Going Beyond Consistency: Target-oriented Multi-view Graph Neural Network

Sujia Huang[1], **Lele Fu**[2], **Shuman Zhuang**[3], **Yide Qiu**[1], **Bo Huang**[1], **Zhen Cui**[4*] and **Tong Zhang**[1*]

[1]School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

[2]School of Systems Science and Engineering, Sun Yat-Sen University, Guangzhou, China

[3]College of Computer and Data Science, Fuzhou University, Fuzhou, China

[4]School of Artificial Intelligence, Beijing Normal University, Beijing, China

hsujia2021@163.com, fulle@mail2.sysu.edu.cn, shumanzhuang@163.com, 121106010824@njust.edu.cn, huangbo@njust.edu.cn, zhen.cui@bnu.edu.cn, tong.zhang@njust.edu.cn

Abstract

Multi-view learning has emerged as a pivotal research area driven by the growing heterogeneity of real-world data, and graph neural network-based models, modeling multi-view data as multi-view graphs, have achieved remarkable performance by revealing its deep semantics. However, by assuming cross-view consistency, most approaches collect not only task-relevant (determinative) semantics but also symbiotic yet task-irrelevant (incidental) factors are collected to obscure model inference. Furthermore, these approaches often lack rigorous theoretical analysis that bridges training data to test data. To address these issues, we propose Target-oriented Graph Neural Network (TGNN), a novel framework that goes beyond traditional consistency by prioritizing task-relevant information, ensuring alignment with the target. Specifically, TGNN employs a class-level dual-objective loss to minimize the classification similarity between determinative and incidental factors, accentuating the former while suppressing the latter during model inference. Meanwhile, to ensure consistency between the learned semantics and predictions in representation learning, we introduce a penalty term that aims to amplify the divergence between these two types of factors. Furthermore, we derive an upper bound on the loss discrepancy between training and test data, providing formal guarantees for generalization to test domains. Extensive experiments conducted on three types of multi-view datasets validate the superiority of TGNN.

1 Introduction

In the real world, data is often collected through diverse methods, enabling each entity to be represented by different feature sets. For instance, in medical diagnosis, imaging scans, pathological reports, and genomic information collectively provide a comprehensive description of a patient. This type of data, which integrates multiple sources, is referred to as multi-view data, offering richer semantic information compared to single-view data and enhancing model decision. Multi-view learning is specifically designed for addressing such data, which extracts essential representations for downstream tasks through fully excavating the intrinsic information embedded across multiple views. Over the past few decades, a wide range of multi-view paradigms have been developed across various domains, including image processing [Tan *et al.*, 2023; Yu *et al.*, 2024] and biomedicine [Wen and Li, 2025; Huang *et al.*, 2023]. Among them, graph-based multi-view algorithms have garnered significant attention. These approaches employ graphs to model inherent relationships between instances, providing a structured framework for complex multi-view data and achieving competitive performance [Zhao *et al.*, 2025].

As a powerful tool for handling graph-structured data, Graph Neural Network (GNN) have driven breakthroughs across numerous domains [Zhou *et al.*, 2024; Guo *et al.*, 2024]. GNN propagates and aggregates features from neighboring nodes along given graph structures to learn rich contextual information. This capability enables GNN suitable for multi-view learning, where diverse interactions between entities in each view can be naturally and intuitively modeled as graphs. Currently, many multi-view GNNs have been proposed, leveraging multi-view data to extract informative representations through various relational structures [Yu *et al.*, 2025; Xie *et al.*, 2020a]. Most multi-view learning models aim to explore the consistency and specificity across views, with GNN-based algorithms building upon this principle by employing diverse strategies to effectively mine these essential characteristics.

Prevailing models in multi-view GNNs can be broadly divided into two types: pre-fusion models and post-fusion models. The former focus on integrating multi-view graphs and features into a unified topology and feature matrix, respectively, before feeding them into a GNN to learn representations [Wang *et al.*, 2024b]. Post-fusion models perform message passing on each view-specific graph and then merge the results [Xia *et al.*, 2023; Qi *et al.*, 2024]. However, they primarily eliminate overtly negative information, while im-

*Corresponding authors.

Figure 1: Captured consistent semantics across views are not always relevant to the labels.

plicit and insignificant factors, such as illumination or environment, persist as part of the learned consistent features. For instance, in scenarios identifying horses from images captured in different settings, features like grass may frequently co-occur with horses across views. In such cases, cross-view alignment may inadvertently cause the model to misidentify grass as feature of the horse, as shown in Figure 1. In fact, features like 'grass', which are consistent across views but spuriously related with labels, are considered incidental factors and should be discarded as noise. An effective model should focus on determinative factors like 'horse', capturing the intrinsic relation between features and labels. Although some studies [Lin *et al.*, 2022; Liang *et al.*, 2024] use contrastive loss and mutual information to obtain task-relevant semantics as much as possible, they remain: 1) constrained by the framework of consistent representation learning; 2) neglecting the influence of incidental predictions on determinative predictions; 3) overlooking the exploration of generalization from training data to test data.

In this paper, we propose a new model, termed Target-oriented Graph Neural Network (TGNN), which goes beyond the consistency across views and focuses on task-oriented, pure representations. In specific, we employs two types of feature extractors on each view to separately learn determinative and incidental features. The determinative factors capture the deep causal relationships between data and labels, while the incidental factors represent irrelevant and external relations introduced by data bias. These transformed features are discriminated via a classifier using a class-level dual objective loss. The loss aims to achieving: minimizing the classification similarity of the two types of features and maximizing the alignment between determinative representations and the labels. Meanwhile, to maintain alignment between semantics and predictions during representation learning, we introduce a penalty term to enhance the semantic separability between the two types of features within each view, ensuring that distinct predictions are not derived from same semantics. Moreover, grounded in theoretical analyses, we establish a tight generalization bound between the losses on training and test data, ensuring the model's capability to generalize effectively to unknown data. By uniformly optimizing the overall framework, a discriminative representation is learned, enabling optimal predictions on test data. Our contributions are concluded as

- Propose TGNN, a novel model that eliminates incidental factors embedded in consistent features across views and focuses on learning determinative factors that capture the intrinsic dependency between data and labels.
- Design a class-level dual objective loss to maximize the classification discrepancy between different features to highlight determinative representations. Moreover, we derive an upper bound on losses between training and test data, enhancing generalization on unknown data.
- Construct extensive experiments on three types of data to validate the superiority of the proposed model. Furthermore, TGNN works robustly when attack presents.

2 Related Work

2.1 Multi-view Learning

Multi-view learning aims to uncover intrinsic representations for decision-making by integrating multiple views of data. Existing paradigms in this domain can be broadly categorized into two technical approaches. On one hand, traditional multi-view algorithms rely on mechanisms such as subspace projection and matrix decomposition to obtain latent consistent representations. For example, [Chang *et al.*, 2024] employs a partitioned strategy to divide the dictionary matrix within the subspace framework, effectively addressing large-scale data. [Wen *et al.*, 2024] introduces a matrix factorization approach with adaptively weighted features to identify significant and discriminative information. On the other hand, due to the powerful expressive capacity of neural networks, deep learning-based multi-view methods have garnered widespread attention. For instance, [Huang *et al.*, 2022] bridges deep learning and matrix decomposition within a unified framework, achieving promising results. Meanwhile, [Wang *et al.*, 2023] applies contrastive learning to better capture the cluster structure, as reflected by the consensus affinity matrix. Although these methods perform well, they typically process affinity matrices that encode node relationships but overlook high-order structural information embedded in attributed graphs. To address the limitation, various multi-view graph neural networks have been developed to capture high-order interactions, thereby enhancing the discrimination of node representations.

2.2 Graph Neural Network

Considering the inherent relationships between entities in real-world scenarios, graph neural network has been used in multi-view learning field, facilitating the modeling of complex instance relationships in multi-view data through message passing mechanisms to achieve impressive performance across various tasks [Yu *et al.*, 2025; Ren *et al.*, 2024]. For instance, [Wang *et al.*, 2024b] proposes a heterogeneous GNN

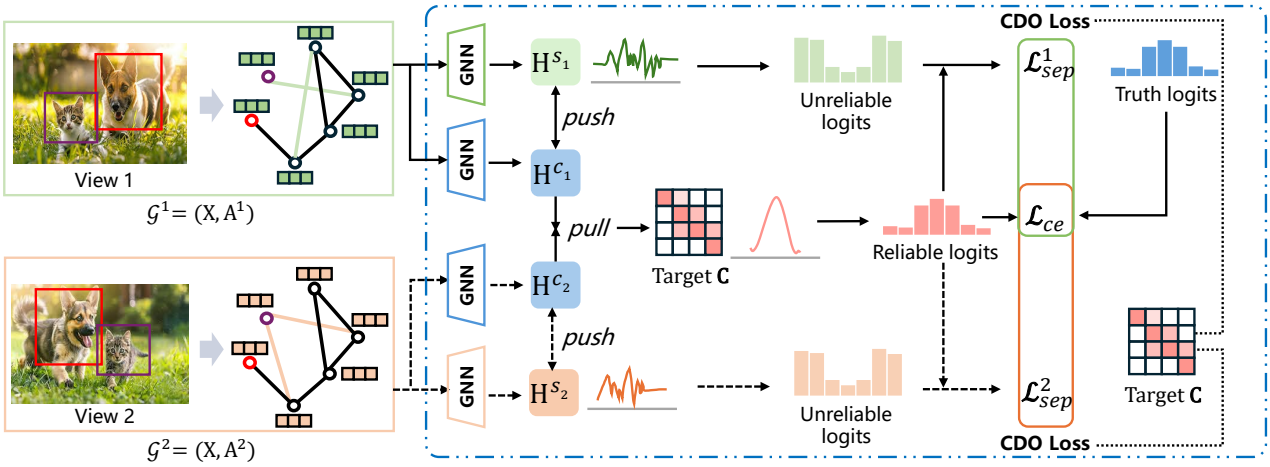


Figure 2: Overview of the proposed framework. The proposed framework begins with feature disentanglement for each view, ensuring the effective capture of target-oriented semantics. Subsequently, the separation loss is employed to amplify the classification discrepancy between the determinative representation $\mathbf{C}$ and the incidental representation $\mathbf{H}^{s_v}$. By jointly optimizing $\mathcal{L}^v_{sep}$ and $\mathcal{L}_{ce}$, the CDO loss effectively enhances the task-relevant features embedded in $\mathbf{C}$.

that treats the multi-view dataset as a unified graph with multiple relation types, where nodes are shared across views and each edge belongs to one view. [Yu *et al.*, 2025] uses a loss based on the Cauchy-Schwarz divergence to improve the feature extraction in a graph convolutional encoder. However, existing models primarily emphasize consistency and specificity across views, overlooking the fundamental objective of establishing meaningful relationships between representations and labels. In contrast, our approach transcends the consistency learning by disentangling multi-view data into components that are related and unrelated to classification tasks. Task-relevant features inherently tied to downstream objectives are defined as determinative features, while those with only spurious associations are termed as incidental features. This distinction enables our model to prioritize the extraction of task-relevant semantics for robust and discriminative representations.

3 Proposed Model

Notations. Given a set of multi-view graphs $\mathcal{G} = \{G^v\}_{v=1}^V$ with multiple correlations between nodes. Here, V represents the number of views, and each graph $G^v = (\mathcal{V}, \mathcal{E}^v)$ consists of a shared set of nodes $\mathcal{V}$ and a view-specific set of edges $\mathcal{E}^v$. The node feature matrix is represented as $\mathbf{X} \in \mathbb{R}^{n \times d}$, where each node is associated with a d-dimensional feature vector. For the graph G^v, its adjacency matrix is denoted by $\mathbf{A}^v \in \mathbb{R}^{n \times n}$, where $A^v_{ij} = 1$ if an edge $e_{ij} \in \mathcal{E}^v$ exists, and $A^v_{ij} = 0$ otherwise. To prevent gradient explosion during training, a renormalization trick is applied to $\mathbf{A}^v$, yielding $\widetilde{\mathbf{A}}^v = \mathbf{D}^{v^{-\frac{1}{2}}}(\mathbf{A}^v + \mathbf{I})\mathbf{D}^{v^{-\frac{1}{2}}}$, where $D^v_{ii} = \sum_j (\mathbf{A}^v + \mathbf{I})_{ij}$. $\mathbf{Y} \in \mathbb{R}^{n \times c}$ is the label matrix with the number of class c, and each row is a one-hot vector.

Preliminary. We partition n samples into n_l labeled instances and n_u unlabeled instances. Let $\mathcal{P} := \mathcal{D}_{tr}(\mathbf{X}, \mathbf{Y})$ be the joint distribution defined over the input space $\mathcal{X} \times \mathcal{Y}$ in the training set. Similarly, the test dataset, which is inaccessible during training, is denoted as $\mathcal{J} := \mathcal{D}_{te}(\mathbf{X}, \mathbf{Y})$. Equivalently, we represent the marginal distribution over variable $\mathbf{X}$ on the training and test data as $\mathcal{D}_{tr}(\mathbf{X})$ and $\mathcal{D}_{te}(\mathbf{X})$, respectively. The goal of multi-view learning is to minimize the prediction error of a model $f^{\mathbf{\Theta}}$ parameterized by $\mathbf{\Theta}$ on the test data, leveraging n_l labeled instances and the discriminative representation $\mathbf{C}$ across V views. We represent the conditional distribution over $\mathbf{C}$ as $\mathcal{D}_{te}(\mathbf{C}|\{\mathbf{A}^v, \mathbf{X}\}_{v=1}^V)$, where $\mathbf{C}$ is computed based on the multi-view graph structures $\{\mathbf{A}^v\}_{v=1}^V$ and the input feature $\mathbf{X}$. Then, mathematically, the expected risk of $f^{\mathbf{\Theta}}$ on the test data can be expressed as:

$$\min_{\mathbf{\Theta}} \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{J}} \left[\mathbb{E}_{\mathbf{c}\sim\mathcal{D}_{te}(\mathbf{C}|\{\mathbf{A}^v,\mathbf{X}=\mathbf{x}\}_{v=1}^V)} \ell(f^{\mathbf{\Theta}}(\mathbf{c}), \mathbf{y}) \right], \quad (1)$$

where $\ell(\cdot,\cdot)$ is the loss function that quantifies the discrepancy between the model prediction $f^{\mathbf{\Theta}}(\mathbf{c})$ and the ground truth label $\mathbf{y}$. The key step of the above procedure is to learn the representation $\mathbf{C}$, which captures critical information for classification tasks, enabling the predictor to produce reliable predictions in the context of complex multi-view data.

3.1 Disentanglement of View-wise Semantics

Multi-view graphs describe instance relationships from multiple aspects, providing richer contexts for learning discriminative representations. However, these diverse view patterns often introduce spurious factors that are incidentally correlated with labels, such as light color, illumination, or other environmental attributes. Most multi-view learning strategies fall into two main technical categories: capturing invariant representations across views [Zhang *et al.*, 2025; Yang *et al.*, 2025] and leveraging adaptive weight learning to aggregate semantics from multiple views [Kant *et al.*, 2024; Liao *et al.*, 2024]. Despite their effectiveness, both approaches risk integrating consistent yet task-irrelevant features into the final representation, which may hinder model performance. To ensure the use of features that are beneficial for the current task while mitigating the impact of task-

irrelevant semantics, view disentanglement is a high-level intuition. In other words, it is necessary to remove incidental factors embedded in the consistent information and focus on extracting determinative features in a target-oriented manner.

Semantic Disentangler. Following the line of thought, we first design a semantic disentangler that consists of two types of feature extractors for each view: $\{\mathrm{SE}^v(\cdot)\}_{v=1}^V$ which are extractors specifically designed for each view, and a unified extractor $\mathrm{CE}(\cdot)$ shared across views. Both $\mathrm{SE}^v(\cdot)$ and $\mathrm{CE}(\cdot)$ are implemented using a GNN-based backbone, formalized as follows:

$$\begin{aligned}\mathbf{H}^{s_v} &= \mathrm{SE}^v\Big(\widetilde{\mathbf{A}}^v, \mathbf{X}; \mathbf{\Phi}^{s_v}\Big) = \mathrm{gCov}\Big(\widetilde{\mathbf{A}}^v\mathbf{X}, \mathbf{\Phi}^{s_v}\Big);\\ \mathbf{H}^{c_v} &= \mathrm{CE}\Big(\widetilde{\mathbf{A}}^v, \mathbf{X}; \mathbf{\Phi}^{c}\Big) = \mathrm{gCov}\Big(\widetilde{\mathbf{A}}^v\mathbf{X}, \mathbf{\Phi}^{c}\Big),\end{aligned} \tag{2}$$

in which $\mathbf{H}^{s_v} \in \mathbb{R}^{n\times m}$ and $\mathbf{H}^{c_v} \in \mathbb{R}^{n\times m}$ denote the learned incidental and determinative features, respectively. There features are disentangled using the proposed loss function in Eq. (8) and further refined by the penalty term described in Section 3.4. "gCov" denotes a standard graph convolution network with L layers followed by ReLU activation. The parameter set $\mathbf{\Phi}^{s_v} = \{[\mathbf{\Phi}^{s_v}]^{(l)}\}_{l=1}^L$ is specific to the v-th network, and $\mathbf{\Phi}^{c} = \{[\mathbf{\Phi}^{c}]^{(l)}\}_{l=1}^L$ is the set of shared learnable parameters across views. Specifically, $\mathrm{SE}^v(\cdot)$ captures incidental features that are spuriously correlated with labels, such as biases introduced by data or superficial statistical relationships. Conversely, the shared extractor $\mathrm{CE}(\cdot)$ focuses on capturing the deep causal relationships between features and labels.

Cross-view Feature Fusion. After obtaining $\{\mathbf{H}^{c_v}\}_{v=1}^V$, we learn the complete representation $\mathbf{C} \in \mathbb{R}^{n\times m}$ by

$$\mathbf{C} = [\mathbf{H}^{c_1}||\cdots||\mathbf{H}^{c_V}]\mathbf{\Omega} + \mathbf{b}, \tag{3}$$

where $||$ is the concatenation operation, $\mathbf{\Omega} \in \mathbf{R}^{Vm\times m}$ is a learnable weight matrix, and $\mathbf{b} \in \mathbf{R}^m$ is a learnable bias term. This fusion mechanism aggregates determinative factors from all views to form a comprehensive representation $\mathbf{C}$ that encapsulates the meaningful semantics across views. To guarantee the integration of view-wise variable into $\mathbf{C}$, we devise a view-shared constraint as below

$$\mathcal{L}_{sha}^v(\mathbf{\Phi}^c, \mathbf{\Omega}) = \frac{\mathcal{L}_{ce}\big(\mathbf{P}, \mathbf{I}\big) + \mathcal{L}_{ce}\big(\mathbf{P}^\top, \mathbf{I}\big)}{2}, \tag{4}$$

where

$$P_{ij} = \frac{\exp(\Gamma_{ij})}{\sum_{k=1}^n \exp(\Gamma_{ik})}, \mathbf{\Gamma} = \exp(\tau)\cdot\frac{\mathbf{C}}{\|\mathbf{C}\|_F}\Big(\frac{\mathbf{H}^{c_v}}{\|\mathbf{H}^{c_v}\|_F}\Big)^\top, \tag{5}$$

where τ is a temperature parameter, $\mathbf{I} \in \mathbb{R}^{n\times n}$ is an identity matrix and $\mathcal{L}_{ce}$ denotes the cross-entropy loss measuring the difference between $\mathbf{P}$ and $\mathbf{I}$, enabling the alignment of $\mathbf{C}$ with $\{\mathbf{H}^{c_v}\}_{v=1}^V$.

3.2 Class-level Dual Objective Loss

Obviously, a predictor with strong classification ability should focus on determinative features while mitigating the influence of incidental factors during inference. In other words, for a classifier $f^{\mathbf{\Theta}}$, the classification results derived from the determinative features $\mathbf{C}$ should differ significantly from those derived from the incidental features $\mathbf{H}^{s_v}$, while the predictions based on $\mathbf{C}$ should align closely with the ground truth labels $\mathbf{Y}$. To this end, we propose the Class-level Dual Objective (CDO) loss $\mathcal{L}_{cdo}$, which encompasses two key objectives: minimizing the classification similarity between determinative and incidental features, and maximizing the alignment between determinative representations and labels.

For simplicity, we denote the conditional distribution over $\mathbf{C}$ on the test set $\mathcal{D}_{te}(\mathbf{C}|\{\mathbf{A}^v, \mathbf{X} = \mathbf{x}\}_{v=1}^V)$ as $\mathcal{D}_{te}(\mathbf{C} \mid \mathcal{G}(\mathbf{x}))$, and $\mathcal{D}_{te}(\mathbf{H}^{s_v}|\mathbf{A}^v, \mathbf{X} = \mathbf{x})$ over $\mathbf{H}^{s_v}$ as $\mathcal{D}_{te}(\mathbf{C} \mid G^v(\mathbf{x}))$. Moreover, we further denote $\mathcal{D}_{te}^{\mathbf{\Phi}^c,\mathbf{\Omega}}(\mathbf{C} \mid \mathcal{G}(\mathbf{x}))$ and $\mathcal{D}_{te}^{\mathbf{\Phi}^{s_v}}(\mathbf{H}^{s_v} \mid G^v(\mathbf{x}))$ as the estimated distributions parameterized by $\mathbf{\Phi}^c, \mathbf{\Omega}$ and $\mathbf{\Phi}^{s_v}$, respectively. Formally, the first goal is defined as minimizing the following prediction separation loss,

$$\begin{aligned}&\mathcal{L}_{sep}^v(\mathbf{\Phi}^c, \mathbf{\Omega}, \mathbf{\Phi}^{s_v}, \mathbf{\Theta}) = \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{J}}\Big[\mathbb{E}_{\mathbf{c}\sim\mathcal{D}_{te}^{\mathbf{\Phi}^c,\mathbf{\Omega}}(\mathbf{C}|\mathcal{G}(\mathbf{x}))}\\ &\mathbb{E}_{\mathbf{h}\sim\mathcal{D}_{te}^{\mathbf{\Phi}^{s_v}}(\mathbf{H}^{s_v}|G^v(\mathbf{x}))}\mathrm{I}\left[f^{\mathbf{\Theta}}(\mathbf{c}) = f^{\mathbf{\Theta}}(\mathbf{h})\right]\Big].\end{aligned} \tag{6}$$

Here, $\mathrm{I}(\cdot)$ is an indicator function[1]. The risk decreases when $\mathbf{C}$ and $\mathbf{H}^{s_v}$ achieve different classified results. Meanwhile, since the learned representation $\mathbf{C}$ should be closely related to the labels, in semi-supervised classification task, we use the cross-entropy loss to minimize the difference between the predictive distribution obtained from $\mathbf{C}$ and $\mathbf{Y}$

$$\begin{aligned}&\mathcal{L}_{ce}(\mathbf{\Phi}^c, \mathbf{\Omega}, \mathbf{\Theta}) =\\ &-\mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\mathcal{J}}\mathbb{E}_{\mathbf{c}\sim\mathcal{D}_{te}^{\mathbf{\Phi}^c,\mathbf{\Omega}}(\mathbf{C}|\mathcal{G}(\mathbf{x}))}\Big[\mathbf{y}\ln\Big(\sigma\big(f^{\mathbf{\Theta}}(\mathbf{c})\big)\Big)\Big],\end{aligned} \tag{7}$$

where $\sigma(\cdot)$ is a non-linear Softmax activation function. Thus, the total separation procedure for each view is accomplished by minimizing the CDO loss combining (6) and (7),

$$\mathcal{L}_{cdo}^v(\mathbf{\Phi}^c, \mathbf{\Omega}, \mathbf{\Phi}^{s_v}, \mathbf{\Theta}) = \mathcal{L}_{ce} + \mathcal{L}_{sep}^v. \tag{8}$$

The flowchart of the above procedure is illustrated in Figure 2.

3.3 Generalization Analysis of CDO Loss

In real-world scenarios, model optimization relies solely on the observable training data, while test samples in $\mathcal{D}_{te}$ are unavailable. Consequently, the CDO loss, defined on $\mathcal{D}_{te}$, cannot be directly employed to evaluate the learned determinative representation. To address this limitation, we conduct a generalization analysis to quantify the gap between the empirical risk of the model on the training data and the expected risk on the test data. We separately discuss the losses $\mathcal{L}_{sep}^v$ and $\mathcal{L}_{ce}$ constituting $\mathcal{L}_{cdo}^v$, and define their corresponding empirical risks as $\widehat{\mathcal{L}}_{sep}^v$ and $\widehat{\mathcal{L}}_{ce}$, respectively. Building on works [Shalev-Shwartz and Ben-David, 2014; Yang *et al.*, 2023; Wang *et al.*, 2024a], we derive an upper bound for the gap between these two losses, which is expressed as follows:

Theorem 1. *Given the learnable parameters $\mathbf{\Phi}^c$, $\mathbf{\Phi}^{s_v}$ and $\mathbf{\Omega}$, for any $\mathbf{\Theta}$: $\mathbf{R}^m \to \mathcal{Y}$, prior distributions π_c*

[1]The function can be instantiated using various common loss functions, such as cross-entropy loss or mean squared error.

and π_{s_v} *that make* $\mathbb{E}_{\mathcal{J}} \mathrm{KL}\left(\mathcal{D}_{tr}^{\mathbf{\Phi}^c,\mathbf{\Omega}}(\mathbf{C} \mid \mathcal{G}(\mathbf{x}))\|\pi_c\right)$ *and* $\mathbb{E}_{\mathcal{J}} \mathrm{KL}\left(\mathcal{D}_{tr}^{\mathbf{\Phi}^{s_v}}(\mathbf{H}^{s_v} \mid G^v(\mathbf{x}))\|\pi_{s_v}\right)$ *both lower than a positive constant* B*, then with probability at least* $1-\epsilon$ *over the training data* $\mathcal{P} := \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^{n_l}$*, the following bound holds:*

$$|\mathcal{L}_{sep}^v(\mathbf{\Phi}^c, \mathbf{\Omega}, \mathbf{\Phi}^{s_v}, \mathbf{\Theta}) - \widehat{\mathcal{L}}_{sep}^v(\mathbf{\Phi}^c, \mathbf{\Omega}, \mathbf{\Phi}^{s_v}, \mathbf{\Theta})| \leq \mathbb{E}_{\mathcal{P}} \mathrm{KL}\left(\mathcal{D}_{tr}^{\mathbf{\Phi}^c,\mathbf{\Omega}}(\mathbf{C} \mid \mathcal{G}(\mathbf{x}))\|\pi_c\right) + \mathbb{E}_{\mathcal{P}} \mathrm{KL}\left(\mathcal{D}_{tr}^{\mathbf{\Phi}^{s_v}}(\mathbf{H}^{s_v} \mid G^v(\mathbf{x}))\|\pi_{s_v}\right) + \frac{\ln \frac{n_l}{\epsilon}}{4(n_l-1)} + B. \tag{9}$$

$$|\mathcal{L}_{ce}(\mathbf{\Phi}^c, \mathbf{\Omega}, \mathbf{\Theta}) - \widehat{\mathcal{L}}_{ce}(\mathbf{\Phi}^c, \mathbf{\Omega}, \mathbf{\Theta})| \leq \mathbb{E}_{\mathcal{P}} \mathrm{KL}\left(\mathcal{D}_{tr}^{\mathbf{\Phi}^c,\mathbf{\Omega}}(\mathbf{C} \mid \mathcal{G}(\mathbf{x}))\|\pi_c\right) + \frac{\ln \frac{n_l}{\epsilon}}{4(n_l-1)} + 2B. \tag{10}$$

Proof. The detailed proof can refer to Appendix **A**. □

The above theorem demonstrates that by decreasing the KL divergence, the gap between the empirical and expected risks becomes smaller. In other words, we can approximate the CDO loss on the test data by evaluating the loss on the training data, ensuring that the learned features are well-suited for the test domain.

3.4 The Overall Objective

To ensure that the learned representations are feasible, and to prevent the occurrence of different predictions for the same semantics during the separation procedure, we introduce an additional penalty term that enforces semantic separability between the learned representation $\mathbf{C}$ and the labels $\mathbf{Y}$. Specifically, for any $\mathcal{D} \in \{\mathcal{D}_{tr}, \mathcal{D}_{te}\}$, if $\mathbf{c} \sim \mathcal{D}(\mathbf{C}|\mathbf{Y} = \mathbf{y})$ and $\mathbf{h} \sim \mathcal{D}(\mathbf{H}|\mathbf{Y} \neq \mathbf{y})$, the inequality $\|\mathbf{c} - \mathbf{h}\|_2 > \theta$ should hold. This term ensures distinguishable semantics between $\mathbf{c}$ and $\mathbf{h}$, thus preventing confusion and improving interpretability during the process of minimizing the CDO loss in representation learning. Thus, combining Eqs. (4), (6), (7), (9), (10), we obtain the following overall objective

$$\min_{\substack{\mathbf{\Omega}, \mathbf{\Phi}^c, \\ \{\mathbf{\Phi}^{s_v}\}_{v=1}^V, \mathbf{\Theta}}} \mathcal{L}_{ce} + \alpha \sum_{v=1}^{V} \left(\mathcal{L}_{sha}^v + \mathcal{L}_{sep}^v + \mathcal{L}_{KL}^v\right), \quad \text{s.t.} \|\mathbf{c} - \mathbf{h}\|_2 > \theta \text{ for any view.} \tag{11}$$

Here, the KL divergence term $\mathcal{L}_{KL}^v = \mathbb{E}_{\mathcal{P}} \mathrm{KL}\left(\mathcal{D}_{tr}^{\mathbf{\Phi}^c,\mathbf{\Omega}}(\mathbf{C} \mid \mathcal{G}(\mathbf{x}))\|\pi_c\right) + \mathbb{E}_{\mathcal{P}} \mathrm{KL}\left(\mathcal{D}_{tr}^{\mathbf{\Phi}^{s_v}}(\mathbf{H}^{s_v} \mid G^v(\mathbf{x}))\|\pi_{s_v}\right)$ and α is a trade-off hyperparameter. The algorithm of TGNN is presented in Appendix **B**. Our code can refer to appendix.

Complexity Analysis. Review that the number of nodes is n, the feature dimension is d and the number of edges of view v is $|\mathcal{E}^v|$. Considering the forward propagation for $\mathbf{C}$ and $\mathbf{H}^{s_v}$, the graph convolution operation for all views costs $\sum_{v=1}^{V} \mathcal{O}(|\mathcal{E}^v|m)$. Moreover, for the fusion step of $\mathbf{C}$ (as shown in Eq. (3)), the complexity is $\mathcal{O}(nVm^2)$. Therefore, the overall time complexity of the model is $\mathcal{O}(nm^2 + \sum_{v=1}^{V} |\mathcal{E}^v|m)$, where $V \ll m$. Code and Appendix refer to https://github.com/huangsuj/TGNN.git.

4 Experiments

4.1 Experimental Setups

Datasets

To evaluate the effectiveness of the proposed TGNN, we conduct comprehensive experiments on three types of multi-view datasets. These include three multi-relational datasets (ACM, DBLP, YELP), three multi-attribute datasets (Animals, HW, MNIST), and three multi-modal datasets (BDGP, esp-game, Flickr). A detailed description of these datasets is provided in Appendix **C.1**.

Compared Methods

To demonstrate the superior performance of TGNN, we compare it against eight state-of-the-art methods designed for multi-relational graphs, including 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], AMOGCN [Chen *et al.*, 2024], and ECMGD [Lu *et al.*, 2024a]. Additionally, we evaluate TGNN against eight methods developed for multi-attribute and multi-modal data, such as Co-GCN [Li *et al.*, 2020], HLR-M^{2}VS [Xie *et al.*, 2020b], ERL-MVSC [Huang *et al.*, 2021], DSRL [Wang *et al.*, 2022], IMvGCN [Wu *et al.*, 2023], PDMF [Jiang *et al.*, 2023], GEGCN [Lu *et al.*, 2024b], and ECMGD [Lu *et al.*, 2024a]. A comprehensive introduction to these competitors is provided in Appendix **C.2**.

Implemented Details

To ensure a fair comparison, we record the mean and standard deviation of all models after performing 5 runs on all datasets. The parameters of TGNN are configured as below: the training epoch is 300, learning rate is 0.001, the hidden dimension is 512, the number of layers is 2, θ and α range in $\{0.1, 0.5, 0.7, 1, 1.3\}$ and $\{0.001, 0.005, 0.01, 0.05, 0.1, 0.5\}$, respectively. The Adam optimizer is adopted with a weight decay of $5e^{-6}$ for the DBLP, Flickr, and HW datasets, and $5e^{-4}$ for the remaining datasets. For node classification tasks on multi-relational graphs, 10% of the samples are used for validation, with the training set size varying across $\{20\%, 40\%\}$, and the remaining data is used for testing. For multi-attribute and multi-modal datasets, we split the data into training/testing/validation sets with a ratio of $10\%/10\%/80\%$.

4.2 Performance Analysis

We evaluate the performance of TGNN through semi-supervised node classification on multiple types of data. For a comprehensive analysis, we use Macro-F1 and Micro-F1 as evaluation metrics for multi-relational graphs, and ACC and Macro-F1 for multi-attribute and multi-modal graphs.

Classification on multi-relational graphs. As shown in Table 1, the results demonstrate that the proposed model consistently achieves superior performance across the multi-relational graphs. Notably, for most models, performance improves as the number of training samples increases, which aligns with expected trends in semi-supervised learning.
Classification on multi-attribute and multi-modal graphs. For these datasets, we construct graphs based on the original features using the k-nearest neighbor algorithm, with k

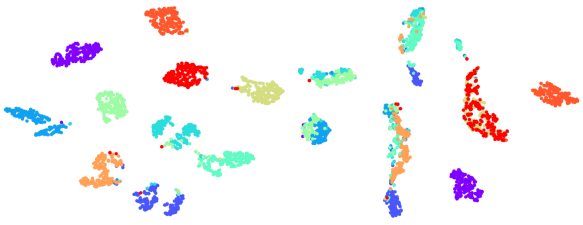
Datasets	Metrics	Training	HAN	DMGI	IGNN	MRGCN	SSDCM	MHGCN	AMOGCN	ECMGD	Ours
ACM	Macro-F1	20%	87.95 (0.42)	66.73 (1.84)	82.90 (0.03)	87.58 (0.19)	84.34 (3.46)	69.04 (1.76)	90.14 (0.48)	92.24 (0.40)	**93.22 (0.09)**
		40%	91.28 (0.33)	71.17 (2.24)	85.01 (1.01)	88.44 (0.20)	85.05 (3.66)	70.74 (1.29)	91.03 (0.50)	92.07 (0.13)	**94.09 (0.09)**
	Micro-F1	20%	87.98 (0.38)	70.43 (1.13)	82.70 (0.03)	87.45 (0.24)	85.23 (2.86)	69.07 (1.40)	90.01 (0.50)	91.24 (0.41)	**93.18 (0.09)**
		40%	91.20 (0.33)	74.06 (1.50)	84.91 (1.01)	88.32 (0.20)	85.77 (3.09)	70.78 (1.22)	90.97 (0.51)	92.07 (0.14)	**94.07 (0.09)**
DBLP	Macro-F1	20%	89.30 (0.21)	75.35 (1.28)	86.81 (0.01)	89.49 (1.61)	55.72 (2.71)	**92.52 (0.29)**	92.27 (0.42)	91.85 (0.55)	91.51 (0.20)
		40%	90.02 (0.34)	81.47 (0.76)	88.40 (0.01)	91.15 (0.08)	79.88 (2.13)	92.00 (0.22)	92.24 (0.29)	92.14 (0.39)	**92.73 (0.15)**
	Micro-F1	20%	90.44 (0.20)	81.21 (0.72)	87.50 (0.01)	90.47 (1.23)	62.82 (3.90)	**92.87 (0.23)**	92.80 (0.37)	91.85 (0.44)	92.06 (0.20)
		40%	90.46 (0.28)	83.77 (0.51)	88.41 (0.01)	91.71 (0.12)	80.64 (2.11)	92.74 (0.25)	92.70 (0.26)	92.14 (0.36)	**93.12 (0.14)**
YELP	Macro-F1	20%	55.39 (4.52)	52.72 (2.27)	71.40 (0.01)	54.35 (0.39)	55.86 (2.99)	60.85 (1.02)	70.77 (2.32)	91.89 (0.23)	**93.94 (0.24)**
		40%	55.59 (4.80)	55.54 (3.24)	73.33 (0.01)	54.74 (0.91)	69.54 (2.04)	60.07 (1.01)	70.97 (1.81)	92.06 (0.36)	**94.06 (0.14)**
	Micro-F1	20%	68.00 (5.03)	69.52 (0.68)	75.01 (0.01)	73.70 (0.46)	68.87 (5.54)	73.28 (0.24)	77.43 (0.36)	91.89 (0.24)	**93.42 (0.26)**
		40%	69.69 (6.25)	72.60 (0.25)	75.91 (0.01)	73.53 (0.50)	75.77 (2.10)	73.01 (0.49)	78.81 (0.19)	92.06 (0.38)	**93.46 (0.16)**

Table 1: Macro-F1 and Micro-F1 (mean and std%) of nine models with various percentages of training samples on multi-relational graphs, in which the optimal results are highlighted in bold.

Datasets	Metrics	HLR-H^{2}VS	Co-GCN	ERL-MVSC	DSRL	IMvGCN	PDMF	GEGCN	ECMGD	Ours
Animals	ACC	72.74 (0.53)	79.51 (1.48)	69.89 (0.42)	80.03 (0.51)	82.53 (0.12)	72.82 (2.18)	79.43 (0.28)	80.52 (0.52)	**85.26 (0.08)**
	Macro-F1	68.12 (0.93)	73.02 (2.11)	65.72 (0.31)	74.26 (1.03)	76.14 (0.27)	67.81 (1.85)	69.79 (0.38)	73.94 (0.65)	**79.35 (0.09)**
HW	ACC	85.33 (0.00)	91.58 (2.65)	85.45 (2.06)	77.92 (0.86)	90.66 (4.13)	89.69 (1.78)	96.01 (0.23)	96.81 (0.48)	**97.24 (0.17)**
	Macro-F1	86.87 (0.00)	91.48 (2.78)	85.26 (1.30)	78.68 (0.73)	90.47 (4.63)	89.73 (1.81)	96.03 (0.22)	96.82 (0.47)	**97.19 (0.18)**
MNIST	ACC	84.94 (0.31)	92.02 (0.53)	91.73 (0.12)	89.17 (0.02)	84.46 (1.37)	85.40 (1.06)	93.21 (0.11)	89.91 (0.29)	**93.81 (0.03)**
	Macro-F1	83.20 (0.29)	91.86 (0.45)	91.63 (0.10)	89.48 (0.03)	83.96 (1.64)	84.72 (1.25)	93.07 (0.16)	89.71 (0.30)	**93.70 (0.04)**
BDGP	ACC	94.31 (1.18)	94.56 (1.73)	93.48 (0.81)	98.03 (1.73)	93.34 (0.45)	90.72 (1.83)	95.62 (0.71)	96.97 (0.05)	**98.98 (0.10)**
	Macro-F1	94.41 (1.07)	94.54 (1.74)	93.52 (0.82)	98.02 (1.76)	93.28 (0.46)	90.73 (1.80)	95.63 (0.75)	96.98 (0.05)	**99.00 (0.10)**
Flickr	ACC	56.11 (0.64)	61.24 (2.59)	59.24 (0.52)	67.47 (8.34)	59.12 (0.82)	63.97 (1.21)	67.21 (0.21)	70.52 (0.01)	**72.66 (0.25)**
	Macro-F1	55.54 (0.67)	61.08 (2.43)	59.01 (0.51)	67.16 (8.45)	58.88 (0.88)	63.05 (1.45)	67.13 (0.21)	70.48 (0.07)	**72.47 (0.22)**
esp-game	ACC	66.97 (0.67)	75.94 (3.51)	68.56 (0.42)	83.75 (6.41)	71.31 (0.74)	67.44 (1.47)	75.49 (0.19)	85.03 (0.08)	**87.97 (0.09)**
	Macro-F1	67.12 (0.53)	75.59 (3.52)	68.47 (0.42)	83.38 (6.49)	70.96 (0.76)	66.69 (1.78)	75.43 (0.17)	84.79 (0.08)	**87.83 (0.08)**

Table 2: ACC and Macro-F1 (mean and std%) of nine models with 10% labeled samples as supervision on multi-attribute and multi-modal data, in which the optimal results are highlighted in bold

set to 20. Table 2 shows the comparison results, highlighting several key observations: 1) GNN-based models, such as Co-GCN, IMvGCN, ECMGD, and TGNN, tend to achieve competitive precision. 2) Among these, TGNN stands out since we capture the essential semantics related to labels from a class-level perspective.



(a) (b)

Figure 3: T-SNE visualization of (a) Determinative representation and (b) Incidental representation on HW.

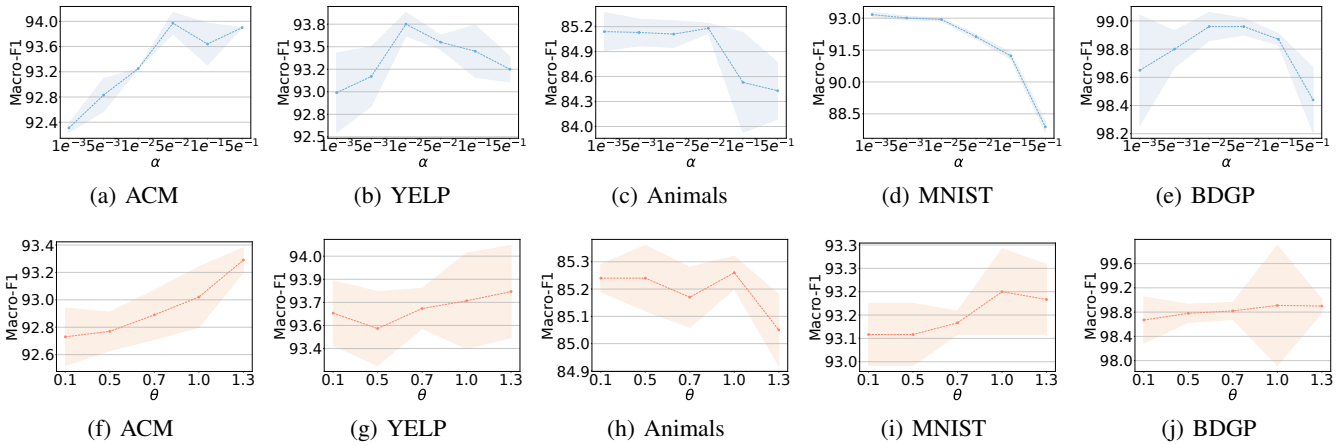
Moreover, to validate that TGNN successfully achieves the classification difference, we visualize the classification results of determinative and incidental representations. We sum the incidental representations of all views to compare them with the unified determinative representation, as shown in Figure 3. This visualization shows that TGNN, optimized by the proposed CDO loss, disentangles each view and separates the determinative and incidental features. Visualizations of other competitors are provided in Appendix **C.3**.

4.3 Parameter Sensitivity

To validate the significance of the trade-off parameter α and the separability parameter θ on model, we construct a sensitivity analysis. Due to page limitations, we display results on 2 multi-relational and 2 multi-attribute datasets, 1 multi-modal dataset, as shown in Figure 4. Results for the remaining datasets are provided in Appendix **C.3**. Figure 4 (a)-(e) plot the performance changes of TGNN w.r.t. α. We observe that the performance fluctuates as α changes, with optimal results achieved when α is relatively large on most datasets, which supports the impact of the proposed losses on model. Moreover, performance declines when α approaches 0.1, indicating that the cross-entropy loss should dominate during model optimization. Figure 4 (f)-(j) presents the influence of θ on the performance, which controls the distance between determinative and incidental features within each view. It is clear that performance peaks when the discrepancy between these two factors is a large value, which validates the importance of sematic separability.

4.4 Robustness Analysis

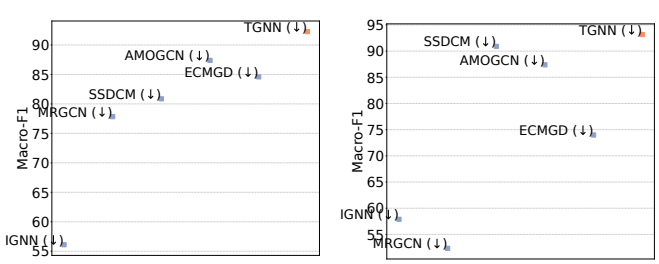
Since the model focuses on capturing task-relevant representations, it exhibits robustness when exposed to adversarial at-



(a) ACM (b) YELP (c) Animals (d) MNIST (e) BDGP

(f) ACM (g) YELP (h) Animals (i) MNIST (j) BDGP

Figure 4: Parameter sensitivity on five multi-view datasets, where (a)-(e) show the performance of TGNN w.r.t. α and (f)-(j) display the performance of TGNN w.r.t. θ.



(a) ACM (b) YELP

Figure 5: Performance of various GNN-based methods after the attack on ACM and YELP.

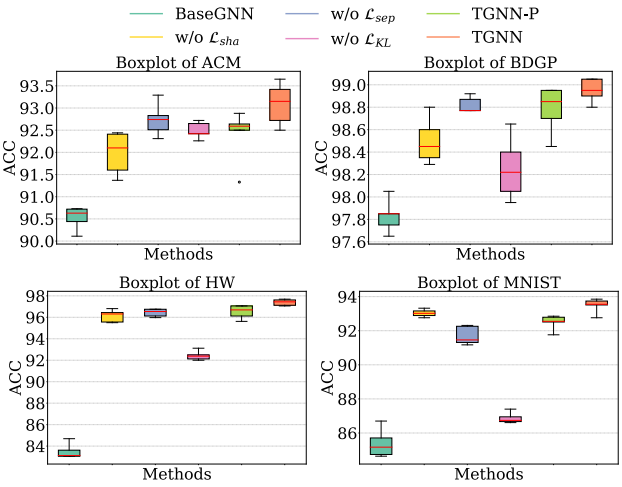


Figure 6: Results of TGNN and its variants on diverse datasets.

tacks on the graph. To verify the robustness, we utilize the state-of-the-art adversarial attack method, Mettack [Zugner and Gunnemann, 2019], which perturbs the graph with a 25% perturbation rate to mislead the classification. Following the approach of [Zhang *et al.*, 2022], we selectively attack one type of edge and evaluate the performance of various methods after the attack, as shown in Figure 5. Obviously, all methods experience a performance drop after the attack, with TGNN being the least affected. A detailed introduction of Mettack and the results on DBLP are provided in Appendix **C.3**.

4.5 Ablation Study

The proposed model emphasizes the design of diverse losses to achieve superior performance on the test set. Retaining the foundational cross-entropy loss $\mathcal{L}_{ce}$, we conduct an ablation study to assess the effectiveness of each loss term, including the view-shared loss $\mathcal{L}_{sha}$, the prediction separation loss $\mathcal{L}_{sep}$, and the KL divergence $\mathcal{L}_{KL}$. Moreover, we define TGNN-P as TGNN without considering the penalty term $\|\mathbf{c} - \mathbf{h}\|_2 > \theta$. As illustrated in Figure 6, we observe the following: 1) BaseGNN performs the worst, and the fusion loss $\mathcal{L}_{sha}$ and the prediction separation loss $\mathcal{L}_{sep}$ have a significant impact on performance, validating the importance of learning determinative representations. 2) The generalization loss $\mathcal{L}_{KL}$ is essential, as it clearly improves the model's performance on the test data. 3) TGNN outperforms all variants, highlighting the importance of each designed loss term in enhancing the model's effectiveness.

5 Conclusion

In this paper, we propose Target-oriented Graph Neural Network, which incorporates a class-level alignment mechanism across views to capture targeted-oriented representations. We begins with disentangling the features of each view into determinative and incidental factors based on their relationships with the labels. To achieve this, we design a CDO loss that distinguishes these two types of features, enabling the model to focus on determinative semantics while mitigating the influence of incidental information. Furthermore, we derive a generalization bound between the losses on the training and test data, which assists the model in fitting unknown data effectively. Extensive experiments are performed on three types of datasets, demonstrating the effectiveness of our model.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grants Nos. 62476133, 62372238), the Natural Science Foundation of Shandong Province (Grant No. ZR2022LZH003).

Contribution Statement

Sujia Huang and Lele Fu contributed equally as co-first authors.

References

[Chang *et al.*, 2024] Wei Chang, Huimin Chen, Feiping Nie, Rong Wang, and Xuelong Li. Tensorized and compressed multi-view subspace clustering via structured constraint. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10434–10451, 2024.

[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*, 174:106225, 2024.

[Gu *et al.*, 2020] Fangda Gu, Heng Chang, Wenwu Zhu, Somayeh Sojoudi, and Laurent El Ghaoui. Implicit graph neural networks. In *Proceedings of the Advances in Neural Information Processing Systems*, pages 1–12, 2020.

[Guo *et al.*, 2024] Hongtai Guo, Zhangbing Zhou, Deng Zhao, and Walid Gaaloul. Egnn: Energy-efficient anomaly detection for iot multivariate time series data using graph neural network. *Future Generation Computer Systems*, 151:45–56, 2024.

[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 *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 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 Transactions on Image Processing*, 30:6997–7011, 2021.

[Huang *et al.*, 2022] Sheng Huang, Yunhe Zhang, Lele Fu, and Shiping Wang. Learnable multi-view matrix factorization with graph embedding and flexible loss. *IEEE Transactions on Multimedia*, 25:3259–3272, 2022.

[Huang *et al.*, 2023] Sujia Huang, Shunxin Xiao, Wenzhe Liu, Jielong Lu, Zhihao Wu, Shiping Wang, and Jagath C Rajapakse. Multi-level knowledge integration with graph convolutional network for cancer molecular subtype classification. In *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine*, pages 1983–1988, 2023.

[Jiang *et al.*, 2023] Qian Jiang, Changyou Chen, Han Zhao, Liqun Chen, Qing Ping, Son Dinh Tran, Yi Xu, Belinda Zeng, and Trishul Chilimbi. Understanding and constructing latent modality structures in multi-modal representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7661–7671, 2023.

[Kant *et al.*, 2024] Yash Kant, Aliaksandr Siarohin, Ziyi Wu, Michael Vasilkovsky, Guocheng Qian, Jian Ren, Riza Alp Guler, Bernard Ghanem, Sergey Tulyakov, and Igor Gilitschenski. Spad: Spatially aware multi-view diffusers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10026–10038, 2024.

[Li *et al.*, 2020] Shu Li, Wen-Tao Li, and Wei Wang. Co-gcn for multi-view semi-supervised learning. In *Proceedings of the AAAI conference on artificial intelligence*, pages 4691–4698, 2020.

[Liang *et al.*, 2024] Paul Pu Liang, Zihao Deng, Martin Q Ma, James Y Zou, Louis-Philippe Morency, and Ruslan Salakhutdinov. Factorized contrastive learning: Going beyond multi-view redundancy. *Advances in Neural Information Processing Systems*, 36, 2024.

[Liao *et al.*, 2024] Jian Liao, Feng Liu, Jianxing Zheng, Suge Wang, Deyu Li, and Qian Chen. A dynamic adaptive multi-view fusion graph convolutional network recommendation model with dilated mask convolution mechanism. *Information Sciences*, 658:120028, 2024.

[Lin *et al.*, 2022] Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4447–4461, 2022.

[Lu *et al.*, 2024a] Jielong Lu, Zhihao Wu, Zhaoliang Chen, Zhiling Cai, and Shiping Wang. Towards multi-view consistent graph diffusion. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 186–195, 2024.

[Lu *et al.*, 2024b] 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 Transactions on Multimedia*, 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 *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1234–1244, 2021.

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

[Qi *et al.*, 2024] Zhuang Qi, Lei Meng, Weihao He, Ruohan Zhang, Yu Wang, Xin Qi, and Xiangxu Meng. Cross-training with multi-view knowledge fusion for heterogenous federated learning. *CoRR*, abs/2405.20046, 2024.

[Ren *et al.*, 2024] Yanjiao Ren, Yimeng Gao, Wei Du, Weibo Qiao, Wei Li, Qianqian Yang, Yanchun Liang, and Gaoyang Li. Classifying breast cancer using multi-view graph neural network based on multi-omics data. *Frontiers in Genetics*, 15:1363896, 2024.

[Shalev-Shwartz and Ben-David, 2014] Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. 2014.

[Tan *et al.*, 2023] Yuze Tan, Yixi Liu, Shudong Huang, Wentao Feng, and Jiancheng Lv. Sample-level multi-view graph clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23966–23975, 2023.

[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 *Proceedings of the world wide web conference*, pages 2022–2032, 2019.

[Wang *et al.*, 2022] Shiping Wang, Zhaoliang Chen, Shide Du, and Zhouchen Lin. Learning deep sparse regularizers with applications to multi-view clustering and semi-supervised classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5042–5055, 2022.

[Wang *et al.*, 2023] Jing Wang, Songhe Feng, Gengyu Lyu, and Zhibin Gu. Triple-granularity contrastive learning for deep multi-view subspace clustering. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 2994–3002, 2023.

[Wang *et al.*, 2024a] Jingyao Wang, Wenwen Qiang, Jiangmeng Li, Lingyu Si, Changwen Zheng, and Bing Su. On the causal sufficiency and necessity of multi-modal representation learning. *CoRR*, abs/2407.14058, 2024.

[Wang *et al.*, 2024b] Shiping Wang, Sujia Huang, Zhihao Wu, Rui Liu, Yong Chen, and Dell Zhang. Heterogeneous graph convolutional network for multi-view semi-supervised classification. *Neural Networks*, page 106438, 2024.

[Wen and Li, 2025] Gang Wen and Limin Li. Mmosurv: meta-learning for few-shot survival analysis with multi-omics data. *Bioinformatics*, 41(1):btae684, 2025.

[Wen *et al.*, 2024] Jie Wen, Gehui Xu, Zhanyan Tang, Wei Wang, Lunke Fei, and Yong Xu. Graph regularized and feature aware matrix factorization for robust incomplete multi-view clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(5):3728–3741, 2024.

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

[Xia *et al.*, 2023] Wei Xia, Tianxiu Wang, Quanxue Gao, Ming Yang, and Xinbo Gao. Graph embedding contrastive multi-modal representation learning for clustering. *IEEE Transactions on Image Processing*, 32:1170–1183, 2023.

[Xie *et al.*, 2020a] Yu Xie, Yuanqiao Zhang, Maoguo Gong, Zedong Tang, and Chao Han. Mgat: Multi-view graph attention networks. *Neural Networks*, 132:180–189, 2020.

[Xie *et al.*, 2020b] Yuan Xie, Wensheng Zhang, Yanyun Qu, Longquan Dai, and Dacheng Tao. Hyper-laplacian regularized multilinear multiview self-representations for clustering and semisupervised learning. *IEEE Transactions on Cybernetics*, 50(2):572–586, 2020.

[Yang *et al.*, 2023] Mengyue Yang, Zhen Fang, Yonggang Zhang, Yali Du, Furui Liu, Jean-Francois Ton, Jianhong Wang, and Jun Wang. Invariant learning via probability of sufficient and necessary causes. *Advances in Neural Information Processing Systems*, 36:79832–79857, 2023.

[Yang *et al.*, 2025] Xuanhao Yang, Hangjun Che, and Man-Fai Leung. Tensor-based unsupervised feature selection for error-robust handling of unbalanced incomplete multi-view data. *Information Fusion*, 114:102693, 2025.

[Yu *et al.*, 2022] Pengyang Yu, Chaofan Fu, Yanwei Yu, Chao Huang, Zhongying Zhao, and Junyu Dong. Multiplex heterogeneous graph convolutional network. In *In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2377–2387, 2022.

[Yu *et al.*, 2024] Zaiyang Yu, Lusi Li, Jinlong Xie, Changshuo Wang, Weijun Li, and Xin Ning. Pedestrian 3d shape understanding for person re-identification via multi-view learning. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–14, 2024.

[Yu *et al.*, 2025] Xuejiao Yu, Yi Jiang, Guoqing Chao, and Dianhui Chu. Deep contrastive multi-view subspace clustering with representation and cluster interactive learning. *IEEE Transactions on Knowledge and Data Engineering*, 37(1):188–199, 2025.

[Zhang *et al.*, 2022] Mengmei Zhang, Xiao Wang, Meiqi Zhu, Chuan Shi, Zhiqiang Zhang, and Jun Zhou. Robust heterogeneous graph neural networks against adversarial attacks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4363–4370, 2022.

[Zhang *et al.*, 2025] Zitong Zhang, Xiaojun Chen, Chen Wang, Ruili Wang, Wei Song, and Feiping Nie. Structured multi-view k-means clustering. *Pattern Recognition*, 160:111113, 2025.

[Zhao *et al.*, 2025] Zihua Zhao, Ting Wang, Haonan Xin, Rong Wang, and Feiping Nie. Multi-view clustering via high-order bipartite graph fusion. *Information Fusion*, 113:102630, 2025.

[Zhou *et al.*, 2024] Xiaokang Zhou, Jiayi Wu, Wei Liang, Kevin I-Kai Wang, Zheng Yan, Laurence T. Yang, and Qun Jin. Reconstructed graph neural network with knowledge distillation for lightweight anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems*, 35(9):11817–11828, 2024.

[Zugner and Gunnemann, 2019] Daniel Zugner and Stephan Gunnemann. Adversarial attacks on graph neural networks via meta learning. In *Proceedings of the Tnternational Conference on Learning Representations*, 2019.