Multi-view Regression Clustering via Low-Rank Manifold Decomposition

Xiaowei Zhao[1,2], **Xinyu Kou**[3], **Yan Chen**[4], **Linrui Xie**[1,2], **Qiang Zhang**[3] and **Liang Du**[1,2*]
[1]School of Artificial Intelligence, Shanxi University, Taiyuan 030006, Shanxi, China
[2]Key Laboratory of Evolutionary Science Intelligence of Shanxi Province, Shanxi University, China
[3]State Key Laboratory of Electromechanical Integrated Manufacturing of High-Performance Electronic Equipment, Xidian University, Xi'an 710071, Shaanxi, China
[4] Taiyuan University of Technology, Taiyuan, Shanxi, China
xiaoweizhao4@gmail.com, xinyukou450@gmail.com, chenyan01@tyut.edu.cn, xielinrui6@gmail.com, qzhang@xidian.edu.cn, duliang@sxu.edu.cn

Abstract

Similarity-graph-based multi-view clustering is effective in capturing non-Gaussian cluster structures, yet it suffers from an inherent conflict between scalability and geometry-consistent cluster assignment. In particular, cluster labels are typically inferred via secondary clustering rather than being learned directly from the underlying data geometry. Although anchor-based extensions alleviate the computational burden, they introduce the challenge of anchor-number selection. To address these limitations, we propose Multi-view Clustering via Manifold Decomposition (MvMD), which directly infers cluster labels from multi-view data without explicit similarity graph construction or anchor selection. MvMD reformulates multi-view clustering as a unified multi-view regression problem, where cluster labels are optimized as model variables. To improve robustness against missing small clusters, a global balance regularization is incorporated. Meanwhile, local geometric consistency and the low-rank structure of cluster assignments are jointly enforced through a symmetric matrix factorization scheme, which is efficiently realized via a truncated SVD-based low-rank approximation. Extensive experiments on benchmark datasets validate the effectiveness and efficiency of the proposed MvMD. The code is available at https://github.com/Vince-Doit/MvMD.

1 Introduction

Clustering is a fundamental problem in unsupervised learning, aiming to partition samples into homogeneous groups such that intra-cluster similarity is maximized while inter-cluster similarity is minimized. Due to its broad applicability, clustering has been extensively studied in data mining, image processing, bioinformatics, and related fields [Zhao *et al.*, 2026; Zhao *et al.*, 2016]. With the rapid diversification of data acquisition modalities, modern datasets are increasingly collected from multiple heterogeneous sources or perspectives. This has given rise to *multi-view clustering* (MvC), which seeks to exploit complementary yet consistent information across views to obtain more accurate and robust clustering results than single-view approaches.

Among existing MvC methods, similarity-graph-based models have demonstrated strong capability in capturing complex, non-Gaussian cluster structures. Typically, view-specific similarity graphs are constructed and fused into a shared graph, from which clustering results are obtained via spectral decomposition. To improve partition quality, adaptive graph-based methods impose rank constraints on the Laplacian matrix, forcing the shared graph to decompose into exactly c connected components, each corresponding to a cluster [Nie *et al.*, 2017a; Wang *et al.*, 2019; Zhao *et al.*, 2026]. More recent studies further refine this framework by disentangling view-specific graphs into consistent and inconsistent components, or by modeling high-order neighborhood relations through tensors, hypergraphs, or diffusion processes [Huang *et al.*, 2022; Lin *et al.*, 2023; Liu *et al.*, 2025; Jiang *et al.*, 2025a].

Despite their effectiveness, graph-based MvC methods suffer from a fundamental scalability bottleneck. Constructing an $n \times n$ similarity graph requires $\mathcal{O}(n^2)$ pairwise computations (where n is the number of samples), followed by a spectral decomposition with $\mathcal{O}(n^3)$ complexity, which limits scalability to large datasets [Jiang *et al.*, 2025b]. More importantly, this computational burden is tightly coupled with a deeper modeling issue: pseudo-labels are not learned directly from the data geometry but are instead obtained through a secondary clustering step applied to spectral embeddings. As a result, pseudo-label inference is inherently decoupled from the original data distribution and the optimization objective, which may lead to suboptimal clustering solutions.

To reduce graph construction costs, anchor-based MvC methods replace the similarity graph with an $n \times m$ bipartite graph ($m \ll n$) linking samples to a compact anchor set. Within this framework, two paradigms have emerged: the first adopts matrix factorization, decomposing each view-specific data matrix into anchor representations and a bipartite coefficient matrix [Wang *et al.*, 2022; Liu *et al.*, 2022b; Zhang *et al.*, 2024]; the second directly models sample-to-anchor relationships, avoiding explicit anchor alignment

*Corresponding author.

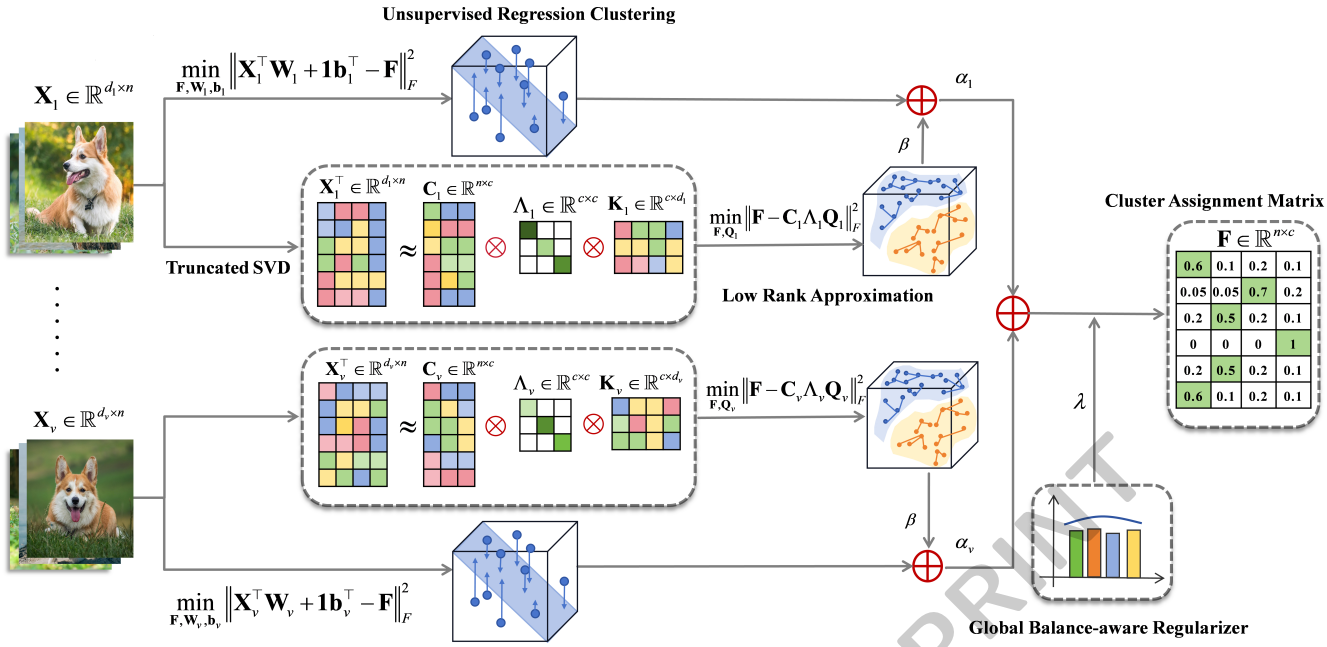

Figure 1: Framework of the proposed **MvMD**. An unsupervised regression-based clustering scheme is applied to each view to directly map samples into the cluster space, assisted by a truncated SVD-based low-rank approximation that enforces low-rank structure and local smoothness without explicit graph construction. A global balance regularizer is imposed on the unified cluster assignment matrix to regulate the overall distribution of cluster assignments across all views.

across views [Qiang *et al.*, 2021; Shi *et al.*, 2023; Chen *et al.*, 2025]. Although these reduce computational complexity, they introduce new challenges.

Specifically, increasing the number of anchors improves representational capacity but inevitably degrades efficiency, leading to an inherent trade-off between clustering accuracy and computational cost. Moreover, determining the optimal number of anchors is an NP-hard problem, making anchor selection highly sensitive and data-dependent. More fundamentally, most anchor-based methods inherit the two-stage pipeline of graph-based clustering, where representation learning and pseudo-label inference are performed separately and final cluster assignments are obtained via post-processing steps such as spectral clustering or K-Means. Consequently, pseudo-label learning remains decoupled from the intrinsic data geometry, and scalability is achieved at the expense of modeling consistency.

In essence, existing graph-based and anchor-based MvC methods are constrained by an inherent conflict between scalability and geometry-consistent pseudo-label learning. Addressing one aspect often exacerbates the other. This observation naturally raises the following question: *Can multi-view clustering be reformulated to directly infer pseudo-labels from multi-view data in a scalable manner, without relying on explicit similarity graph construction or anchor selection?*

To answer this question, we propose **Multi-view Clustering via Manifold Decomposition (MvMD)**. Fig. 1 illustrates the overall framework of MvMD, which reformulates multi-view clustering as a regression problem, where the cluster assignment matrix $\mathbf{F}$ is directly optimized rather than obtained via a post hoc clustering step. Within linear-time complexity, MvMD preserves the low-rank structure of $\mathbf{F}$, enforces manifold-based smoothness, and mitigates the risk that improper initialization fails to identify small clusters commonly observed in real-world data. The main contributions are summarized as follows:

- We propose a novel multi-view clustering framework by reformulating MvC as a unified multi-view regression problem, which enables direct inference of cluster assignments and alleviates the tendency to miss small clusters through a global balance regularization.
- We introduce a symmetric matrix factorization on similarity matrices to enforce the low-rank property and enhance the local smoothness of cluster labels, which is efficiently realized via truncated SVD-based low-rank approximation without explicit graph construction.
- We develop an efficient alternating optimization algorithm. Extensive experiments on benchmark datasets demonstrate that the proposed method consistently outperforms state-of-the-art approaches in terms of both clustering accuracy and computational efficiency.

Throughout this paper, matrices are denoted by bold Roman uppercase letters (e.g., $\mathbf{M}$), vectors by bold Roman lowercase letters (e.g., $\mathbf{m}$), and scalars by regular lowercase letters (e.g., m). The operators $\mathbf{M}^\top$, $Tr(\mathbf{M})$, and $rank(\mathbf{M})$ represent the transpose, trace, and rank of a matrix, respectively. The Frobenius norm of $\mathbf{M}$ is defined as $\|\mathbf{M}\|_F$. The symbols $\mathbf{1}$, $\mathbf{0}$, and $\mathbf{I}$ represent a column vector of ones, a zero matrix, and an identity matrix, respectively.

2 Related Work

2.1 Unsupervised Regression Clustering

As a label-free technique, unsupervised regression clustering captures linear relationships between samples and their pseudo-labels, making it well suited for clustering. Early efforts mainly focus on enhancing robustness and discriminability by incorporating graph regularization [Hou *et al.*, 2014], non-negativity and sparsity constraints [Li and Tang, 2015], or sparse residual modeling to suppress outliers [Shi *et al.*, 2014]. However, most of these methods decouple similarity graph construction from regression optimization, typically relying on Gaussian kernels or k-NN graphs built in a preprocessing step, thereby overlooking their intrinsic interaction. To address this issue, Li et al. [Li *et al.*, 2018] proposed a unified framework that jointly learns the similarity graph and regression model, enabling adaptive neighborhood structures. Peng et al. [Peng *et al.*, 2019] further improved robustness by imposing low-rank constraints on the learned graph and sparsity on feature selection. Despite these advances, existing approaches still depend on explicit graph construction, which limits scalability on large-scale data.

2.2 Similarity-Driven Multi-View Clustering

Similarity graph learning constitutes a core component of most multi-view clustering methods. The dominant paradigm follows a two-stage process: constructing view-specific graphs and integrating them for partitioning. Early methods use fixed similarity measures like Gaussian kernels, which are efficient but often insufficient for complex structures. To improve flexibility, recent studies focus on adaptive and unified graph learning. Chen et al. [Chen *et al.*, 2021] proposed a low-rank tensor graph framework that jointly learns representations and affinity matrices, enhancing robustness. Jiang et al. [Jiang *et al.*, 2022] incorporated latent-space feature enhancement and self-expressiveness to preserve inter-view consistency. Along this line, Nie et al. introduced adaptive view weighting to jointly optimize graph construction and clustering, inspiring a series of subsequent works [Li *et al.*, 2019; Shi *et al.*, 2024; Zhao *et al.*, 2025]. Nevertheless, these methods still rely on spectral decomposition of similarity graphs, posing scalability challenges.

3 Methodology

Regression-based clustering directly treats cluster assignments as optimization variables, enabling one-step inference of clustering results rather than relying on post hoc partitioning [Liu *et al.*, 2017; Zhang *et al.*, 2020]. This property motivates us to revisit multi-view clustering from a regression perspective and develop the proposed model as follows:

$$\min_{\substack{\mathbf{F},\mathbf{W}_v,\\ \boldsymbol{b}_v,\alpha_v}} \sum_{v=1}^{p} \alpha_v^2 \left\| \mathbf{X}_v^\top \mathbf{W}_v + \mathbf{1}\boldsymbol{b}_v^\top - \mathbf{F} \right\|_F^2 + \lambda Tr(\mathbf{F}^\top \mathbf{1}\mathbf{1}^\top \mathbf{F})$$
$$\text{s.t.} \quad \mathbf{F} \geq \mathbf{0}, \mathbf{F}\mathbf{1} = \mathbf{1}, \mathbf{W}_v^\top \mathbf{W}_v = \mathbf{I}, \alpha_v \geq 0, \sum_{v=1}^{p} \alpha_v = 1 \tag{1}$$

where p is the number of views. For the v-th view, $\mathbf{X}_v \in \mathbb{R}^{d_v \times n}$ denotes the data matrix, whose columns are normalized to unit length with non-negative features; $\mathbf{W}_v \in \mathbb{R}^{d_v \times c}$ is the projection matrix ensuring uncorrelated subspace dimensions; and $\boldsymbol{b}_v \in \mathbb{R}^{1\times c}$ denotes the bias. $\mathbf{F}$ denotes cluster assignment matrix, with f_{ij} indicating the probability of assigning the i-th sample to the j-th cluster. $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p]^\top$ denotes the view-weight vector. λ is the regularization parameter. The first term promotes cross-view consistency in the clustering space by minimizing the weighted regression error between the projected features of each view and the unified cluster assignment matrix. The second term prevents the optimal solution of $\mathbf{F}$ from converging to a hard clustering [Liu *et al.*, 2022a], while ensuring a more balanced assignment distribution, thus mitigating the issue where improper initialization may fail to detect small clusters.

Symmetric nonnegative matrix factorization (SNMF) has proven effective for clustering data with nonlinear structure [Kuang *et al.*, 2015]. To further enhance the local smoothness of cluster assignments, we improve SNMF by relaxing part of the nonnegativity constraint while enforcing stronger low-rank properties on $\mathbf{F}$. For the normalized data $\mathbf{X}_v$ in the v-th view, the symmetric nonnegative similarity matrix is constructed as:

$$\mathbf{S}_v = \mathbf{X}_v^\top \mathbf{X}_v \tag{2}$$

where $\mathbf{S}_v \geq \mathbf{0}$ represents the pairwise similarity between samples. After performing truncated SVD on $\mathbf{X}_v^\top$, we obtain the top c left singular vectors $\mathbf{C}_v \in \mathbb{R}^{n\times c}$, and the corresponding singular value matrix $\Lambda_v \in \mathbb{R}^{c\times c}$. Consequently, the low-rank version of $\mathbf{S}_v$ is expressed as:

$$\widehat{\mathbf{S}}_v = (\mathbf{C}_v \Lambda_v \mathbf{Q}_v)(\mathbf{C}\Lambda_v \mathbf{Q}_v)^\top \tag{3}$$

where $\mathbf{Q}_v \in \mathbb{R}^{c\times c}$ is an arbitrary orthonormal matrix. It is obvious that $\mathbf{Q}_v^\top \mathbf{Q}_v = \mathbf{Q}_v \mathbf{Q}_v^\top = \mathbf{I}$. According to the Eckart-Young-Mirsky theorem, $\widehat{\mathbf{S}}_v$ is the optimal rank-c approximation of $\mathbf{S}_v$ under both the Frobenius and spectral norms. We thus formulate the low-rank symmetric matrix factorization for the v-th view as:

$$\min_{\mathbf{F}\geq\mathbf{0},\mathbf{F}\mathbf{1}=\mathbf{1},\mathbf{Q}_v^\top\mathbf{Q}_v=\mathbf{I}} \| \mathbf{F}\mathbf{F}^\top - (\mathbf{C}_v \Lambda_v \mathbf{Q}_v)(\mathbf{C}\Lambda_v \mathbf{Q}_v)^\top \|_F^2 \tag{4}$$

To avoid the quadratic complexity of Eq. (4), we derive a tractable surrogate as follows:

$$\min_{\mathbf{F}\geq\mathbf{0},\mathbf{F}\mathbf{1}=\mathbf{1},\mathbf{Q}_v^\top\mathbf{Q}_v=\mathbf{I}} \| \mathbf{F} - \mathbf{C}_v \Lambda_v \mathbf{Q}_v \|_F^2 \tag{5}$$

By applying the sub-multiplicative property of the Frobenius norm and the triangle inequality, we obtain $\|\mathbf{F}\mathbf{F}^\top - (\mathbf{C}_v\mathbf{\Lambda}_v\mathbf{Q}_v)(\mathbf{C}_v\mathbf{\Lambda}_v\mathbf{Q}_v)^\top\|_F \leq (\|\mathbf{F}\|_F + \|\mathbf{C}_v\mathbf{\Lambda}_v\mathbf{Q}_v\|_F)\|\mathbf{F} - \mathbf{C}_v\mathbf{\Lambda}_v\mathbf{Q}_v\|_F$ where $\|\mathbf{F}\|_F^2 \leq n$. Under the common low-rank assumption in clustering, i.e., $c \ll n$ and the dominant energy is captured by the leading singular values, $\|\mathbf{C}_v\mathbf{\Lambda}_v\mathbf{Q}_v\|_F$ is bounded and grows significantly slower than n. Consequently, the objective function in Eq. (4) can be upper-bounded by a scaled version of $\|\mathbf{F} - \mathbf{C}_v\mathbf{\Lambda}_v\mathbf{Q}_v\|_F$. Furthermore, the low-rank structure of $\mathbf{C}_v\mathbf{\Lambda}_v\mathbf{Q}_v$ induces an approximate low-rank property in $\mathbf{F}$.

To achieve high-quality multi-view clustering, Eqs.(1) and (5) can be integrated into a unified framework. Consequently, the MvMD model is formulated as:

$$\begin{aligned}\min_{\substack{\alpha_v,\mathbf{F},\mathbf{Q}_v,\\ \mathbf{W}_v,b_v}} & \sum_{v=1}^{p}\alpha_v^2\left\|\mathbf{X}_v^\top\mathbf{W}_v+\mathbf{1}\boldsymbol{b}_v^\top-\mathbf{F}\right\|_F^2 \\ & +\beta\sum_{v=1}^{p}\alpha_v^2\left\|\mathbf{F}-\mathbf{C}_v\Lambda_v\mathbf{Q}_v\right\|_F^2+\lambda\,\mathrm{Tr}\left(\mathbf{F}^\top\mathbf{1}\mathbf{1}^\top\mathbf{F}\right) \\ \text{s.t.}\quad & \mathbf{F}\geq\mathbf{0},\ \mathbf{F}\mathbf{1}=\mathbf{1},\ \mathbf{W}_v^\top\mathbf{W}_v=\mathbf{I},\ \mathbf{Q}_v^\top\mathbf{Q}_v=\mathbf{I}, \\ & \alpha_v\geq 0,\ \sum_{v=1}^{p}\alpha_v=1.\end{aligned} \tag{6}$$

where the regularization parameter β controls the trade-off between local manifold regularization and regression fidelity.

4 Optimization Algorithm

In this subsection, we propose an alternating optimization algorithm to solve Eq. (6) by cyclically updating each variable while keeping the others fixed.

(1) Optimize $\boldsymbol{b}_v$

When $\mathbf{Q}_v$, $\mathbf{F}$, $\mathbf{W}_v$, and α_v are fixed, $\boldsymbol{b}_v$ can be updated by:

$$\boldsymbol{b}_v=(\frac{1}{n}\mathbf{F}^\top\mathbf{1}-\mathbf{W}_v^\top\mathbf{X}_v\mathbf{1}) \tag{7}$$

(2) Optimize $\mathbf{W}_v$

When $\mathbf{Q}_v$, $\mathbf{F}$, $\boldsymbol{b}_v$, and α_v are fixed, the Eq. (6) reduces to:

$$\min_{\mathbf{W}_v^\top\mathbf{W}_v=\mathbf{I}} Tr(\mathbf{W}_v^\top\mathbf{P}_v\mathbf{W}_v-2\mathbf{W}_v^\top\mathbf{X}_v(\mathbf{F}-\mathbf{1}b_v^\top)) \tag{8}$$

where $\mathbf{P}_v=\mathbf{X}_v\mathbf{X}_v^\top\in\mathbb{R}^{d_v\times d_v}$. It is evident that Eq. (8) can be equivalently reformulated as:

$$\max_{\mathbf{W}_v^\top\mathbf{W}_v=\mathbf{I}} Tr(\mathbf{W}_v^\top\widetilde{\mathbf{P}_v}\mathbf{W})+2Tr(\mathbf{W}_v^\top\mathbf{X}_v(\mathbf{F}-\mathbf{1}\boldsymbol{b}_v^\top)) \tag{9}$$

where $\widetilde{\mathbf{P}}_v=\delta\mathbf{I}_{d_v}-\mathbf{P}_v$, and δ is a regularization parameter introduced to ensure the positive definiteness of $\widetilde{\mathbf{P}}_v$. To solve this optimization problem, we adopt the following power iteration method proposed in [Nie *et al.*, 2017b].

(3) Optimize $\mathbf{Q}_v$

When $\mathbf{W}_v$, $\mathbf{F}$, b_v, and α_v are fixed, the Eq. (6) reduces to:

$$\begin{aligned}\min_{\mathbf{Q}_v}\quad & Tr\left(\mathbf{Q}_v^\top(\mathbf{C}_v\Lambda_v)^\top\mathbf{C}_v\Lambda_v\mathbf{Q}_v\right)-2Tr\left(\mathbf{Q}_v^\top(\mathbf{C}_v\Lambda_v)^\top\mathbf{F}\right) \\ \text{s.t.}\quad & \mathbf{Q}_v^\top\mathbf{Q}_v=\mathbf{I}\end{aligned} \tag{10}$$

Since Eq. (10) shares the same structure as Eq. (9), the corresponding minimization problem can similarly be reformulated as a maximization problem. Unlike $\mathbf{W}_v$ in Eq. (9), the matrix $\mathbf{Q}_v$ satisfies both $\mathbf{Q}_v\mathbf{Q}_v^\top=\mathbf{I}$ and $\mathbf{Q}_v^\top\mathbf{Q}_v=\mathbf{I}$. Therefore, the Eq. (10) can be further simplified as:

$$\max_{\mathbf{Q}_v^\top\mathbf{Q}_v=\mathbf{I}} Tr(\mathbf{Q}_v^\top(\mathbf{C}_v\Lambda_v)^\top\mathbf{F}) \tag{11}$$

This subproblem reduces to an orthogonal Procrustes problem, and the optimal solution of $\mathbf{Q}_v$ can be optimized by performing SVD on $(\mathbf{C}_v\Lambda_v)^\top\mathbf{F}$.

(4) Optimize $\mathbf{F}$

When $\mathbf{W}_v$, $\mathbf{Q}_v$, $\boldsymbol{b}_v$, and α_v are fixed, the Eq. (6) can be reformulated as follows:

$$\min_{\mathbf{F}\geq\mathbf{0},\mathbf{F}\mathbf{1}=\mathbf{1}}\|\mathbf{F}-\mathbf{A}\|_F^2+\lambda Tr(\mathbf{F}^\top\mathbf{1}\mathbf{1}^\top\mathbf{F}) \tag{12}$$

where $\mathbf{A}=\frac{\sum_{v=1}^{p}\alpha_v^2(\mathbf{X}_v^\top\mathbf{W}_v+\mathbf{1}\boldsymbol{b}_v^\top+\beta\mathbf{C}_v\Lambda_v\mathbf{Q}_v)}{\sum_{v=1}^{p}\alpha_v^2(1+\beta)}$. By introducing the auxiliary variable $\mathbf{B}$, and setting $\mathbf{B}=\mathbf{F}$, the corresponding augmented Lagrangian function is formulated as:

$$\begin{aligned}\min_{\substack{\mathbf{F},\mathbf{B},\\ \mu,\Psi}} & \|\mathbf{F}-\mathbf{A}\|_F^2+\lambda Tr(\mathbf{F}^\top\mathbf{1}\mathbf{1}^\top\mathbf{B})+\frac{\mu}{2}\left\|\mathbf{F}-\mathbf{B}+\frac{1}{\mu}\Psi\right\|_F^2 \\ \text{s.t.}\quad & \mathbf{F}\geq\mathbf{0},\mathbf{F}\mathbf{1}=\mathbf{1}\end{aligned} \tag{13}$$

where Ψ is the Lagrange multiplier matrix, and μ is the penalty parameter.

The subproblem for $\mathbf{B}$ is expressed as:

$$\min_{\mathbf{B}} Tr(\mathbf{F}^\top\mathbf{1}\mathbf{1}^\top\mathbf{B})+\frac{\mu}{2\lambda}\left\|\mathbf{F}-\mathbf{B}+\frac{1}{\mu}\Psi\right\|_F^2 \tag{14}$$

Setting the derivative of the objective function in Eq. (14) with respect to $\mathbf{B}$ to zero, we have:

$$\mathbf{B}=\mathbf{F}+\frac{1}{\mu}\Psi-\frac{\lambda}{\mu}\mathbf{1}\mathbf{1}^\top\mathbf{F} \tag{15}$$

The subproblem for $\mathbf{F}$ is formulated as:

$$\min_{\mathbf{F}\geq\mathbf{0},\mathbf{F}\mathbf{1}=\mathbf{1}}\left\|\mathbf{F}-\frac{\mathbf{A}-\frac{\lambda}{2}\mathbf{1}\mathbf{1}^\top\mathbf{B}-\frac{\mu}{2}(-\mathbf{B}+\frac{1}{\mu}\Psi)}{1+\mu/2}\right\|_F^2 \tag{16}$$

Notably, each row of $\mathbf{F}$ can be solved independently through the method proposed in [Nie *et al.*, 2014].

Finally, the Lagrange multipliers are updated as follows:

$$\Psi=\Psi+\mu(\mathbf{F}-\mathbf{B}) \tag{17}$$

$$\mu=\rho\mu \tag{18}$$

where ρ is the learning rate. The complete optimization procedure for Eq. (12) is summarized in Algorithm 1.

(5) Optimize α_v

When $\mathbf{W}_v$, $\mathbf{Q}_v$ and $\mathbf{F}$ are fixed, the Eq. (6) reduces to:

$$\min_{\mathbf{1}^\top\boldsymbol{\alpha}=1,\alpha_v\geq 0}\sum_{v=1}^{p}\alpha_v^2\mu_v \tag{19}$$

where $\mu_v=\left\|\mathbf{X}_v^\top\mathbf{W}_v+\mathbf{1}\boldsymbol{b}_v^\top-\mathbf{F}\right\|_F^2+\beta\left\|\mathbf{C}_v\Lambda_v\mathbf{Q}_v-\mathbf{F}\right\|_F^2$. By introducing the Lagrange multiplier ξ, the corresponding Lagrangian function of Eq. (19) can be formulated as:

$$\mathcal{J}(\boldsymbol{\alpha})=\sum_{v=1}^{p}\alpha_v^2\mu_v-\xi(\sum_{v=1}^{p}\alpha_v-1) \tag{20}$$

By taking the derivative of Eq. (20) with respect to ξ and setting it to zero, we obtain:

$$\frac{\partial\mathcal{J}}{\partial\boldsymbol{\alpha}}=2\alpha_v\mu_v-\xi=0 \tag{21}$$

Algorithm 1 Efficient Algorithm for Eq. (12)

Input: Normalized data matrix $\mathbf{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_p\}$, initial cluster assignment matrix $\mathbf{F}$, projection matrices $\mathbf{W}_v$, $\mathbf{Q}_v$, truncated SVD factors $\mathbf{C}_v\Lambda_v$, bias vectors $\boldsymbol{b}_v$, weights α_v, and regularization parameters β and λ.
Output: The cluster assignment matrix $\mathbf{F}$.

Initialize

$$\mathbf{A} = \frac{\sum_{v=1}^{p} \alpha_v^2 \left(\mathbf{X}_v^\top \mathbf{W}_v + \mathbf{1}\boldsymbol{b}_v^\top + \beta \mathbf{C}_v \Lambda_v \mathbf{Q}_v \right)}{\sum_{v=1}^{p} \alpha_v^2 (1+\beta)};$$

$\mathbf{B} = \mathbf{F}$;
$\mu = 0.01$; $\rho = 1.1$; $\Psi = \mathbf{0}$;
while not converge **do**
    Update $\mathbf{B}$ by Eq.(15);
    Update $\mathbf{F}$ by solving Eq.(16);
    Update Ψ by Eq.(17);
    Update μ by Eq.(18);

By applying the Karush-Kuhn-Tucker (KKT) conditions under the constraint $\mathbf{1}^\top \boldsymbol{\alpha} = 1$, the optimal solution for each α_v can be derived as:

$$\alpha_v = \frac{\mu_v^{-1}}{\sum_{v=1}^{p} \mu_v^{-1}} \tag{22}$$

The complete optimization procedure for Eq. (6) is summarized in Algorithm 2.

5 Complexity Analysis

The time complexity of **MvMD** consists of an initialization stage and an iterative optimization stage. In the initialization stage, for each view, the matrix $\mathbf{C}_v\boldsymbol{\Lambda}_v$ is constructed from $\mathbf{X}_v$ via truncated singular value decomposition, resulting in a time complexity of $\mathcal{O}(nd_vc + nc^2)$. In the iterative stage, updating all $\boldsymbol{b}_v$ incurs $\mathcal{O}(ndc)$, where $d = \sum_{v=1}^{p} d_v$. Updating $\mathbf{W}_v$ and $\mathbf{Q}_v$ across all views requires $\mathcal{O}(c\sum_{v=1}^{p} d_v^2 + ndc + dc^2)$; updating $\mathbf{F}$ involves computing the auxiliary matrix $\mathbf{A}$ with complexity $\mathcal{O}(ndc)$ followed by a row-wise optimization costing $\mathcal{O}(nc\log c)$; and updating the view weights $\boldsymbol{\alpha}$ requires $\mathcal{O}(ndc + pnc^2)$.

Given that $c \ll d \ll n$ and $p \ll n$, which typically holds for large-scale datasets, the overall computational complexity of **MvMD** can be approximated as $\mathcal{O}(ndcT)$, where T denotes the number of iterations in Algorithm 2.

6 Experiments

6.1 Experimental Setup

MvMD is compared with nine representative multi-view clustering algorithms: Graph-based Multi-view Clustering (**GMC**) [Wang *et al.*, 2019], Structured Graph Learning for Scalable Subspace Clustering: From Single View to Multi-view (**SGL**) [Kang *et al.*, 2021], Efficient One-Pass Multi-view Subspace Clustering with Consensus Anchors (**EOMSC**) [Liu *et al.*, 2022c], Efficient Multi-view K-Means Clustering with Multiple Anchor Graphs (**EMKMC**) [Yang *et al.*, 2023], Auto-Weighted Multi-view Clustering for Large-Scale Data (**AWMvC**) [Wan *et al.*, 2023], Efficient Multi-view Graph Clustering with Local and Global Structure Preservation (**EMvGC**) [Wen *et al.*, 2023], Efficient Multi-view Clustering via Unified and Discrete Bipartite Graph Learning (**UDBGL**) [Fang *et al.*, 2024], Dual-Constraint Multi-view Fuzzy Clustering with Scalable Anchor Graph Learning (**DMFC**) [Cui *et al.*, 2025], and Fast Multiview Anchor-Graph Clustering (**FMAGC**) [Yang *et al.*, 2025].

[1]https://predictioncenter.org/
[2]http://lig-membres.imag.fr/grimal/data.html#
[3]https://www.cs.tau.ac.il/~wolf/ytfaces/

Algorithm 2 : Efficient Algorithm for Eq. (6)

Input: Normalized data matrix $\{\mathbf{X}_1, \ldots, \mathbf{X}_p\}$, and regularization parameters β and λ.
Output: The cluster assignment matrix $\mathbf{F}$.

Initialize $\mathbf{F}$;
Initialize the projection matrices $\mathbf{W}_v$ and $\mathbf{Q}_v$ to satisfy $\mathbf{W}_v^\top \mathbf{W}_v = \mathbf{I}$ and $\mathbf{Q}_v^\top \mathbf{Q}_v = \mathbf{I}$, respectively;
Initialize $\alpha_v = 1/p$ for all views;
repeat
    Update $\boldsymbol{b}_v$ by $\boldsymbol{b}_v = (\frac{1}{n}\mathbf{F}^\top \mathbf{1} - \mathbf{W}_v^\top \mathbf{X}_v \mathbf{1})$;
    Update $\mathbf{W}_v$ by solving Eq. (9);
    Update $\mathbf{Q}_v$ by solving Eq. (11);
    Update $\mathbf{F}$ via Algorithm 1;
    Update α_v according to Eq.(22);
until *Convergence*.

Datasets	Samples	Views	Clusters
MSRC [Winn and Jojic, 2005]	210	5	7
Prokaryotic [Brbić *et al.*, 2016]	551	3	4
ProteinFold[1]	649	12	27
WikipediaArticles[2]	693	2	10
BDGP [Cai *et al.*, 2012]	2500	2	5
YouTubeFace[3]	12512	11	31
CIFAR10 [Lu *et al.*, 2024]	60000	7	10

Table 1: Description of multi-view datasets

The datasets used in our experiments are shown in Table 1. For each view, samples are normalized to unit length and have non-negative feature values. To evaluate clustering performance, we adopt three widely used metrics: Accuracy (ACC), F1-score, and Normalized Mutual Information (NMI), where higher values indicate better performance. For our **MvMD**, the regularization parameters β and λ are selected from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}, 10^{4}\}$. The initial cluster assignment matrix is obtained by applying K-means++ to the concatenated multi-view features. For other methods, the number of anchors m is selected from $\{2c, 3c, 4c, 5c\}$; other hyper-parameters follow their original papers. Each algorithm is executed 10 times, and the average best results are reported. All experiments are conducted in MATLAB R2021b on Linux with an Intel Core i7-12700KF CPU and 64GB RAM.

Methods	MSRC	Prokaryotic	ProteinFold	WikipediaArticles	BDGP	YouTubeFace	CIFAR10
				ACC			
GMC	76.19	54.45	23.63	35.93	74.64	26.40	N/A
SGL	58.33±1.75	60.36±3.41	23.03±0.76	59.33±1.00	74.60±6.85	26.85±7.66	27.01±2.14
EOMSC	54.29	51.36	34.58	58.01	93.72	26.93	26.16
EMKMC	67.86±8.75	53.87±5.36	23.97±0.12	58.87±3.57	75.50±10.56	9.06±4.69	27.90±1.91
AWMvC	31.90	42.11	21.90	57.58	92.56	23.20	27.00
EMvGC	58.81±9.94	55.79±0.38	27.33±0.78	43.55±4.54	60.75±1.19	20.98±0.82	24.88±2.30
UDBGL	67.62	**62.79**	17.00	59.74	71.40	28.88	30.30
DMFC	53.81	57.71	27.38	37.52	47.72	44.06	27.78
FMAGC	75.29±5.41	57.68±4.64	33.50±1.57	61.02±2.28	68.97±9.16	48.82±1.62	N/A
Our	**82.71±1.21**	61.92±4.50	**37.57±0.63**	**61.45±0.31**	**96.22±0.68**	**55.20±0.69**	**30.54±0.98**
				F1-score			
GMC	68.55	54.21	11.31	25.76	74.58	16.17	N/A
SGL	46.40±0.45	52.38±5.50	14.34±1.68	50.04±0.17	62.51±2.78	16.64±2.78	17.52±0.47
EOMSC	40.81	49.36	20.67	50.00	88.06	16.60	19.49
EMKMC	60.39±10.81	52.94±3.45	13.32±1.70	48.97±0.95	63.98±7.24	5.84±3.09	17.86±0.76
AWMVC	49.50	47.44	12.05	51.48	84.86	9.69	18.73
EMVGC	41.98±9.25	45.93±1.52	12.47±0.59	33.10±8.36	36.23±1.40	12.91±0.41	17.34±1.26
UDBGL	57.78	57.06	9.93	48.98	53.02	17.07	22.22
DMFC	39.62	**99.57**	17.83	28.34	43.01	27.56	18.23
FMAGC	64.39±5.26	49.84±3.37	22.05±1.34	53.38±2.17	54.87±5.88	24.51±0.59	N/A
Our	**70.27±2.03**	57.89±2.32	**22.34±0.94**	**53.56±0.12**	**92.76±1.29**	**25.98±0.38**	**23.46±0.01**
				NMI			
GMC	**78.76**	4.44	29.46	35.80	76.08	6.59	N/A
SGL	52.30±2.41	35.04±5.26	32.12±0.07	51.62±0.72	61.93±3.84	14.96±6.65	14.44±1.30
EOMSC	44.33	27.56	41.26	54.93	85.66	9.79	16.48
EMKMC	63.00±8.75	13.58±3.79	31.74±0.89	53.13±2.04	62.85±8.431	6.59±3.05	14.62±1.06
AWMvC	26.90	23.82	28.60	53.67	80.75	**88.81**	16.61
EMvGC	54.53±10.63	20.45±1.15	35.72±0.01	32.21±5.08	39.36±3.59	16.64±0.27	14.35±1.78
UDBGL	62.40	28.62	15.01	56.65	60.41	11.55	17.47
DMFC	39.55	9.32	37.71	32.61	31.88	69.30	16.12
FMAGC	67.99±4.43	34.31±3.10	44.35±0.85	56.73±1.51	55.83±5.12	52.82±0.73	N/A
Our	71.18±0.35	**35.29±3.74**	**44.78±0.96**	**57.61±0.37**	**90.39±1.74**	55.59±0.45	**17.91±0.12**

Table 2: Clustering performance (%) of multi-view clustering methods on benchmark datasets. The best result is shown in **bold**, and the second-best is underlined. 'N/A' means out of memory.

6.2 Clustering Results

Table 2 reports the clustering performance of ten MvC methods, where variances are not applicable for non-randomized (deterministic) algorithms. We observe: (1) **FMAGC** typically outperforms most anchor-based clustering methods. A possible explanation is that it derives discrete pseudo-labels directly through similarity graph decomposition, thereby avoiding secondary clustering. This highlights the benefit of decomposing similarity graphs for clustering. By comparison, **GMC** constructs a complete large-scale similarity graph, but its reliance on secondary clustering tends to introduce instability. (2) The proposed **MvMD** consistently achieves the best performance across most datasets. For example, on BDGP, **MvMD** surpasses the second-best method by 2.5% (ACC), 4.7% (F1-score), and 4.73% (NMI). This advantage stems from its ability to jointly capture global and local structural information. Specifically, the balance-constrained unsupervised regression directly produces cluster labels, promoting globally consistent assignments, while low-rank approximation preserves local label smoothness. These complementary components contribute to the superior performance of **MvMD**.

6.3 Running Time

To evaluate the efficiency of **MvMD**, Fig. 2 presents its running time alongside baseline methods across all datasets. **MvMD** achieves a runtime comparable to the most efficient baselines, consistent with the theoretical analysis indicating linear complexity in sample size.

6.4 Parameter Sensitivity and Convergence

MvMD involves two hyper-parameters, β and λ. In unsupervised settings, we select parameters based on ACC. Fig. 3 reports the sensitivity results, where good performance is consistently achieved at $\beta = 100$ and $\lambda = 100$.

To examine the convergence of Algorithm 2, the objective value is tracked across iterations. As shown in Fig. 3, it

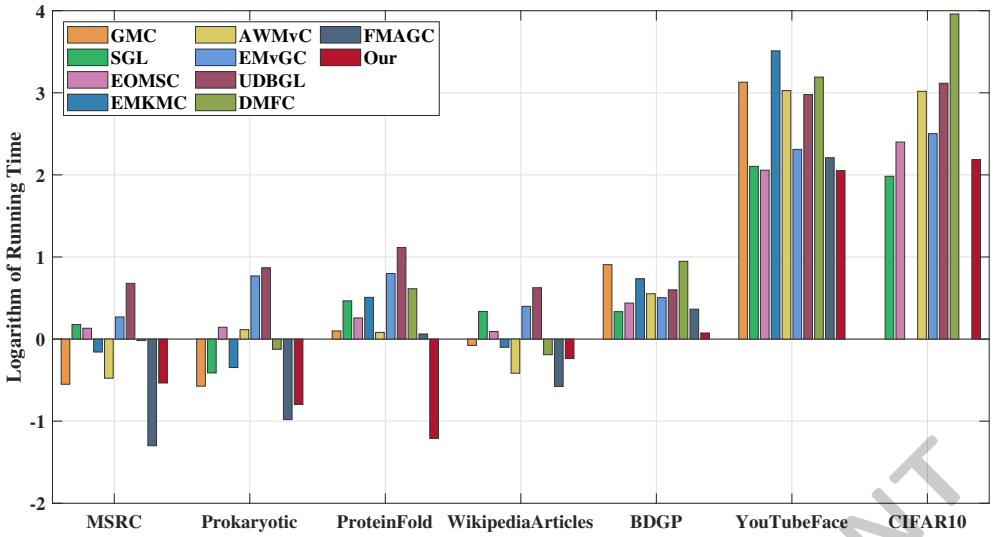

Figure 2: The running time (in seconds) of various algorithms on benchmark datasets. The time is scaled logarithmically.

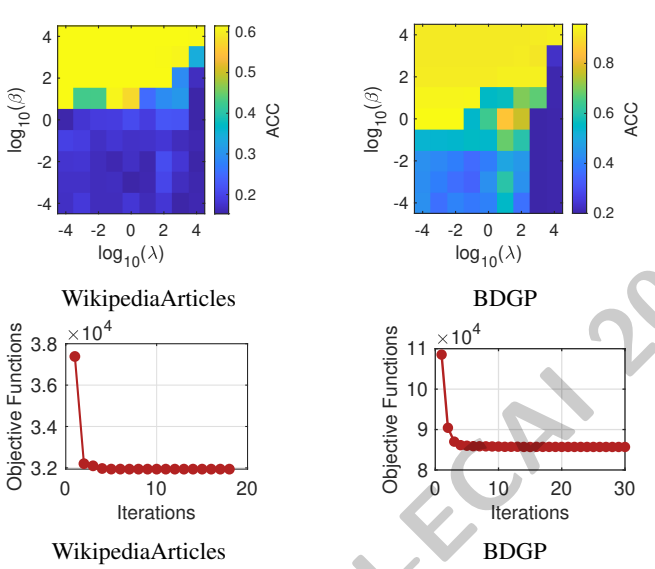

Figure 3: Sensitivity analysis and convergence behavior.

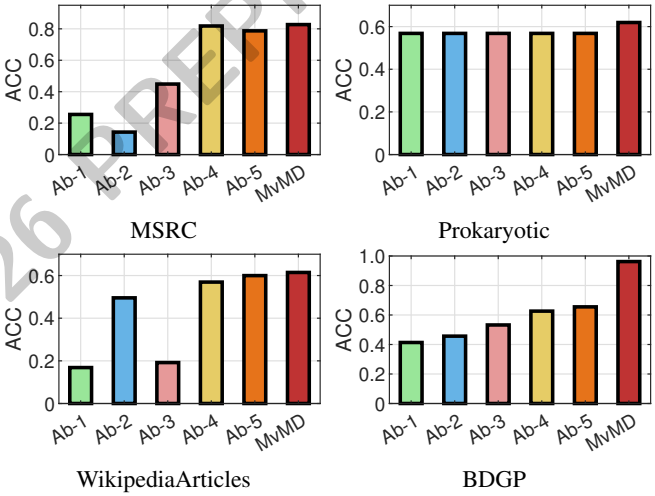

Figure 4: Ablation experiment results.

decreases monotonically and converges within 20 iterations, confirming empirical convergence.

6.5 Ablation Experiments

Five ablation variants are designed to evaluate the contribution of each module in **MvMD**. Ab-1 to Ab-5 respectively retain: (i) only the regression term, (ii) only the low-rank approximation term, (iii) the regression term with balance regularization, (iv) the regression term with low-rank approximation, and (v) the low-rank approximation with balance regularization. The ACC results are reported in Fig. 4.

In most cases, combining any two components improves clustering performance over each individual component. Moreover, all ablation variants underperform the complete **MvMD** model, highlighting the necessity of jointly integrating all components. This suggests that neither low-rank manifold decomposition nor balance-aware regularization alone suffices to fully exploit unsupervised regression clustering. Their joint integration with proper weighting is crucial for superior performance.

7 Conclusion

In this paper, we proposed **MvMD**, a scalable multi-view clustering framework that directly learns cluster assignments from multi-view data. By formulating clustering as a unified multi-view regression problem with balance-aware and geometry-preserving regularization, **MvMD** avoids explicit similarity graph construction and anchor selection. Extensive experiments on benchmark datasets demonstrate that **MvMD** consistently outperforms SOTA methods in both clustering accuracy and efficiency.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant Nos. 62506279 and 62541326), and the Major Scientific and Technological Innovation Project of Xianyang (Grant No. L2025-ZDKJ-ZDGG-RGZN-001).

References

[Brbić *et al.*, 2016] Maria Brbić, Matija Piškorec, Vedrana Vidulin, Anita Kriško, Tomislav Šmuc, and Fran Supek. The landscape of microbial phenotypic traits and associated genes. *Nucleic Acids Research*, 44(21):10074—-10090, 2016.

[Cai *et al.*, 2012] Xiao Cai, Hua Wang, Heng Huang, and Chris Ding. Joint stage recognition and anatomical annotation of drosophila gene expression patterns. *Bioinformatics*, 28(12):i16–i24, 2012.

[Chen *et al.*, 2021] Yongyong Chen, Xiaolin Xiao, Chong Peng, Guangming Lu, and Yicong Zhou. Low-rank tensor graph learning for multi-view subspace clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(1):92–104, 2021.

[Chen *et al.*, 2025] Yawei Chen, Huibing Wang, Jinjia Peng, and Yang Wang. Anchor learning with potential cluster constraints for multi-view clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(15):15939–15947, 2025.

[Cui *et al.*, 2025] Luyan Cui, Huibing Wang, Yawei Chen, Mingze Yao, Xianping Fu, and Jiqing Zhang. Dual-constraint multi-view fuzzy clustering with scalable anchor graph learning. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 7748–7756, 2025.

[Fang *et al.*, 2024] Siguo Fang, Dong Huang, Xiaosha Cai, Changdong Wang, Chaobo He, and Yong Tang. Efficient multi-view clustering via unified and discrete bipartite graph learning. *IEEE Transactions on Neural Networks and Learning Systems*, 35(8):11436–11447, 2024.

[Hou *et al.*, 2014] Chenping Hou, Feiping Nie, Xuelong Li, Dongyun Yi, and Yi Wu. Joint embedding learning and sparse regression: A framework for unsupervised feature selection. *IEEE Transactions on Cybernetics*, 44(6):793–804, 2014.

[Huang *et al.*, 2022] Shudong Huang, Ivor W. Tsang, Zenglin Xu, and Jiancheng Lv. Measuring diversity in graph learning: A unified framework for structured multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 34(12):5869–5883, 2022.

[Jiang *et al.*, 2022] Guangqi Jiang, Jinjia Peng, Huibing Wang, Zetian Mi, and Xianping Fu. Tensorial multi-view clustering via low-rank constrained high-order graph learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(8):5307–5318, 2022.

[Jiang *et al.*, 2025a] 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. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(17):17617–17625, 2025.

[Jiang *et al.*, 2025b] Bingbing Jiang, Chenglong Zhang, Zhongli Wang, Xinyan Liang, Peng Zhou, Liang Du, Qinghua Zhang, Weiping Ding, and Yi Liu. Scalable fuzzy clustering with collaborative structure learning and preservation. *IEEE Transactions on Fuzzy Systems*, 33(9):3047–3060, 2025.

[Kang *et al.*, 2021] Zhao Kang, Zhiping Lin, Xiaofeng Zhu, and Wenbo Xu. Structured graph learning for scalable subspace clustering: From single view to multi-view. *IEEE Transactions on Cybernetics*, 52(9):8976–8986, 2021.

[Kuang *et al.*, 2015] Da Kuang, Sangwoon Yun, and Haesun Park. Symnmf: nonnegative low-rank approximation of a similarity matrix for graph clustering. *Journal of Global Optimization*, 62(3):545–574, 2015.

[Li and Tang, 2015] Zechao Li and Jinhui Tang. Unsupervised feature selection via nonnegative spectral analysis and redundancy control. *IEEE Transactions on Image Processing*, 24(12):5343–5355, 2015.

[Li *et al.*, 2018] Xuelong Li, Han Zhang, Rui Zhang, Yun Liu, and Feiping Nie. Generalized uncorrelated regression with adaptive graph for unsupervised feature selection. *IEEE Transactions on Neural Networks and Learning Systems*, 30(5):1587–1595, 2018.

[Li *et al.*, 2019] Xuelong Li, Mulin Chen, and Qi Wang. Adaptive consistency propagation method for graph clustering. *IEEE Transactions on Knowledge and Data Engineering*, 32(4):797–802, 2019.

[Lin *et al.*, 2023] Zhiping Lin, Zhao Kang, Lizong Zhang, and Ling Tian. Multi-view attributed graph clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(2):1872–1880, 2023.

[Liu *et al.*, 2017] Hanyang Liu, Junwei Han, Feiping Nie, and Xuelong Li. Balanced clustering with least square regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, pages 2231–2237, 2017.

[Liu *et al.*, 2022a] Chaodie Liu, Feiping Nie, Rong Wang, and Xuelong Li. Scalable fuzzy clustering with anchor graph. *IEEE Transactions on Knowledge and Data Engineering*, 35(8):8503–8514, 2022.

[Liu *et al.*, 2022b] Suyuan Liu, Xinwang Liu, Siwei Wang, Xin Niu, and En Zhu. Fast incomplete multi-view clustering with view-independent anchors. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):8705–8717, 2022.

[Liu *et al.*, 2022c] Suyuan Liu, Siwei Wang, Pei Zhang, Kai Xu, Xinwang Liu, Changwang Zhang, and Feng Gao. Efficient one-pass multi-view subspace clustering with consensus anchors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 7576–7584, 2022.

[Liu *et al.*, 2025] Jiyuan Liu, Xinwang Liu, Chuankun Li, Xinhang Wan, Hao Tan, Yi Zhang, Weixuan Liang, Qian

Qu, Yu Feng, Renxiang Guan, and Ke Liang. Large-scale multi-view tensor clustering with implicit linear kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20727–20736, June 2025.

[Lu *et al.*, 2024] Jitao Lu, Feiping Nie, Rong Wang, and Xuelong Li. Fast multiview clustering by optimal graph mining. *IEEE Transactions on Neural Networks and Learning Systems*, 35(9):13071–13077, 2024.

[Nie *et al.*, 2014] Feiping Nie, Xiaoqian Wang, and Heng Huang. Clustering and projected clustering with adaptive neighbors. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 977–986, 2014.

[Nie *et al.*, 2017a] Feiping Nie, Jing Li, and Xuelong Li. Self-weighted multiview clustering with multiple graphs. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2564–2570, Melbourne, Australia, 2017. ijcai.org.

[Nie *et al.*, 2017b] Feiping Nie, Rui Zhang, and Xuelong Li. A generalized power iteration method for solving quadratic problem on the stiefel manifold. *Science China Information Sciences*, 60:112101:1–112101:10, 2017.

[Peng *et al.*, 2019] Yong Peng, Leijie Zhang, Wanzeng Kong, Feiping Nie, and Andrzej Cichocki. Joint structured graph learning and unsupervised feature selection. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3572–3576. IEEE, 2019.

[Qiang *et al.*, 2021] Qianyao Qiang, Bin Zhang, Fei Wang, and Feiping Nie. Fast multi-view discrete clustering with anchor graphs. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)*, pages 9360–9367, 2021.

[Shi *et al.*, 2014] Lei Shi, Liang Du, and Yi-Dong Shen. Robust spectral learning for unsupervised feature selection. In *IEEE International Conference on Data Mining (ICDM)*, pages 977–982. IEEE, 2014.

[Shi *et al.*, 2023] Shaojun Shi, Feiping Nie, Rong Wang, and Xuelong Li. Fast multi-view clustering via prototype graph. *IEEE Transactions on Knowledge and Data Engineering*, 35(1):443–455, 2023.

[Shi *et al.*, 2024] Long Shi, Lei Cao, Jun Wang, and Badong Chen. Enhanced latent multi-view subspace clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(12):12480–12495, 2024.

[Wan *et al.*, 2023] Xinhang Wan, Xinwang Liu, Jiyuan Liu, Siwei Wang, Yi Wen, Weixuan Liang, En Zhu, Zhe Liu, and Lu Zhou. Auto-weighted multi-view clustering for large-scale data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(8):10078–10086, Jun. 2023.

[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.*, 2022] Siwei Wang, Xinwang Liu, Suyuan Liu, Jiaqi Jin, Wenxuan Tu, Xinzhong Zhu, and En Zhu. Align then fusion: Generalized large-scale multi-view clustering with anchor matching correspondences. In *Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS 2022)*, pages 5882–5895, Red Hook, NY, USA, 2022. Curran Associates, Inc.

[Wen *et al.*, 2023] Yi Wen, Suyuan Liu, Xinhang Wan, Siwei Wang, Ke Liang, Xinwang Liu, Xihong Yang, and Pei Zhang. Efficient multi-view graph clustering with local and global structure preservation. In *Proceedings of the ACM International Conference on Multimedia*, pages 3021–3030. Association for Computing Machinery, 2023.

[Winn and Jojic, 2005] John M. Winn and Nebojsa Jojic. Locus: Learning object classes with unsupervised segmentation. In *The IEEE International Conference on Computer Vision*, volume 1, pages 1–8. IEEE, 2005.

[Yang *et al.*, 2023] Ben Yang, Xuetao Zhang, Zhongheng Li, Feiping Nie, and Fei Wang. Efficient multi-view k-means clustering with multiple anchor graphs. *IEEE Transactions on Knowledge and Data Engineering*, 35(7):6887–6900, 2023.

[Yang *et al.*, 2025] Ben Yang, Xuetao Zhang, Jinghan Wu, Feiping Nie, Zhiping Lin, Fei Wang, and Badong Chen. Fast multiview anchor-graph clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 36(3):4947–4958, 2025.

[Zhang *et al.*, 2020] Rui Zhang, Xuelong Li, Tong Wu, and Yi Zhao. Data clustering via uncorrelated ridge regression. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):450–456, 2020.

[Zhang *et al.*, 2024] Chao Zhang, Ding Xu, Xiaojie Jia, Chao Chen, and Hongzhi Li. Continual multi-view clustering with consistent anchor guidance. In *Proceedings of the 33rd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5434–5442. International Joint Conferences on Artificial Intelligence Organization, 2024.

[Zhao *et al.*, 2016] Haifeng Zhao, Zheng Wang, and Feiping Nie. Orthogonal least squares regression for feature extraction. *Neurocomputing*, 216:200–207, 2016.

[Zhao *et al.*, 2025] Mingyu Zhao, Feiping Nie, Cong Wang, and Xuelong Li. Balanced and discrete multi-view clustering with adaptive graph learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.

[Zhao *et al.*, 2026] Xiaowei Zhao, Liuyun Guo, Xiaojun Chang, Jun Guo, Feiping Nie, and Qiang Zhang. Efficient co-clustering via bipartite graph factorization. *IEEE Transactions on Knowledge and Data Engineering*, 38(3):1695–1709, 2026.