Multi-view Clustering via Multi-granularity Ensemble

Jie Yang[1] , **Wei Chen**[2] , **Feng Liu**[3] , **Peng Zhou**[4] , **Zhongli Wang**[5] , **Xinyan Liang**[6*] , **Bingbing Jiang**[5*]

[1]University of Technology Sydney, NSW, Australia
[2]The University of Sydney, NSW, Australia
[3]The University of Melbourne, VIC, Australia
[4]Anhui University, Hefei, China
[5]Hangzhou Normal University, Hangzhou, China
[6]Shanxi University, Taiyuan, China

jie.yang.uts@gmail.com, wei.chenbme@sydney.edu.au, fengliu.ml@gmail.com, zhoupeng@ahu.edu.cn, wangzhongli1@stu.hznu.edu.cn, liangxinyan48@163.com, jiangbb@hznu.edu.cn

Abstract

Multi-view clustering aims to integrate complementary information from multiple views to improve clustering performance. However, existing ensemble-based methods suffer from information loss due to their reliance on single-granularity labels, limiting the discriminative capability of learned representations. Meanwhile, representation and graph fusion-based approaches face challenges such as explicit view alignment and manual weight tuning, making them less effective for heterogeneous views with varying data distributions. To address these limitations, we propose a novel multi-view clustering framework via Multi-granularity Ensemble (MGE), fully using the multi-granularity information across diverse views for accurate and consistent clustering. Specifically, MGE first modifies the hierarchical clustering and then leverages it on each view (including the fused view) to achieve multi-granularity labels. Moreover, the cross-view and cross-granularity fusion strategy is designed to learn a robust co-association similarity matrix, which effectively preserves the fine-grained and coarse-grained structures of multi-view data and facilitates subsequent clustering. Therefore, MGE can provide a comprehensive representation of local and global patterns within data, eliminating the requirement for view alignment and weight tuning. Experiments demonstrate that MGE consistently outperforms state-of-the-art methods across multiple datasets, validating its effectiveness and superiority in handling heterogeneous views.

1 Introduction

Multi-view clustering integrates complementary information from various modalities (e.g., images, texts, sensor readings) to produce more accurate and robust clustering [Chao *et al.*, 2021; Jiang *et al.*, 2025b]. However, effectively leveraging multi-view information remains challenging due to inconsistencies in data quality, feature distributions, structural heterogeneity among views, and issues such as missing or redundant views [Jiang *et al.*, 2025a]. Multi-view clustering can be broadly categorized based on their fusion strategies: (1) representation and graph-based fusion methods [Xu *et al.*, 2023; Wang *et al.*, 2024b; Huang *et al.*, 2022; Huang *et al.*, 2024], which construct a unified feature representation or similarity graph for clustering (commonly referred to as early fusion), and (2) ensemble-based methods [Pfeifer *et al.*, 2023; Zheng *et al.*, 2024; Xu *et al.*, 2024], which combine clustering labels from different views to form a consensus result (referred to as late fusion). Some hybrid approaches combine feature-level and label-level integration for improved performance. Despite substantial progress, these categories face challenges such as information loss due to reliance on single-granularity labels [Wang *et al.*, 2025], sensitivity to heterogeneous view inconsistencies, and the need for manual tuning of view weights[Zhang *et al.*, 2024a; Zhang *et al.*, 2024b; Sun *et al.*, 2025].

Representation and graph-based fusion methods aim to learn a shared representation or construct a consensus graph by aggregating information from multiple views [Chen *et al.*, 2024b; Wen *et al.*, 2024; Chen *et al.*, 2024a; Huang *et al.*, 2023b]. For instance, the consensus graph learning framework builds a robust consensus similarity graph by integrating spectral embeddings with a weighted tensor-based low-rank representation [Li *et al.*, 2021]. The summarized multi-view clustering approach reduces redundancy and enhances inter-view consistency by leveraging an information-theoretic variational lower bound [Cui *et al.*, 2024]. Similarly, the robust multi-view clustering with noisy correspondence method employs a noise-tolerant contrastive loss to learn embeddings that remain robust even with misaligned views [Sun *et al.*, 2024]. The multi-level feature learning framework for contrastive multi-view clustering independently learns low-level and high-level features, avoiding direct feature fusion and ensuring that private, view-specific information does not interfere with shared representations [Xu *et al.*, 2022]. However,
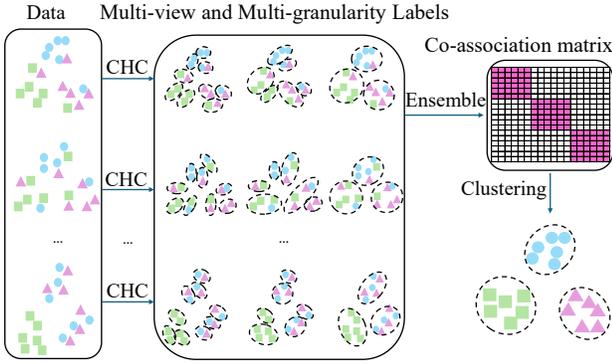
*Corresponding author.

Figure 1: The proposed MGE framework exploits the constrained hierarchical clustering (CHC) on each view to generate multi-granularity labels. These labels progressively capture information from fine-grained to coarse-grained structures (as the number of clusters decreases), guiding the cross-view and cross-granularity fusion to learn a robust co-association matrix for clustering.

these methods often involve complex representation or graph alignment processes, which pose challenges for highly heterogeneous views, and they rely heavily on hyperparameter tuning, such as adjusting view weights and embedding dimensions, to achieve optimal performance [Wu *et al.*, 2024; Lou *et al.*, 2024; Liang *et al.*, 2024; Wu *et al.*, 2025].

Ensemble-based multi-view clustering methods aim to combine the strengths of multiple base clustering results to achieve robust and consistent consensus outcomes [Liang *et al.*, 2025]. For example, the Parea hierarchical clustering ensemble framework leverages late-stage fusion to integrate clustering solutions from heterogeneous biomedical datasets, enhancing disease subtype discovery [Pfeifer *et al.*, 2023]. The low-rank and sparse decomposition approach formulates ensemble clustering as a tensor decomposition problem, capturing high-order correlations across views [Zhang *et al.*, 2023]. Similarly, the hybrid multi-view clustering ensemble method employs diverse view transformations and hybrid subspace learning to enhance the diversity of base clusterings [Yu *et al.*, 2020]. The Fast Multi-view Ensemble Clustering (FMVEC) approach introduces a hybrid early-late fusion strategy with random view groups, achieving near-linear time complexity for large-scale datasets [Huang *et al.*, 2023a]. However, most ensemble-based methods rely on single-granularity clustering labels, which often fail to capture critical hierarchical structures and limit their ability to learn more discriminative representations. Additionally, directly aggregating clustering labels into a final consensus label set can amplify errors, particularly when individual view-specific labels are noisy, undermining their robustness and effectiveness [Wang *et al.*, 2024a].

In this paper, we propose a novel multi-view clustering framework called Multi-granularity Ensemble (MGE), which addresses the limitations of both representation/graph fusion-based and ensemble-based methods by incorporating multi-granularity clustering labels and constructing a co-association similarity matrix that encodes rich multi-view and multi-level information. Specifically, a constrained hierarchical clustering is introduced and exploited to each view and their

fused view to generate multi-granularity labels that capture both fine-grained and coarse-grained cluster structures. The cross-view and cross-granularity fusion is then performed on these labels to construct a co-association matrix, encoding discriminative representations across local and global, multi-granularity, and multi-view levels. As a result, MGE effectively avoids the challenges of heterogeneous view alignment and view-weight hyperparameter tuning that are typically required in representation/graph fusion-based methods. Moreover, it mitigates the impact of noisy labels from individual views by avoiding direct label aggregation and instead employing secondary clustering on the collaborative representation. Compared to existing ensemble-based methods, MGE captures richer local and global clustering structures by integrating multi-granularity information, ensuring more discriminative representation learning. Figure 1 illustrates the basic framework of MGE and the main contributions of this paper are summarized as follows:

- We propose a novel multi-view clustering framework, i.e., Multi-Granularity Ensemble (MGE), which effectively addresses the inherent information loss caused by reliance on single-granularity labels in existing methods, achieving more robust and accurate clustering results.

- MGE tackles the challenge of heterogeneous views existing in multi-view data, which effectively explore the structure information from varying data distributions without requiring explicit view alignment, enabling more accurate and robust clustering.

- We enhance representation learning via the cross-view and cross-granularity fusion that seamlessly integrates fine-grained and coarse-grained clustering structures, such that the local and global patterns across multiple views are simultaneously considered, further improving the discrimination and robustness of representations.

2 The Multi-granularity Ensemble Framework

2.1 Multi-view Multi-granularity Label Generation

To comprehensively capture fine-grained and coarse-grained clustering structures across multiple views, the proposed MGE framework introduced the Constrained Hierarchical Clustering (CHC) [Yang and Lin, 2024] to accommodate multi-view data, generating multi-granularity labels for each view and the fused view. Let $\mathbf{X} = \{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_v\}$ be the multi-view dataset, where $\mathbf{X}_i$ represents the i-th view, and v denotes the number of views. The fused view $\mathbf{X}_f$ is defined based on the average of the similarity matrices $\mathbf{S}_i$ derived from each view:

$$\mathbf{S}_f = \frac{1}{v} \sum_{i=1}^{v} \mathbf{S}_i, \tag{1}$$

where $\mathbf{S}_i$ is the similarity matrix corresponding to the i-th view. The matrix S_f serves as the similarity matrix for the fused view X_f, preserving cross-view consensus information.

To generate multi-granularity label sets for each view $\mathbf{X}_i$ (including $\mathbf{X}_f$), CHC is applied to produce $\mathcal{L}_i = \{L_{i,1}, L_{i,2}, \ldots, L_{i,k_i}\}$, where $L_{i,j}$ represents the clustering
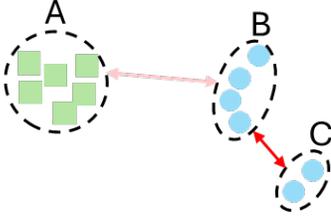
Figure 2: Constraints in CHC facilitate the generation of high-purity multi-granularity labels.

labels at the j-th granularity level for view i. The labels are ordered such that $L_{i,1}$ corresponds to the finest partition (with the most clusters), and L_{i,k_i} represents the coarsest partition (with fewer clusters). The total number of clustering levels k_i is determined based on a granularity termination parameter λ (discussed below).

Constrained Hierarchical Clustering
Hierarchical clustering (HC) constructs dendrograms to generate multi-granularity labels, enabling flexible exploration of data structures. However, traditional HC methods, such as single, complete, and average linkage, rely solely on 1-nearest neighbor statistics, often merging sub-clusters from different categories, leading to low-purity multi-granularity labels. CHC addresses these limitations by introducing adjacency constraints, prioritizing sub-clusters that align with ground-truth structures, and using larger clusters to guide smaller ones. As illustrated in Figure 2, for the current clusters A, B, and C, a parallel merge based on 1-nearest neighbor relationships would result in A merging with B, and B merging with C. In contrast, CHC applies the constraint that the size of A is larger than that of B, no merge them; whereas the size of C is smaller than B, merge them, avoiding A merging with B while allowing B merging with C. This perfectly matches the ground truth and significantly improves the purity of the merging process at each iteration.

For each view $\mathbf{X}_i$ (including $\mathbf{X}_f$), CHC initializes each sample as its cluster. A graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ is constructed, where $\mathbf{V}$ represents the set of clusters (i.e., nodes), and an edge $(C_i, C_j) \in \mathbf{E}$ exists if cluster C_j is the 1-nearest neighbor (denoting as $1NN$) of cluster C_i and $|C_i| \leq |C_j|$. The adjacency matrix $\mathbf{A}$ is defined as:

$$A_{ij} = \boldsymbol{I}(1NN(C_i) = C_j \cap |C_i| \leq |C_j|) \qquad (2)$$

where $\boldsymbol{I}(\cdot)$ is the indicator function that returns 1 if the condition is true and 0 otherwise. The connected components of the graph $\mathbf{G}$ after each iteration represent the clusters at a specific granularity level. By iteratively merging clusters based on adjacency constraints, CHC produces hierarchical partitions corresponding to various levels of granularity.

When a target cluster number K is specified, CHC constructs a hierarchy and removes the $K-1$ strongest edges from the graph $\mathbf{G}$. The weight of each edge is computed as:

$$w(C_i, C_j) = d^2(C_i, C_j) \times |C_i| \times |C_j|, \qquad (3)$$

where $d^2(C_i, C_j)$ denotes the squared distance between clusters C_i and C_j, and $|C_i|$ represents the sizes of the clusters.

The $K-1$ edges with the highest weights are removed to disconnect the graph into K connected components, resulting in precisely K clusters.

Granularity Termination Hyperparameter λ
For each view $\mathbf{X}_i$ (including $\mathbf{X}_f$), CHC generates m_i clustering partitions ranging from fine-grained to coarse-grained labels. The granularity termination parameter k_i for view i determines the number of labels selected from the hierarchy:

$$k_i = \max\left(1, \lfloor \lambda m_i + 0.5 \rfloor\right), \qquad (4)$$

where $\lambda \in (0, 1]$ is a shared hyperparameter across all views, ensuring consistency in the number of clustering levels. Here, $\lfloor \cdot \rfloor$ denotes the rounding operation to the nearest integer. A smaller λ results in fewer clustering partitions (i.e., lower k_i), corresponding to a coarser analysis. Conversely, a larger λ produces more clustering partitions (i.e., higher k_i), corresponding to a finer breakdown with more detailed clustering levels.

By performing CHC to each view $\mathbf{X}_i$ and the fused view $\mathbf{X}_f$, the proposed framework generates multi-granularity label sets $\mathcal{L}_i = \{L_{i,1}, L_{i,2}, \ldots, L_{i,k_i}\}$, where L_{i,k_i} corresponds to the clustering partition at the granularity level determined by k_i. This ensures that both fine-grained and coarse-grained clustering structures are captured, enabling richer multi-view and multi-level representations for the subsequent integration process.

2.2 Cross-view and Cross-granularity Fusion
Previous ensemble methods, such as the Ensemble Learning via Propagation of Cluster-wise Similarities (ELPCS) [Huang *et al.*, 2018], are restricted to single-view contexts, integrating labels of the same granularity from different algorithms. This design overlooks the complementary information inherent in multiple views and the enriched discriminative power provided by multi-granularity labels, thereby limiting clustering performance. To overcome these limitations, we extend ELPCS to support multi-view and multi-granularity data by introducing a cross-view and cross-granularity fusion strategy. This strategy facilitates a seamless global interaction between views and granularities, synergizing complementary insights and hierarchical structures to extract the most discriminative representations. The resulting co-association matrix effectively captures intricate relationships across views and granularities, serving as a robust foundation for accurate and reliable consensus clustering.

Cluster-wise Similarity Graph Construction
Given the multi-granularity label $\mathcal{L}_i = \{L_{i,1}, \cdots, L_{i,k_i}\}$ for each view $\mathbf{X}_i$ (including the fused view $\mathbf{X}_f$), we construct a combined cluster-wise similarity graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, where: $\mathbf{V}$ represents the clusters across all views and granularities, and $\mathbf{E}$ represents edges weighted by the Jaccard coefficient:

$$J(C_p, C_q) = \frac{|C_p \cap C_q|}{|C_p \cup C_q|}, \qquad (5)$$

where C_p and C_q are clusters from different views and/or different granularities, reflecting the proportion of overlapping data points. In this combined graph, nodes correspond

Algorithm 1 MGE framework

Input: Multi-view data $\mathbf{X} = \{\mathbf{X}_1, \cdots, \mathbf{X}_v\}$, and the cluster number c, and the granularity control parameter λ;

1: **for** each view $\mathbf{X}_i$ (including the fused view $\mathbf{X}_f$);
2:     Apply CHC to $\mathbf{X}_i$ to generate the multi-granularity label set $\mathcal{L}_i$ by Eqs. (1)-(4);
3: **end for**
4: Construct the cluster-wise similarity graph using all label sets $\mathcal{L}_i$ from all views by Eq. (5);
5: Construct the transition probability matrix $\mathbf{P}$ by Eq. (6);
6: Propagate cluster-wise similarities by Eq. (7);
7: Compute the cluster-wise similarity matrix $\mathbf{Z}$ by Eq. (8);
8: Construct the co-association matrix $\mathbf{B}$ by Eq. (9);
9: Apply CHC to $\mathbf{B}$ to obtain the final clustering label set $\mathcal{L}$ by Eqs. (2) and (3);

Output: The clustering result $\mathcal{L}$.

---

to clusters from different views and granularities, and edges represent the similarity between these clusters based on the overlap of their constituent points.

Propagation of Cluster-wise Similarities

To capture both direct and indirect connections among clusters, a random walk propagation process is applied to the combined graph G. Let $\mathbf{P}$ denote the transition probability matrix for G, where each entry $p_{p,q}$ represents the probability of transitioning from cluster C_p to C_q in one step:

$$p_{p,q} = \frac{e_{p,q}}{\sum_{r \in \mathbf{V}} e_{p,r}}, \tag{6}$$

where $e_{p,q}$ is the Jaccard similarity between clusters C_p and C_q. The random walk propagation matrix $\mathbf{P}^{(t)}$ after t steps captures indirect relationships:

$$\mathbf{P}^{(t)} = (\mathbf{P})^t. \tag{7}$$

The cluster-wise similarity matrix $\mathbf{Z}$ is computed by taking the cosine similarity of the random walk trajectories:

$$z_{p,q} = \frac{\langle \mathbf{P}^{(1:t)}(p,:), \mathbf{P}^{(1:t)}(q,:) \rangle}{\|\mathbf{P}^{(1:t)}(p,:)\| \cdot \|\mathbf{P}^{(1:t)}(q,:)\|}, \tag{8}$$

where $\mathbf{P}^{(1:t)}(p,:)$ represents the trajectory of cluster C_p during the propagation process. The value of t is chosen to ensure clustering stability and efficiency, guided by practices commonly observed in related works.

Construction of the Co-association Matrix

To map the cluster-level similarities back to the object level, we construct the co-association matrix $\mathbf{B}$ as follows:

$$b_{x,y} = \frac{1}{v+1} \sum_{i=1}^{v+1} \frac{1}{k_i} \sum_{j=1}^{k_i} z_{p,q}^{(j)}, \tag{9}$$

where v is the number of original views, and "+1" accounts for the fused view $\mathbf{X}_f$, k_i is the number of granularities in the i-th view. $z_{p,q}^{(j)}$ represents the propagated similarity between clusters C_p and C_q at the j-th granularity level of the i-th

Dataset	Classes	Data size	Feature size
100Leaves	100	1600	192(64/64/64)
UCI	10	2000	356(76/216/64)
COIL20	20	1440	11078(1024/3304/6750)
Handwritten	10	2000	316(76/240)
CMU-PIE	68	2856	90(30/30/30)
ORL	40	400	1689(512/59/864/254)

Table 1: The detailed information on multi-view datasets.

view. $x \in C_p$ and $y \in C_q$ indicate that data points x and y belong to these clusters.

Figure 1 illustrates the proposed MGE framework. For a multi-view dataset, the CHC is applied to each view (including the fused view) to generate multi-view, multi-granularity labels. These labels are then integrated through an ensemble method that performs cross-view and cross-granularity fusion, resulting in the co-association similarity matrix $\mathbf{B}$. As a result, the co-association matrix $\mathbf{B}$ averages the cluster-wise similarities from all views and all granularity levels, capturing direct and indirect object-level relationships across all views and granularities, and it serves as input for the secondary clustering to generate the final clustering results. Algorithm 1 presents the procedures of the MGE framework.

2.3 Computational Complexity Analysis

The computational complexity of MGE is primarily determined by three stages: multi-view multi-granularity label generation, cross-view and cross-granularity fusion, and secondary clustering. In the first stage, applying CHC to each view (including the fused view) incurs a complexity of $\mathcal{O}(n^2)$, where v is the number of views and n is the number of data points. In the second stage, constructing the cluster-wise similarity graph requires computing pairwise Jaccard coefficients between clusters, resulting in a complexity of $\mathcal{O}(K_{\text{total}}^2)$, where K_{total} is the total number of clusters across all views and granularities. The subsequent random walk propagation step involves matrix multiplications with a complexity of $\mathcal{O}(tK_{\text{total}}^2)$, where t is the number of propagation steps. Finally, the secondary clustering step applies CHC to the co-association matrix, contributing an additional complexity of $\mathcal{O}(n^2)$. Therefore, the computational complexity of MGE is $\mathcal{O}(n^2 + tK_{\text{total}}^2)$. By utilizing approximate methods such as kd-tree to accelerate the 1-nearest neighbor search in CHC, the complexity can be further reduced to $\mathcal{O}(n \log n + tK_{\text{total}}^2)$.

3 Experiments

In this section, we present the experimental studies of the proposed MGE on a synthetic data and six real-world datasets, in which three views of the synthetic data are shown in Figures 3 (a)-(c), and the detailed information of real-world multi-view datasets are reported in Table 1.

3.1 Experimental Settings

MGE is compared with the state-of-the-art competitors, including three ensemble-based methods: Fast Multi-View Ensemble Clustering (FMVEC) [Huang *et al.*, 2023a], Matrix Multi-View Ensemble Clustering (MMEC) [Zhang *et al.*,

	Methods	100Leaves	UCI	COIL20	Handwritten	CMU-PIE	ORL	Average
	GMC	0.8237	0.8495	0.7910	0.8300	0.7048	0.6325	0.7719
	TMMSC	0.8356	0.9024	0.8042	0.9105	0.7953	0.7825	0.8384
	V3H	0.8237	0.9051	0.6012	0.8669	0.7231	0.7412	0.7769
	LMVSC	0.6575	0.8935	0.7569	0.9005	0.4769	0.6300	0.7192
	AWP	0.7856	0.8670	0.6757	<u>0.9325</u>	<u>0.8120</u>	0.6900	0.7938
ACC	CoReg	0.8456	<u>0.9560</u>	<u>0.8472</u>	0.9110	0.7507	0.8200	<u>0.8550</u>
	WMSC	<u>0.8769</u>	0.8410	0.8465	0.8335	0.6590	<u>0.8300</u>	0.8145
	MMEC	0.6770	0.6996	0.6494	0.8578	0.3548	0.5860	0.6374
	FMVEC	0.7981	0.7770	0.7979	0.8760	0.6604	0.7675	0.7794
	CDEC	0.5162	0.7490	0.7208	0.8155	0.5014	0.5650	0.6447
	MGE	**0.9481**	**0.9725**	**1.0000**	**0.9825**	**0.9783**	**0.8525**	**0.9557**
	GMC	0.9296	0.9013	0.9410	0.8767	0.8892	0.8590	0.8995
	TMMSC	0.9248	0.8885	0.9190	<u>0.9190</u>	0.9072	0.7800	0.8898
	V3H	0.9096	0.8118	0.7639	0.7425	0.8667	0.8633	0.8263
	LMVSC	0.8504	0.8321	0.8404	0.8366	0.6916	0.8246	0.8127
	AWP	0.8968	0.8949	0.9148	0.9072	<u>0.9296</u>	0.8529	0.8989
NMI	CoReg	0.9346	<u>0.9188</u>	<u>0.9548</u>	0.8811	0.8791	0.9011	<u>0.9116</u>
	WMSC	<u>0.9481</u>	0.8839	0.9486	0.8772	0.8571	0.8985	0.9022
	MMEC	0.8939	0.7177	0.8001	0.7814	0.7066	0.7984	0.7830
	FMVEC	0.9235	0.8894	0.9433	0.9008	0.8242	<u>0.9029</u>	0.8974
	CDEC	0.7123	0.6745	0.7754	0.7240	0.6224	0.7182	0.7045
	MGE	**0.9712**	**0.9406**	**1.0000**	**0.9604**	**0.9846**	**0.9245**	**0.9636**

Table 2: The ACC and NMI of multi-view clustering methods, where the best and second are in bold and underlined, respectively.
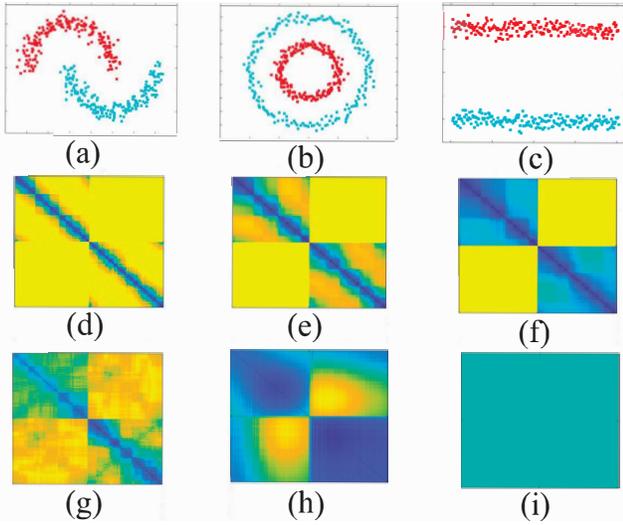

Figure 3: Results on the synthetic data. (a)-(c) depict the original data distributions with three views, (d)-(f) show the co-association matrices learned by MGE with different granularities (i.e. 3 granularities, 6 granularities and 9 granularities), (g)-(i) illustrate the results of FMVEC, MMEC, and GMC, respectively.

Figure 4: Runtime of MGE and other representative methods.

2023], and Coordinate Descent Ensemble Clustering (CDEC) [Li *et al.*, 2024], as well as seven graph/representation fusion-based methods, such as Graph-based Multi-View Clustering (GMC) [Wang *et al.*, 2019], Multi-view Subspace Clustering on Topological Manifold (TMMSC) [Huang *et al.*, 2022], View Variation and View Heredity for Multi-View Clustering (V3H) [Fang *et al.*, 2020], Large-Scale Multi-View Subspace Clustering (LMVSC) [Kang *et al.*, 2020], Multi-View Clustering via Adaptively Weighted Procrustes (AWP) [Nie
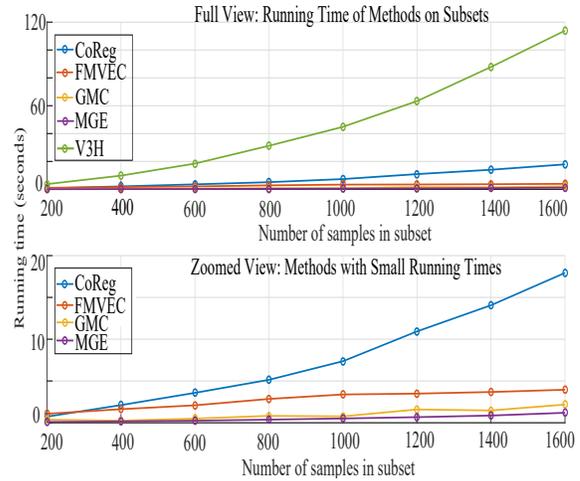
et al., 2018], Co-Regularized Multi-View Spectral Clustering (CoReg) [Kumar *et al.*, 2011], and Weighted Multi-View Spectral Clustering (WMSC) [Zong *et al.*, 2018]. We evaluate the performance using two widely used external validation metrics: Accuracy (ACC), and Normalized Mutual Information (NMI) [Strehl and Ghosh, 2002]. For all methods, parameters are tuned according to the default settings in their original publications to ensure optimal performance.

3.2 Experiments on Synthetic Dataset

To intuitively demonstrate the advantages of MGE, we generated a synthetic dataset with three heterogeneous views, as illustrated in Figures 3(a)-(c). Differing from other ensemble-based methods, MGE integrates multi-granularity labels, cap-

turing both fine-grained and coarse-grained structures to encode local and global patterns within data. This manner facilitates the learning of more discriminative ensemble representations. As shown in Figures 3(d)-(f), increasing the number of granularities (and λ is set to 0.3, 0.6, and 0.9, progressively covering granularities from fine to coarse) from three to nine progressively clarifies the clustering structures in the co-association matrices, providing a stronger foundation for subsequent clustering. Figures 3(g)-(i) further demonstrate that FMVEC and MMEC produce weaker representations due to their reliance on single-granularity label integration, while GMC fails to fuse these heterogeneous views, manifesting the limitations of graph fusion-based methods in such scenarios.

It should be pointed out that MGE can effectively leverage its advantages to enhance the clusterability of the co-association matrix even when a single view exhibits near-perfect clustering potential, as seen in View 3. By employing a constrained 1NN merging strategy, the multi-granularity labels generated by MGE accurately capture proximity relationships across both local and global scales. In contrast, the single-granularity methods (e.g., FMVEC and MMEC) and the fusion-based methods fail to utilize the multi-layered, near-ground-truth information provided by such views, resulting in co-association matrices with significantly diminished discrimination.

3.3 Experiments on Real-world Datasets

The results of the proposed MGE and other state-of-the-art methods across six datasets are presented in Table 2, from which we can find that MGE achieves the best performance across all datasets on the ACC and NMI, Specifically, MGE achieves improvements in the average ACC of 17.62% and 10.07% over the best ensemble-based method (i.e., FMVEC) and the best graph fusion-based (i.e., CoReg), respectively. For the average NMI metric, MGE likewise achieves 7.38% and 5.70% improvements over the best ensemble-based method (i.e., FMVEC) and the best graph fusion-based (i.e., CoReg), respectively. The superior performance of MGE can be attributed to its innovative clustering mechanism that addresses key limitations of existing methods. Unlike ensemble-based approaches like FMVEC and MMEC, which rely on single-granularity label aggregation and are sensitive to noisy labels, MGE employs multi-granularity clustering to preserve fine-grained and coarse-grained structures. Its secondary clustering stage further mitigates the impact of noisy labels, ensuring robust results. Compared to graph fusion-based methods like GMC and CoReg, which require explicit view alignment and extensive parameter tuning, MGE leverages multi-granularity information for consistent and scalable clustering across heterogeneous views. Additionally, MGE surpasses representation fusion-based methods like TMMSC and LMVSC by utilizing clustering labels and encoding both local and global patterns, delivering comprehensive and discriminative representations.

Figure 4 shows the runtime of MGE with several representative methods on the 100Leaves dataset, in which the number of samples varies from 200 to 1600. We note that MGE consistently exhibits the shortest runtime across different data sizes, attributed to its efficient two-stage process, i.e.,

generating multi-granularity clustering labels first and then clustering on the co-association matrix. This manner minimizes computational overhead by avoiding iterative optimization and graph alignment, which demonstrates the scalability of MGE and positions it as an effective solution for relatively large-scale clustering tasks.

3.4 Visualization

To demonstrate the advantages of the cross-view and cross-granularity fusion method in MGE, we compare its co-association similarity matrix with those generated by two ensemble-based methods, MMEC and FMVEC, on the 100Leaves dataset. As shown in Figure 5, the first row presents the similarity matrices of the original three views, and the second row displays the co-association matrices learned by MGE, MMEC, and FMVEC. Specifically, the matrix of MGE exhibits a much clearer and more coherent clustering structure than those of MMEC and FMVEC. Moreover, the matrix of MMEC shows indistinct clustering patterns with significant noises, while the matrix of FMVEC, though clearer than MMEC, likewise suffers from low intra-cluster similarity (indicated by lighter diagonal regions), resulting in less compact clusters. This indicates that making full use of multi-granularity information facilitates capturing both fine-grained and coarse-grained structures as well as learning a more comprehensive representation for clustering.

3.5 Ablation Study

The MGE framework involves two clustering processes: the first generates multi-view, multi-granularity clustering labels, and the second performs secondary clustering on the co-association matrix to produce the final result. In both stages, CHC is employed as the clustering method. To assess the impact of this choice, we replaced CHC with other hierarchical clustering algorithms, including single-linkage and average-linkage, and named the corresponding variants MGE-1 and MGE-2, respectively. As shown in Figure 6, the ACC scores of MGE-1 and MGE-2 across three datasets were significantly lower than that of MGE, indicating the importance of using CHC in both stages of the proposed MGE framework.

3.6 Hyperparameter Sensitivity Analysis

The MGE framework contains only one hyperparameter λ, which controls the granularity of the clustering labels generated for all views, including the fused view. Figure 7 illustrates the ACC scores of MGE on three datasets when adjusting λ within the range of [0.1:0.1:0.9]. It can be observed that setting λ around 0.5 achieves the best average performance across the three datasets. A larger λ produces an excessive number of coarse-grained multi-view labels, reducing the discriminative ability between different classes in the ground truth. Conversely, a smaller λ generates insufficient types of multi-granularity labels, leading to significant information loss and diminishing the ability of the co-association matrix to distinguish ground-truth classes.

4 Conclusion

In this paper, we introduce Multi-view Clustering via Multi-granularity Ensemble (MGE), a novel framework designed to

(a) View1

(b) View2

(c) View3

(d) Ensemble (FMVEC)

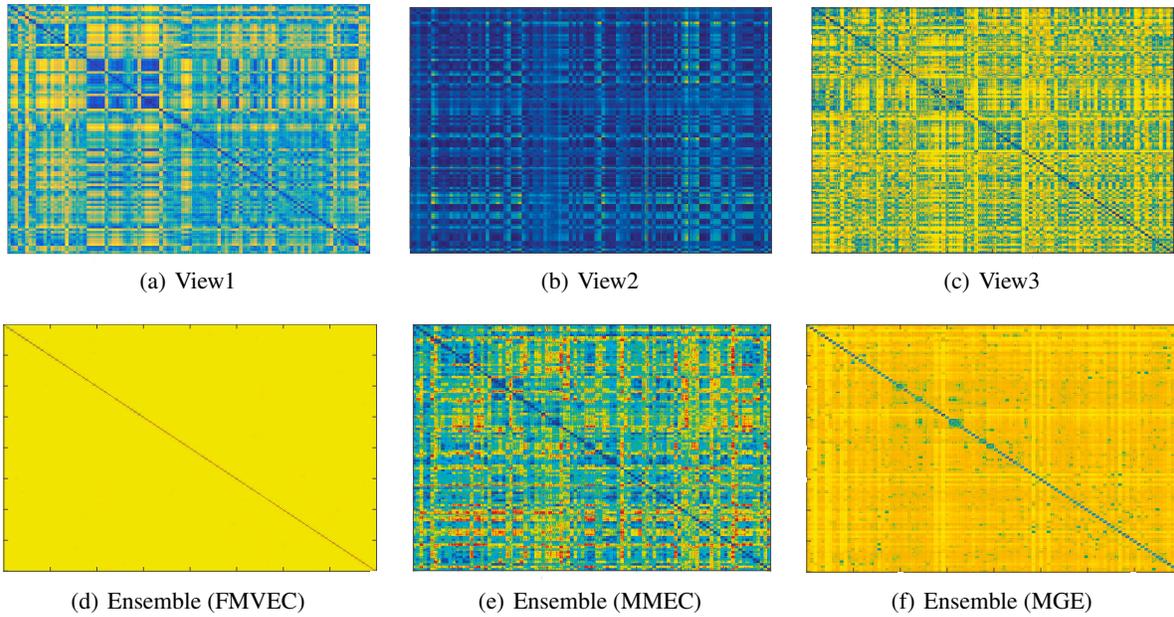(e) Ensemble (MMEC)

(f) Ensemble (MGE)

Figure 5: Visualization of the learned matrices of MGE and other ensemble-based clustering methods on the 100Leaves dataset.
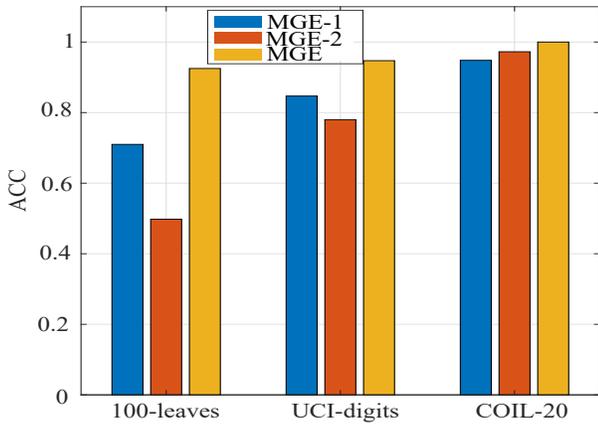

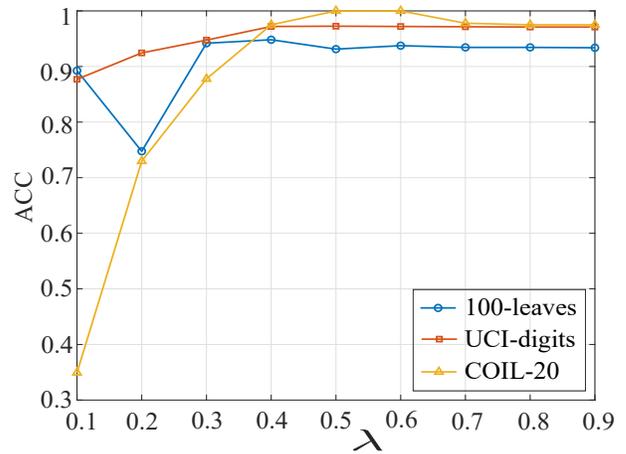Figure 6: Comparison of MGE with its variants.

Figure 7: Performance of MGE with different λ values.

address the challenges posed by view-specific inconsistencies and single-granularity limitations in ensemble-based multi-view clustering methods. The proposed framework leverages the CHC to generate multi-granularity clustering labels for each view, including the fused view, capturing both fine-grained and coarse-grained cluster structures. By performing cross-view and cross-granularity fusion, we constructed a co-association similarity matrix that integrates local and global clustering patterns across views, enabling robust secondary clustering. Extensive experiments conducted on a synthetic dataset with heterogeneous views and real-world multi-view datasets demonstrated the superior performance of MGE compared to state-of-the-art multi-view clustering algorithms across key evaluation metrics (ACC and NMI). Notably, MGE consistently achieved the highest clustering accuracy, highlighting its resilience to noisy or low-quality views and its ability to provide comprehensive representations of complex data structures. In future work, we aim to extend MGE to scenarios with incomplete or missing views and explore its scalability for large-scale, high-dimensional datasets. Additionally, we plan to investigate adaptive granularity control mechanisms to further enhance its generalization across various domains.

Acknowledgments

References

[Chao *et al.*, 2021] Guoqing Chao, Shiliang Sun, and Jinbo Bi. A survey on multiview clustering. *IEEE Transactions on Artificial Intelligence*, 2(2):146–168, 2021.

[Chen *et al.*, 2024a] Man-Sheng Chen, Zhi-Yuan Li, Jia-Qi Lin, Chang-Dong Wang, and Dong Huang. Sparse graph tensor learning for multi-view spectral clustering. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(5):3534–3543, 2024.

[Chen *et al.*, 2024b] Mulin Chen, Bocheng Wang, and Xuelong Li. Deep contrastive graph learning with clustering-oriented guidance. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 11364–11372, 2024.

[Cui *et al.*, 2024] Chenhang Cui, Yazhou Ren, Jingyu Pu, Jiawei Li, Xiaorong Pu, Tianyi Wu, Yutao Shi, and Lifang He. A novel approach for effective multi-view clustering with information-theoretic perspective. *Advances in Neural Information Processing Systems*, 36:44847–44859, 2024.

[Fang *et al.*, 2020] Xiang Fang, Yuchong Hu, Pan Zhou, and Dapeng Oliver Wu. V3h: View variation and view heredity for incomplete multiview clustering. *IEEE Transactions on Artificial Intelligence*, 1(3):233–247, 2020.

[Huang *et al.*, 2018] Dong Huang, Chang-Dong Wang, Hongxing Peng, Jianhuang Lai, and Chee-Keong Kwoh. Enhanced ensemble clustering via fast propagation of cluster-wise similarities. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(1):508–520, 2018.

[Huang *et al.*, 2022] Shudong Huang, Hongjie Wu, Yazhou Ren, Ivor Tsang, Zenglin Xu, Wentao Feng, and Jiancheng Lv. Multi-view subspace clustering on topological manifold. *Advances in Neural Information Processing Systems*, 35:25883–25894, 2022.

[Huang *et al.*, 2023a] Dong Huang, Chang-Dong Wang, and Jian-Huang Lai. Fast multi-view clustering via ensembles: Towards scalability, superiority, and simplicity. *IEEE Transactions on Knowledge and Data Engineering*, 35(11):11388–11402, 2023.

[Huang *et al.*, 2023b] Shudong Huang, Yixi Liu, Hecheng Cai, Yuze Tan, Chenwei Tang, and Jiancheng Lv. Smooth representation learning from multi-view data. *Information Fusion*, 100:101916, 2023.

[Huang *et al.*, 2024] Shudong Huang, Hecheng Cai, Hao Dai, Wentao Feng, and Jiancheng Lv. Adaptive instance-wise multi-view clustering. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 5299–5307, 2024.

[Jiang *et al.*, 2025a] Bingbing Jiang, Jun Liu, Zidong Wang, Chenglong Zhang, Jie Yang, Yadi Wang, Weiguo Sheng, and Weiping Ding. Semi-supervised multi-view feature selection with adaptive similarity fusion and learning. *Pattern Recognition*, 159:111159, 2025.

[Jiang *et al.*, 2025b] Bingbing Jiang, Chenglong Zhang, Xinyan Liang, Peng Zhou, Jie Yang, Xingyu Wu, Junyi Guan, Weiping Ding, and Weiguo Sheng. Collaborative similarity fusion and consistency recovery for incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 17617–17625, 2025.

[Kang *et al.*, 2020] Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4412–4419, 2020.

[Kumar *et al.*, 2011] Abhishek Kumar, Piyush Rai, and Hal Daume. Co-regularized multi-view spectral clustering. *Proceedings of the International Conference on Neural Information Processing Systems*, pages 1413–1421, 2011.

[Li *et al.*, 2021] Zhenglai Li, Chang Tang, Xinwang Liu, Xiao Zheng, Wei Zhang, and En Zhu. Consensus graph learning for multi-view clustering. *IEEE Transactions on Multimedia*, 24:2461–2472, 2021.

[Li *et al.*, 2024] Taiyong Li, Xiaoyang Shu, Jiang Wu, Qingxiao Zheng, Xi Lv, and Jiaxuan Xu. Adaptive weighted ensemble clustering via kernel learning and local information preservation. *Knowledge-Based Systems*, 294:111793, 2024.

[Liang *et al.*, 2024] Xinyan Liang, Pinhan Fu, Qian Guo, Keyin Zheng, and Yuhua Qian. Dc-nas: Divide-and-conquer neural architecture search for multi-modal classification. In *Proceedings of the AAAI conference on Artificial Intelligence*, pages 13754–13762, 2024.

[Liang *et al.*, 2025] Xinyan Liang, Pinhan Fu, Yuhua Qian, Qian Guo, and Guoqing Liu. Trusted multi-view classification via evolutionary multi-view fusion. In *Proceedings of the 13th International Conference on Learning Representations*, pages 1–14, 2025.

[Lou *et al.*, 2024] Zhengzheng Lou, Chaoyang Zhang, Hang Xue, Yangdong Ye, Qinglei Zhou, and Shizhe Hu. Self-supervised weighted information bottleneck for multi-view clustering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 4643–4650, 2024.

[Nie *et al.*, 2018] Feiping Nie, Lai Tian, and Xuelong Li. Multiview clustering via adaptively weighted procrustes. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2022–2030, 2018.

[Pfeifer *et al.*, 2023] Bastian Pfeifer, Marcus D Bloice, and Michael G Schimek. Parea: multi-view ensemble clustering for cancer subtype discovery. *Journal of Biomedical Informatics*, 143:104406, 2023.

[Strehl and Ghosh, 2002] Alexander Strehl and Joydeep Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3(12):583–617, 2002.

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

[Sun *et al.*, 2025] Bohang Sun, Yongjian Deng, Yuena Lin, Qiuru Hai, Zhen Yang, and Gengyu Lyu. Graph consistency and diversity measurement for federated multi-view clustering. In *Proceedings of the AAAI conference on Artificial Intelligence*, pages 20663–20671, 2025.

[Wang *et al.*, 2019] Hao Wang, Yan Yang, and Bing Liu. Gmc: Graph-based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 32(6):1116–1129, 2019.

[Wang *et al.*, 2024a] Jing Wang, Songhe Feng, Gengyu Lyu, and Jiazheng Yuan. Surer: Structure-adaptive unified graph neural network for multi-view clustering. In *Proceedings of the AAAI conference on Artificial Intelligence*, pages 15520–15527, 2024.

[Wang *et al.*, 2024b] Siwei Wang, Xinwang Liu, Suyuan Liu, Wenxuan Tu, and En Zhu. Scalable and structural multi-view graph clustering with adaptive anchor fusion. *IEEE Transactions on Image Processing*, 33:4627– 4639, 2024.

[Wang *et al.*, 2025] Zhongli Wang, Jie Yang, Junyi Guan, Chenglong Zhang, Xinyan Liang, Bingbing Jiang, and Weiguo Sheng. Enhanced density peak clustering for high-dimensional data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 21411–21419, 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.*, 2024] Danyang Wu, Zhenkun Yang, Jitao Lu, Jin Xu, Xiangmin Xu, and Feiping Nie. Ebmgc-gnf: Efficient balanced multi-view graph clustering via good neighbor fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):7878–7892, 2024.

[Wu *et al.*, 2025] Danyang Wu, Penglei Wang, Jitao Lu, Zhanxuan Hu, Hongming Zhang, and Feiping Nie. Triangle topology enhancement for multi-view graph clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2025.

[Xu *et al.*, 2022] Jie Xu, Huayi Tang, Yazhou Ren, Liang Peng, Xiaofeng Zhu, and Lifang He. Multi-level feature learning for contrastive multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16051–16060, 2022.

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

[Xu *et al.*, 2024] Jiaxuan Xu, Taiyong Li, and Lei Duan. Enhancing ensemble clustering with adaptive high-order topological weights. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 16184–16192, 2024.

[Yang and Lin, 2024] Jie Yang and Chin-Teng Lin. Enhanced adjacency-constrained hierarchical clustering using fine-grained pseudo labels. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(3):2481–2492, 2024.

[Yu *et al.*, 2020] Zhiwen Yu, Daxing Wang, Xian-Bing Meng, and CL Philip Chen. Clustering ensemble based on hybrid multiview clustering. *IEEE Transactions on Cybernetics*, 52(7):6518–6530, 2020.

[Zhang *et al.*, 2023] Xuanqi Zhang, Qiangqiang Shen, Yongyong Chen, Guokai Zhang, Zhongyun Hua, and Jingyong Su. Multi-view ensemble clustering via low-rank and sparse decomposition: From matrix to tensor. *ACM Transactions on Knowledge Discovery from Data*, 17(7):1–19, 2023.

[Zhang *et al.*, 2024a] Chenglong Zhang, Yang Fang, Xinyan Liang, Xingyu Wu, Bingbing Jiang, et al. Efficient multi-view unsupervised feature selection with adaptive structure learning and inference. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*, pages 5443–5452, 2024.

[Zhang *et al.*, 2024b] Chenglong Zhang, Xinyan Liang, Peng Zhou, Zhaolong Ling, Yingwei Zhang, Xingyu Wu, Weiguo Sheng, and Bingbing Jiang. Scalable multi-view unsupervised feature selection with structure learning and fusion. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 5479–5488, 2024.

[Zheng *et al.*, 2024] Dacheng Zheng, Zhiwen Yu, Wuxing Chen, Weiwen Zhang, Qiying Feng, Yifan Shi, and Kaixiang Yang. Multiview ensemble clustering of hypergraph p-laplacian regularization with weighting and denoising. *Information Sciences*, 681:121187, 2024.

[Zong *et al.*, 2018] Linlin Zong, Xianchao Zhang, Xinyue Liu, and Hong Yu. Weighted multi-view spectral clustering based on spectral perturbation. In *Proceedings of the AAAI conference on Artificial Intelligence*, pages 4621–4628, 2018.