Incomplete Multi-View Clustering via Structure Exploration and Missing-view Inference

Ziyu Wang[a], Lusi Li[a,*], Xin Ning[b], Wenkai Tan[c], Yongxin Liu[c], Houbing Song[d]

[a]*Department of Computer Science, Old Dominion University, Norfolk, VA 23529 USA*
[b]*Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083 China*
[c]*Department of Electrical Engineering and Computer Science, Embry-Riddle Aeronautical University, FL 32114 USA*
[d]*Department of Information Systems, University of Maryland, Baltimore County, MD 21250 USA*

Abstract

Incomplete multi-view clustering (IMVC) aims to boost clustering performance by capturing complementary information from incomplete multi-views, where partial data samples in one or more views are missing. Current IMVC methods mostly impute missing samples at the guidance of the global/local structure or directly learn a common representation without imputation using subspace or graph learning techniques. However, the consistent and inconsistent structures across views are often ignored during imputation, leading to the introduction of noise and biases. Additionally, lacking the handling of missing samples would mislead the learning methods and degrade clustering performance. To this end, we propose a novel approach called Structure Exploration and Missing-view Inference (SEMI) for IMVC. Specifically, SEMI explores the underlying multi-structures of data, including global, local, consistent, and inconsistent structures, by jointly modeling self-expression subspace, graph, and clustering-oriented partition learning. This enables the capture of consistent and discriminative information and fuses it into a unified coefficient matrix. The learned coefficient matrix with the explored multi-structures then guides the inference

*Corresponding author.
Email addresses: zwang007@odu.edu (Ziyu Wang), lusili@cs.odu.edu (**Lusi Li**), ningxin@semi.ac.cn (Xin Ning), tanw1@my.erau.edu (Wenkai Tan), liuy11@erau.edu (Yongxin Liu), and songh@umbc.edu (Houbing Song).

of missing views, facilitating the alleviation of the influence of existing noise and biases and the mitigation of the introduction of further noise and biases. These two components are seamlessly integrated and mutually improved through an efficient alternating optimization strategy. Experimental results demonstrate the effectiveness and superior performance of the proposed method.
Keywords: Data imputation, multi-structure exploration, consistent and discriminative information, guidance, noise, and biases.

1. Introduction

Multi-view clustering (MVC) aims to partition multi-view data samples into different clusters by learning a common representation and then employing clustering on this representation. In real-world applications, multi-view data consists of features collected from multiple sources (or views), providing rich and complementary information. It is crucial and valuable to build a robust MVC model that can fuse this information. Many MVC models, such as co-training-based [1], kernel-based [2, 3, 4], subspace-based [5, 6], graph-based [7, 8], and tensor-based [9, 10] approaches, have been proposed to improve performance in various tasks [11]. Most of them have achieved some success because they operate under the rigorous assumption of multi-view data completeness. However, in practice, it's not uncommon to encounter incomplete data due to sensor failures or human errors. Applying these MVC methods directly to such data can significantly degrade clustering performance or even cause failures. Addressing missing views remains a key challenge in incomplete multi-view clustering (IMVC).

To address incomplete views, several IMVC methods have been introduced, encompassing two primary technical strategies: omitting unavailable views and imputing missing data using available information. For the first strategy, the majority of methods emphasize clustering while disregarding the missing views to prevent the introduction of noise and biases. This is because ensuring the quality of imputed views is challenging [12, 13]. For instance, [14, 15] intro-

duced a method to learn a latent subspace via non-negative matrix factorization (NMF) using the data available in two incomplete views. To minimize the adverse effects of missing views in datasets with more than two views, [16, 17, 18] devised a weighting mechanism that assigns lower weights to samples with missing views. However, the absence of handling for missing views can misdirect learning methods, leading to decreased clustering performance. For the second strategy, efforts have been made to recover missing samples either in the original data space or a shared subspace [19, 20, 21]. For example, [22] suggested a method that jointly undertakes self-representation learning and missing data imputation in the original data space. This approach considers the guidance of the global structure but overlooks that of the local structure. [23] proposed alternating between kernel refining and embedding updating. Meanwhile, [24] integrated the completion of the incomplete basis matrix with later clustering. However, a shared drawback of these kernel-based IMVC methods is the necessity for substantial prior knowledge to choose a suitable kernel function for discriminative feature extraction. [25, 26] aimed to complete incomplete view graphs and integrate them into a consensus graph. Yet, each view graph serves as a sub-graph of the global graph structure, potentially containing biased view-specific information. This can adversely influence the final clustering outcome in the absence of consistent structure guidance across views.

Previous MVC studies have underscored the significant improvement in clustering performance resulting from the preservation of a consistent structure [27, 28]. To capture this consistent structure, it's essential to consider potential instances of inconsistent structures. These can arise from factors such as noise, errors, biases, corruption, or view-specific characteristics embedded within the dataset. For instance, the authors in [29, 30] proposed two distinct graph-based MVC methods. Both methods addressed the inconsistent components by decomposing each view's similarity matrix into a combination of a consistent similarity matrix and an inconsistent matrix.

A recent endeavor in this area is the incorporation of tensor completion techniques to learn the view-consensus graph [9, 10]. For instance, a tensor

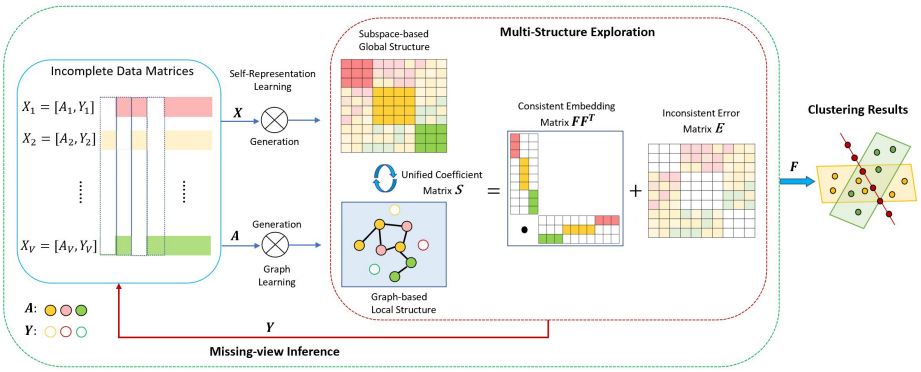

Figure 1: The architecture of our proposed SEMI method. Given an incomplete multi-view data $X = [A, Y]$ with un-missing data A and missing data Y, SEMI learns a unified coefficient matrix S by jointly performing the exploration of global, local, consistent, and inconsistent structures as well as the inference of missing views Y. At last, the clustering results are produced from the consistent embedding matrix F.

Schatten p-norm-based completion technique has been proposed to leverage both the spatial structure and the complementary information contained within the similarity matrices of different views [10]. While this method ensures that the resulting overarching graph not only captures the inter-view similarity structure but also aligns with the intra-view similarity, it has its own limitations. For high-dimensional datasets, the method struggles to effectively explore both global and local structures simultaneously. Moreover, in its effort to construct the graph, it does not delve deeply into the consistency and correlations between views, potentially missing crucial nuances.

Turning back to the decomposition approaches, they predominantly focus on the inconsistencies intrinsic to individual views and can achieve comparable results when the data is complete. However, in scenarios involving missing data across views, exploring the consistency structure becomes even more critical [31]. Incomplete data heightens the risk of divergent clustering outcomes from different views since certain non-missing samples are interpreted as inconsistent instances. This introduces challenges in fusing view information. Thus, effectively leveraging the observable information presented across different views to

explore both consistent and inconsistent structures is indispensable for achieving efficient and meaningful results, especially when faced with the diverse complexities of incomplete data.

To address these challenges, we introduce a novel approach named Structure Exploration and Missing-view Inference (SEMI) that jointly explores the underlying global, local, and consistent structures, infers the missing views under the guidance of the captured multi-structures, and performs clustering. The overall architecture is depicted in Fig. 1. Specifically, SEMI first uncovers the global structure using a self-representation learning technique and procures a unified coefficient matrix, which captures the relationships among all data samples across views. Subsequently, SEMI allows the unified coefficient matrix to capture the local structure using the available non-missing data samples. This aids in ensuring genuine relationship information among non-missing samples and mitigating the introduction of new noise and biases. To diminish the effects of specific information and examine the discriminative information within the consistent structure, we decompose the coefficient matrix into two parts: a consistent section represented by an embedding matrix and an inconsistent section denoted by an error matrix. Following this, the missing views are inferred under the guidance of the coefficient matrix with multiple structures. The final cluster assignments are derived from the embedding matrix. In summary, our contributions are as follows:

- We propose a novel unified approach named SEMI, which jointly integrates the exploration of the global, local, consistent, and inconsistent structures to capture consistent and discriminative information across views.
- We introduce the inference of missing views under the structure guidance. The captured multi-structures can guide the missing-view inference, and the completed data can also facilitate structure exploration.
- We exploit an efficient alternation optimization method to iteratively optimize the involved variables. SEMI can achieve highly competitive clustering results on various multi-view datasets with different missing ratios.

The remainder of the paper is organized as follows: Section 2 presents related work, encompassing both complete and incomplete multi-view clustering methods. In Section 3, we introduce our SEMI approach. The optimization strategy, as well as the convergence and complexity analysis of SEMI, are detailed in Section 4. Section 5 discusses the experiments and their results. Finally, Section 6 concludes the paper.

2. Related Works

Before we present the related work, the main notations are as follows. Throughout the paper, we denote scalars, vectors, and matrices using the following conventions: italic lowercase letters (e.g., x) for scalars, lowercase letters (e.g., x) for vectors, and italic capital letters (e.g., X) for matrices. Let us consider a data matrix with n data samples, denoted as $X \in \mathbb{R}^{d\times n}$. Thereinto, $\mathrm{x}^{(\mathrm{i})} \in \mathbb{R}^{\mathrm{d}\times 1}$ represents the i-th column vector, and $x^{(ij)} \in \mathbb{R}^{1\times 1}$ represents the (i, j)-th entry. For any matrix X, we define the trace operation as $Tr(X)$, the transpose as X^T, and the Frobenius norm as $||X||_F$. The Lp, q-norm of X is denoted as $||X||p, q = \sum i = 1^n((\sum_{j=1}^{d} |x^{(ij)}|^p)^{q/p})^{1/q}$. Additionally, $\mathbf{I}$ represents the identity matrix, and $\mathbf{1}$ represents a vector with all entries equal to one.

Given a complete multi-view data set $\{X_v\}_{v=1}^{V}$ with V views and c clusters, $X_v = [\mathrm{x}_v^{(1)}, \ldots, \mathrm{x}_v^{(n)}] \in \mathbb{R}^{d_v\times n}$ represents the v-th view data, where its dimension is d_v and its number of data samples is n. Partly data samples of one or more views are missing for an incomplete data setting.

In this section, we briefly review two topics related to this work, including self-expression subspace multi-view clustering and graph-based multi-view clustering.

2.1. Self-Expression Subspace Multi-View Clustering

In a single-view setting, given a data set $X \in \mathbb{R}^{d\times n}$, self-expression subspace clustering aims to learn a coefficient matrix S by reconstructing each data sample using a linear combination of the other samples, which can be formulated

into

$$\min_{S} \mathfrak{L}(X - XS) + \lambda \mathfrak{R}(S), \quad s.t., \Phi(S), \tag{1}$$

where the coefficient matrix $S \in \mathbb{R}^{n \times n}$ contains the underlying relationships across all data samples, which can be considered as the global structure of the data; $\mathfrak{L}$ can be any kind of norms [32, 33, 34, 35], such as l_1-norm, l_2-norm, etc; $\mathfrak{R}$ can be any kind of regularization for S like a graph regularization [36, 37], aiming to introduce constraints or penalties to the learning of S in order to impose specific structures on the learned models; Φ denotes the feasible regions of the values of S; λ is a trade-off parameter to balance the two terms. With the learned S, the cluster assignments can be computed.

In a complete multi-view setting, given a complete data set $\{X_v\}_{v=1}^V$, it is natural to extend self-expression subspace clustering in Eq. (1) into self-expression subspace MVC [38, 39] as follows

$$\begin{aligned} &\min_{S} \sum_{v=1}^{V} \mathfrak{L}(X_v - X_v S_v) + \alpha \sum_{v=1}^{V} \mathfrak{L}'(S_v - S) + \lambda \mathfrak{R}(\{S_v\}_{v=1}^V, S), \\ &s.t., \Phi(\{S_v\}_{v=1}^V, S), \end{aligned} \tag{2}$$

where $S_v \in \mathbb{R}^{n \times n}$ for each view v; $\mathfrak{L}'$ can be any kind of norms to force each S_v to S; $\mathfrak{R}$ indicates a regularization for both $\{S_v\}_{v=1}^V$ and S; α and λ are trade-off parameters to balance the three terms. It can be observed that coefficient matrices $\{S_v\}_{v=1}^V$ are fused into a common matrix S, which refers to the overarching relationships across all views, indicating the global structure of the multi-view data. In an incomplete multi-view setting, each S_v is learned from the available data samples, and a unified S is obtained by fusing these individual S_v matrices [14, 17]. However, missing views are ignored, resulting in an inadequate exploration of the global structure [18, 20, 22]. With increased missing ratios, the global structural quality becomes compromised, leading to the potential introduction of more noise and biases.

2.2. Graph-based Multi-View Clustering

A graph is a crucial data structure used to depict the connections between data samples. In a graph, each sample is represented by a node, while the

relationships between two samples are represented by edges [40, 41, 42, 43]. Graph-based approaches exhibit robust adaptability when handling incomplete multi-view data. They can construct graph structures leveraging available view information without significant perturbation from missing data. This endows them with enhanced robustness in practical applications, especially in scenarios with data absence in certain views. In a single-view setting, a popular method is to transform the data X with similarities $S \in \mathbb{R}^{n\times n}$ into a graph through

$$\min_{S} \sum_{i,j=1}^{n} \mathfrak{L}(x^{(i)} - x^{(j)}) s^{(ij)} + \lambda \mathfrak{R}(S), \;\; s.t., \Phi(S), \tag{3}$$

where each entry $s^{(ij)}$ in S represents the similarity probability between $x^{(i)}$ and $x^{(j)}$ in which a larger $s^{(ij)}$ means $x^{(i)}$ is closer to $x^{(j)}$; $\mathfrak{L}$, $\mathfrak{R}$, and Φ can be instanced with multiple norms, regularization, and regions. Generally, the graph is often constrained into a k-nearest neighbor graph to capture the local structure of the data, where $x^{(i)}$ and $x^{(j)}$ are connected with the probability $s^{(ij)}$ only if $x^{(j)}$ belongs to the k-nearest neighbors of $x^{(i)}$. The learned graph S can be partitioned to obtain the cluster assignments by a partitioning strategy like spectral clustering.

Similarly, we can have graph-based MVC by extending Eq. (3) into the following problem

$$\begin{aligned} &\min_{S} \sum_{v=1}^{V} \sum_{i,j=1}^{n} \mathfrak{L}(x_v^{(i)} - x_v^{(j)}) s_v^{(ij)} + \alpha \sum_{v=1}^{V} \mathfrak{L}'(S_v - S) \; + \lambda \mathfrak{R}(\{S_v\}_{v=1}^{V}, S), \\ &s.t., \Phi(\{S_v\}_{v=1}^{V}, S), \end{aligned} \tag{4}$$

where $\mathfrak{L}$ and $\mathfrak{L}'$ are instanced with multiple norms; $\mathfrak{R}$ and Φ are instanced with multiple regularizations and feasible regions; The consensus graph S across all views can be formed by fusing $\{S_v\}_{v=1}^{V}$, which contain the learned local structure in each view. However, each view graph S_v is unable to completely encapsulate the local structure in an incomplete multi-view setting [26, 44, 45, 46], leading to the emergence of distinct view graphs, and ultimately causing a decline in the quality of the fused graph.

To mitigate the challenges, our proposed method aims to jointly explore

the global, local, and consistent structures, while also accounting for potential inconsistencies. Subsequently, the learned structure is employed to guide the imputation of missing samples. This process of structure exploration and missing-view inference is designed to be mutually reinforcing.

3. Proposed Method

With incomplete multi-view data, the authentic un-missing data samples are denoted by $\{A_v\}_{v=1}^V \in \mathbb{R}^{d_v \times n_v}$ and the missing data samples are denoted by $\{Y_v\}_{v=1}^V \in \mathbb{R}^{d_v \times (n-n_v)}$ for v-th view, where X_v is the horizontal concatenation of A_v and Y_v, i.e., $X_v = [A_v, Y_v]$.

Our proposed SEMI method aims to group the incomplete data $\{X_v = [A_v, Y_v]\}_{v=1}^V$ into c clusters by jointly imputing the missing data samples for each view Y_v and preserving the structures of un-missing data samples. Specifically, SEMI models global-structure learning, authentic local-structure learning, structure-consistency learning, and missing data imputation into a unified framework.

3.1. Global-Structure Learning

Given complete multi-view data $\{X_v\}_{v=1}^V$, we utilize a self-expression subspace learning technique to explore the global structure of different views. Unlike Eq. (2), each data sample can be represented by a linear combination of other samples within the same view through a unified coefficient matrix S, not S_v. Hence, S can be constructed by utilizing the full data samples from multiple views as below:

$$\begin{aligned} &\min_{S} \sum_{v=1}^{V} ||X_v - X_v S||_F^2 + \lambda ||S||_F^2, \\ &s.t., \forall i, s^{(ii)} = 0, \mathbf{1}^T \mathrm{s}^{(\mathrm{i})} = 1, \mathrm{S} \geq 0, \end{aligned} \tag{5}$$

where the coefficient matrix $S \in \mathcal{R}^{n \times n}$ contains the relationships among all data samples across views and can be used to represent the global structure of multi-view data; a l_2 norm is employed for $\mathfrak{L}$ and $\mathfrak{R}$, respectively; a diagonal constraint

is imposed on S to avoid a trivial solution, which only uses the sample itself to reconstruct each data sample; $s^{(ij)} \geq 0$ is to make S non-negative; $\mathbf{1}^T \mathrm{s}^{(\mathrm{i})} = 1$ can be considered as a sparse constraint on S; λ is a balancing parameter.

Unlike MVC on the complete data, for IMVC on incomplete data, each view data consists of authentic un-missing data and missing data. We aim to learn the global structure across views contained in S as follows:

$$\begin{aligned} &\min_{S} \sum_{v=1}^{V} ||X_v - X_v S||_F^2 + \lambda ||S||_F^2 \\ &s.t., \forall v, X_v = [A_v, Y_v],\ \forall i, s^{(ii)} = 0, \mathbf{1}^T \mathrm{s}^{(\mathrm{i})} = 1, \mathrm{S} \geq 0, \end{aligned} \tag{6}$$

where $A_v \in \mathbb{R}^{d_v \times n_v}$ refers to the un-missing data in X_v; $Y_v \in \mathbb{R}^{d_v \times (n - n_v)}$ indicates the missing data in X_v.

3.2. Missing-view Inference

It can be noted that learning global structure in S is difficult on an incomplete data set. The more missing samples in each X_v, the more challenging learning S is. Hence, we propose to complete the data set by performing an imputation for each view with missing samples while learning the global structure. Specifically, we can first utilize a zero-filling strategy to impute each missing sample with an all-zero vector to learn the coefficient matrix S and then iteratively use the improved S to update Y_v for each view. Hence, we add a variable Y in Eq. (6) and have

$$\begin{aligned} &\min_{S,Y} \sum_{v=1}^{V} ||X_v - X_v S||_F^2 + \lambda ||S||_F^2 \\ &s.t., \forall v, X_v = [A_v, Y_v],\ \forall i, s^{(ii)} = 0, \mathbf{1}^T \mathrm{s}^{(\mathrm{i})} = 1, \mathrm{S} \geq 0, \end{aligned} \tag{7}$$

where $Y = \{Y_v\}_{v=1}^{V}$ represents a set of missing data for all the views. In this way, the imputations of missing samples Y can benefit from the guidance of the learned global structure S. To alleviate the influence of view-specific information and the introduction of biases, we would further enable the coefficient matrix S to contain local and consistent structures.

3.3. Local-Structure Learning

With global structure learning and missing-view inference, each data sample in v-th view can be linearly reconstructed by the entire data set X_v that contains the authentic un-missing samples A_v and the imputed samples Y_v. The S learned from Eq. (6) can explore the global information of multiple views with A_v and Y_v but ignores the local structure among un-missing samples. Coexisting with completing the incomplete view data, we also need to guarantee authentic relationship information between the un-missing samples in S. It is conducive to achieving a more reliable imputation and eliminating the possible introduced negative impacts by the imputed missing views. Generally, two different data samples are closer and are likely to have more similar representations. Unlike the graph-based MVC in Eq. (4), we simplify the learning of each view graph and directly make the unified coefficient matrix S involve the overall local structure. Hence, we integrate graph learning into Eq. (7) to simultaneously capture the local structure of each view with only authentic un-missing samples.

$$\begin{aligned} &\min_{S,Y} \sum_{v=1}^{V} ||X_v - X_v S||_F^2 + \lambda ||S||_F^2 + \gamma \sum_{v=1}^{V} \sum_{i,j=1}^{n_v} ||a_v^{(i)} - a_v^{(j)}||_F^2 z_v^{(ij)} \\ &s.t., \forall v, X_v = [A_v, Y_v], Z_v = G_v S G_v^T, \ \forall i, s^{(ii)} = 0, \mathbf{1}^T \mathrm{s}^{(\mathrm{i})} = 1, \mathrm{S} \geq 0, \end{aligned} \tag{8}$$

where γ is a balancing parameter and n_v is the number of un-missing data samples in v-th view; $Z_v \in \mathcal{R}^{n_v \times n_v}$ and $Z_v = G_v S G_v^T$ represents the authentic relationship matrix; $G_v \in \mathcal{R}^{n_v \times n}$ is a prior index matrix, where the entries related to the missing samples are zero. The matrix G_v can be defined as below.

$$g_v^{(ij)} = \begin{cases} 1, & \text{if } \mathrm{a_v^{(i)}} \text{ is the original } \mathrm{x_v^{(j)}} \\ 0, & otherwise \end{cases}$$

3.4. Consistent and Inconsistent Structure Learning

With the global- and local-structure learning, the obtained unified coefficient matrix S can capture the global and local information across views. To further reduce the influence of unreliable and irrelevant information in completed multi-view data, we propose to model the consistent and inconsistent

structures jointly. Specifically, the coefficient matrix S can be decomposed into a consistent part and an inconsistent part, respectively.

Since this task is a clustering-oriented partition learning and the number of clusters is known to be c, we can push the consistent part towards FF^T in which $F \in \mathcal{R}^{n\times c}$ is an orthogonal partition matrix across views. The inconsistent part is expected to involve noise, errors, view-specific information in the un-missing samples, and view-specific information introduced by the imputed samples, which highly affect the quality of the learned structure. Hence, the separation between the consistent and inconsistent structures is necessary. Given S, we decompose it into these two parts:

$$S = FF^T + E. \tag{9}$$

Then we can get F and E by

$$\min_{F,E} ||E||_1,\ s.t., S = FF^T + E, F^T F = I_c, \tag{10}$$

where $E \in \mathcal{R}^{n\times n}$ is used to model the inconsistent information by utilizing ι_1 norm on it, which can enforce E to capture sparse inconsistent structure across different views.

3.5. Overall Objective Function

By integrating the consistent-structure learning Eq. (10) into Eq. (8), the learned coefficient matrix S involves the information of the global, local, and consistent structures, which will be more conducive to the inference of missing views. The overall objective function can be found as follows:

$$\begin{aligned}
&\min_{S,Y,F,E} \sum_{v=1}^{V} ||X_v - X_v S||_F^2 + \lambda ||S||_F^2 + \gamma \sum_{v=1}^{V} \sum_{i,j=1}^{n_v} ||a_v^{(i)} - a_v^{(j)}||_F^2 z_v^{(ij)} + \beta ||E||_1 \\
&s.t., \forall v, X_v = [A_v, Y_v], Z_v = G_v S G_v^T,\ \forall i, s^{(ii)} = 0, \mathbf{1}^T \mathrm{s}^{(\mathrm{i})} = 1, \\
&S \geq 0,\ S = FF^T + E, F^T F = I_c,
\end{aligned} \tag{11}$$

where λ, γ, and β are trade-off balancing parameters; there are four variables S, Y, F, and E that need to be optimized.

4. Optimization Strategy

Since the variables in Eq. (11) are interdependent, obtaining an optimized solution for each variable is a challenging task. To tackle this challenge, the alternating direction method of multipliers (ADMM) is utilized to update the variables iteratively. We can first convert Eq. (11) into the following augmented Lagrangian function:

$$
\begin{aligned}
&L(S,Y,F,E) = \sum_{v=1}^{V} ||X_v - X_v S||_F^2 + \lambda ||S||_F^2 \\
&+ \gamma \sum_{v=1}^{V} \sum_{i,j=1}^{n_v} ||a_v^{(i)} - a_v^{(j)}||_F^2 z_v^{(ij)} + \beta ||E||_1 + \frac{\alpha}{2} ||S - FF^T - E + \frac{P}{\alpha}||_F^2 \\
&s.t., \forall v, X_v = [A_v, Y_v], Z_v = G_v S G_v^T, \ \forall i, s^{(ii)} = 0, \mathbf{1}^T \mathrm{s^{(i)}} = 1, \ \mathrm{S} \geq 0, \mathrm{F^T F} = \mathrm{I_c},
\end{aligned}
\tag{12}
$$

where the matrix P is the Lagrange multiplier and α is a penalty parameter. When updating one variable, the other three variables updated before can be treated as constants in the current step. We focus on optimizing the four variables S, Y, F, and E, and one Lagrange multiplier P. The updated rules are described in the subsequent sections.

4.1. Update F

When we fix S, Y, E, and P, updating F is to solve the augmented Lagrangian function

$$
L(F) = ||S - FF^T - E + \frac{P}{\alpha}||_F^2 \ s.t., F^T F = I_c, \tag{13}
$$

Based on the trace operator's properties, we denote $H = S - E + \frac{P}{\alpha}$, and the term can be written as below

$$\begin{aligned}
&||S - FF^T - E + \frac{P}{\alpha}||_F^2 \\
&= ||H - FF^T||_F^2 \\
&= Tr((H - FF^T)^T(H - FF^T)) \\
&= Tr(H^TH - H^TFF^T - FF^TH + FF^TFF^T) \\
&= Tr(H^TH) - Tr(H^TFF^T) - Tr(FF^TH) + Tr(FI_cF^T) \\
&= Tr(H^TH) - Tr(F^TH^TF) - Tr(F^THF) + Tr(FI_cF^T) \\
&= Tr(H^TH) - Tr(F^T(H^T + H)F) + Tr(FI_cF^T)
\end{aligned} \tag{14}$$

Notice that $Tr(H^TH)$ is a constant with respect to F. To update F, we can drop this term and reduce Eq. (13) into the following problem for simplicity

$$F \approx \underset{F^TF=I_c}{\operatorname{argmin}} Tr(F^T(H^T + H)F) \tag{15}$$

The approximated optimal solution for F is formed by the c eigenvectors of $(H^T + H)$ corresponding to the largest c eigenvalues.

4.2. Update S

When F, Y, E, and P are fixed, X_v is fixed and the augmented Lagrangian function in terms of S can be

$$\begin{aligned}
&L(S) = \sum_{v=1}^{V} ||X_v - X_vS||_F^2 + \lambda||S||_F^2 + \gamma \sum_{v=1}^{V} \sum_{i,j=1}^{n_v} ||a_v^{(i)} - a_v^{(j)}||_F^2 z_v^{(ij)} \\
&+ \frac{\alpha}{2}||S - FF^T - E + \frac{P}{\alpha}||_F^2 \\
&s.t., \forall v, Z_v = G_vSG_v^T;\ \forall i, s^{(ii)} = 0, \mathbf{1}^T\mathrm{s}^{(\mathrm{i})} = 1, \mathrm{S} \geq 0
\end{aligned} \tag{16}$$

A two-step approximation strategy is utilized to optimize S. Specifically, we can first use $\widehat{S}$ denote S without considering the constraints on S and then push $\widehat{S}$ to approximate S by adding these constraints. The problem in Eq. (16) can be

split into two problems: Eq. (17) and Eq. (18) as follows.

$$
\begin{aligned}
&L(\widehat{S}) = \sum_{v=1}^{V} ||X_v - X_v\widehat{S}||_F^2 + \lambda||\widehat{S}||_F^2 + \gamma\sum_{v=1}^{V}\sum_{i,j=1}^{n_v} ||a_v^{(i)} - a_v^{(j)}||_F^2 z_v^{(ij)} \\
&+ \frac{\alpha}{2}||\widehat{S} - FF^T - E + \frac{P}{\alpha}||_F^2 \\
&s.t., \forall v, Z_v = G_v\widehat{S}G_v^T
\end{aligned}
\tag{17}
$$

$$
S = \underset{\forall i, s^{(ii)}=0, \mathbf{1}^T \mathrm{s}^{(\mathrm{i})}=1, \mathrm{S}\geq 0}{\operatorname{argmin}} ||S - \widehat{S}||_F^2 \tag{18}
$$

To solve Eq. (17), we take the derivative of $L(\widehat{S})$ as zero and then have

$$
\widehat{S} = (Q + (\lambda + \frac{\alpha}{2})\mathbf{I}_n)^{-1}(Q + \frac{\alpha}{2}C - \gamma\frac{1}{2}\sum_{v=1}^{V} G_v^T B_v^T G_v) \tag{19}
$$

where $Q = \sum_{v=1}^{V} X_v^T X_v$, $C = FF^T + E - \dfrac{P}{\alpha}$, and $b_v^{(ij)} = ||a_v^{(i)} - a_v^{(j)}||_F^2$ is the ij-th element of $B_v \in \mathcal{R}^{n_v \times n_v}$.

With the obtained $\widehat{S}$, S can be updated by solving the constrained quadratic optimization problem in Eq. (18) with the Karush–Kuhn–Tucker (KKT) condition. For each row,

$$
s^{(i)} = \underset{\forall i, s^{(ii)}=0, \mathbf{1}^T \mathrm{s}^{(\mathrm{i})}=1, \mathrm{S}\geq 0}{\operatorname{argmin}} ||s^{(i)} - \widehat{s^{(i)}}||_F^2 \tag{20}
$$

Its Lagrangian function can be found as follows

$$
L(s^{(i)}, \mu, \sigma) = \frac{1}{2}||s^{(i)} - \widehat{s^{(i)}}||_F^2 - \mu(\mathbf{1}^T \mathrm{s}^{(\mathrm{i})} = 1) - \sigma \mathrm{s}^{(\mathrm{i})} \tag{21}
$$

Based on the KKT conditions, we have

$$
\begin{cases}
\forall j, \quad s^{(ij)} - \widehat{s^{(ij)}} - \mu - \sigma_j = 0 \\
\forall j, \quad s^{(ij)} \leq 0 \\
\forall j, \quad \sigma_j \leq 0 \\
\forall j, \quad \sigma_j s^{(ij)} = 0
\end{cases}
$$

Hence we can have the solution of $s^{(i)}$ as

$$
\forall j, s^{(ij)} = (\widehat{s^{(ij)}} + \mu)_+, \tag{22}
$$

where the optimal μ is the root of $\sum_j (\widehat{s^{(ij)}} + \mu)_+ - 1$.

4.3. Update Y

Fixing F, S, E and P, the function of Eq. (12) can be transformed into

$$L(Y) = \sum_{v=1}^{V} ||X_v - X_v S||_F^2 \ s.t., X_v = [A_v, Y_v] \tag{23}$$

where

$$\begin{aligned} &||X_v - X_v S||_F^2 = Tr((X_v - X_v S)^T (X_v - X_v S)) \\ &= Tr(X_v^T X_v) - 2Tr(S X_v^T X_v) + Tr(S^T X_v^T X_v S) \\ &= Tr(X_v^T X_v) - 2Tr(X_v^T S X_v) + Tr(S^T X_v^T X_v S) \\ &= Tr(X_v^T X_v) - Tr(X_v^T S^T X_v) - Tr(X_v^T S X_v) + Tr(X_v^T S S^T X_v) \\ &= Tr(X_v^T (\mathbf{I}_n - S^T - S + S S^T) X_v). \end{aligned} \tag{24}$$

From Eq. (23), each Y_v in Y can be updated independently by solving

$$L(Y_v) = Tr([A_v, Y_v]^T (\mathbf{I}_n - S^T - S + S S^T)[A_v, Y_v]) \tag{25}$$

Let $Z = (\mathbf{I}_n - S^T - S + S S^T)$. Then we can convert Z into a block matrix since it is a square matrix as below

$$Z = \begin{bmatrix} Z_{aa} & Z_{ay} \\ Z_{ya} & Z_{yy} \end{bmatrix}$$

where $Z_{aa} \in \mathcal{R}^{n_v \times n_v}$ refers to the block corresponding to the relationships between available samples; $Z_{ay} \in \mathcal{R}^{n_v \times (n-n_v)}$ is the block matrix corresponding to the relationships between available samples and missing samples; $Z_{ya} \in \mathcal{R}^{(n-n_v) \times n_v}$ is the block matrix corresponding to the relationships between missing samples and available samples; $Z_{yy} \in \mathcal{R}^{(n-n_v) \times (n-n_v)}$ refers to the block corresponding to the relationships between missing samples. Adding the block matrix Z into Eq. (25), we can have

$$L(Y_v) = Tr(Y_v Z_{yy} Y_v^T + A_v (Z_{ay} + Z_{ya}^T) Y_v^T) \tag{26}$$

Through making the derivative of Eq. (26) in terms of Y_v as zero, we can get

$$Y_v = -A_v Z_{ay} Z_{yy}^{-1} \tag{27}$$

where $Z_{yy} = (\mathbf{I}_n - S)_y^T (\mathbf{I}_n - S)_y$ is positive-definite.

4.4. Update E

When fixing F, S, Y, and P, the sub-problem to solve E can be found below

$$L(E) = \beta||E||_1 + \frac{\alpha}{2}||S - FF^T - E + \frac{P}{\alpha}||_F^2 \tag{28}$$

We have $\Gamma = S - FF^T + \frac{P}{\alpha}$. The closed-form solution of Eq. (28) can be obtained by applying the shrinkage operator [47] to each element of H

$$E = shrink(\Gamma, \frac{\beta}{\alpha}) \tag{29}$$

4.5. Update P

The Lagrange multiplier matrix P can be updated via

$$P = P + \alpha(S - FF^T - E) \tag{30}$$

where the penalty parameter $\alpha = \min(\rho\alpha, \alpha_0)$ in which ρ and α_0 are constants.

Algorithm 1 summarizes the overall procedure presented above. After obtaining the partition matrix F, we employ the k-means algorithm on it to obtain the final cluster assignments.

4.6. Convergence and Complexity Analysis

1) The convergence of SEMI. SEMI is optimized by the ADMM framework, which encompasses four sub-problems concerning F, S, Y, and E. Unfortunately, establishing a theoretical proof for the convergence of ADMM in multi-block optimization [48] poses challenges. Nevertheless, by following similar procedures [44], the convergence can be guaranteed given that the Lagrangian function in Equation (12) is Lipschitz differentiable and all sub-problems exhibit strong convexity. Additionally, in Section V, convergence analysis experiments are performed to demonstrate the convergence property.

The primary convergence criterion is the stability of the objective function across consecutive iterations. Specifically, the algorithm monitors the absolute relative change in the objective function value. If the change between two successive iterations is smaller than a predefined threshold, ϵ (set to 1×10^{-4},

Algorithm 1 SEMI Optimization Method

Input: Incomplete multi-view dataset of V views $\{X_v\}_{v=1}^V$ with $X_v = [A_v, Y_v] \in \mathbb{R}^{d_v \times n}$, the number of clusters c, parameters λ, γ, β, $\alpha = 0.01$, $\alpha_0 = 10^5$, $\rho = 1.1$, and $\epsilon = 1 \times 10^{-4}$.

Output: The partition matrix F

Initialize the index matrices $\{G_v\}_{v=1}^V$.
Initialize the missing matrices $\{Y_v\}_{v=1}^V$ with a zero filling strategy.
Initialize the coefficient matrix S.
repeat
  Fix S, $\{Y_v\}_{v=1}^V$, E, and P, update F by Eq. (15).
  Fix F, $\{Y^v\}_{v=1}^m$, E, and P, update S by Eq. (22).
  Fix F, S, E, and P, update $\{Y^v\}_{v=1}^m$ by Eq. (27).
  Fix F, S, $\{Y^v\}_{v=1}^m$, and P, update E by Eq. (29).
  Fix F, S, $\{Y^v\}_{v=1}^m$, and E, update P by Eq. (30).
until $\tau < \epsilon$
Calculate the cluster assignments with F.

the algorithm is considered to have converged. Mathematically, the condition is given by:

$$\tau = \left| \frac{\texttt{obj}(t-1) - \texttt{obj}(t)}{\texttt{obj}(t)} \right| < \epsilon \tag{31}$$

An alternative is to set a maximum iteration limit.

2) The complexity of SEMI. From Algorithm 1, the computational complexity of our proposed SEMI method includes five parts, which are the initialization and updates of our variables. To be more specific, updating F takes $O(cn^2)$, where c is the number of required clusters and n is the number of data objects. The update of S takes $O(n^3)$. Updating $\{Y^v\}_{v=1}^V$ takes $O(Vn_v(n-n_v)^2 + Vd_vn_v(n-n_v))$. Updating E needs to cost $O(n^2)$. Overall, the computational complexity of SEMI can be simplified into $O(n^3)$. For the space complexity, SEMI takes $O(n^2)$ to store F, S, E, P, and other auxiliary variables and takes $O(V(n-n_v)^2)$ to store $\{Y^v\}_{v=1}^V$.

5. Experiments

Extensive experiments are carried out in this section to validate the effectiveness of our proposed SEMI approach.

5.1. Experiment Setting

5.1.1. Datasets

To evaluate the performance of our proposed SEMI method, we use 5 complete benchmark multi-view datasets to generate incomplete multi-view datasets as follows.

(1) 100leaves[1]: The one hundred plant leaves (100leaves) dataset is a benchmark dataset widely used in the fields of computer vision and pattern recognition. This dataset has 3 views with 1600 samples in each view. Each sample has 64 features and belongs to one of one hundred plant species.

(2) BBCSports[2]: The BBCSport dataset is a text classification resource gathered from the BBC Sport website. It encompasses 5 categories to classify these texts into soccer, rugby, cricket, tennis, and athletics. This dataset has a total of 116 texts in four views. Each sample in different views has different features.

(3) Caltech7[3]: The Caltech7 dataset is a classic image classification dataset that contains images from 7 object categories, which are bread, cookies, candy, coffee cups, beverage cans, flowers, and apples. This dataset contains 1,474 color images, each with different image sizes. This dataset is also a widely used benchmark dataset for image classification, data enhancement, and deep learning.

(4) ORL[4]: The Olivetti Research Laboratory (ORL) face dataset is a commonly used face recognition dataset, provided by Olivetti Labs. The dataset

[1] https://data.world/uci/one-hundred-plant-species-leaves-data-set
[2] http://mlg.ucd.ie/datasets/bbc.html
[3] http://www.vision.caltech.edu/Image_Datasets/Caltech101
[4] https://paperswithcode.com/dataset/orl

contains 400 images from different people and is widely used for research and experiments in the field of face recognition.

(5) Mfeat[5]: The Mfeat handwritten digit dataset contains 10 handwritten number categories (0 to 9) from the UCI repository. There are a total of 2000 data samples, and each sample has 6 types of features.

Table 1: Statistics of the Benchmark Datasets

Dataset	# of Views	# of Clusters	# of Samples	# of Features in Views
100leaves	3	100	1600	64/64/64
BBCSport	4	5	116	1991/2063/2113/2158
Caltech7	6	7	1474	48/40/254/1984/512/928
ORL	2	40	400	1024/288
Mfeat	6	10	2000	216/76/64/6/240/47

The summarised statistics can be found in Table 1. The incomplete datasets are created from the five complete datasets by randomly eliminating a certain percentage of instances from each view, including 10%, 20%, 30%, 40%, and 50%, which are five missing ratios from 0.1 to 0.5.

5.1.2. Baseline Methods

We adopt 7 state-of-the-art IMVC methods as baselines to compare the performance, including

(1) IMSC_AGL [12] integrates the graph learning and spectral clustering techniques to learn a common representation and then uses the k-means to obtain final partitions.

(2) IMSR [22] first leverages self-representation subspace learning to obtain the global structure and then fills the missing samples under its guidance.

(3) PIMVC [46] method for IMVC that utilizes a graph regularized projective

[5]http://archive.ics.uci.edu/ml/datasets/Multiple+Features

consensus representation in a unified low-dimensional subspace.

(4) LSIMVC [49] is an IMVC approach based on matrix factorization, which aims to learn a consensus latent representation by incorporating a sparse constraint and a graph embedding term.

(5) AGC_IMC [26] simultaneously performs graph completion and consensus representation learning while utilizing an adaptive scale vector to mitigate the negative impact of information imbalance.

(6) UEAF [21] infers missing views using a locality-preserved reconstruction, aligns all views naturally, embeds an adaptively learned consensus graph, and employs an adaptive weighting strategy to gauge the importance of various views.

(7) FLSD [31] employs a graph-regularized matrix factorization model to maintain local geometric similarities and uses a semantic consistency constraint for a unified representation.

To evaluate the baselines, we acquire their source codes from the authors' websites and follow the settings outlined in the original papers to fine-tune the hyperparameters and report the best experimental results. Regarding our proposed SEMI method, the balancing parameters λ, γ, and β are determined via grid searching within the range of $2^{\{-10,-8,\ldots,10\}}$. The initial value of α is set to 0.01. The constants α_0 and ρ are assigned values of 10^5 and 1.1, respectively.

5.1.3. Evaluation Metrics

We utilize 3 commonly used metrics to assess the performance of all the methods: accuracy (ACC) [50], Normalized Mutual Information (NMI) [51], and purity [52]. The values of these metrics range from 0 to 1, with higher values indicating superior clustering performance.

To ensure a fair comparison, we execute all the methods ten times with different missing ratios and then calculate the average values as well as the standard deviations of the metrics.

5.1.4. Experimental Environment

We conducted the experiments in the Matlab2023a environment. All our experiments were done on a desktop computer configured with an Intel i9-12900f processor and an NVIDIA RTX 3080 graphics card with 32 GB of RAM.

5.2. Experimental Results

The Tables 2, 3, and 4 present the average experimental results, aggregated over the 5 missing ratios $[0.1, 0.2, 0.3, 0.4, 0.5]$, for 8 methods on 5 incomplete multi-view datasets. The results include the mean metric values and standard deviations denoted as mean±standard-deviation from the ten best experimental runs to ensure the stability and reliability of our experimental outcomes. The average best results are highlighted in bold and the results in the table marked with an asterisk (*) indicate the second-highest results. From the tables, we have the following observations.

In Table 2, SEMI outperforms other methods in terms of ACC across all datasets, notably excelling in the BBCSports dataset. This highlights SEMI's robustness and versatility. Its strength lies in its information aggregation strategy, effectively capturing complex data patterns by blending local, global, and consistent information. This integration leads to superior performance. Moreover, SEMI can achieve consensus results, as seen from its reduced standard deviations compared to other baselines. Among the baselines, IMSC_AGL performs better than the other methods, chiefly due to its emphasis on graph learning, highlighting the significance of local information. While SEMI jointly explores multi-structure information, providing a distinct advantage. Additionally, UEAF, which aims to learn a consensus graph, underperforms in our simulations. Its graph quality may be influenced by the noise or outliers caused by incomplete multi-view data. However, SEMI imputes the missing views under the guidance of the learned structures, alleviating the introduction of noise and outliers.

In Table 3, SEMI performs better than the other methods in terms of average NMI across all datasets with different missing ratios. While IMSR and

Table 2: ACCs with mean ± standard-deviation of different methods on the five datasets across all missing ratios

Datasets	IMSC_AGL	IMSR	PIMVC	LSIMVC	AGC_IMC	UEAF	FLSD	SEMI
100Leaves	*0.619±0.039	0.476±0.024	0.591±0.018	0.356±0.037	0.579±0.010	0.267±0.039	0.318±0.032	**0.646±0.014**
BBCSports	0.720±0.017	0.455±0.033	*0.737±0.036	0.725±0.026	0.574±0.005	0.677±0.027	0.723±0.051	**0.803±0.015**
Caltech7	0.487±0.053	0.434±0.046	0.465±0.026	0.501±0.017	*0.520±0.022	0.447±0.068	0.448±0.048	**0.668±0.015**
ORL	*0.615±0.028	0.603±0.021	0.607±0.036	0.530±0.018	0.511±0.035	0.435±0.031	0.535±0.066	**0.683±0.019**
Mfeat	*0.814±0.021	0.626±0.026	0.798±0.020	0.756±0.019	0.741±0.034	0.658±0.051	0.694±0.019	**0.835±0.009**

FLSD achieve notable results, they cannot outperform SEMI. Specifically, IMSR learns a unified coefficient matrix by self-representation subspace learning and leverages it to guide the data imputation. However, the coefficient matrix solely encompasses global information, resulting in imprecise imputation. This is primarily due to the difficulty in accurately capturing global information, especially as the missing ratios increase. In contrast, SEMI integrates both global and local structures, allowing it to capture more complex and diverse patterns. FLSD employs a graph-regularized matrix factorization model for IMVC, prioritizing local geometry and the importance of different views. However, it ignores inconsistencies, making it challenging to learn a consistent structure. In the face of biased or skewed data, SEMI stands out with its design that considers both consistent and inconsistent structures.

Table 3: NMIs with mean ± standard-deviation of different methods on the five datasets across all missing ratios

Datasets	IMSC_AGL	IMSR	PIMVC	LSIMVC	AGC_IMC	UEAF	FLSD	SEMI
100Leaves	*0.781±0.043	0.707±0.063	0.769±0.059	0.623±0.015	0.774±0.015	0.542±0.062	0.588±0.055	**0.801±0.016**
BBCSports	*0.633±0.032	0.248±0.023	0.614±0.025	0.603±0.005	0.422±0.012	0.552±0.043	0.605±0.048	**0.720±0.011**
Caltech7	0.394±0.038	0.387±0.041	*0.438±0.026	0.421±0.007	0.433±0.003	0.379±0.06	0.401±0.042	**0.475±0.012**
ORL	*0.767±0.027	0.752±0.007	0.753±0.049	0.710±0.014	0.691±0.020	0.604±0.031	0.708±0.028	**0.798±0.019**
Mfeat	*0.767±0.015	0.570±0.017	0.752±0.027	0.721±0.008	0.765±0.011	0.598±0.046	0.654±0.037	**0.788±0.014**

Table 4 shows the purities of various methods across datasets with all missing ratios. SEMI excels in most datasets, with the exception of Caltech7. A hallmark of SEMI is its integration of both consistency and inconsistency across

views, distinguishing it from methods that focus solely on consistency. Leveraging the inherent correlations and complementarities between views, our method markedly separates the inconsistencies and consequently improves clustering purity. On the Caltech7 dataset, PIMVC performs the best while our SEMI method can still achieve comparable results. When considering all three evaluation metrics for Caltech7, our method can surpass the other baselines in most cases.

Table 4: Purities with mean ± standard-deviation of different methods on the five datasets across all missing ratios

Datasets	IMSC_AGL	IMSR	PIMVC	LSIMVC	AGC_IMC	UEAF	FLSD	SEMI
100Leaves	*0.660±0.024	0.498±0.056	0.618±0.037	0.398±0.038	0.604±0.004	0.305±0.035	0.343±0.064	**0.668±0.013**
BBCSports	0.820±0.035	0.541±0.016	0.832±0.025	0.823±0.013	0.656±0.024	0.762±0.049	*0.838±0.035	**0.893±0.011**
Caltech7	0.836±0.014	0.835±0.029	**0.862±0.030**	*0.849±0.026	0.848±0.002	0.830±0.058	0.845±0.031	0.844±0.025
ORL	0.641±0.017	0.649±0.019	*0.650±0.029	0.575±0.029	0.534±0.030	0.478±0.025	0.568±0.052	**0.727±0.018**
Mfeat	*0.818±0.021	0.622±0.028	0.802±0.015	0.761±0.001	0.771±0.014	0.667±0.023	0.698±0.042	**0.832±0.015**

To systematically analyze the performance changes of different methods under various missing ratios, we present metric curves on multiple incomplete multi-view datasets. A notable observation becomes evident: as the missing ratios increase, the performance in terms of three evaluation metrics generally declines, as shown in Figs. 2, 3, 4, 5, and 6. This underscores the challenges of missing views. Yet, even amid these challenges, SEMI's resilience stands out, consistently outperforming baselines, especially in terms of ACC and NMI metrics. Instead of a singular focus, SEMI integrates local, global, and consistent structures, capturing the depth and breadth of each view. This holistic approach, enabling the detection of both broad patterns and intricate details, is key to SEMI's edge. In essence, SEMI's unique representation approach and adaptability to data incompleteness set it as a leading method in incomplete multi-view clustering.

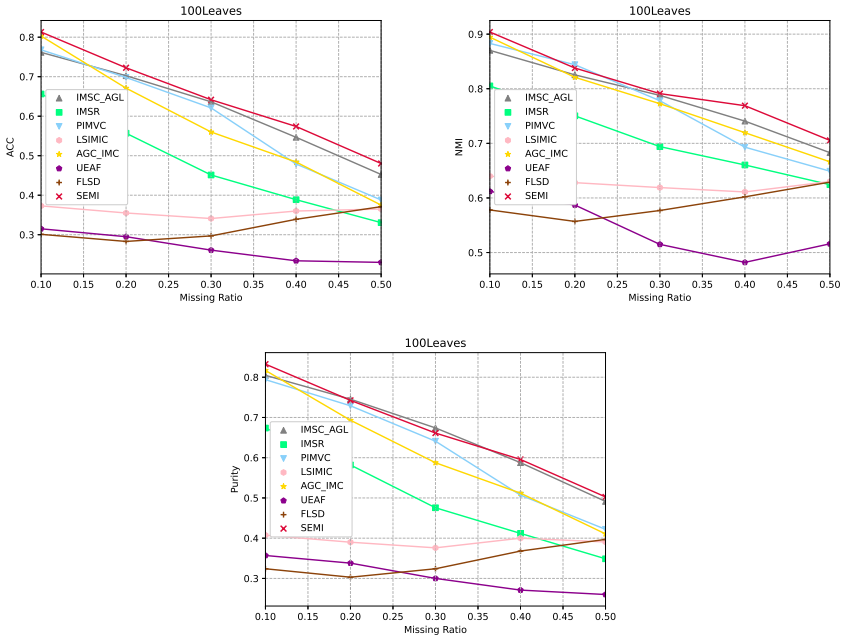

Figure 2: Clustering performance in terms of ACC, NMI, and purity on 100 leaves dataset with different missing ratios.

5.3. Time Complexity Analysis

Table 5 presents the average running time of different methods on multiple datasets at various missing ratios. Notably, our proposed SEMI algorithm can achieve comparable running time in most cases and rank third place on average. Specifically, SEMI, IMSC_AGL, and PIMVC all have the same time complexity of $O(n^3)$. However, a distinguishing feature of SEMI is its incorporation of a consistency term in the optimization process. This term, designed to capture both consistent and inconsistent structures, leverages the consistent structure to guide the imputation of missing views. The explicit computation of inconsistency, represented by matrix E, introduces an additional computational burden of $O(n^2)$. For instance, on the BBCsports dataset, SEMI converges in a mere 0.167 seconds, trailing slightly behind IMSC_AGL, which takes 0.121 seconds. However, by incorporating the exploration of inconsistencies, our SEMI algo-

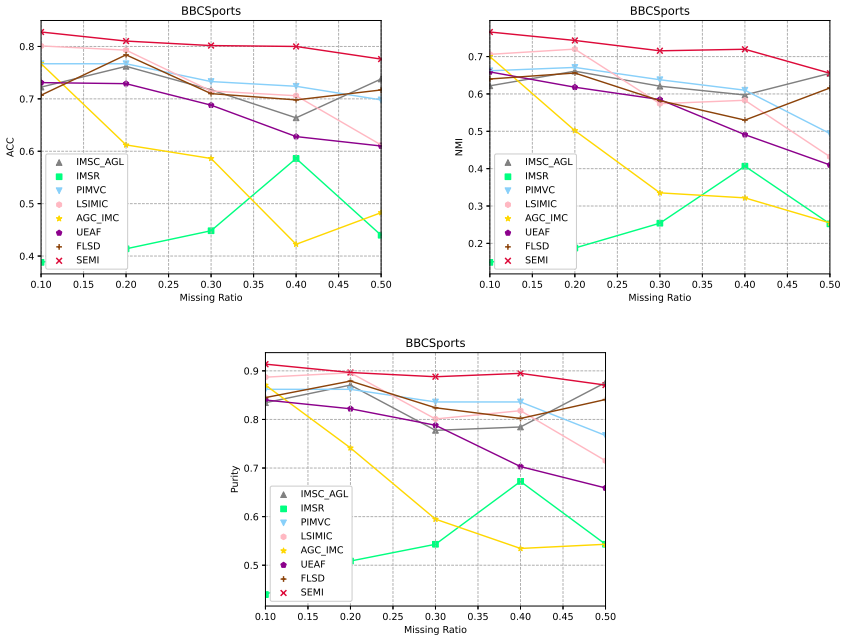

Figure 3: Clustering performance in terms of ACC, NMI, and purity on BBCSports dataset with different missing ratios.

rithm can achieve superior clustering performance, illustrating the existence of a trade-off between performance and complexity.

Table 5: Average elapsed time across all missing ratios for various methods on different datasets

Dataset	IMSC_AGL	IMSR	PIMVC	LSIMVC	AGC_IMC	UEAF	FLSD	SEMI
100Leaves	5.62s	27.9s	0.609s	14.0s	9.71s	8.50s	11.3s	2.31s
BBCsports	0.121s	1.38s	1.23s	12.1s	0.231s	7.24s	4.47s	0.167s
Caltech7	2.54s	24.9s	0.472s	7.68s	7.45s	32.8s	12.6s	2.53s
ORL	0.549s	2.06s	0.171s	27.0s	1.48s	1.45s	0.745s	3.99s
Mfeat	5.07s	50.5s	0.422s	8.62s	13.8s	58.7s	14.9s	5.89s
Average	*2.78s	21.4s	**0.581s**	13.9s	6.53s	21.7s	8.80s	2.97s

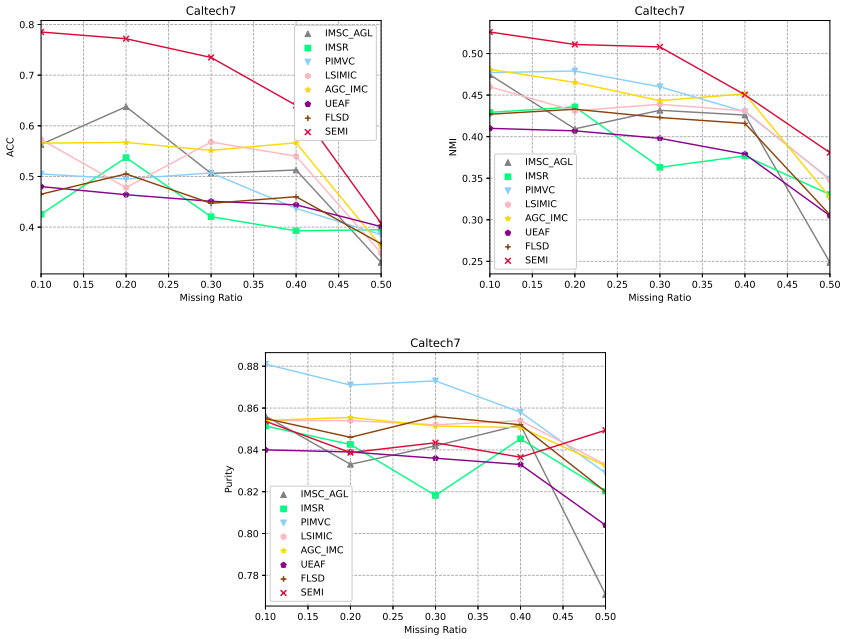

Figure 4: Clustering performance in terms of ACC, NMI, and purity on Caltech7 dataset with different missing ratios.

5.4. Convergence Analysis

Convergence analysis plays a crucial role in assessing the effectiveness and stability of a model since it enables us to evaluate the training efficiency of the model. In light of this, we designed a series of experiments to thoroughly analyze convergence. Our model computes the value of the objective function at each iteration according to Eq. (12). We perform 50 iterations and record the objective function value for each iteration on BBCSport and ORL datasets, plotting these values as a line figure for an intuitive visualization of the model's convergence behavior. The experimental results are depicted in Figure 7, with the x-axis representing the number of iterations and the y-axis representing the objective function value. The figures demonstrate that SEMI exhibits rapid convergence, indicating that it provides an efficient and optimized solution.

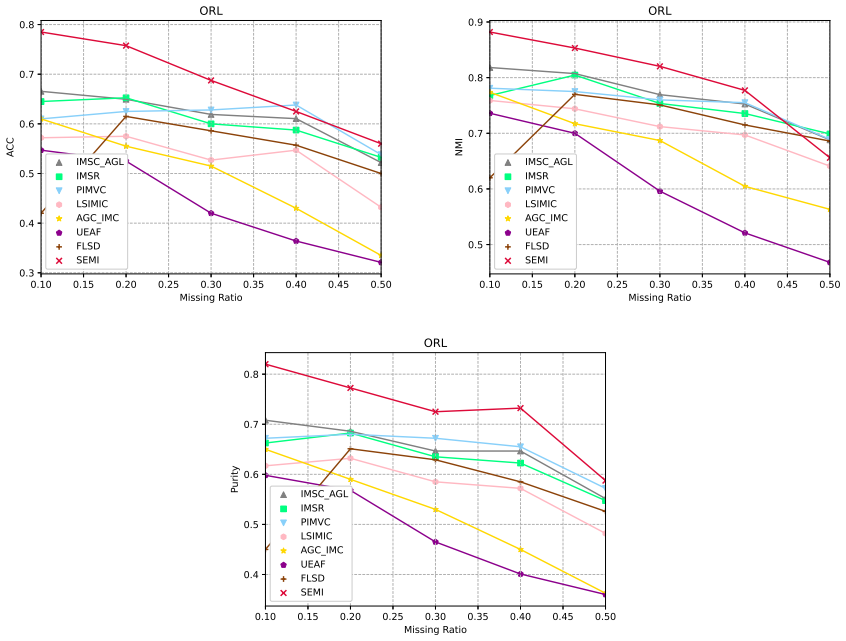

Figure 5: Clustering performance in terms of ACC, NMI, and purity on ORL dataset with different missing ratios.

5.5. Hyperparameter Sensitivity Analysis

To evaluate the sensitivity of SEMI to hyper-parameter configurations, we perform a sensitivity analysis, focusing on two critical hyper-parameters: the balancing parameter λ and γ. We designate a range of values for each hyper-parameter, specifically 2^{-10}, 2^{-5}, 2^{0}, 2^{5}, and 2^{10}. In our experiments, the values of λ and γ are adjusted to examine their influence on the model's performance on the Mfeat and 100Leaves datasets with a 0.1 missing ratio. For each combination of λ and γ, we show the relationships of the ACCs and the two parameters. We visualize the experimental results using 3D graphics in Fig. 8. It can be observed that SEMI can obtain stable and satisfied ACCs when they are in some feasible ranges. On the Mfeat dataset, the ranges of λ and γ are $[2^5, 2^{10}]$ and $[2^{-10}, 2^0]$. On the 100Leaves dataset, the ranges of λ and γ are $[2^5, 2^{10}]$ and $[2^{-10}, 2^5]$. It demonstrates that SEMI is insensitive to the two parameters to some extent

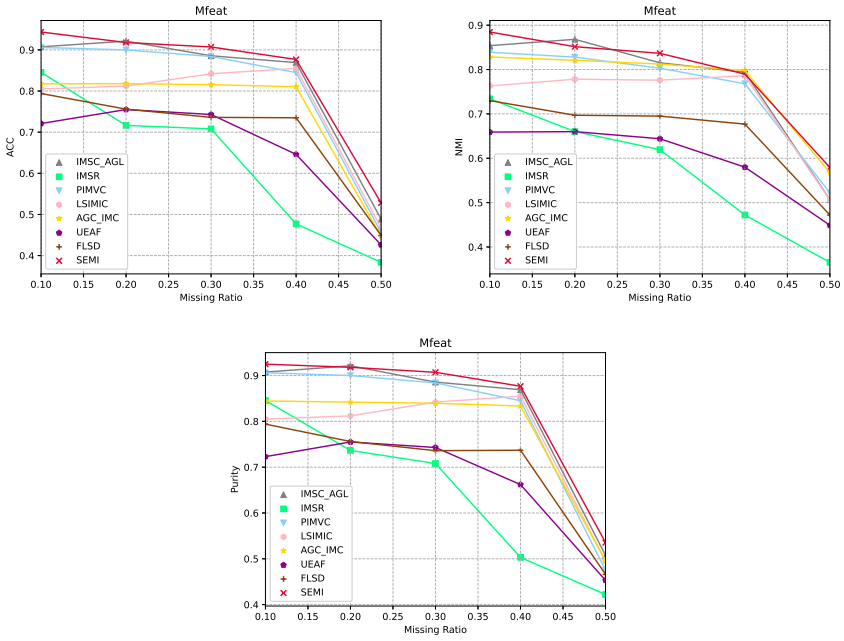

Figure 6: Clustering performance in terms of ACC, NMI, and purity on Mfeat dataset with different missing ratios.

when λ is within the range of $[2^5, 2^{10}]$ and γ is within the range of $[2^{-10}, 2^0]$. Through this sensitivity analysis, we are able to identify the optimal hyper-parameter settings for each dataset and gain a better understanding of how these hyper-parameters influence the model's performance, which is valuable for fine-tuning the model to achieve optimal results in practical applications.

5.6. Ablation Study

In this subsection, we conduct a series of ablation experiments to assess the individual contributions of each component within our proposed model concerning performance. Ablation experiments involve sequentially removing crucial components from the model and observing the subsequent changes in performance. To maintain experimental fairness, we train and evaluate both the complete model and its ablated versions on the Mfeat dataset with a 0.1 missing ratio and experimental conditions. We employ the three metrics to gauge

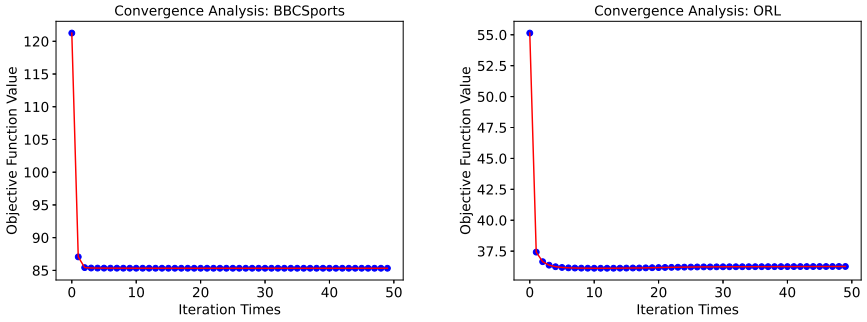

Figure 7: Convergence curves on BBCSport and ORL datasets with a 0.1 missing ratio, respectively.

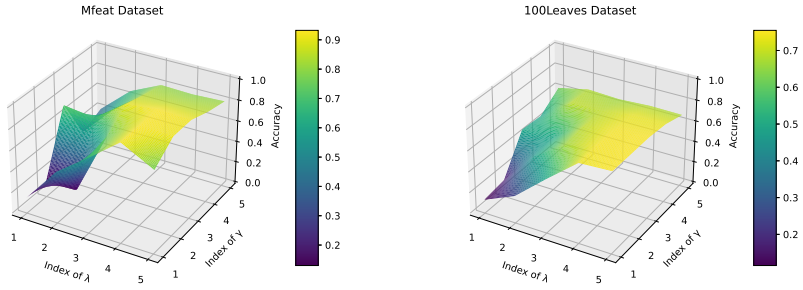

Figure 8: ACCs versus parameters λ and γ on Mfeat and 100Leaves datasets with a 0.1 missing ratio, respectively.

the performance of each variant model. We have SEMI and its three variants as follows.

- The full model of SEMI;
- No consistent structure consideration in SEMI, where the learned coefficient matrix S is not split into FF^T and E and the terms (4)-(5) are removed from Eq. (12);
- No local structure consideration in SEMI, where the learned coefficient matrix S does not contain the local structure exploration and the term (3) is removed from Eq. (12);
- No data imputation in SEMI, where the missing views Y are not updated.

- No global structure in SEMI, where $\sum_{v=1}^{V} ||X_v - X_v S||_F^2$ is removed from Eq.(12) ;

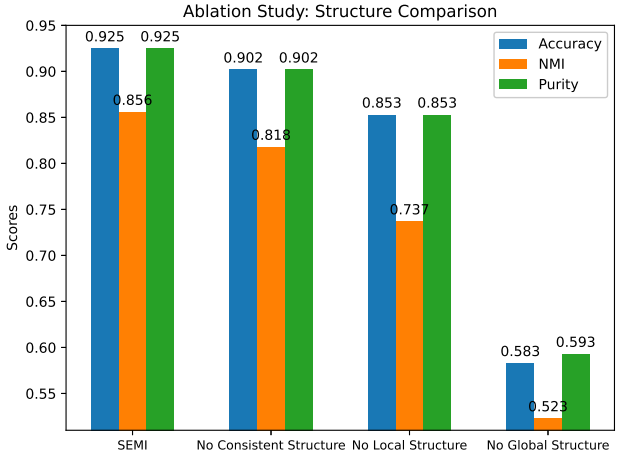

Figure 9: Ablation study histogram on Mfeat dataset with a 0.1 missing ratio.

The results of the ablation study experiments are shown in Fig. 9. From the results, we observe the following trends: the performance across all metrics experiences a decline when the consistent structure component is omitted, signifying its pivotal role and contribution to the enhanced performance. Additionally, when the local structure component is excluded, the performance deterioration becomes more pronounced, indicating that the local structure also substantially impacts the model's performance. Without the data imputation, the missing views would degrade the performance severely. Notably, the most substantial impact on performance was observed when the global structure component was removed. The absence of global structure had a marked negative effect on all evaluated metrics, signifying its pivotal role in enhancing algorithmic performance. This highlights the indispensable contribution of global structure to our algorithm. In summary, our experimental findings underscore the essential nature of each component in the SEMI algorithm. These components reinforce each other during the iterative optimization process. This ablation experiment provides valuable insights into the internal structure of the algorithm and the

interplay among its components, aiding in further optimization and understanding of our approach.

6. Conclusion and Future work

In this paper, we proposed a novel incomplete multi-view clustering method by exploring multiple structures and imputing missing views. We unified the self-expression learning, graph learning, and clustering-oriented learning to capture the underlying global, local, and consistent structures, respectively. In the guiding of the captured structure, the missing samples are inferred and then reinforce the exploration of the multiple structures during an alternative optimization process. We conducted a series of experiments to evaluate the efficacy of our proposed method across a diverse range of scenarios, including various ratios of missing data, multiple data distributions, and distinct numbers of views. Experimental results show that our method can achieve better performance than the baselines in most cases, demonstrating the effectiveness in handling incomplete multi-view datasets. For future work, we would enhance the proposed approach by integrating deep learning techniques to investigate the nonlinear relationships among data samples and impute missing views in an adversarial manner.

7. Acknowledgement

This work was supported by the National Science Foundation under Grant 2245918.

References

[1] U. Fang, M. Li, J. Li, L. Gao, T. Jia, Y. Zhang, A comprehensive survey on multi-view clustering, IEEE Transactions on Knowledge and Data Engineering (2023).

[2] J. Liu, F. Cao, X.-Z. Gao, L. Yu, J. Liang, A cluster-weighted kernel k-means method for multi-view clustering, in: Proceedings of the aaai conference on artificial intelligence, Vol. 34, 2020, pp. 4860–4867.

[3] C. Tang, K. Sun, C. Tang, X. Zheng, X. Liu, J.-J. Huang, W. Zhang, Multi-view subspace clustering via adaptive graph learning and late fusion alignment, Neural Networks (2023).

[4] C. Tang, Z. Li, W. Yan, G. Yue, W. Zhang, Efficient multiple kernel clustering via spectral perturbation, in: Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 1603–1611.

[5] H. Zhang, J. Yang, F. Shang, C. Gong, Z. Zhang, Lrr for subspace segmentation via tractable schatten-p norm minimization and factorization, IEEE transactions on cybernetics 49 (5) (2018) 1722–1734.

[6] H. Zhang, F. Qian, P. Shi, W. Du, Y. Tang, J. Qian, C. Gong, J. Yang, Generalized nonconvex nonsmooth low-rank matrix recovery framework with feasible algorithm designs and convergence analysis, IEEE Transactions on Neural Networks and Learning Systems, DOI: 10.1109/TNNLS.2022.3183970 (2022).

[7] S. Shi, F. Nie, R. Wang, X. Li, Multi-view clustering via nonnegative and orthogonal graph reconstruction, IEEE Transactions on Neural Networks and Learning Systems (2021).

[8] S.-G. Fang, D. Huang, X.-S. Cai, C.-D. Wang, C. He, Y. Tang, Efficient multi-view clustering via unified and discrete bipartite graph learning, IEEE Transactions on Neural Networks and Learning Systems (2023).

[9] J. Wen, Z. Zhang, Z. Zhang, L. Zhu, L. Fei, B. Zhang, Y. Xu, Unified tensor framework for incomplete multi-view clustering and missing-view inferring, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 35, 2021, pp. 10273–10281.

[10] W. Xia, Q. Gao, Q. Wang, X. Gao, Tensor completion-based incomplete multiview clustering, IEEE Transactions on Cybernetics 52 (12) (2022) 13635–13644.

[11] J. Liu, M. Gong, K. Qin, P. Zhang, A deep convolutional coupling network for change detection based on heterogeneous optical and radar images, IEEE transactions on neural networks and learning systems 29 (3) (2016) 545–559.

[12] J. Wen, Y. Xu, H. Liu, Incomplete multiview spectral clustering with adaptive graph learning, IEEE transactions on cybernetics 50 (4) (2018) 1418–1429.

[13] J. Wen, Z. Zhang, Z. Zhang, Z. Wu, L. Fei, Y. Xu, B. Zhang, Dimc-net: Deep incomplete multi-view clustering network, in: Proceedings of the 28th ACM international conference on multimedia, 2020, pp. 3753–3761.

[14] S.-Y. Li, Y. Jiang, Z.-H. Zhou, Partial multi-view clustering, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 28, 2014.

[15] H. Zhao, H. Liu, Y. Fu, Incomplete multi-modal visual data grouping., in: IJCAI, 2016, pp. 2392–2398.

[16] W. Shao, L. He, P. S. Yu, Multiple incomplete views clustering via weighted nonnegative matrix factorization with regularization, in: Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2015, Porto, Portugal, September 7-11, 2015, Proceedings, Part I, Springer, 2015, pp. 318–334.

[17] W. Shao, L. He, C.-t. Lu, S. Y. Philip, Online multi-view clustering with incomplete views, in: 2016 IEEE International conference on big data (Big Data), IEEE, 2016, pp. 1012–1017.

[18] M. Hu, S. Chen, One-pass incomplete multi-view clustering, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 33, 2019, pp. 3838–3845.

[19] Y. Lin, Y. Gou, Z. Liu, B. Li, J. Lv, X. Peng, Completer: Incomplete multi-view clustering via contrastive prediction, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 11174–11183.

[20] X. Liu, X. Zhu, M. Li, C. Tang, E. Zhu, J. Yin, W. Gao, Efficient and effective incomplete multi-view clustering, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 4392–4399.

[21] J. Wen, Z. Zhang, Y. Xu, B. Zhang, L. Fei, H. Liu, Unified embedding alignment with missing views inferring for incomplete multi-view clustering, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 33, 2019, pp. 5393–5400.

[22] J. Liu, X. Liu, Y. Zhang, P. Zhang, W. Tu, S. Wang, S. Zhou, W. Liang, S. Wang, Y. Yang, Self-representation subspace clustering for incomplete multi-view data, in: Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 2726–2734.

[23] X. Liu, X. Zhu, M. Li, L. Wang, E. Zhu, T. Liu, M. Kloft, D. Shen, J. Yin, W. Gao, Multiple kernel k k-means with incomplete kernels, IEEE transactions on pattern analysis and machine intelligence 42 (5) (2019) 1191–1204.

[24] X. Liu, M. Li, C. Tang, J. Xia, J. Xiong, L. Liu, M. Kloft, E. Zhu, Efficient and effective regularized incomplete multi-view clustering, IEEE transactions on pattern analysis and machine intelligence 43 (8) (2020) 2634–2646.

[25] H. Wang, L. Zong, B. Liu, Y. Yang, W. Zhou, Spectral perturbation meets incomplete multi-view data, in: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, International Joint Conferences on Artificial Intelligence Organization, 2019, pp. 3677–3683. doi:10.24963/ijcai.2019/510.

[26] J. Wen, K. Yan, Z. Zhang, Y. Xu, J. Wang, L. Fei, B. Zhang, Adaptive

graph completion based incomplete multi-view clustering, IEEE Transactions on Multimedia 23 (2020) 2493–2504.

[27] Q. Wang, Z. Ding, Z. Tao, Q. Gao, Y. Fu, Partial multi-view clustering via consistent gan, in: 2018 IEEE International Conference on Data Mining (ICDM), IEEE, 2018, pp. 1290–1295.

[28] S. Luo, C. Zhang, W. Zhang, X. Cao, Consistent and specific multi-view subspace clustering, in: Proceedings of the AAAI conference on artificial intelligence, Vol. 32, 2018.

[29] Y. Liang, D. Huang, C.-D. Wang, Consistency meets inconsistency: A unified graph learning framework for multi-view clustering, in: 2019 IEEE International Conference on Data Mining (ICDM), IEEE, 2019, pp. 1204–1209.

[30] M. Horie, H. Kasai, Consistency-aware and inconsistency-aware graph-based multi-view clustering, in: 2020 28th European Signal Processing Conference (EUSIPCO), IEEE, 2021, pp. 1472–1476.

[31] J. Wen, Z. Zhang, Z. Zhang, L. Fei, M. Wang, Generalized incomplete multiview clustering with flexible locality structure diffusion, IEEE transactions on cybernetics 51 (1) (2020) 101–114.

[32] H. Zhang, J. Qian, B. Zhang, J. Yang, C. Gong, Y. Wei, Low-rank matrix recovery via modified schatten-p norm minimization with convergence guarantees, IEEE Transactions on Image Processing 29 (2020) 3132–3142.

[33] H. Zhang, C. Gong, J. Qian, B. Zhang, C. Xu, J. Yang, Efficient recovery of low-rank matrix via double nonconvex nonsmooth rank minimization, IEEE Transactions on Neural Networks and Learning Systems 30 (10) (2019) 2916–2925.

[34] H. Zhang, F. Qian, B. Zhang, W. Du, J. Qian, J. Yang, Incorporating linear regression problems into an adaptive framework with feasible optimizations,

IEEE Transactions on Multimedia, DOI: 10.1109/TMM.2022.3171088 (2022).

[35] L. Li, H. Jiang, H. He, Graph-based multi-view learning for cooperative spectrum sensing, in: 2021 International Joint Conference on Neural Networks (IJCNN), IEEE, 2021, pp. 1–7.

[36] H. Wang, Y. Yang, B. Liu, Gmc: Graph-based multi-view clustering, IEEE Transactions on Knowledge and Data Engineering 32 (6) (2019) 1116–1129.

[37] H. Zhang, J. Yang, J. Xie, J. Qian, B. Zhang, Weighted sparse coding regularized nonconvex matrix regression for robust face recognition, Information Sciences 394 (2017) 1–17.

[38] Z. Kang, X. Zhao, C. Peng, H. Zhu, J. T. Zhou, X. Peng, W. Chen, Z. Xu, Partition level multiview subspace clustering, Neural Networks 122 (2020) 279–288.

[39] C. Zhang, H. Fu, Q. Hu, X. Cao, Y. Xie, D. Tao, D. Xu, Generalized latent multi-view subspace clustering, IEEE transactions on pattern analysis and machine intelligence 42 (1) (2018) 86–99.

[40] J. Liu, M. Gong, Q. Miao, X. Wang, H. Li, Structure learning for deep neural networks based on multiobjective optimization, IEEE transactions on neural networks and learning systems 29 (6) (2017) 2450–2463.

[41] H. Wang, Y. Yang, B. Liu, H. Fujita, A study of graph-based system for multi-view clustering, Knowledge-Based Systems 163 (2019) 1009–1019.

[42] L. Li, H. He, Bipartite graph based multi-view clustering, IEEE transactions on knowledge and data engineering 34 (7) (2020) 3111–3125.

[43] W. Liang, X. Liu, S. Zhou, J. Liu, S. Wang, E. Zhu, Robust graph-based multi-view clustering, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, 2022, pp. 7462–7469.

[44] X.-L. Li, M.-S. Chen, C.-D. Wang, J.-H. Lai, Refining graph structure for incomplete multi-view clustering, IEEE Transactions on Neural Networks and Learning Systems (2022).

[45] L. Li, Z. Wan, H. He, Incomplete multi-view clustering with joint partition and graph learning, IEEE Transactions on Knowledge and Data Engineering 35 (1) (2021) 589–602.

[46] S. Deng, J. Wen, C. Liu, K. Yan, G. Xu, Y. Xu, Projective incomplete multi-view clustering, IEEE Transactions on Neural Networks and Learning Systems (2023).

[47] E. J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis?, Journal of the ACM (JACM) 58 (3) (2011) 1–37.

[48] C. Chen, B. He, Y. Ye, X. Yuan, The direct extension of admm for multi-block convex minimization problems is not necessarily convergent, Mathematical Programming 155 (1-2) (2016) 57–79.

[49] C. Liu, Z. Wu, J. Wen, Y. Xu, C. Huang, Localized sparse incomplete multi-view clustering, IEEE Transactions on Multimedia (2022).

[50] K. A. Nazeer, M. Sebastian, Improving the accuracy and efficiency of the k-means clustering algorithm, in: Proceedings of the world congress on engineering, Vol. 1, Association of Engineers London London, UK, 2009, pp. 1–3.

[51] T. O. Kvålseth, On normalized mutual information: measure derivations and properties, Entropy 19 (11) (2017) 631.

[52] D. Marutho, S. H. Handaka, E. Wijaya, et al., The determination of cluster number at k-mean using elbow method and purity evaluation on headline news, in: 2018 international seminar on application for technology of information and communication, IEEE, 2018, pp. 533–538.