boldsymbol{z}_j^{v'} | j \in \mathcal{N}_i^{v'}\}$ respectively. Naturally, all non-neighbors of sample i across view v and v' serve as intra-view and inter-view negatives. Expanding positive samples through neighborhood structures harnesses the full topological information, especially the consensual edges, to capture the consistency and complementarity both within and across views.

Since any data view can be considered a semantic augmentation of another, we randomly choose one view as the pivot view, termed $\boldsymbol{z}^p$. The rationale behind constructing the loss in this manner is detailed in Appendix C. The overall loss is then implemented as the sum of the losses between the pivot view and the other views, as follows:

$$\mathcal{L} = \frac{1}{2V}\sum_{v=1,v\neq p}^{V} \ell_{f_\Theta}(\boldsymbol{z}_i^v, \boldsymbol{z}_i^p), \tag{19}$$

$$\ell_{f_\Theta}(\boldsymbol{z}_i^v, \boldsymbol{z}_i^p) = \frac{1}{2N}\sum_{i}^{N}(\ell_{f_\Theta}(\boldsymbol{z}_i^v), \ell_{f_\Theta}(\boldsymbol{z}_i^p)). \tag{20}$$

The algorithm procedure of CAMEL is given in Appendix A.

4 Experiments

4.1 Experimental Settings

Dataset. To assess the performance of CAMEL, we conduct experiments on three types of datasets, including four multi-relational datasets (ACM, DBLP, IMDB, YELP), two multi-attribute datasets (COIL20, NoisyMNIST), and two multi-modality datasets (Iaprtc12, NUS-Wide).

Compared Methods. We compare CAMEL with twenty-one representative methods, categorized as follows: (1) *Baseline methods:* GCN [Kipf and Welling, 2017] and SGC [Wu *et al.*, 2019]; (2) *Heterogeneous methods:* HAN [Wang *et al.*, 2019], DMGI [Park *et al.*, 2020], IGNN [Gu *et al.*, 2020], MRGCN [Huang *et al.*, 2020], SSDCM [Mitra *et al.*, 2021], MHGCN [Yu *et al.*, 2022], and AMOGCN [Chen *et al.*, 2024]; (3) *Multi-view classification methods:* ERL-MVSC [Huang *et al.*, 2021], Dynamics [Han *et al.*, 2022], PDMF [Xu *et al.*, 2023], IMvGCN [Wu *et al.*, 2023a], LGCN-FF [Chen *et al.*, 2023a], and RCML [Xu *et al.*, 2024]; (4) *Multi-view clustering methods:* OMSC [Chen *et al.*, 2022], SDSNE [Liu *et al.*, 2022], DSMVC [Tang and Liu, 2022], FSMSC [Chen *et al.*, 2023b], RCAGL [Liu *et al.*, 2024a], and SMVAGC [Wang *et al.*, 2024].

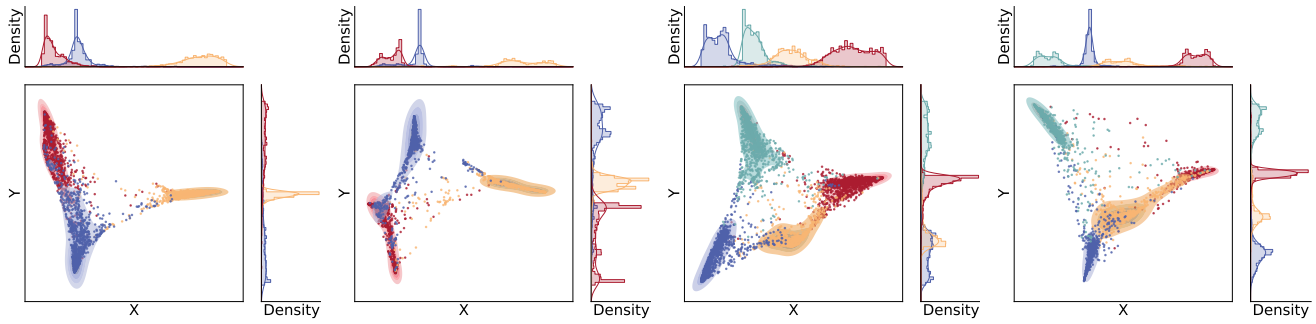
Experimental Settings. The implementation details and parameter settings are introduced in Appendix D.

4.2 Node Classification Performance

In this section, we explore semi-supervised classification tasks, evaluating the efficacy of CAMEL in processing complex data across three types of multi-view graphs: (1) **Multi-relational Graphs.** Classification results for four multi-relational graphs, with a training ratio of 20%, are detailed in Table 1. From these results, we can see that CAMEL surpasses most methods designed for multi-relational data, affirming its effectiveness. Furthermore, CAMEL considerably outperforms baseline models like GCN and SGC, illustrating that static architectures are ill-equipped to manage the heterogeneity inherent in complex datasets. (2) **Multi-attribute**

Datasets	COIL20		NoisyMNIST		Iaprtc12		NUS-Wide	
#Nodes	1440		15000		7855		20000	
Task	Node Classification							
Metrics	Macro F1	Micro F1	Macro F1	Micro F1	Macro F1	Micro F1	Macro F1	Micro F1
ERL-MVSC	87.1 (1.5)	87.4 (1.4)	89.4 (0.0)	89.3 (0.0)	**59.8 (0.8)**	**60.1 (1.2)**	51.2 (0.2)	51.3 (0.1)
Dynamics	70.7 (0.1)	71.6 (0.1)	70.4 (0.6)	72.8 (0.4)	50.1 (0.5)	50.8 (0.3)	60.5 (0.1)	62.7 (0.1)
PDMF	63.6 (1.7)	64.9 (1.7)	66.6 (1.9)	69.7 (1.3)	47.6 (2.6)	46.2 (2.2)	37.1 (1.5)	40.5 (1.3)
IMvGCN	71.3 (0.3)	70.9 (0.4)	86.3 (0.6)	86.0 (0.8)	52.9 (1.0)	53.6 (0.7)	51.9 (1.5)	53.2 (1.2)
LGCN-FF	**89.7 (0.4)**	**89.4 (0.3)**	**89.6 (0.5)**	**89.8 (0.5)**	59.3 (1.4)	57.8 (1.3)	OM	OM
RCML	64.0 (0.3)	66.3 (0.3)	84.5 (0.1)	85.2 (0.1)	39.4 (0.2)	40.7 (0.2)	**54.2 (0.7)**	**56.8 (0.4)**
CAMEL	**92.5 (0.3)**	**92.4 (0.3)**	**91.9 (1.7)**	**92.2 (1.6)**	**61.6 (0.5)**	**60.7 (0.3)**	**69.2 (0.6)**	**69.6 (0.2)**
Task	Node Clustering							
Metrics	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
OMSC	52.0 (2.1)	71.4 (2.4)	49.6 (0.9)	44.3 (1.7)	**44.9 (1.4)**	22.9 (0.6)	36.3 (1.1)	21.6 (1.4)
SDSNE	64.2 (3.0)	77.8 (2.3)	40.0 (11.7)	51.1 (16.7)	36.3 (2.9)	19.4 (3.7)	25.9 (5.9)	19.0 (6.9)
DSMVC	66.7 (3.9)	79.3 (1.5)	33.5 (3.2)	26.6 (3.3)	35.1 (1.9)	14.3 (1.4)	29.7 (1.9)	13.9 (1.4)
FSMSC	**74.7 (1.4)**	**82.8 (0.7)**	**55.5 (0.7)**	**51.3 (1.1)**	44.6 (0.8)	22.2 (0.6)	37.3 (0.2)	21.7 (0.1)
RCAGL	66.8 (0.9)	81.2 (0.5)	55.2 (1.4)	49.8 (1.7)	43.6 (0.6)	**24.7 (0.5)**	**40.2 (0.7)**	**23.5 (0.4)**
SMVAGC	65.0 (4.2)	79.1 (1.7)	53.9 (2.4)	46.2 (1.1)	39.2 (0.8)	17.7 (0.4)	36.1 (2.3)	22.9 (0.9)
CAMEL	**78.7 (0.3)**	**86.2 (0.2)**	**84.8 (1.4)**	**81.7 (2.5)**	**45.3 (1.3)**	**25.4 (1.1)**	**42.4 (0.6)**	**26.1 (1.2)**

Table 2: Macro F1 and Micro F1 scores for the node classification task, along with ACC and NMI for the node clustering task on multi-attribute and multi-modality datasets, where the best results are highlighted in **orange** and the second-best results are highlighted in **blue**.



(a) $\text{ACM}_{\text{AMOGCN}}$ (b) $\text{ACM}_{\text{CAMEL}}$ (c) $\text{DBLP}_{\text{AMOGCN}}$ (d) $\text{DBLP}_{\text{CAMEL}}$

Figure 2: t-SNE visualization of the learned representations in ACM and DBLP dataset. Each node is colored according to its label.

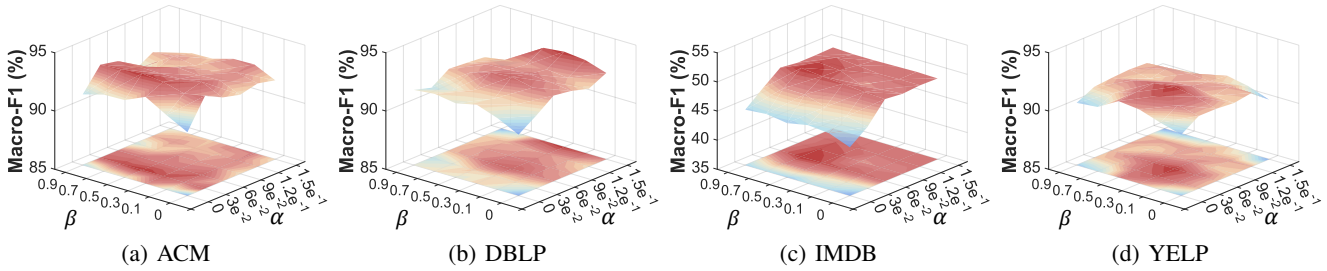
and multi-modality Graphs. We run six multi-view classification methods with a training rate of 10%, and the results are presented in Table 2. It can be observed that graph-based methods dominate the performance landscape, underscoring the pivotal role of leveraging relational information between multi-view samples for performance improvement. Among these, CAMEL achieves superior results through its adept adaptation to the heterogeneity of multi-view data, supported by the advanced contrastive strategy and model architecture.

4.3 Node Clustering Performance

In this section, we assessed node clustering performance on multi-attribute datasets COIL20 and NoisyMNIST, along with multi-modality datasets Iaprtc12 and NUS-wide. Table 2 indicates that CAMEL substantially exceeds the performance of state-of-the-art clustering methods on both ACC and NMI metrics, confirming the efficacy of the proposed contrastive paradigm in producing distinguishable representations. The superior clustering outcomes stem primarily from CAMEL enhancing connectivity density among node pairs distanced from estimated decision boundaries, facilitating the formation of compact, well-defined clusters.

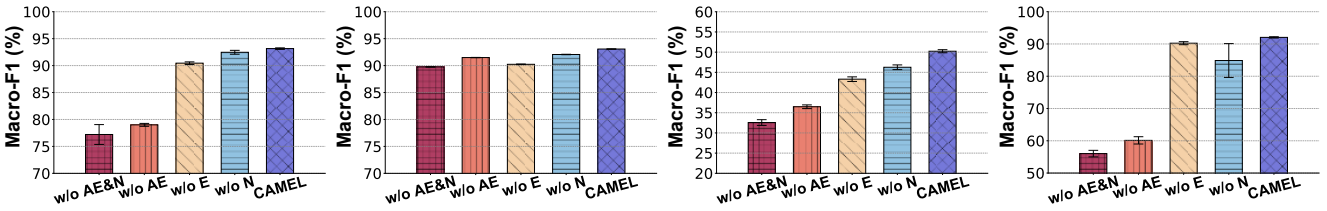
4.4 Visualization

Beyond quantitative analysis, we also visualize the representations derived by the competitive baseline algorithm AMOGCN and the proposed method CAMEL in Figures 2 on the ACM and DBLP datasets using t-SNE [Van der Maaten L, 2008]. Contour lines within colored shaded areas indicate the density levels of different classes. Moreover, the histograms



(a) ACM (b) DBLP (c) IMDB (d) YELP

Figure 3: Parameter sensitivity across all multi-relational datasets.



(a) ACM (b) DBLP (c) IMDB (d) YELP

Figure 4: Ablation study (mean% and standard deviation%) .

positioned at the top and right sides of each plot illustrate the distribution of t-SNE-transformed data along the X and Y axes, respectively. From the visualizations, it is observed that the embeddings generated by CAMEL show distinctly clearer class boundaries and tighter intra-class cohesion than those by AMOGCN. This distinction is observable in the density plots, where CAMEL's embeddings show reduced class overlap and more distinct peaks within each class's distribution. The histograms complement these findings by presenting a more noticeable separation along both axes. These results show that CAMEL excels in learning high-quality node representations, as evidenced by its ability to produce representations with well-defined and separate class clusters.

4.5 Hyperparameter Analysis

In this subsection, we conduct parameter sensitivity to explore the impact of α in Eq. (11) and β in Eq. (15) on model performance. Specifically, α modulates the adaptive edge-dropping probability, while β regulates the balance between self-features and neighbor messages in the learnable GNN encoder. Generally, α shows peak performance at mid-range values across the datasets, indicating that moderate edge-dropping efficiently captures critical semantic information and promotes consistent learning in multi-view data. As α and β values vary, the performance exhibits stable fluctuations across most datasets, reflecting the CAMEL framework's insensitivity to parameter changes. Notably, both α and β yield suboptimal performance when set to zero, especially in the IMDB, highlighting the pivotal role of the adaptive edge-dropping architecture in optimizing model efficacy.

4.6 Ablation Study

To evaluate the importance of each component in CAMEL, we conducted ablation experiments by removing key elements: edge-dropping (denoted as E), the adaptive edge-dropping architecture (denoted as AE), and the neighborhood consistency multi-view contrastive objective (denoted as N). Specifically, to evaluate the adaptive edge-dropping architecture, we replaced the encoder with a vanilla GCN. For the contrastive objective evaluation, we used the contrastive loss from the classic GCL method GRACE [Zhu *et al.*, 2020]. The results of these experiments are shown in Figure 4. Notably, removing the adaptive edge-dropping architecture (w/o AE) leads to a sharp performance drop, highlighting the superiority of the learnable architecture in adapting to heterogeneous views. Subsequently, the absence of either the edge-dropping preprocessing or the contrastive objective results in a certain decrease in performance. Overall, CAMEL consistently outperforms each variant, validating its ability to effectively leverage consistency and complementarity across views.

5 Conclusion

In this paper, we reveal that multi-view graph contrastive learning requires preprocessing, with the edge-dropping strategy promoting consistency supported by mathematical derivation. Inspired by these findings, we introduce a multi-view contrastive paradigm called CAMEL, which synergistically enhances both strategy and architecture to learn consistent representations. On the strategic side, the proposed adaptive edge-dropping preprocessing strategy, guided by dynamic topology, along with the neighborhood consistency contrastive objective, effectively extracts consistency information from multi-view data. On the architectural side, the learnable GNN encoder boosts the adaptability to diverse views. Extensive experiments on various datasets validate the superiority of CAMEL.

Acknowledgments

This work was partially supported by the National Key Research and Development Program of China (2023YFE0111100) and the National Natural Science Foundation of China under Grant Nos. 62202104 and U2441239.

References

[Bu *et al.*, 2024] Weixin Bu, Xiaofeng Cao, Yizhen Zheng, and Shirui Pan. Improving augmentation consistency for graph contrastive learning. *Pattern Recognition*, 148:110182, 2024.

[Chen *et al.*, 2022] Man-Sheng Chen, Chang-Dong Wang, Dong Huang, Jian-Huang Lai, and Philip S Yu. Efficient orthogonal multi-view subspace clustering. In *SIGKDD*, pages 127–135, 2022.

[Chen *et al.*, 2023a] Zhaoliang Chen, Lele Fu, Jie Yao, Wenzhong Guo, Claudia Plant, and Shiping Wang. Learnable graph convolutional network and feature fusion for multi-view learning. *Information Fusion*, 95:109–119, 2023.

[Chen *et al.*, 2023b] Zhe Chen, Xiao-Jun Wu, Tianyang Xu, and Josef Kittler. Fast self-guided multi-view subspace clustering. *IEEE TIP*, 32:6514–6525, 2023.

[Chen *et al.*, 2024] Zhaoliang Chen, Zhihao Wu, Luying Zhong, Claudia Plant, Shiping Wang, and Wenzhong Guo. Attributed multi-order graph convolutional network for heterogeneous graphs. *Neural Networks*, page 106225, 2024.

[Fu *et al.*, 2023] Lele Fu, Zhaoliang Chen, Yongyong Chen, and Shiping Wang. Unified low-rank tensor learning and spectral embedding for multi-view subspace clustering. *IEEE TMM*, 25:4972–4985, 2023.

[Gu *et al.*, 2020] Fangda Gu, Heng Chang, Wenwu Zhu, Somayeh Sojoudi, and Laurent El Ghaoui. Implicit graph nn. In *NeurIPS*, 2020.

[Han *et al.*, 2022] Zongbo Han, Fan Yang, Junzhou Huang, Changqing Zhang, and Jianhua Yao. Multimodal dynamics: Dynamical fusion for trustworthy multimodal classification. In *CVPR*, pages 20707–20717, 2022.

[Huang *et al.*, 2020] Zhichao Huang, Xutao Li, Yunming Ye, and Michael K. Ng. MR-GCN: multi-relational graph convolutional networks based on generalized tensor product. In *IJCAI*, pages 1258–1264, 2020.

[Huang *et al.*, 2021] Aiping Huang, Zheng Wang, Yannan Zheng, Tiesong Zhao, and Chia-Wen Lin. Embedding regularizer learning for multi-view semi-supervised classification. *IEEE TIP*, 30:6997–7011, 2021.

[Huang *et al.*, 2023] Zongmo Huang, Yazhou Ren, Xiaorong Pu, Shudong Huang, Zenglin Xu, and Lifang He. Self-supervised graph attention networks for deep weighted multi-view clustering. In *AAAI*, volume 37, pages 7936–7943, 2023.

[In *et al.*, 2023] Yeonjun In, Kanghoon Yoon, and Chanyoung Park. Similarity preserving adversarial graph contrastive learning. In *SIGKDD*, pages 867–878, 2023.

[Kipf and Welling, 2017] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.

[Koromilas *et al.*, 2024] Panagiotis Koromilas, Giorgos Bouritsas, Theodoros Giannakopoulos, Mihalis Nicolaou, and Yannis Panagakis. Bridging mini-batch and asymptotic analysis in contrastive learning: From infonce to kernel-based losses. In *ICML*, 2024.

[Li *et al.*, 2023] Wen-Zhi Li, Chang-Dong Wang, Jian-Huang Lai, and S Yu Philip. Towards effective and robust graph contrastive learning with graph autoencoding. *IEEE TKDE*, 36(2):868–881, 2023.

[Liang *et al.*, 2022] Weixuan Liang, Xinwang Liu, Sihang Zhou, Jiyuan Liu, Siwei Wang, and En Zhu. Robust graph-based multi-view clustering. In *AAAI*, volume 36, pages 7462–7469, 2022.

[Liu *et al.*, 2022] Chenghua Liu, Zhuolin Liao, Yixuan Ma, and Kun Zhan. Stationary diffusion state neural estimation for multiview clustering. In *AAAI*, volume 36, pages 7542–7549, 2022.

[Liu *et al.*, 2023a] Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Qing Liao, and Yuanqing Xia. Contrastive multi-view kernel learning. *IEEE TPAMI*, 45(8):9552–9566, 2023.

[Liu *et al.*, 2023b] Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Qing Liao, and Yuanqing Xia. Contrastive multi-view kernel learning. *IEEE TPAMI*, 45(8):9552–9566, 2023.

[Liu *et al.*, 2024a] Suyuan Liu, Qing Liao, Siwei Wang, Xinwang Liu, and En Zhu. Robust and consistent anchor graph learning for multi-view clustering. *IEEE TKDE*, 2024.

[Liu *et al.*, 2024b] Yixin Liu, Shiyuan Li, Yu Zheng, Qingfeng Chen, Chengqi Zhang, and Shirui Pan. Arc: A generalist graph anomaly detector with in-context learning. In *NeurIPS*, volume 37, pages 50772–50804, 2024.

[Lu *et al.*, 2024] Jielong Lu, Zhihao Wu, Luying Zhong, Zhaoliang Chen, Hong Zhao, and Shiping Wang. Generative essential graph convolutional network for multi-view semi-supervised classification. *IEEE TMM*, 26:7987–7999, 2024.

[Mitra *et al.*, 2021] Anasua Mitra, Priyesh Vijayan, Sanasam Ranbir Singh, Diganta Goswami, Srinivasan Parthasarathy, and Balaraman Ravindran. Semi-supervised deep learning for multiplex networks. In *SIGKDD*, pages 1234–1244, 2021.

[Oord *et al.*, 2018] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[Park *et al.*, 2020] Chanyoung Park, Donghyun Kim, Jiawei Han, and Hwanjo Yu. Unsupervised attributed multiplex network embedding. In *AAAI*, pages 5371–5378, 2020.

[Shen *et al.*, 2023] Xiao Shen, Dewang Sun, Shirui Pan, Xi Zhou, and Laurence T Yang. Neighbor contrastive learning on learnable graph augmentation. In *AAAI*, volume 37, pages 9782–9791, 2023.

[Su *et al.*, 2024] Peng Su, Yixi Liu, Shujian Li, Shudong Huang, and Jiancheng Lv. Robust contrastive multi-view kernel clustering. In *IJCAI*, pages 4938–4945, 2024.

[Sun *et al.*, 2024] Yuan Sun, Yang Qin, Yongxiang Li, Dezhong Peng, Xi Peng, and Peng Hu. Robust multi-view clustering with noisy correspondence. *IEEE TKDE*, 36(12):9150–9162, 2024.

[Tang and Liu, 2022] Huayi Tang and Yong Liu. Deep safe multi-view clustering: Reducing the risk of clustering performance degradation caused by view increase. In *CVPR*, pages 202–211, 2022.

[Tu *et al.*, 2025] Wenxuan Tu, Sihang Zhou, Xinwang Liu, Zhiping Cai, Yawei Zhao, Yue Liu, and Kunlun He. Wage: Weight-sharing attribute-missing graph autoencoder. *IEEE TPAMI*, 47(7):5760–5777, 2025.

[Van der Maaten L, 2008] Hinton G Van der Maaten L. Visualizing data using t-sne. *Journal of Machine Learning Research*, 2008.

[Velickovic *et al.*, 2019] Petar Velickovic, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax. In *ICLR*, 2019.

[Wan *et al.*, 2024a] Xinhang Wan, Jiyuan Liu, Hao Yu, Qian Qu, Ao Li, Xinwang Liu, Ke Liang, Zhibin Dong, and En Zhu. Contrastive continual multiview clustering with filtered structural fusion. *IEEE TNNLS*, pages 1–14, 2024.

[Wan *et al.*, 2024b] Xinhang Wan, Bin Xiao, Xinwang Liu, Jiyuan Liu, Weixuan Liang, and En Zhu. Fast continual multi-view clustering with incomplete views. *IEEE TIP*, 33:2995–3008, 2024.

[Wang *et al.*, 2019] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In *WWW*, pages 2022–2032, 2019.

[Wang *et al.*, 2021] Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao, and Yun Fu. Generative partial multi-view clustering with adaptive fusion and cycle consistency. *IEEE TIP*, 30:1771–1783, 2021.

[Wang *et al.*, 2023] Zehong Wang, Qi Li, Donghua Yu, Xiaolong Han, Xiao-Zhi Gao, and Shigen Shen. Heterogeneous graph contrastive multi-view learning. In *SDM*, pages 136–144, 2023.

[Wang *et al.*, 2024] Siwei Wang, Xinwang Liu, Suyuan Liu, Wenxuan Tu, and En Zhu. Scalable and structural multi-view graph clustering with adaptive anchor fusion. *IEEE TIP*, 2024.

[Wu *et al.*, 2019] Felix Wu, Amauri H. Souza Jr., Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Q. Weinberger. Simplifying graph convolutional networks. In *ICML*, volume 97, pages 6861–6871, 2019.

[Wu *et al.*, 2022] Chuhan Wu, Fangzhao Wu, and Yongfeng Huang. Rethinking infonce: How many negative samples do you need? In *IJCAI*, pages 2509–2515, 2022.

[Wu *et al.*, 2023a] Zhihao Wu, Xincan Lin, Zhenghong Lin, Zhaoliang Chen, Yang Bai, and Shiping Wang. Interpretable graph convolutional network for multi-view semi-supervised learning. *IEEE TMM*, 25:8593–8606, 2023.

[Wu *et al.*, 2023b] Zhihao Wu, Zhao Zhang, and Jicong Fan. Graph convolutional kernel machine versus graph convolutional networks. In *NeurIPS*, volume 36, pages 19650–19672, 2023.

[Wu *et al.*, 2024] Zhihao Wu, Zhaoliang Chen, Shide Du, Sujia Huang, and Shiping Wang. Graph convolutional network with elastic topology. *Pattern Recognition*, 151:110364, 2024.

[Xu *et al.*, 2023] Cai Xu, Wei Zhao, Jinglong Zhao, Ziyu Guan, Yaming Yang, Long Chen, and Xiangyu Song. Progressive deep multi-view comprehensive representation learning. In *AAAI*, volume 37, pages 10557–10565, 2023.

[Xu *et al.*, 2024] Cai Xu, Jiajun Si, Ziyu Guan, Wei Zhao, Yue Wu, and Xiyue Gao. Reliable conflictive multi-view learning. In *AAAI*, volume 38, pages 16129–16137, 2024.

[Xu *et al.*, 2025a] Shilin Xu, Yuan Sun, Xingfeng Li, Siyuan Duan, Zhenwen Ren, Zheng Liu, and Dezhong Peng. Noisy label calibration for multi-view classification. In Toby Walsh, Julie Shah, and Zico Kolter, editors, *AAAI*, pages 21797–21805, 2025.

[Xu *et al.*, 2025b] Yanchen Xu, Siqi Huang, Hongyuan Zhang, and Xuelong Li. Why does dropping edges usually outperform adding edges in graph contrastive learning? In *AAAI*, volume 39, pages 21824–21832, 2025.

[Yu *et al.*, 2022] Pengyang Yu, Chaofan Fu, Yanwei Yu, Chao Huang, Zhongying Zhao, and Junyu Dong. Multiplex heterogeneous graph convolutional network. In *SIGKDD*, pages 2377–2387, 2022.

[Zhang *et al.*, 2025] Chao Zhang, Deng Xu, Chunlin Chen, and Huaxiong Li. Semi-supervised multi-view clustering with active constraints. In *SIGKDD*, pages 1903–1912, 2025.

[Zhu *et al.*, 2020] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Deep graph contrastive representation learning. *arXiv preprint arXiv:2006.04131*, 2020.

[Zhuang *et al.*, 2024] Shuman Zhuang, Sujia Huang, Wei Huang, Yuhong Chen, Zhihao Wu, and Ximeng Liu. Enhancing multi-view graph neural network with cross-view confluent message passing. In *ACM MM*, pages 10065–10074, 2024.

[Zhuang *et al.*, 2025] Shuman Zhuang, Zhihao Wu, Zhaoliang Chen, Hong-Ning Dai, and Ximeng Liu. Refine then classify: Robust graph nn with reliable neighborhood contrastive refinement. In *AAAI*, volume 39, pages 13473–13482, 2025.