Online Multi-Relational Clustering with Dominant View Mining

Zhengzhong Zhu[1], Pei Zhou[1], Dongsheng Wang[2], Li Cheng[1], Jiangping Zhu[1]*

[1]College of Computer Science, Sichuan University, Chengdu, China
[2]University of Electronic Science and Technology of China, Chengdu, China
dswang1996@163.com, {zjp16,zhoupei}@scu.edu.cn, {zhengzhongzhu,chengli}@stu.scu.edu.cn

Abstract

Multi-relational graph clustering aims to uncover complex node interactions by leveraging multiple relational views, yet existing methods often suffer from two key limitations: they assume equal importance across views and decouple representation learning from clustering, both of which hinder overall performance. To address these issues, we propose OMC-DVM, a novel end-to-end Online Multi-Relational Graph Clustering With Dominant View Mining framework. OMC-DVM introduces two core innovations: (1) A unsupervised dominant view mining module that dynamically identifies the dominant view using Maximum Mean Discrepancy (MMD) and adaptively aligns other views to it, mitigating view imbalance. (2) An online ,multi-relational clustering process that unifies representation learning and clustering into a single stage. By performing clustering-level contrastive learning , OMC-DVM directly generates cluster assignments in an end-to-end manner. Extensive experiments on both real-world and synthetic benchmark datasets demonstrate that OMC-DVM not only achieves state-of-the-art clustering performance but also effectively alleviates the view imbalance problem in multi-relational graphs.

Introduction

Multi-relational graphs are fundamental data structures that consist of nodes connected by multiple types of relationships(Qu et al. 2017). These graphs provide a powerful representation for complex systems and have broad applications in various domains, such as citation networks, social networks, and knowledge graphs (Pan and Kang 2023; Wu et al. 2023).

Recently, there has been increasing interest in exploring the inherent patterns within multi-relational graphs, particularly in the context of Multi-View Graph Clustering (MVGC)(Fan et al. 2020; Park et al. 2020; Qian, Li, and Kang 2024; Jing, Park, and Tong 2021). With the development of deep representation learning, a new wave of contrastive multi-relational graph clustering methods has emerged. For example, HDMI (Jing, Park, and Tong 2021) conducts contrastive learning by maximizing the high-order mutual information among node features; MGDCR (Mo et al. 2023a) reduces the impact of noisy information within each graph and captures the common information among different graphs by maximizing mutual information; BTGF (Qian, Li, and Kang 2024) incorporates the Barlow Twins loss to achieve alignment of representations and prevent dimensional collapse; and BGMP (Shen, He, and Kang 2024) computes the Aggregation Class Distance (ACD) between views, aligning other views to the one with the smallest ACD through contrastive learning. Despite these successes achieved by contrastive MVGC methods, they still face two questions , **Question (1):Representation degradation** : different views are not always equally important, and an imbalance often exists. Simply forcing the alignment of different views may degrade the final performance. **Question(2) : Clustering performance degradation**: they decouple representation learning and clustering into two separate stages, which leads to a misalignment between the learned representations and the clustering task. In practice, multi-relational graphs are often constructed from heterogeneous data sources with varying reliability and semantic richness—for example, social networks combining friendship, interaction, and co-location relations, where some views are inherently noisy. Failing to account for view imbalance risks overfitting to spurious patterns and undermining

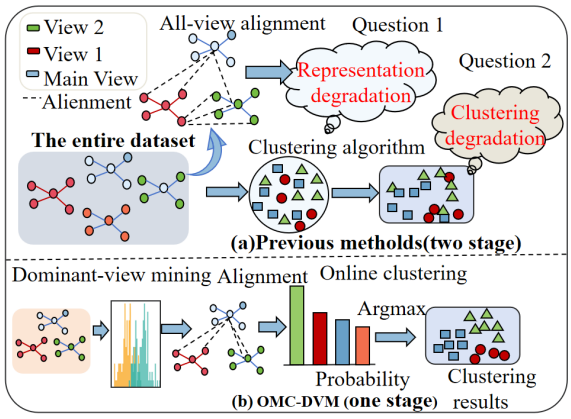


Figure 1: Comparison between previous methods and ours. Previous methods aligned all views equally and decoupled representation learning from clustering, whereas our method aligns the other views to the identified dominant view and integrates clustering and representation learning into a single stage.

*Corresponding author

the utility of informative views(Han et al. 2022; Liu et al. 2018; Wang, Zhang, and Zhou 2025a,b). Moreover, the two-stage optimization separates representation learning from clustering, which weakens their alignment and reduces clustering quality(Chen et al. 2023; Li et al. 2022; Chao, Jiang, and Chu 2024; Sun et al. 2025). It is also inefficient because it requires processing the entire dataset multiple times before producing results. As a result, clustering outcomes are delayed until both stages are complete.

To address these issues, we propose a novel end-to-end framework, Online Multi-Relational Clustering with Dominant View Mining (OMC-DVM). Specifically, OMC-DVM faces two key research challenges: **(C1): How can the dominant view be identified in an unsupervised manner to guide multi-relational graph clustering?** Early studies in multimodal learning have observed the phenomenon of dominant views in imbalanced data (Wang, Tran, and Feiszli 2020; Zhou et al. 2022, 2024). **(C2): How can unsupervised representation learning and clustering be unified into a single stage?** Unlike other multiview data, the discrepancies among views in multi-relational graphs originate from their distinct topological structures. In order to address C1, given the difficulty of directly identifying the dominant view in an unsupervised setting, OMC-DVM propose a dominant-view mining method based on Maximum Mean Discrepancy (MMD) (Wu et al. 2024; Gretton et al. 2006). This method maps each view's features into a high-dimensional Reproducing Kernel Hilbert Space (RKHS) and measures the distance between their mean embeddings. Views with smaller distances to the global features exhibit higher consistency and are thus deemed more representative. We then align the remaining views to this dominant view to enhance representation quality. The effectiveness of this strategy is validated through empirical studies on real-world datasets. To address C2, OMC-DVM introduce a simple yet effective cluster-level contrastive learning mechanism for online clustering, which bridges the performance gap caused by the decoupling of representation learning and clustering. We first perform self-weighted fusion of features from all views to obtain global features, and construct sample pairs by adding Gaussian noise. Cluster-level contrastive learning is then applied in the clustering space, directly producing the final clustering results. This unified, end-to-end framework,OMC-DVM , seamlessly integrates dominant-view alignment and cluster-level contrastive learning, jointly tackling representation degradation and enabling efficient online clustering. Fig. 1 presents a comparison between our method and existing approaches. In total , our contributions are as follows:

- We propose an end-to-end contrastive multi-view clustering framework that jointly addresses view imbalance and representation-clustering misalignment.
- We introduce a novel unsupervised dominant-view mining strategy based on MMD, which identifies the most representative view and aligns others accordingly, effectively mitigating feature degradation.
- We design a cluster-level contrastive learning mechanism that operates directly in the clustering space, unifying representation learning and clustering in a single stage.
- Extensive experiments demonstrate the superior performance, scalability, and generalization ability of our method over state-of-the-art baselines.

Related Work

Multiview graph clustering (MVGC) has garnered increasing attention in recent years. Early methods, such as MvAGC (Lin et al. 2021) and MCGC (Pan and Kang 2021), combine graph filtering with self-expression to exploit both attribute and structural information. With the advent of deep learning, numerous MVGC approaches (Fan et al. 2020; Park et al. 2020; Jing, Park, and Tong 2021; Liu et al. 2022a; Mo et al. 2023a; Qian, Li, and Kang 2024; Ling et al. 2023; Peng, Wang, and Zhu 2023; Mo et al. 2023b) learn node representations via GNNs and perform clustering with k-means (Hartigan and Wong 1979). O2MAC (Fan et al. 2020) pioneers GNN-based MVGC by selecting the most informative view to reconstruct all graphs. Subsequent methods maximize mutual information between local and global embeddings (Park et al. 2020; Jing, Park, and Tong 2021) or employ contrastive losses to align multi-view representations and prevent collapse (Liu et al. 2022a; Mo et al. 2023a; Qian, Li, and Kang 2024). Others incorporate supervised or semi-supervised signals to weigh view importance (Wang et al. 2019; Sadikaj et al. 2023), but their reliance on labels limits unsupervised applicability. BMGC (Mo et al. 2023a) proposes an aggregation-based distance for aligning dominant view representations, yet it decouples representation learning and clustering into two separate stages, lacking joint optimization, which leads to suboptimal clustering performance.

Notation

We consider a multi-relational graph defined as:

$$\mathcal{G} = \{\mathcal{V}, \mathcal{E}_1, \cdots, \mathcal{E}_v, \cdots, \mathcal{E}_V, X\},$$

where $\mathcal{V}$ is a set of N nodes, and each $\mathcal{E}_v$ corresponds to the edge set of the v-th view, with $v \in \{1, \ldots, V\}$ and $V > 1$ representing the total number of views. The feature matrix is denoted as $X \in \mathbb{R}^{N \times d_f}$, d_f is the dimension of features ,where each column $x \in \mathbb{R}^N$ is interpreted as a graph signal over the node set. We use $\tilde{A}^v$ to denote the unnormalized adjacency matrix for the v-th view, and D^v to represent its corresponding degree matrix. The symmetrically normalized adjacency matrix for view v is computed as:

$$A^v = (D^v)^{-\frac{1}{2}} \tilde{A}^v (D^v)^{-\frac{1}{2}}.$$

It is known that the eigenvalues of such normalized matrices A^v are bounded within the interval $[-1, 1]$. To incorporate self-loops, the adjusted normalized adjacency matrix for the v-th view is formulated as:

$$\tilde{A}^v = (D^v + I)^{-\frac{1}{2}} \left(\tilde{A}^v + I\right) (D^v + I)^{-\frac{1}{2}}.$$

Where I is the identity matrix, ensuring each node connects to itself. Lastly, let C denote the number of node classes, and $y \in \mathbb{R}^N$ be the label vector containing the ground-truth class labels for each node in the graph.

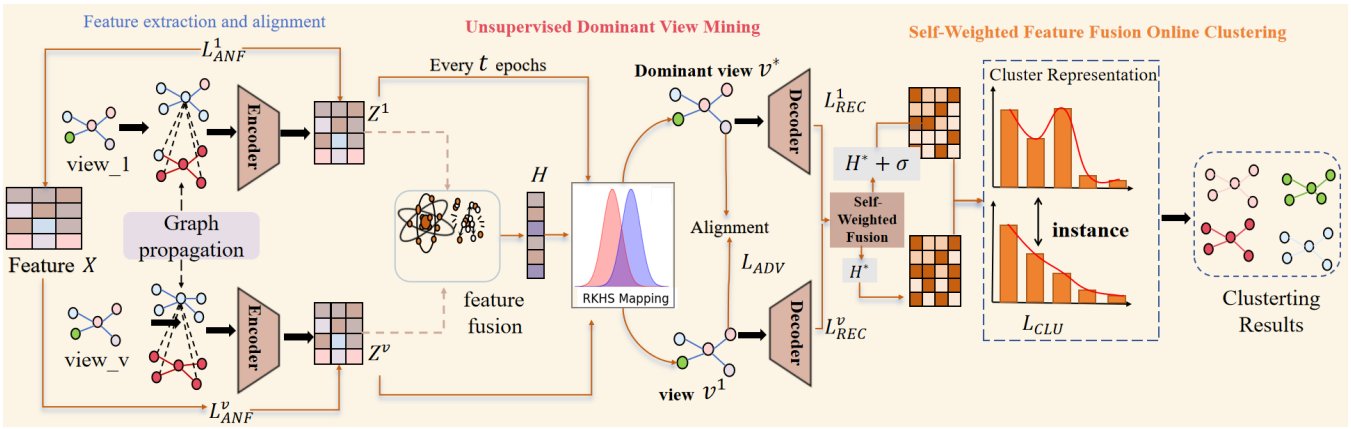


Figure 2: Illustration of our proposed framework OMC-DVM . Firstly, node representations for each view are obtained through scalable graph encoding. Then, unsupervised dominant view mining and self-weighted feature fusion online clustering synergistically facilitate model training.

Method

Scalable Graph Encoding

In contrast to most graph neural network -based approaches (Mo et al. 2023a,b), we aim to enhance scalability by decoupling the graph propagation and dimensionality reduction processes. Our approach begins by performing separate propagation on the node features for each view, yielding view-specific aggregated features. Following the approximate personalized propagation technique in (Chien et al. 2020; Shen, He, and Kang 2024), we use the original node features as a teleport vector at each propagation layer:

$$X^{v,0} = X, \quad X^{v,k+1} = (1-\alpha)\hat{A}^v X^{v,k} + \alpha X, \quad (1)$$

in this equation, X serves as both the initial feature matrix and the teleport set for each view. The hyper-parameter $\alpha \in [0,1]$ controls the teleport probability, and $k \in [0, K-1]$ represents the number of propagation steps. After K steps, the aggregated features are denoted by $X^v = X^{v,K}$. These aggregated features are then passed through a shared encoder for dimensionality reduction:

$$Z^v = f_\theta(X^v), \quad (2)$$

where $Z^v \in \mathbb{R}^{N \times d_r}$ denotes the reduced node representations for the v-th view. This decoupled approach eliminates the need for time-consuming graph convolution operations during training. Next, the representations for each view are passed through a shared decoder to reconstruct the view-specific aggregated features. To ensure effective training, we optimize the reconstruction using the cosine error between the original and the reconstructed features:

$$\hat{X}^v = g_\Theta(Z^v), \quad (3)$$

$$\mathcal{L}_{REC} = \frac{1}{NV}\sum_{v=1}^{V}\sum_{i=1}^{N}\left(1 - \frac{(X_i^v)^\top \hat{X}_i^v}{\|X_i^v\|\|\hat{X}_i^v\|}\right). \quad (4)$$

In this formulation, the encoder $f_\theta(\cdot)$ and decoder $g_\Theta(\cdot)$ are both implemented using Multilayer Perceptrons (MLPs).

Unsupervised Dominant View Mining

To compute the weights of different views, we measure the discrepancy between view features using the maximum mean discrepancy (MMD). MMD is a non-parametric metric that does not assume any specific distribution, directly evaluates discrepancies based on sample space, and performs well with complex, nonlinear distributions. Given two sample sets $\{p_i\}_{i=1}^{n_p}$ and $\{q_j\}_{j=1}^{n_q}$ from distributions P and Q, respectively, MMD is defined as:

$$MMD(P,Q) = \sup_{f\in\Omega}\left(\mathbb{E}_{p\sim P}[f(p)] - \mathbb{E}_{q\sim Q}[f(q)]\right), \quad (5)$$

where Ω is a function class in a reproducing kernel Hilbert space (RKHS). In practice, MMD is computed via kernels:

$$MMD^2(P,Q) = \frac{1}{n_p^2}\sum_{i=1}^{n_p}\sum_{i'=1}^{n_p}k(p_i,p_{i'}) + \frac{1}{n_q^2}\sum_{j=1}^{n_q}\sum_{j'=1}^{n_q}k(q_j,q_{j'}) - \frac{2}{n_p n_q}\sum_{i=1}^{n_p}\sum_{j=1}^{n_q}k(p_i,q_j), \quad (6)$$

we estimate the discrepancy between each view's features Z^ν and the global features H as $D(Z^\nu, H)$. Using a linear kernel $k(x,y) = x^T y$ in RKHS, and given that Z^ν and H have the same dimensionality, the MMD distance is computed as:

$$D_{MMD}(Z^\nu, H) = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}k(Z_i^\nu, Z_j^\nu) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}k(H_i,H_j) - \frac{2}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}k(Z_i^\nu, H_j). \quad (7)$$

To assign higher weights to views with lower discrepancies,

we apply a normalized softmax-based weight function:

$$W^{\nu} = P(D(Z^{\nu}, H)) = \text{Softmax}(-D(Z^{\nu}, H)) = \frac{e^{-D_{MMD}(Z^{\nu}, H)}}{\sum_{\nu'} e^{-D_{MMD}(Z^{\nu'}, H)}}, \tag{8}$$

finally, the dominant view v^* is determined as the one with the highest weight:

$$v^* = \arg\max(W^{\nu}). \tag{9}$$

Representation Alignment and Node Feature Consistency

According to Equation (9), we identify the dominant view, and the representations of other views are aligned to this dominant view through contrastive learning. The contrastive loss for a pair of representations is defined as:

$$\ell(Z_i^v, Z_i^{v*}) = -\log \frac{e^{\text{sim}(\tilde{Z}_i^v, \tilde{Z}_i^{v*})/\tau_I}}{\sum_{j=1}^{N} e^{\text{sim}(\tilde{Z}_i^v, \tilde{Z}_j^{v*})/\tau_I}}, \tag{10}$$

where $\tilde{Z}_i^v$ is the non-linear projection of Z_i^v, $\text{sim}(\cdot)$ denotes cosine similarity, and τ_I is a temperature parameter. The overall alignment loss with respect to the dominant view is computed as:

$$\mathcal{L}_{ADV} = \frac{1}{2N(V-1)} \sum_{\substack{v=1 \\ v \neq v^*}}^{V} \sum_{i=1}^{N} \left(\ell(Z_i^v, Z_i^{v*}) + \ell(Z_i^v, Z_i^v) \right). \tag{11}$$

To further enforce consistency between representations and node features, we introduce an auxiliary loss that aligns the similarity distributions of node features and learned representations:

$$\mathcal{L}_{ANF} = \frac{1}{N^2 V} \sum_{v=1}^{V} \left\| XX^{\top} - Z^v (Z^v)^{\top} \right\|_F^2, \tag{12}$$

finally, the total loss for co-aligned representation learning is given by:

$$\mathcal{L}_{CAL} = \mathcal{L}_{ADV} + \mathcal{L}_{ANF}. \tag{13}$$

Self-Weighted Feature Fusion Online Clustering

Based on W^{ν}, we can obtain the weight for each view. Considering that the importance of different views varies, we integrate the features from each view according to their weights to obtain the fuison feature H^*, where $H^* = \sum_{v=1}^{V} W^{\nu} Z^v$. Then, we apply random Gaussian noise (Zhang et al. 2022) to the fusion features H^*, as formulated below:

$$\hat{H} = H^* + N \quad \text{where} \quad N \sim \mathcal{N}(0, \sigma), \tag{14}$$

where N is sampled from the Gaussian distribution $\mathcal{N}(0, \sigma)$.

Next, we use a two-layer Multilayer Perceptron (MLP) $g_C(\cdot)$ to project the feature matrix into an M-dimensional space. This results in a soft label $y_a = g_C(\hat{h}_i)$, $y_b = g_C(h_i^*)$,where y_i^a represents the soft label of sample x_i^a (the i-th row of y^a). For clarity, let $\hat{y}_i^a$ denote the i-th column of y_a, representing the cluster assignment for cluster i under the first data augmentation. Similarly, we combine $\hat{y}_i^a$ with $\hat{y}_i^b$ to form a positive cluster pair $\{\hat{y}_i^a, \hat{y}_i^b\}$, while all other $2M-2$ pairs are treated as negative pairs, where $\hat{y}_i^b$ is the second augmented representation of cluster i. To measure the similarity between cluster pairs, we use cosine distance:

$$s(\hat{y}_i^k, \hat{y}_j^k) = \frac{\hat{y}_i^{k\top} \hat{y}_j^k}{\|\hat{y}_i^k\| \|\hat{y}_j^k\|}, \tag{15}$$

where $k \in \{a, b\}$ and $i, j \in [1, M]$. Without loss of generality, the following loss function is used to distinguish cluster $\hat{y}_i^a$ from all other clusters except $\hat{y}_i^b$:

$$\hat{\ell}_i^a = -\log \left(\frac{\exp(s(\hat{y}_i^a, \hat{y}_i^b)/\tau_C)}{\sum_{j=1}^{M} \left[\exp(s(\hat{y}_i^a, \hat{y}_j^a)/\tau_C) + \exp(s(\hat{y}_i^a, \hat{y}_j^b)/\tau_C) \right]} \right), \tag{16}$$

where τ_C is the cluster-level temperature parameter that controls the softness of the similarity function. By considering all clusters, the cluster-level contrastive loss is computed as:

$$\mathcal{L}_{CLU} = \frac{1}{2M} \sum_{i=1}^{M} \left(\hat{\ell}_i^a + \hat{\ell}_i^b \right) - M(Y), \tag{17}$$

$$M(Y) = -\sum_{i=1}^{M} \left[P(\hat{y}_i^a) \log P(\hat{y}_i^a) + P(\hat{y}_i^b) \log P(\hat{y}_i^b) \right]. \tag{18}$$

The entropy of the cluster assignment probabilities:$P(\hat{y}_i^k) = \frac{1}{N} \sum_{t=1}^{N} Y_{ti}^k, \quad k \in \{a, b\}$ within a mini-batch under each data augmentation. This entropy term helps avoid the trivial solution where most instances are assigned to the same cluster (Li et al. 2022, 2025, 2023; Liu et al. 2024; Yang et al. 2024). The overall objective function is defined as:

$$\mathcal{L} = \mathcal{L}_{CAL} + \mathcal{L}_{CLU} + \mathcal{L}_{REC}. \tag{19}$$

While it is possible to introduce trade-off hyperparameters to adjust the relative influence of each loss term, we empirically found that assigning equal weights to three losses leads to strong performance.

Experiments

Datasets and Metrics

Datasets. We conduct experiments on five widely used real-world benchmark datasets, along with a large-scale multi-relational dataset. Specifically, ACM (Fan et al. 2020), ACM2 (Fu et al. 2020), and DBLP are academic citation networks, while Yelp (Zhao et al. 2020) and Amazon (Lu et al. 2019) are derived from user review platforms. In addition, MAG (Wang et al. 2020) is a large-scale citation graph and currently serves as the largest dataset for multi-relational graph clustering tasks.

Metrics. We evaluate clustering performance using four widely adopted metrics: Accuracy (ACC), Normalized Mutual Information (NMI), F1-score, and Adjusted Rand Index (ARI)(Liu et al. 2022b; Tu et al. 2021; Chen et al. 2025, 2024; Ma et al. 2025; Ye et al. 2025). In all cases, higher metric values reflect better clustering quality.

Datasets	Metric	HAN*	VGAE	DGI	O2MAC	DMGI	MvAGC	HDMI	MCGC	MGDCR	DuaLGR	DMG	BTGF	BMGC	Ours
Amazon	NMI	0.4037	0.0163	0.0532	0.1344	0.2623	0.2322	0.3702	0.2149	0.0318	0.2767	0.1218	0.3853	**0.5768**	0.5731
	ARI	0.4241	0.0129	0.0202	0.0898	0.2605	0.1141	0.275	0.1056	0.0055	0.2715	0.0283	0.2829	0.5626	**0.5816**
	ACC	0.7437	0.3194	0.3762	0.4428	0.5188	0.5188	0.5251	0.4683	0.3489	0.6123	0.3887	0.6603	0.7856	**0.7987**
	F1	0.7433	0.2725	0.2859	0.4424	0.5463	0.5072	0.5448	0.4804	0.2039	0.6215	0.3441	0.6612	0.7851	**0.7958**
ACM	NMI	0.6864	0.491	0.6364	0.6923	0.6441	0.6735	0.645	0.7126	0.721	0.7328	0.7561	0.7581	0.7841	**0.7887**
	ARI	0.7489	0.54848	0.6822	0.7394	0.6729	0.7212	0.674	0.7627	0.6496	0.7942	0.8033	0.8085	0.8329	**0.8384**
	ACC	0.9088	0.8228	0.8816	0.9042	0.8724	0.8975	0.874	0.9147	0.919	0.9271	0.9302	0.9322	0.9413	**0.9431**
	F1	0.9085	0.8231	0.8829	0.9053	0.8709	0.8986	0.872	0.9155	0.8678	0.927	0.9306	0.9331	0.9416	**0.9433**
DBLP	NMI	0.6998	0.6934	0.6168	0.7294	0.7489	0.7723	0.6361	0.6561	0.7559	0.75559	0.7907	0.6027	0.8013	**0.8175**
	ARI	0.7641	0.413	0.5653	0.7783	0.8032	0.828	0.6145	0.7088	0.8168	0.8168	0.8384	0.6534	0.8539	**0.8650**
	ACC	0.9015	0.8868	0.7446	0.9071	0.9159	0.9284	0.7832	0.8752	0.9242	0.9242	0.9344	0.8509	0.9401	**0.9486**
	F1	0.8966	0.8748	0.7392	0.901	0.9075	0.9231	0.7372	0.8186	0.918	0.918	0.9303	0.8456	0.9364	**0.9455**
ACM2	NMI	0.6435	0.4507	0.5779	0.4223	0.574	0.1819	0.5902	0.5307	0.5988	0.5988	0.6341	0.6483	0.7285	**0.7406**
	ARI	0.6979	0.4347	0.5174	0.4451	0.6243	0.1879	0.5472	0.4396	0.6399	0.6399	0.6726	0.6776	0.7601	**0.7787**
	ACC	0.8955	0.7358	0.8114	0.7537	0.8148	0.5949	0.8258	0.7129	0.8676	0.8676	0.8796	0.8853	0.9185	**0.9253**
	F1	0.8955	0.7101	0.8261	0.7418	0.8267	0.4484	0.8386	0.5809	0.8653	0.8653	0.8773	0.8887	0.9215	**0.9269**
Yelp	NMI	0.6762	0.3919	0.3942	0.3902	0.3729	0.2439	0.3912	0.3825	0.6621	0.6621	0.391	0.4135	0.7173	**0.7297**
	ARI	0.7205	0.4257	0.4262	0.4253	0.4253	0.2925	0.3922	0.3517	0.6847	0.6847	0.4261	0.3564	0.7381	**0.7531**
	ACC	0.9082	0.6507	0.6529	0.6507	0.6507	0.6314	0.6452	0.6561	0.8948	0.8948	0.6512	0.7192	0.9151	**0.9208**
	F1	0.9163	0.5674	0.5679	0.5674	0.5674	0.567	0.5874	0.5749	0.9051	0.9051	0.5679	0.7307	0.9246	**0.9305**

Table 1. Clustering results on real-world datasets. The best and second-place results are highlighted using bold and underline, respectively. The asterisk (*) denotes the supervised baseline. The baseline experimental results we reported are from BMGC.

Baselines

We evaluate OMC-DVM against a broad set of baseline methods, covering both single-view and multi-view graph clustering approaches. Specifically, we include the supervised multi-view graph method HAN (Wang et al. 2019) as a reference, along with unsupervised single-view methods such as VGAE (Kipf and Welling 2016) and DGI (Veličković et al. 2018). For multi-view graph clustering, we consider several state-of-the-art unsupervised models, including O2MAC (Fan et al. 2020), DMGI (Park et al. 2020), MvAGC (Lin et al. 2021), HDMI (Jing, Park, and Tong 2021), MCGC (Pan and Kang 2021), MGDCR (Mo et al. 2023a), DuaLGR (Ling et al. 2023), DMG (Mo et al. 2023b), BGMP (Shen, He, and Kang 2024), and BTGF (Qian, Li, and Kang 2024). Among these, only HAN is a supervised method, while all others operate in an unsupervised setting.

Parameter Setting

Our model is optimized using the Adam optimizer(Kingma and Ba 2014) with a learning rate selected from the range $[5 \times 10^{-3}, 1 \times 10^{-2}]$ and a weight decay of 1×10^{-4}. The dominant view is updated at fixed intervals, specifically every 10 epochs, with the interval parameter t set accordingly. For the ACM2 dataset, we set the representation dimension d_r to 64, while for the remaining datasets, d_r is set to 10. The temperature parameter τ is consistently fixed at 1 across all experiments. We choose the graph propagation radius $K = 3$, and the teleport probability α is selected from $[0, 0.3, 0.5]$. All experiments are implemented on the PyTorch framework, running on an AMD EPYC 9654 CPU paired with an NVIDIA A100 GPU with 80GB of memory.

Main Results

Experiments were conducted on real-world, large-scale, and synthetic datasets(Chien et al. 2021), with the synthetic results reported in Appendix C due to space constraints.

Evaluation on Real-world Datasets

To comprehensively evaluate the effectiveness of OMC-DVM , we conducted experiments on five real-world datasets, comparing it against a variety of baseline methods. For the supervised method HAN, we first extracted node embeddings from the test set and then applied k-means clustering to obtain clustering results. For single-view methods, each view was evaluated independently, and the best result among them was reported. Overall, OMC-DVM consistently achieved superior performance across all datasets on four standard evaluation metrics. Specifically, as is shown in Table 1, OMC-DVM demonstrated significant performance gains on datasets such as DBLP and ACM2, outperforming both unsupervised and supervised baselines, including HAN. Even on the Amazon dataset, where its NMI ranked second overall, OMC-DVM still improved the ACC and ARI scores by 1.73% and 2.0%, respectively, showcasing its effectiveness and competitiveness. Further analysis reveals that multi-view clustering methods generally outperform single-view methods such as VGAE and DGI, highlighting the benefits of leveraging information from multiple relational views. However, on the YELP dataset, many multi-view baselines underperformed compared to single-view models, likely due to their inability to effectively address view imbalance, which compromised model stability. Notably, OMC-DVM maintained its advantage even under such challenging conditions, underscoring its robustness and adaptability. Moreover, compared with O2MAC, which retains only a single dominant view while discarding others, OMC-DVM preserves multi-view information by aligning all views to the identified dominant one. This design enhances the model's ability to capture complex relational structures and results in significantly better performance than O2MAC. Finally, although BMGC attempts to mitigate the view imbalance issue, it fails to bridge the gap between representation learning and clustering objectives. In contrast, OMC-DVM integrates the two tasks through a unified learning framework, achieving more consistent and

superior results across all datasets.

Evaluation on Large-scale Dataset

To demonstrate the advantages of our online clustering approach, we conduct experiments on the large-scale multi-relational graph MAG. For the MAG dataset, the representation dimension is set to 64, with a batch size of 5000, and the final embedding dimension set to 128. Unlike traditional two-stage methods, OMC-DVM integrates representation learning and clustering into a unified optimization process, which not only improves clustering performance but also significantly reduces training time. Specifically, although BMGC also addresses view imbalance, it overlooks the performance degradation and computational overhead caused by decoupling representation learning and clustering into separate stages. On the MAG dataset, OMC-DVM achieves superior clustering performance while requiring only 76% of BMGC's training time. This efficiency gain is attributed to OMC-DVM 's online clustering paradigm, which produces clustering results in a single forward pass as data arrive, eliminating the iterative full-dataset processing inherent to offline methods. Table 2 reports the results, including standard deviations and training times, demonstrating that OMC-DVM consistently achieves the best performance with the shortest training time among all evaluated methods.

Methods	NMI	ARI	ACC	F1	Time
k-means	42.04	32.34	58.63	59.81	-
DGI	53.56±0.48	42.60±0.83	59.89±1.10	57.17±1.88	36
DMGI	49.71±1.37	38.91±1.35	53.57±0.54	49.59±1.39	118
HDMI	48.15±0.98	34.92±1.27	51.78±1.37	49.80±2.04	105
MGDCR	54.43±1.17	43.98±1.16	61.37±2.46	60.53±3.19	39
DMG	44.04±3.32	36.97±2.86	57.65±2.03	55.32±2.53	95
BMGC	<u>57.01±0.19</u>	<u>47.84±0.27</u>	<u>65.31±1.25</u>	<u>63.68±1.84</u>	<u>25</u>
Ours	**58.13±0.17**	**48.22±0.25**	**66.18±1.32**	**64.18±1.74**	**19**

Table 2. Quantitative results with standard deviation ($\% \pm \sigma$) and execution time (seconds) on MAG.

t-SNE Visualization

Fig.3 shows the t-SNE(Van der Maaten and Hinton 2008) visualization results of BMGC (the state-of-the-art method for Multi-Relational Graph clustering) and our proposed OMC-DVM on the ACM and Yelp datasets. Compared to BMGC, OMC-DVM yields more compact and well-separated clusters with fewer outliers. This improvement benefits from two key designs in OMC-DVM : First, the dominant view mining module dynamically identifies the dominant view and aligns other views accordingly, mitigating noise from uninformative or imbalanced views; Second, the self-weighted feature fusion online clustering process unifies representation learning and clustering into a single stage, ensuring that embeddings are directly optimized for clustering objectives. These designs jointly enhance the quality and discriminability of the learned representations, resulting in clearer and more coherent cluster structures.

Ablation Study

To assess the individual contributions of each component within our model, we conduct an ablation study by comparing the full OMC-DVM framework with three of its simplified variants. Based on the results shown in Table 3, several key observations can be made. First, OMC-DVM consistently achieves superior performance compared to all of its ablated versions, which underscores the importance of each module in enhancing the overall effectiveness of the model. Among these components, the alignment loss with the dominant view $\mathcal{L}_{ADV}$ proves to be the most critical, exerting the largest impact on performance across datasets. In contrast, the alignment loss with node features $\mathcal{L}_{ANF}$ has a particularly notable effect on the Amazon dataset. This could be attributed to the relatively lower quality of graph structures in Amazon, thereby increasing the reliance on node attribute information. Finally, the clustering-oriented dominant assignment loss $\mathcal{L}_{CLU}$ plays a vital role in improving clustering accuracy, further validating the design of our joint optimization framework.

Effects of augmented Sample Generation We studied two other ways to generate positive samples in online clustering: feature masking and dropout (Gao, Yao, and Chen 2021; Zhu et al. 2025). As shown in Table 4, our feature perturbation method always gave the best results, showing it is better at creating good positive samples. This means adding random, fine-grained noise to features works better than just dropping neurons. In contrast, feature masking performed the worst, likely because it randomly hides important features.

Weight Analysis

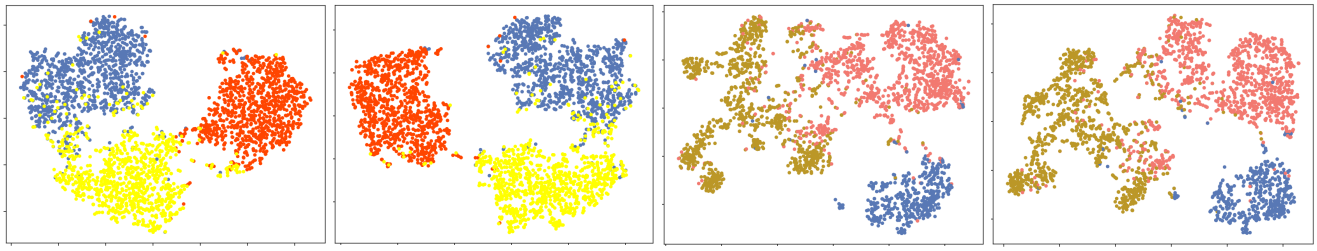
As shown in Fig. 4, on the ACM dataset, View 1 is the high-quality perspective, while View 3 is the low-quality perspective. Two key patterns emerge: (1) At the early stage of training, the model assigns distinct weights to different views at both the cluster and feature levels—high-quality views are given larger weights to emphasize their positive contribution to contrastive learning, while low-quality views are down-weighted to reduce their adverse effect on alignment. (2) As training progresses, the weights of all views gradually converge. This convergence indicates that multi-view contrastive learning effectively narrows the semantic gap among views, progressively aligning them to a consensus representation.

Convergence Analysis

We evaluated the convergence behavior of the proposed method on the ACM2 dataset by tracking the loss values and the corresponding clustering performance over training epochs. As shown in Fig.5 the total loss decreases steadily and converges within 100 epochs. This demonstrates the strong convergence property of our model, which not only ensures training stability but also reduces computational cost by achieving optimal clustering results in fewer iterations. Such efficient convergence is particularly beneficial for large-scale or time-sensitive applications.

Conclusion

In this work, we proposed OMC-DVM, an end-to-end online multi-relational graph clustering framework with dominant view mining. OMC-DVM overcomes two key limitations of

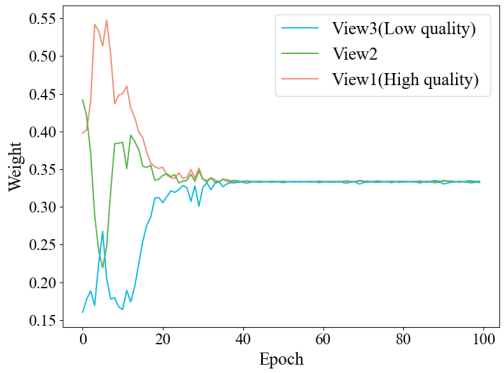


BMGC(ACM) Ours(ACM) BMGC(Yelp) Ours(Yelp)

Figure 3: We use t-SNE to visualize the cluster distribution. on ACM(left) and Yelp(right) datasets

Variants	Amazon		ACM		DBLP		ACM2		Yelp	
	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC
OMC-DVM	0.5731	0.7987	0.7887	0.9431	0.8175	0.9486	0.7406	0.9253	0.7297	0.9208
w/o $\mathcal{L}_{ADV}$	0.5332	0.7548	0.7415	0.9261	0.7773	0.9314	0.6028	0.8326	0.6853	0.8879
w/o $\mathcal{L}_{ANF}$	0.4127	0.6521	0.7371	0.9168	0.7937	0.9144	0.6812	0.8916	0.7126	0.9124
w/o $\mathcal{L}_{CLU}$	0.5594	0.7698	0.7768	0.9231	0.7657	0.9228	0.7063	0.7852	0.6946	0.9048
w/o $\mathcal{L}_{REC}$	0.5621	0.7752	0.7781	0.9342	0.7842	0.9212	0.7342	0.9037	0.7152	0.9127

Table 3. Performance of OMC-DVM and its variants.



(a) ACM

Figure 4: View weighting analysis is conducted on the ACM datasets.

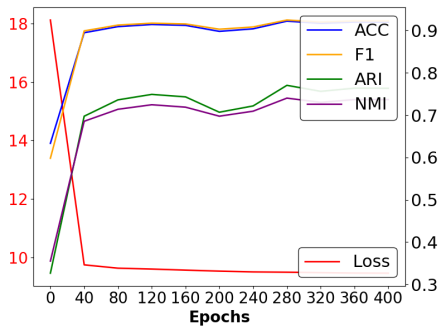


Figure 5: Convergence analysis of OMC-DVM on the ACM2 dataset.

Strategy	ACC	NMI
Feature Mask	0.7731	0.5425
Dropout	0.7892	0.5561
Feature perturbation (Ours)	**0.7987**	**0.5731**

Table 4. Performance of different positive example generation techniques on the Amazon dataset. Specifically, the feature mask strategy randomly sets part of the node features to zero, while the dropout strategy applies dropout twice to produce two representations.

prior methods—equal view weighting and decoupled representation and clustering—by integrating a dynamic dominant view mining module and a unified clustering process. The unsupervised dominant view mining mitigates view imbalance by aligning other views to the dominant one, while the joint contrastive learning enhances representations and cluster assignments simultaneously. Experiments on real-world and synthetic datasets show that OMC-DVM achieves superior clustering performance and effectively alleviates view imbalance, demonstrating its effectiveness. In the future, we aim to improve the scalability of OMC-DVM for large-scale datasets and explore its application to heterogeneous data sources.

Acknowledgments

This work was supported by the National natural science foundation of China (No. 62571355)

References

Chao, G.; Jiang, Y.; and Chu, D. 2024. Incomplete contrastive multi-view clustering with high-confidence guiding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 11221–11229.

Chen, J.; Mao, H.; Woo, W. L.; and Peng, X. 2023. Deep multiview clustering by contrasting cluster assignments. In *Proceedings of the IEEE/CVF international conference on computer vision*, 16752–16761.

Chen, Y.; Du, L.; Zhou, P.; Duan, L.; and Qian, Y. 2024. Multiple kernel clustering with local kernel reconstruction and global heat diffusion. *Information Fusion*, 105: 102219.

Chen, Y.; Jiang, B.; Zhou, P.; Duan, L.; Qian, Y.; and Du, L. 2025. Balanced Multiple Kernel Clustering with Discrete Partition Entropy Auto Regularization. In *Proceedings of the 33rd ACM International Conference on Multimedia*, 2197–2206.

Chien, E.; Peng, J.; Li, P.; and Milenkovic, O. 2020. Adaptive universal generalized pagerank graph neural network. *arXiv preprint arXiv:2006.07988*.

Chien, E.; Peng, J.; Li, P.; and Milenkovic, O. 2021. ADAPTIVE UNIVERSAL GENERALIZED PAGERANK GRAPH NEURAL NETWORK. In *9th International Conference on Learning Representations, ICLR 2021*.

Fan, S.; Wang, X.; Shi, C.; Lu, E.; Lin, K.; and Wang, B. 2020. One2multi graph autoencoder for multi-view graph clustering. In *proceedings of the web conference 2020*, 3070–3076.

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

Gao, T.; Yao, X.; and Chen, D. 2021. SimCSE: Simple Contrastive Learning of Sentence Embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 6894–6910.

Gretton, A.; Borgwardt, K.; Rasch, M.; Schölkopf, B.; and Smola, A. 2006. A kernel method for the two-sample-problem. *Advances in neural information processing systems*, 19.

Han, Z.; Zhang, C.; Fu, H.; and Zhou, J. T. 2022. Trusted multi-view classification with dynamic evidential fusion. *IEEE transactions on pattern analysis and machine intelligence*, 45(2): 2551–2566.

Hartigan, J. A.; and Wong, M. A. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1): 100–108.

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

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

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

Li, Y.; Hu, P.; Peng, D.; Lv, J.; Fan, J.; and Peng, X. 2023. Image clustering with external guidance. *arXiv preprint arXiv:2310.11989*.

Li, Y.; Li, H.; Lin, Y.; Zhang, D.; Peng, D.; Liu, X.; Xie, J.; Hu, P.; Chen, L.; Luo, H.; et al. 2025. MetaQ: fast, scalable and accurate metacell inference via single-cell quantization. *Nature Communications*, 16(1): 1205.

Li, Y.; Yang, M.; Peng, D.; Li, T.; Huang, J.; and Peng, X. 2022. Twin contrastive learning for online clustering. *International Journal of Computer Vision*, 130(9): 2205–2221.

Lin, Z.; Kang, Z.; Zhang, L.; and Tian, L. 2021. Multi-view attributed graph clustering. *IEEE Transactions on knowledge and data engineering*, 35(2): 1872–1880.

Ling, Y.; Chen, J.; Ren, Y.; Pu, X.; Xu, J.; Zhu, X.; and He, L. 2023. Dual label-guided graph refinement for multi-view graph clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 8791–8798.

Liu, H.; Hu, P.; Zhang, C.; Li, Y.; and Peng, X. 2024. Interactive deep clustering via value mining. *Advances in Neural Information Processing Systems*, 37: 42369–42387.

Liu, L.; Kang, Z.; Ruan, J.; and He, X. 2022a. Multilayer graph contrastive clustering network. *Information Sciences*, 613: 256–267.

Liu, X.; Zhu, X.; Li, M.; Wang, L.; Tang, C.; Yin, J.; Shen, D.; Wang, H.; and Gao, W. 2018. Late fusion incomplete multi-view clustering. *IEEE transactions on pattern analysis and machine intelligence*, 41(10): 2410–2423.

Liu, Y.; Tu, W.; Zhou, S.; Liu, X.; Song, L.; Yang, X.; and Zhu, E. 2022b. Deep graph clustering via dual correlation reduction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, 7603–7611.

Lu, Y.; Shi, C.; Hu, L.; and Liu, Z. 2019. Relation structure-aware heterogeneous information network embedding. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 4456–4463.

Ma, X.; Wang, Y.; Yi, S.; Ju, W.; Wu, B.; Qiao, Z.; Tang, C.; and Lv, J. 2025. PALA: Class-imbalanced graph domain adaptation via prototype-anchored learning and alignment. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, 3198–3207.

Mo, Y.; Chen, Y.; Lei, Y.; Peng, L.; Shi, X.; Yuan, C.; and Zhu, X. 2023a. Multiplex graph representation learning via dual correlation reduction. *IEEE Transactions on Knowledge and Data Engineering*, 35(12): 12814–12827.

Mo, Y.; Lei, Y.; Shen, J.; Shi, X.; Shen, H. T.; and Zhu, X. 2023b. Disentangled multiplex graph representation learning. In *International conference on machine learning*, 24983–25005. PMLR.

Pan, E.; and Kang, Z. 2021. Multi-view contrastive graph clustering. *Advances in neural information processing systems*, 34: 2148–2159.

Pan, E.; and Kang, Z. 2023. High-order multi-view clustering for generic data. *Information Fusion*, 100: 101947.

Park, C.; Kim, D.; Han, J.; and Yu, H. 2020. Unsupervised attributed multiplex network embedding. In *Proceedings of*

the AAAI conference on artificial intelligence, volume 34, 5371–5378.

Peng, L.; Wang, X.; and Zhu, X. 2023. Unsupervised multiplex graph learning with complementary and consistent information. In *Proceedings of the 31st ACM International Conference on Multimedia*, 454–462.

Qian, X.; Li, B.; and Kang, Z. 2024. Upper bounding barlow twins: A novel filter for multi-relational clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, 14660–14668.

Qu, M.; Tang, J.; Shang, J.; Ren, X.; Zhang, M.; and Han, J. 2017. An attention-based collaboration framework for multi-view network representation learning. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 1767–1776.

Sadikaj, Y.; Rass, J.; Velaj, Y.; and Plant, C. 2023. Semi-supervised embedding of attributed multiplex networks. In *Proceedings of the ACM Web Conference 2023*, 578–587.

Shen, Z.; He, H.; and Kang, Z. 2024. Balanced multi-relational graph clustering. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 4120–4128.

Sun, Y.; Li, Y.; Ren, Z.; Duan, G.; Peng, D.; and Hu, P. 2025. ROLL: Robust Noisy Pseudo-label Learning for Multi-View Clustering with Noisy Correspondence. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 30732–30741.

Tu, W.; Zhou, S.; Liu, X.; Guo, X.; Cai, Z.; Zhu, E.; and Cheng, J. 2021. Deep fusion clustering network. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 9978–9987.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).

Veličković, P.; Fedus, W.; Hamilton, W. L.; Liò, P.; Bengio, Y.; and Hjelm, R. D. 2018. Deep graph infomax. *arXiv preprint arXiv:1809.10341*.

Wang, K.; Shen, Z.; Huang, C.; Wu, C.-H.; Dong, Y.; and Kanakia, A. 2020. Microsoft academic graph: When experts are not enough. *Quantitative Science Studies*, 1(1): 396–413.

Wang, W.; Tran, D.; and Feiszli, M. 2020. What makes training multi-modal classification networks hard? In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12695–12705.

Wang, X.; Ji, H.; Shi, C.; Wang, B.; Ye, Y.; Cui, P.; and Yu, P. S. 2019. Heterogeneous graph attention network. In *The world wide web conference*, 2022–2032.

Wang, X.; Zhang, Y.; and Zhou, Y. 2025a. Bidirectional Probabilistic Multi-graph Learning and Decomposition for Multi-view Clustering. *IEEE Transactions on Image Processing*, 34: 3609–3621.

Wang, X.; Zhang, Y.; and Zhou, Y. 2025b. Pseudo-Supervision Affinity Propagation for Efficient and Scalable Multiview Clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 36(8): 15282–15293.

Wu, L.; Chen, Y.; Shen, K.; Guo, X.; Gao, H.; Li, S.; Pei, J.; Long, B.; et al. 2023. Graph neural networks for natural language processing: A survey. *Foundations and Trends® in Machine Learning*, 16(2): 119–328.

Wu, S.; Zheng, Y.; Ren, Y.; He, J.; Pu, X.; Huang, S.; Hao, Z.; and He, L. 2024. Self-weighted contrastive fusion for deep multi-view clustering. *IEEE Transactions on Multimedia*.

Yang, R.; Hu, P.; Peng, X.; Liu, X.; and Li, Y. 2024. DiFiC: Your Diffusion Model Holds the Secret to Fine-Grained Clustering. *arXiv preprint arXiv:2412.18838*.

Ye, Y.; Zhang, Y.; Peng, L.; Li, R.; Liu, C.; Wu, S.; and Wong, H.-S. 2025. Cross-View Neighborhood Contrastive Multi-View Clustering with View Mixup Feature Learning. In *2025 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6.

Zhang, D.; Xiao, W.; Zhu, H.; Ma, X.; and Arnold, A. 2022. Virtual Augmentation Supported Contrastive Learning of Sentence Representations. In *Findings of the Association for Computational Linguistics: ACL 2022*, 864–876.

Zhao, J.; Wang, X.; Shi, C.; Liu, Z.; and Ye, Y. 2020. Network schema preserving heterogeneous information network embedding. In *International joint conference on artificial intelligence (IJCAI)*.

Zhou, P.; Chen, J.; Du, L.; and Li, X. 2022. Balanced spectral feature selection. *IEEE Transactions on Cybernetics*, 53(7): 4232–4244.

Zhou, P.; Li, R.; Ling, Z.; Du, L.; and Liu, X. 2024. Fair clustering ensemble with equal cluster capacity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhu, Z.; Zhou, P.; Li, Z.; Chen, K.; and Zhu, J. 2025. Multi-label text classification with label attention aware and correlation aware contrastive learning. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, 8420–8428.