Received: 20 January 2021 | Revised: 18 February 2021 | Accepted: 21 February 2021

DOI: 10.1002/int.22409

RESEARCH ARTICLE

WILEY

# Partial multiview clustering with locality graph regularization

**Huiqiang Lian**[1] | **Huiying Xu**[2] | **Siwei Wang**[3] | **Miaomiao Li**[4] | **Xinzhong Zhu**[2,5] | **Xinwang Liu**[3]

[1]University of Chinese Academy of Sciences, Beijing, China

[2]College of Mathematics and Computer Science, Zhejiang Normal University, Jinhua, China

[3]College of Computer, National University of Defense Technology, Changsha, China

[4]College of Electronic Information and Electrical Engineering, Changsha University, Changsha, China

[5]Research Institute of Ningbo Cixing Co. Ltd, Cixi City, China

**Correspondence**
Huiying Xu, College of Mathematics and Computer Science, Zhejiang Normal University, 321004 Jinhua, China.
Email: xyh@zjnu.edu.cn

**Funding information**
Key Technologies Research and Development Program, Grant/Award Number: 2020AAA0107100

**Abstract**

Multiview clustering (MVC) collects complementary and abundant information, which draws much attention in machine learning and data mining community. Existing MVC methods usually hold the assumption that all the views are complete. However, multiple source data are often incomplete in real-world applications, and so on sensor failure or unfinished collection process, which gives rise to incomplete multiview clustering (IMVC). Although enormous efforts have been devoted in IMVC, there still are some urgent issues that need to be solved: (i) The locality among multiple views has not been utilized in the existing mechanism; (ii) Existing methods inappropriately force all the views to share consensus representation while ignoring specific structures. In this paper, we propose a novel method termed partial MVC with locality graph regularization to address these issues. First, followed the traditional IMVC approaches, we construct weighted semi-nonnegative matrix factorization models to handle incomplete multiview data. Then, upon the consensus representation matrix, the locality graph is constructed for regularizing the shared feature matrix. Moreover, we add the coefficient regression term to constraint the various base matrices among views. We incorporate the three aforementioned processes into a unified framework, whereas they can negotiate

  

with each other serving for learning tasks. An effective iterative algorithm is proposed to solve the resultant optimization problem with theoretically guaranteed convergence. The comprehensive experiment results on several benchmarks demonstrate the effectiveness of the proposed method.

**KEYWORDS**
incomplete data clustering, multiple view clustering

## 1 | INTRODUCTION

Clustering has been intensively studied by combing information to categorize unlabelled data items into appropriate groups.[1-16] With increasing data collection, many data in the real-world could be presented from different perspectives, which are termed multiview data in the literature. Multiview data could provide more abundant information than the traditional single-view feature representation to reveal the intrinsic structure of the data. Existing multiview clustering (MVC) approaches can be summarized into the following four strategies. The co-training approaches for MVC iteratively product multiple clustering results that can provide predicted clustering labels for other views. In this way, besides extracting the specific cluster information from the corresponding view, the clustering results are forced to be consistent across views. The second method bases on the assumption that high-dimensional data points are drawn from the low-dimensional spaces and therefore each cluster is formed from one of them. By following the multiple kernel learning framework, multiple kernel clustering (MKC) seeks an optimal kernel similarity matrix through linear combination of predefined kernels. The last category extends semi-nonnegative matrix factorization (SNMF) to incorporate multiple source information to learn consensus latent representation in Reference [1]. The essential idea is to optimize a united representation for downstream clustering task with the respective linear transformation.[2] In this paper, we focus on the multiview SNMF clustering methods.

Despite many MVC algorithms have been proposed and achieved great achievements in various applications, traditional MVC algorithms cannot effectively deal with multiview data with incomplete features. However, incompleteness happens quite common in real-world applications. For example, questionnaire forms take missing values due to unfinished processes and medica examines can be partially done for the lack of financial support. All these factors could contribute to incomplete multiview data. Therefore, the incomplete multiview clustering algorithms (IMVC) have attracted extensive attention in recent years. Followed the multiview SNMF methods, NMF-based IMVC methods consider the incomplete data items with zero weight and therefore can be regarded as a special weighted version for NMF MVC. Most of them take the strategy of combining the view-specific and common representations into a unified one.[7-10,17-28] They usually accomplish the missing features with mean values, and then use the weighted NMF to reduce the weight of the missing samples to obtain a consistent representation.[29] These weighted NMF methods commonly impose different constraints on the basis matrix or consensus representation for capturing different properties. Among these methods, $\ell_{2,1}$, $\ell_1$, and $\ell_F$ norm constraints are adopted on the latent representations or basis

matrices, respectively. However, it is limited to the following two points: (i) The locality among multiple views have not been utilized in existing mechanism. They usually neglect the intrinsic local structure of views, which could degrade clustering performance. (ii) Existing methods inappropriately force all the views share consensus representation while ignoring specific structures.

The graph-based IMVC methods utilize the geometric structure of data.[18,30,31] However, it is naturally difficult to construct an accurate graph similarity due to the lack of partial samples.[30] First estimates the missing features and then applies with complete MVC mechanism. Zhao et al. obtained consistent representations to guide the generation of graphs and therefore contain the local structure.[18] However, when the missing rate is high, the filling strategy will dominate the learning of the representation, resulting in the filled samples being connected with each other. Moreover, there graphs are constructed in high-dimensional space and therefore cause huge time complexity.

To address the above issues, this paper proposes a novel IMVC method to fuse the low-dimensional representation learning and locality graph. First, followed the traditional IMVC approaches, we construct weighted semi-NMF models to handle incomplete multiview data. Then, upon the consensus representation matrix, the locality graph is constructed for regularizing the shared feature matrix obtained from multiple sources. Moreover, we add the base coefficient regression $\ell_{2,1}$-norm regularization term to constraint the various base matrices among views. This term is here introduced to constraint sparseness in rows. We incorporates the three aforementioned processes into a unified framework, whereas they can negotiate with each other serving for learning task. The final learned graph not only considers the incomplete data features but also ensures locality among all views.

Compared with existing IMVC methods, the main advantages of the proposed partial multiview clustering with locality graph regularization (PMVC-LGR) can be included as follows:

- We induce the $\ell_{2,1}$ regularization term into traditional NMF based incomplete MVC approaches. This regularization not only ensures sparsity in various base matrices but also performs feature selection in the original model.
- The constructed graph is utilized for preserving the locality in the consensus representation while also encourages the view-consistency and degrades the intra-view disagreements. These factors greatly promote the clustering performance of our proposed method.
- Comprehensive experiment are conducted on several benchmarks data sets. The experimental results clearly demonstrate the effectiveness of the proposed method.

The rest of this paper is organized as follows. Section 2 introduces some notations and related work in our paper. Section 3 presents the proposed optimization objective and the three-step alternate algorithms. Section 5 shows the experimental results with evaluation. Section 6 concludes the paper.

## 2 | PRELIMINARIES AND RELATED WORK

In this section, we introduce some necessary notations and preliminaries in our paper. Throughout this paper, we use boldface lowercase letters to denote matrices. The $(i, j)$th elements of a matrix $U$ is referred as $U_{ij}$. The notations are summarized in Table 1.

**TABLE 1** Main notations used in the paper

| Notation | Meaning |
|---|---|
| MVC | Multiview clustering |
| IMVC | Incomplete multiview clustering |
| NMF | Nonnegative matrix factorization |
| $n$ | The number of samples |
| $m$ | The number of or views |
| $k$ | The number of clusters |
| $\mathrm{d}^{(v)}$ | The dimension of $v$th view |
| $\mathbf{U}^{(v)} \in \mathbb{R}^{d^{(v)} \times k}$ | The $v$th view's base matrix |
| $\mathbf{X}^{(v)}$ | The $v$th view's data matrix |
| $\mathbf{V}$ | The shared representation matrix |
| $\mathbf{S}$ | The learned local graph |
| $\mathbf{B}^{(v)}$ | The regression matrix for $v$th view |

## 2.1 | Semi-NMF for single view and multiple views

The semi-NMF proposed in Reference [32] behaves as a constrained extension of NMF, which under the constraint that the coefficient matrix should be nonnegative. Specifically, given the single-view data matrix $\mathbf{X} \in \mathbb{R}^{d \times N}$, the semi-NMF utilizes base matrix $\mathbf{U} \in \mathbb{R}^{d \times k}$ and the nonnegative coefficient matrix $\mathbf{V} \in \mathbb{R}^{k \times N}$ to reconstruct the original matrix $\mathbf{X}$.

The optimization goal of SNMF can be mathematically formulated as follows:

$$\min_{\mathbf{U},\mathbf{V}} \|\mathbf{X} - \mathbf{U}\mathbf{V}\|_F^2, \quad \text{s.t.} \quad \mathbf{V} \geq 0. \tag{1}$$

As can be seen, Equation (1) seamlessly performs dimension reduction from original space to lower dimension.[32] Further proposes an iterative optimization algorithm to solve Equation (1). The updating strategy can be rewritten as follows:

(i) With $\mathbf{V}$ being fixed, $\mathbf{U}$ can be updated by

$$\mathbf{U} = \mathbf{X}\mathbf{V}^{\top}(\mathbf{V}\mathbf{V}^{\top})^{-1}. \tag{2}$$

(ii) With $\mathbf{U}$ being fixed, $\mathbf{V}$ can be iteratively updated by

$$\mathbf{V}_{ij} \leftarrow \mathbf{V}_{ij} \sqrt{\frac{(\mathbf{X}^{\top}\mathbf{U})_{ij}^{+} + [\mathbf{V}^{\top}(\mathbf{U}^{\top}\mathbf{U})^{-}]_{ij}}{(\mathbf{X}^{\top}\mathbf{U})_{ij}^{-} + [\mathbf{V}^{\top}(\mathbf{U}^{\top}\mathbf{U})^{+}]_{ij}}}, \tag{3}$$

where the positive and negative elements of matrix $\mathbf{M}$ are denoted as: $M_{ij}^{+}$ and $M_{ij}^{-}$.

For multiview data setting, the fundamental optimization formulation can be easily extended as follows:

$$\min_{\mathbf{U}^{(v)},\mathbf{V}} \sum_{v=1}^{m} \|\mathbf{X}^{(v)} - \mathbf{U}^{(v)}\mathbf{V}\|_F^2 + \Omega(\mathbf{V}), \quad \text{s.t.} \quad \mathbf{V} \geq 0, \tag{4}$$

where $\Omega(\mathbf{V})$ denotes various regularization terms for the consensus representation matrix $\mathbf{V}$. They take the assumption that multiview data could share the same latent representation in low-dimensional space. Among these methods, $\ell_{2,1}$, $\ell_1$, and $\ell_F$ norm constraint are adopted into Equation (4) to regularize the representation matrix.

## 2.2 | Weighted NMF for IMVC

To handle IMVC with samples obtaining missing features, the natural way to extend Equation (4) is first to define a weight matrix $W$ which indicates the existence of complete samples in views. The weight matrix $W$ can be formulated as follows:

$$\mathbf{W}_{jj}^{(i)} = \begin{cases} 1 & \text{if } j\text{ - th instance is in the } i\text{ - th view}, \\ 0 & \text{otherwise}. \end{cases} \tag{5}$$

With the weight matrix $W$, Equation (4) can be extended as the following equation:

$$\min_{\mathbf{U}^{(i)},\mathbf{V}} \sum_{i=1}^{m} \|(\mathbf{X}^{(i)} - \mathbf{U}^{(i)}\mathbf{V}^T)\mathbf{W}^{(i)}\|_F^2, \quad \text{s.t.} \quad \mathbf{V} \geq 0. \tag{6}$$

By solving Equation (6), we can get the shared representation for multiple incomplete-view samples.

# 3 | PROPOSED METHOD

Although Equation (6) could cope with the basic IMVC, the various basis matrices for multiview have not been utilized. Despite sharing the same representation, the basis for each view should be orthogonal which indicates differences in clustering centers. Thus, we design to align the basis matrices of respective views to approximate the identity matrix to ensure orthogonality, which can be fulfilled as follows:

$$\min_{\mathbf{U}^{(v)},\mathbf{V},\mathbf{B}^{(v)},} \sum_{v=1}^{m} \|\mathbf{B}^{(v)\top}\mathbf{U}^{(v)} - \mathbf{I}\|_F^2 + \|\mathbf{B}^{(v)}\|_{2,1}. \tag{7}$$

Based on Equation (7), the learned basis centers can be disjoint with each others which makes them more discriminate. Further, the $\ell_{2,1}$ norm on $\mathbf{B}^{(v)}$ simultaneously performs dimension reduction form original dimension $d_v$ to $k$ dimension. The learning process will automatically selects top $k$ vital dimension features to represent data items.

Moreover, after obtaining the consensus representation $V$, the local structures hidden among views should be preserved and we adopt the local graph $S$ to preserve locality, which is formed by minimizing $\sum_{i,j=1}^{n} \mathbf{S}_{ij} \|v_i - v_j\|_2^2$. To be specific, a large value of $S_{ij}$ indicates with high probability that they are in the same cluster. By introducing local graph regularization term, the model preserves the neighbourhood relationships in the unified space.

Integrating the above parts into an unified objective function, we formulate the final formula of our proposed as follows:

$$\begin{aligned}\min_{\substack{\mathbf{U}^{(v)},\mathbf{V},\\ \mathbf{B}^{(v)},\mathbf{S}}} \quad & \sum_{v=1}^{m} \|(\mathbf{X}^{(v)} - \mathbf{U}^{(v)}\mathbf{V}^{\top})\mathbf{W}^{(v)}\|_F^2 + \frac{1}{2}\sum_{i,j=1}^{n} \mathbf{S}_{ij} \|v_i - v_j\|_2^2 \\ & + \alpha\left(\|\mathbf{B}^{(v)\top}\mathbf{U}^{(v)} - \mathbf{I}\|_F^2 + \|\mathbf{B}^{(v)}\|_{2,1}\right) + \lambda \|\mathbf{S}\|_F^2, \\ \text{s.t.} \quad & \mathbf{V} \geq 0, \quad \mathbf{S} \geq 0, \quad \mathbf{S}^{\top}\mathbf{1} = \mathbf{1},\end{aligned} \tag{8}$$

where $S$ is the learned local graph and $V$ is the respective consensus representation matrix. The first term performs the incomplete NMF with multimodality data. The following part in Equation (8) ensures the orthogonality of the basis matrices for each view and also preserves the local structure for the consensus representation.

Regarding of Equation (8), our proposed model has the following merits: (i) The induced $\ell_{2,1}$ regularization term not only ensures sparsity in various base matrices but also performs feature selection in the original model. (ii) The constructed graph is utilized for preserving the locality in the consensus representation while also encourages the view-consistency and degrades the intra-view disagreements. These factors greatly promote the clustering performance of our proposed method.

## 4 | OPTIMIZATION

It is difficult to simultaneously optimize Equation (8) over all the variables. In this section, we propose an alternate algorithm to solve this optimization problem.

### 4.1 | Update $\mathbf{U}^{(v)}$

With $\mathbf{V}$, $\mathbf{B}^{(v)}$ and $\mathbf{S}$ fixed, for each $\mathbf{U}^{(v)}$, we need to solve the following problem in Equation (9):

$$\sum_{v=1}^{m} \min_{\mathbf{U}^{(v)}} \|(\mathbf{X}^{(v)} - \mathbf{U}^{(v)}\mathbf{V}^{\top})\mathbf{W}^{(v)}\|_F^2 + \alpha \|\mathbf{B}^{(v)\top}\mathbf{U}^{(v)} - \mathbf{I}\|_F^2. \tag{9}$$

Setting the derivation of the objective function with respect to $\mathbf{U}^{(v)}$, we can get the following equation:

$$(\mathbf{U}^{(v)}\mathbf{V}^{\top} - \mathbf{X}^{(v)})\mathbf{W}^{(v)}\mathbf{V} + \alpha\mathbf{B}^{(v)}(\mathbf{B}^{(v)\top}\mathbf{U}^{(v)} - \mathbf{I}) = 0, \tag{10}$$

where $\mathbf{W}^{(v)} = \mathbf{W}^{(v)}\mathbf{W}^{(v)\top}$. According to the form of the Sylvester equation, Equation (10) can be further reformulated into Equation (11) utilizing Kronecker product and vectorization operator as follows:

$$(\mathbf{I}_k \otimes (\alpha\mathbf{B}^{(v)}\mathbf{B}^{(v)\top}) + (\mathbf{V}^{\top}\mathbf{W}^{(v)}\mathbf{V}) \otimes \mathbf{I}_{d_v})\mathrm{vec}(\mathbf{U}^{(v)}) = \mathrm{vec}(\mathbf{X}^{(v)}\mathbf{W}^{(v)}\mathbf{V} + \alpha\mathbf{B}^{(v)}). \tag{11}$$

In this paper, we employ the *lyap* function following the work in Reference [19] to solve this problem.

## 4.2 | Update V

Fixing $\mathbf{U}^{(v)}$, $\mathbf{B}^{(v)}$, and $\mathbf{S}$, the minimum problem for optimizing $\mathbf{V}$ is reduced into the following problem:

$$\begin{aligned}\min_{\mathbf{V}} \ & \sum_{v=1}^{m} \|(\mathbf{X}^{(v)} - \mathbf{U}^{(v)}\mathbf{V}^{\top})\mathbf{W}^{(v)}\|_F^2 \\ & + \frac{1}{2}\sum_{i,j=1}^{n} \mathbf{S}_{ij} \ \|v_i - v_j\|_2^2 \ , \quad \text{s.t.} \quad \mathbf{V} \geq 0.\end{aligned} \tag{12}$$

The second term can be rewritten into the form of matrix according to Theorem 1.

**Theorem 1.** *Note that* $\mathbf{L} = \mathbf{D} - \mathbf{S}$. $\mathbf{L}$ *and* $\mathbf{D}$ *are the Laplacian matrix and degree matrix of the symmetric similarity matrix* $\mathbf{S}$*, respectively. Then we can get the following conclusion:*

$$\frac{1}{2}\sum_{i,j=1}^{n} \mathbf{S}_{ij} \ \|v_i - v_j\|_2^2 \ = \mathrm{Tr}(\mathbf{V}^{\top}\mathbf{L}\mathbf{V}), \tag{13}$$

*where* $v_i$ *and* $v_j$ *are the* $i$*th and* $j$*th rows of matrix* $\mathbf{V}$, $S_{i,j}$ *is the element of* $i$*th row and* $j$*th column of* $\mathbf{S}$.

*Proof.* We can transform into the following equation mathematically:

$$\begin{aligned}\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{S}_{ij} \ \|v_i - v_j\|_2^2 &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{S}_{ij}\left(v_i^{\top}v_i - 2v_i^{\top}v_j + v_j^{\top}v_j\right) \\ &= \frac{1}{2}\left(2\sum_{i=1}^{n} d_{ii}v_i^{\top}v_i - 2\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{S}_{ij}v_i^{\top}v_j\right) \\ &= \mathrm{Tr}(\mathbf{V}^{\top}\mathbf{D}\mathbf{V}) - \mathrm{Tr}(\mathbf{V}^{\top}\mathbf{S}\mathbf{V}) = \mathrm{Tr}(\mathbf{V}^{\top}\mathbf{L}\mathbf{V}).\end{aligned} \tag{14}$$

The proof is completed. □

Thus, we can transform Equation (12) into minimizing the following problem:

$$\min_{\mathbf{V}} \sum_{v=1}^{m} \|(\mathbf{X}^{(v)} - \mathbf{U}^{(v)}\mathbf{V}^{\top})\mathbf{W}^{(v)}\|_F^2 \ + \mathrm{Tr}(\mathbf{V}^{\top}\mathbf{L}\mathbf{V}), \quad \text{s.t.} \quad \mathbf{V} \geq 0. \tag{15}$$

The partial derivation of $\mathcal{L}(\mathbf{V})$ with respect to $\mathbf{V}$ is

$$\frac{\partial \mathcal{L}(\mathbf{V})}{\partial \mathbf{V}} = \mathbf{W}^{(v)}\mathbf{V}\mathbf{U}^{(v)\top}\mathbf{U}^{(v)} + \mathbf{L}\mathbf{V} - \mathbf{W}^{(v)}\mathbf{X}^{(v)\top}\mathbf{U}^{(v)}. \tag{16}$$

According to the optimization of semi-NMF[32] and the KKT condition, we can get

$$(\mathbf{W}^{(v)}\mathbf{V}\mathbf{U}^{(v)\top}\mathbf{U}^{(v)} + \mathbf{L}\mathbf{V} - \mathbf{W}^{(v)}\mathbf{X}^{(v)\top}\mathbf{U}^{(v)})_{jk}\mathbf{V}_{jk} = 0. \tag{17}$$

Based on this, we can get the updating rule by Following the work in Reference [19], we normalize the solution $\mathbf{V}$ and $\mathbf{U}^{(v)}$ by $\mathbf{V} \leftarrow \mathbf{V}\mathbf{H}^{-1}$ and $\mathbf{U}^{(v)} \leftarrow \mathbf{U}^{(v)}\mathbf{H}^{-1}$. Consequently, $\mathbf{V}\mathbf{H}^{-1}$ and

$\mathbf{U}^{(v)}\mathbf{H}^{-1}$ forms another solution of Equation (12), where $\mathbf{H}$ is an invertible matrix and is defined as $\mathbf{H} = diag\left(\sum_i^n H_{i1}, \sum_i^n H_{i2}, ..., \sum_i^n H_{ik}\right)$.

## 4.3 | Update $\mathbf{B}^{(v)}$

Given $\mathbf{U}^{(v)}, \mathbf{V}$ and $\mathbf{S}$, the optimization for updating $\mathbf{B}^{(v)}$ can be reduced to

$$\min_{\mathbf{B}^{(v)}} \|\mathbf{B}^{(v)\top}\mathbf{U}^{(v)} - \mathbf{I}\|_F^2 + \|\mathbf{B}^{(v)}\|_{2,1}. \tag{18}$$

Similarly, through setting the derivation of $\mathcal{L}(\mathbf{B}^{(v)})$ w.r.t $\mathbf{B}^{(v)}$ to zero, we can obtain the close-form solution for updating $\mathbf{B}^{(v)}$.

$$\mathbf{B}^{(v)} = (2\mathbf{U}^{(v)}\mathbf{U}^{(v)\top} + \mathbf{P})^{-1}2\mathbf{U}^{(v)}, \tag{19}$$

where $\frac{\partial \|\mathbf{B}^{(v)}\|_{2,1}}{\partial \mathbf{B}^{(v)}} = \mathbf{P}^{(v)}\mathbf{B}^{(v)}$ and $\mathbf{P}^{(v)}$ is diagonal matrix with its $i$th diagonal element being $\frac{1}{\|\mathbf{B}_{i,:}^{(v)}\|_2}$.

## 4.4 | Update S

When $\mathbf{U}^{(v)}, \mathbf{V}$, and $\mathbf{B}^{(v)}$ are fixed, the optimization for $\mathbf{S}$ can be simplified as:

$$\min_{\mathbf{S}} \frac{1}{2}\sum_{i,j=1}^{n} S_{ij} \|v_i - v_j\|_2^2 + \lambda \|\mathbf{S}\|_F^2, \quad \text{s.t.} \quad \mathbf{S} \geq 0, \quad \mathbf{S}^\top\mathbf{1} = \mathbf{1}. \tag{20}$$

**Theorem 2.** *$\mathbf{L}$ is the Laplacian matrix of symmetry similarity matrix $\mathbf{S}$ and the $i,j$th element of $\mathbf{Q}$ is denoted as $Q_{ij} = \|v_i - v_j\|_2^2$, where $v_i$ is the $i$th row of matrix $\mathbf{V}$. Then we can obtain the equivalent representation.*[33]

$$\mathrm{Tr}(\mathbf{V}^\top\mathbf{L}\mathbf{V}) = \mathrm{Tr}(\mathbf{S}\mathbf{Q}). \tag{21}$$

*Proof.* We can achieve this equation by mathematical derivation as following:

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij} \|v_i - v_j\|_2^2 = \frac{1}{2}\mathbf{1}^\top(\mathbf{Q} \odot \mathbf{S})\mathbf{1} = \frac{1}{2}\mathrm{Tr}(\mathbf{S}\mathbf{Q}). \tag{22}$$

This finishes the proof. □

On the basis of Theorem 2, we optimize $\mathbf{S}$ by transforming into Equation (23)

$$\min_{\mathbf{S}} \frac{1}{2}\mathrm{Tr}(\mathbf{S}\mathbf{Q}) + \lambda \|\mathbf{S}\|_F^2, \quad \text{s.t.} \quad \mathbf{S} \geq 0, \quad \mathbf{S}^\top\mathbf{1} = \mathbf{1}. \tag{23}$$

Analogous to Reference [34], we adopt a two-step algorithm to gain a approximate result efficiently. In the first step, we neglect the constraints and straightly acquire a close-form solution Equation (24) by setting the derivation w.r.t $\mathbf{S}$ to zero.

$$\hat{\mathbf{S}} = -\frac{1}{2\lambda}\mathbf{Q}^\top. \tag{24}$$

In the second step, we minimize the following problem to learn a approximate optimal solution under a constrained condition:

$$\min_{\mathbf{S}} \frac{1}{2} \|\mathbf{S} - \hat{\mathbf{S}}\|_F^2, \quad s.t. \quad \mathbf{S} \geq 0, \quad \mathbf{S}^\top \mathbf{1} = \mathbf{1}. \tag{25}$$

For each column $S_{:,j}$, the Lagrange function can be formulated as

$$\mathcal{L} = \frac{1}{2} \|S_{:,j} - \hat{S}_{:,j}\|_2^2 - \tau_i \left( S_{:,j}^\top \mathbf{1} - 1 \right) - \gamma_i^\top S_{:,j}. \tag{26}$$

where $\tau_j$ and $\gamma_j$ are Lagrange multipliers. Therefore, combining the result of taking the derivation with respect to $S_{:,j}$ and the KKT condition, we get

$$S_{:,j} = max(\hat{S}_{:,j} + \tau_j \mathbf{1}, 0). \tag{27}$$

Considering the constraint $\mathbf{S}^\top \mathbf{1} = \mathbf{1}$, that is, $\sum_{i=1}^{n} (\hat{S}_{ij} + \tau_j \mathbf{1}) = 1$, we can obtain the update rule for $\tau_j$ that

$$\tau_j = \frac{1 + \hat{S}_{:,j}^\top \mathbf{1}}{n - 1}. \tag{28}$$

We can update $S_{:,j}$ by Equation (27) after gaining $\tau_j$. Then we perform $\mathbf{S} = \frac{\mathbf{S} + \mathbf{S}^\top}{2}$ to keep the symmetry of graph $\mathbf{S}$.

The alternate optimization algorithm terminates when objective function converges or reaches the maximum number of iterations. Then we get an approximate optimal solution of Equation (20) in an efficient fashion.

### 4.5 | Convergence analysis

In the iterative optimization process, when one variable is updated and the other variables are kept fixed, these two subproblems are strictly convex. When optimizing a variable and fixing other variables in each iteration, the goal of our algorithm will decrease monotonically. Until the lower bound of the entire optimization function approaches 0, the algorithm will converge. Also, the results of the evolution of the objective value on the benchmark data sets demonstrates the convergence, as shown in Figures 2, 3, and 4.

## 5 | EXPERIMENTS

In this section, we introduce the data sets used in this paper and the setting of generating incomplete samples. Then, we evaluate the effectiveness of our proposed method and compared state-of-the-art algorithms on the six natural or manual incomplete data sets.

### 5.1 | Data sets

The data sets used in our experiments are BUAA-visnir face database (BUAA)*, orlRnSp†, 3 Sources‡, BBC.[35]

*BUAA*. The BUAA data set contains 1350 samples, belonging to 150 clusters. Each of sample are composed with visible images (VIS) and near infrared images (NIR). This makes the data set naturally been a two-view data set.

*orlRnSp*. orlRnSp data set has 400 face images. We construct two views one based on raw pixel values and the other comprising of HOG features.

*3Sources*. This data set is composed of 948 news articles, which are collected from BBC, Reuters, and Guardian. We select a subset containing 169 stories reported in the above three sources. The 169 stories have six different subject and therefore can be group into six classes.

*BBC*. The BBC data set contains 2225 documents about the sport news articles from the BBC Sport website. These documents are described by 2–4 views and categorized into five classes.

The detailed information of these data sets is shown in Table 2. The first four data sets are complete multiview data sets. We generate different incomplete data sets under two incomplete setting, which we introduce in next section. The last two data sets are naturally incomplete.

## 5.2 | Experiment setting

In this section, we will introduce our experimental setting for handmade incomplete data sets.

*One-complete setting*. We randomly select one view as the complete view and the rest of the views suffer incomplete ratio (IR) equally from 10% to 70%. In this way, we generate the incomplete data sets under one-complete setting.

*Random incomplete setting*. Under random incomplete setting, we randomly remove 10%, 20%, 30%, 40%, and 50% samples in each view, but each sample has at least one existing view.

*Natural incomplete setting*. Data sets 3Sources and BBC are natural incomplete. The missing samples in each view are denoted as NaN.

## 5.3 | Compared algorithm

In this section, we list the compared algorithms as follows,

- Best single view (BSV): BSV method first fills the missing samples with the average of other existing samples. Then gaussian affinity matrices are constructed on each view to perform spectral clustering. The best performance is reported as BSV.
- Direct concatenation (Concate): Similar to BSV, Concate first fills the missing instance. Then all features across views are concatenated into a single view representation. The clustering results are obtained by performing $K$-means on the connected feature.

**TABLE 2** Description of the benchmark multiview data set

| Data set | Sample | View | Cluster | Feature dimension | | | |
|---|---|---|---|---|---|---|---|
| orlRnSp | 400 | 2 | 40 | 1024 | 288 | – | – |
| BUAA | 1350 | 2 | 150 | 100 | 100 | – | – |
| 3Sources | 169 | 3 | 6 | 3560 | 3610 | 3068 | – |
| BBC | 2225 | 4 | 5 | 4659 | 4633 | 4665 | 4684 |

- Incomplete multimodal visual data grouping (IMG).[18] IMG proposes to use the latent representation to generate a complete graph, which establishes a connection between missing data from different views.
- IMVC via weighted NMF with $\ell_{2,1}$ regularization (MIC).[29] This paper first fills the missing instances with average values of features and then learns a $\ell_{2,1}$ regularized latent subspace by weighted NMF.
- Partial multiview clustering (PVC).[17] Based on the NMF, this method integrates the common and view-specific representations in the latent space to form a unified representation.
- Incomplete multiview spectral clustering with adaptive graph learning (INMF-AGL)[36] induces a coregularization term to learn the common representation, which integrates the graph learning and spectral clustering.
- Doubly aligned incomplete multiview clustering (DAIMC).[19] The proposed method first aligns the samples into a common representation by semi-NMF and then aligns the base matrices with the help of $\ell_{2,1}$ regularized regression modal.

Note that PVC and IMG only work for two-view data. Following work in Reference [19], we evaluate PVC and IMG on all two-view combinations and report the best results. MIC, IMSC-AGL, DAIMC, and our proposed method can be applied for arbitrary incomplete multiview data. Then we use these methods as suggusted.

## 5.4 | Experimental results and analysis

We generated the incomplete data sets under various settings according to the introduction in experiment setting. Therefore, we obtain the BUAA data set with one view that is complete and the random missing orlRnSp data set. Data set 3Source and BBC are natural incomplete. The experimental results of different data set at different settings are shown in Tables 3–6. We report five metrics of accuracy (ACC), normalized mutual information (NMI), Purity, F-score and precision to compare the clustering performance. The best results are presented as bold numbers.

**TABLE 3** The clustering performance of different methods on natural incomplete data set 3Source

| Method | ACC | NMI | Purity | F-score | Precision |
|---|---|---|---|---|---|
| BSV | 0.2716 | 0.0500 | 0.3101 | 0.2008 | 0.1933 |
| Concate | 0.3966 | 0.2176 | 0.4303 | 0.3114 | 0.3026 |
| IMG | 0.3510 | 0.2032 | 0.3675 | 0.2947 | 0.1961 |
| MIC | 0.4313 | 0.3601 | 0.4678 | 0.3273 | 0.2554 |
| PVC | 0.5859 | 0.4720 | 0.6326 | 0.5083 | 0.4924 |
| IMSC-AGL | 0.8173 | 0.6678 | 0.8173 | 0.7122 | 0.6949 |
| DAIMC | 0.8125 | 0.6751 | 0.8125 | 0.6790 | 0.6924 |
| Ours | **0.8510** | **0.7073** | **0.8510** | **0.7440** | **0.7621** |

*Note*: Bold numbers denote the best results and the underlined numbers denote the second-best results.

Abbreviations: ACC, accuracy; BSV, best single view; Concate, direct concatenation; DAIMC, doubly aligned incomplete multiview clustering; IMG, Incomplete multimodal visual data grouping; INMF-AGL, incomplete multiview spectral clustering with adaptive graph learning; NMI, normalized mutual information.

**TABLE 4** The clustering performance of different methods on natural incomplete data set BBC

| Method | ACC | NMI | Purity | F-score | Precision |
|---|---|---|---|---|---|
| BSV | 0.2679 | 0.0184 | 0.2724 | 0.2511 | 0.2155 |
| Concate | 0.2710 | 0.0361 | 0.3065 | 0.2338 | 0.2262 |
| IMG | 0.3741 | 0.1852 | 0.3757 | 0.3430 | 0.2392 |
| MIC | 0.4951 | 0.3572 | 0.5183 | 0.4114 | 0.3145 |
| PVC | 0.6149 | 0.4272 | 0.6394 | 0.5181 | 0.4871 |
| IMSC-AGL | 0.9020 | 0.7391 | 0.9020 | 0.8214 | 0.8178 |
| DAIMC | 0.8503 | 0.6785 | 0.8503 | 0.7311 | 0.7208 |
| Ours | **0.9267** | **0.7919** | **0.9267** | **0.8641** | **0.8601** |

*Note*: Bold numbers denote the best results and the underlined numbers denote the second-best results.

Abbreviations: ACC, accuracy; BSV, best single view; Concate, direct concatenation; DAIMC, doubly aligned incomplete multiview clustering; IMG, Incomplete multimodal visual data grouping; INMF-AGL, Incomplete multiview spectral clustering with adaptive graph learning; NMI, normalized mutual information.

Tables 3 and 4 show the comparison of the data sets 3Source and BBC with other comparison algorithms under the natural incomplete setting with five evaluation metrics, respectively.

It can be seen from the two tables that our proposed algorithm achieves the best results under all metrics compared to other comparison algorithms. Our results outperform DAIMC because the graphs we learn contain not only the consistency information in the consensus representation but also the local structural information between samples. The latest IMVC algorithms based on graph and matrix decomposition or subspace (our proposed method and IMSC-AGL) outperform those based on matrix decomposition alone, which illustrates the importance of mining local information between samples for the clustering task.

Tables 5 and 6 show the comparative experimental results for BUAA and orlRnSp under one-complete and random incomplete settings, respectively. In the first one-complete setting, the two view data set BUAA suffers 10%–70% missing samples in one view with another view is complete. Compared with other methods, our proposed method has the best performance no matter under different evaluation criteria or different incomplete ratio. The ACC obtained by our method exceeds the corresponding DAIMC results by 18.67%, 14.08%, 16.37%, 12.08%, 11.19%, 9.19%, and 7.11% from the deletion rate of 10%–70%, respectively. This illustrates the effectiveness of the local structure extract between the consensus representation.

Under the random incomplete setting in Figure 7, our proposed method outperforms IMSC-AGL by 20%, 20%, 24%, 16.25%, and 9.25% in terms of ACC, respectively. Although IMSC-AGL combines subspace learning and consensus representation learning, the high-dimensional representation could lead to inaccurate distance measurement. Our algorithm incorporates the NMF and LGR, learning of consistent representation and acquisition of local structure in a more low-dimensional and compact feature space, which contributes to our excellent performance.

To show the comparison between different methods more clearly, we draw the ACC of different methods under different missing rates as a line graph, as shown in Figure 1. From this picture, we can find that our method (red line) is always higher than the other methods and can keep relatively stable as the incomplete ratio increases.

**TABLE 5** The clustering performance on BUAA under one-complete setting

| IR | Metric | BSV | Concate | IMG | MIC | PVC | IMSC-AGL | AGC-IMC | DAIMC | Ours |
|---|---|---|---|---|---|---|---|---|---|---|
| 10% | ACC | 0.2889 | 0.3570 | 0.3064 | 0.0193 | 0.2699 | 0.2607 | 0.4593 | 0.3585 | **0.5452** |
| | NMI | 0.6438 | 0.6728 | 0.5622 | 0.2215 | 0.6253 | 0.6219 | 0.7329 | 0.6857 | **0.7770** |
| | Purity | 0.3281 | 0.3733 | 0.3516 | 0.1170 | 0.2865 | 0.2800 | 0.4756 | 0.3822 | **0.5733** |
| | F-score | 0.1271 | 0.1663 | 0.0322 | 0.0132 | 0.0953 | 0.0930 | 0.2680 | 0.1869 | **0.3746** |
| | Precision | 0.0925 | 0.1424 | 0.0168 | 0.0067 | 0.0802 | 0.0770 | 0.2472 | 0.1550 | **0.3504** |
| 20% | ACC | 0.2852 | 0.3333 | 0.2881 | 0.0193 | 0.2318 | 0.2541 | 0.4400 | 0.3222 | **0.4630** |
| | NMI | 0.6419 | 0.6602 | 0.5355 | 0.2215 | 0.5936 | 0.6201 | 0.7267 | 0.6575 | **0.7372** |
| | Purity | 0.3222 | 0.3578 | 0.3276 | 0.1170 | 0.2510 | 0.2763 | 0.4563 | 0.3504 | **0.4859** |
| | F-score | 0.1264 | 0.1394 | 0.0292 | 0.0132 | 0.0665 | 0.0896 | 0.2572 | 0.1417 | **0.2840** |
| | Precision | 0.0917 | 0.1173 | 0.0152 | 0.0067 | 0.0497 | 0.0731 | 0.2394 | 0.1169 | **0.2642** |
| 30% | ACC | 0.2852 | 0.2844 | 0.2729 | 0.0193 | 0.2381 | 0.2548 | 0.4170 | 0.2985 | **0.4622** |
| | NMI | 0.6419 | 0.6422 | 0.5168 | 0.2215 | 0.5945 | 0.6188 | 0.7152 | 0.6446 | **0.7287** |
| | Purity | 0.3222 | 0.3163 | 0.3119 | 0.1170 | 0.2522 | 0.2733 | 0.4363 | 0.3252 | **0.4881** |
| | F-score | 0.1264 | 0.1096 | 0.0248 | 0.0132 | 0.0625 | 0.0880 | 0.2357 | 0.1166 | **0.2692** |
| | Precision | 0.0917 | 0.0886 | 0.0129 | 0.0067 | 0.0483 | 0.0732 | 0.2225 | 0.0953 | **0.2484** |
| 40% | ACC | 0.3007 | 0.2763 | 0.2541 | 0.0193 | 0.2291 | 0.2474 | 0.4126 | 0.2970 | **0.4178** |
| | NMI | 0.6509 | 0.6320 | 0.5039 | 0.2215 | 0.5796 | 0.6146 | 0.7082 | 0.6450 | **0.7110** |
| | Purity | 0.3415 | 0.3007 | 0.2927 | 0.1170 | 0.2440 | 0.2667 | 0.4259 | 0.3230 | **0.4348** |
| | F-score | 0.1365 | 0.1024 | 0.0247 | 0.0132 | 0.0477 | 0.0826 | 0.2219 | 0.1223 | **0.2380** |
| | Precision | 0.1008 | 0.0817 | 0.0128 | 0.0067 | 0.0341 | 0.0681 | 0.2107 | 0.0964 | **0.2203** |
| 50% | ACC | 0.2852 | 0.2763 | 0.2347 | 0.0193 | 0.2088 | 0.2474 | 0.4074 | 0.3111 | **0.4230** |
| | NMI | 0.6419 | 0.6268 | 0.4804 | 0.2215 | 0.5778 | 0.6128 | 0.7077 | 0.6511 | **0.7169** |
| | Purity | 0.3222 | 0.2948 | 0.2685 | 0.1170 | 0.2221 | 0.2637 | 0.4222 | 0.3348 | **0.4452** |
| | F-score | 0.1264 | 0.0973 | 0.0232 | 0.0132 | 0.0436 | 0.0840 | 0.2238 | 0.1342 | **0.2413** |
| | Precision | 0.0917 | 0.0735 | 0.0121 | 0.0067 | 0.0314 | 0.0696 | 0.2118 | 0.1048 | **0.2235** |
| 60% | ACC | 0.2852 | 0.2719 | 0.2304 | 0.0193 | 0.1971 | 0.2467 | 0.3904 | 0.3244 | **0.4163** |
| | NMI | 0.6419 | 0.6256 | 0.4808 | 0.2215 | 0.5500 | 0.6153 | 0.6982 | 0.6529 | **0.7059** |
| | Purity | 0.3222 | 0.2919 | 0.2632 | 0.1170 | 0.2078 | 0.2667 | 0.4067 | 0.3474 | **0.4385** |
| | F-score | 0.1264 | 0.0971 | 0.0231 | 0.0132 | 0.0300 | 0.0837 | 0.2040 | 0.1369 | **0.2187** |
| | Precision | 0.0917 | 0.0737 | 0.0120 | 0.0067 | 0.0199 | 0.0707 | 0.1890 | 0.1063 | **0.2004** |
| 70% | ACC | 0.2889 | 0.2615 | 0.1996 | 0.0193 | 0.1699 | 0.2452 | 0.3830 | 0.3311 | **0.4022** |
| | NMI | 0.6438 | 0.6146 | 0.4661 | 0.2215 | 0.5016 | 0.6122 | 0.6984 | 0.6653 | **0.7050** |
| | Purity | 0.3281 | 0.2830 | 0.2344 | 0.1170 | 0.1836 | 0.2622 | 0.4030 | 0.3504 | **0.4244** |
| | F-score | 0.1271 | 0.0940 | 0.0214 | 0.0132 | 0.0216 | 0.0806 | 0.2025 | 0.1596 | **0.2165** |

(Continues)

**TABLE 5** (Continued)

| IR | Metric | BSV | Concate | IMG | MIC | PVC | IMSC-AGL | AGC-IMC | DAIMC | Ours |
|---|---|---|---|---|---|---|---|---|---|---|
| | Precision | 0.0925 | 0.0707 | 0.0111 | 0.0067 | 0.0125 | 0.0674 | 0.1924 | 0.1275 | **0.2015** |

*Note*: Bold numbers denote the best results.

Abbreviations: ACC, accuracy; BSV, best single view; Concate, direct concatenation; DAIMC, doubly aligned incomplete multiview clustering; IMG, Incomplete multimodal visual data grouping; INMF-AGL, incomplete multiview spectral clustering with adaptive graph learning; NMI, normalized mutual information.

**TABLE 6** The clustering performan on orlRnSp under thet random incomplete setting

| IR | Metric | BSV | Concate | IMG | MIC | PVC | IMSC-AGL | AGC-IMC | DAIMC | Ours |
|---|---|---|---|---|---|---|---|---|---|---|
| 10% | ACC | 0.2625 | 0.4675 | 0.4853 | 0.5535 | 0.5091 | 0.5150 | 0.6625 | 0.6675 | **0.715** |
| | NMI | 0.4972 | 0.6692 | 0.6885 | 0.7212 | 0.7004 | 0.7103 | 0.806 | 0.8162 | **0.8478** |
| | Purity | 0.2925 | 0.5325 | 0.555 | 0.581 | 0.5466 | 0.5475 | 0.6900 | 0.7050 | **0.755** |
| | F-score | 0.0918 | 0.2261 | 0.1783 | 0.3435 | 0.3193 | 0.3688 | 0.5468 | 0.5476 | **0.5926** |
| | Precision | 0.0535 | 0.1557 | 0.108 | 0.2794 | 0.2628 | 0.3449 | 0.5225 | 0.5052 | **0.5657** |
| 20% | ACC | 0.2325 | 0.4100 | 0.4348 | 0.4680 | 0.4616 | 0.4925 | 0.5800 | 0.6425 | **0.6925** |
| | NMI | 0.4438 | 0.6160 | 0.6275 | 0.6496 | 0.6449 | 0.6824 | 0.7632 | 0.7920 | **0.8423** |
| | Purity | 0.2600 | 0.4550 | 0.496 | 0.5075 | 0.4978 | 0.5200 | 0.6275 | 0.6750 | **0.7375** |
| | F-score | 0.0719 | 0.1901 | 0.1210 | 0.2007 | 0.1942 | 0.3240 | 0.4611 | 0.5081 | **0.5862** |
| | Precision | 0.0433 | 0.1388 | 0.0694 | 0.1384 | 0.1346 | 0.2995 | 0.4254 | 0.4547 | **0.5373** |
| 30% | ACC | 0.2425 | 0.4000 | 0.3820 | 0.3255 | 0.4061 | 0.4575 | 0.5575 | 0.6125 | **0.6975** |
| | NMI | 0.4621 | 0.5911 | 0.5514 | 0.5090 | 0.5787 | 0.6569 | 0.7184 | 0.7644 | **0.8259** |
| | Purity | 0.2600 | 0.4350 | 0.4330 | 0.3515 | 0.4380 | 0.4875 | 0.5825 | 0.6425 | **0.7375** |
| | F-score | 0.0852 | 0.1793 | 0.0778 | 0.0974 | 0.1133 | 0.2891 | 0.4038 | 0.4508 | **0.5783** |
| | Precision | 0.0742 | 0.1354 | 0.0426 | 0.0631 | 0.0689 | 0.2630 | 0.3751 | 0.3947 | **0.5301** |
| 40% | ACC | 0.2425 | 0.3475 | 0.3343 | 0.2720 | 0.3505 | 0.3775 | 0.4125 | 0.4925 | **0.5400** |
| | NMI | 0.4824 | 0.5764 | 0.4886 | 0.4917 | 0.5220 | 0.5862 | 0.6279 | 0.6912 | **0.7396** |
| | Purity | 0.2600 | 0.3875 | 0.3678 | 0.2935 | 0.3790 | 0.3900 | 0.4450 | 0.5350 | **0.5925** |
| | F-score | 0.0897 | 0.1602 | 0.0614 | 0.1024 | 0.0742 | 0.1929 | 0.2423 | 0.3218 | **0.3825** |
| | Precision | 0.0806 | 0.1173 | 0.0330 | 0.0777 | 0.0420 | 0.1791 | 0.2282 | 0.2799 | **0.3448** |
| 50% | ACC | 0.2650 | 0.3550 | 0.1285 | 0.2930 | 0.3533 | 0.3175 | 0.4075 | 0.3900 | **0.4100** |
| | NMI | 0.5001 | 0.576 | 0.2714 | 0.5214 | 0.5988 | 0.5542 | 0.6372 | 0.6228 | **0.6408** |
| | Purity | 0.2800 | 0.3850 | 0.1510 | 0.3110 | 0.3863 | 0.3375 | 0.4425 | 0.4275 | **0.4575** |
| | F-score | 0.0948 | 0.1541 | 0.0454 | 0.1257 | 0.1895 | 0.1498 | 0.2370 | 0.2237 | **0.2472** |
| | Precision | 0.0897 | 0.1115 | 0.0250 | 0.1023 | 0.1650 | 0.1380 | 0.2227 | 0.1969 | **0.2361** |

*Note*: The bold number demotes the best result.

Abbreviations: ACC, accuracy; BSV, best single view; Concate, direct concatenation; DAIMC, doubly aligned incomplete multiview clustering; IMG, Incomplete multimodal visual data grouping; INMF-AGL, incomplete multiview spectral clustering with adaptive graph learning; NMI, normalized mutual information.

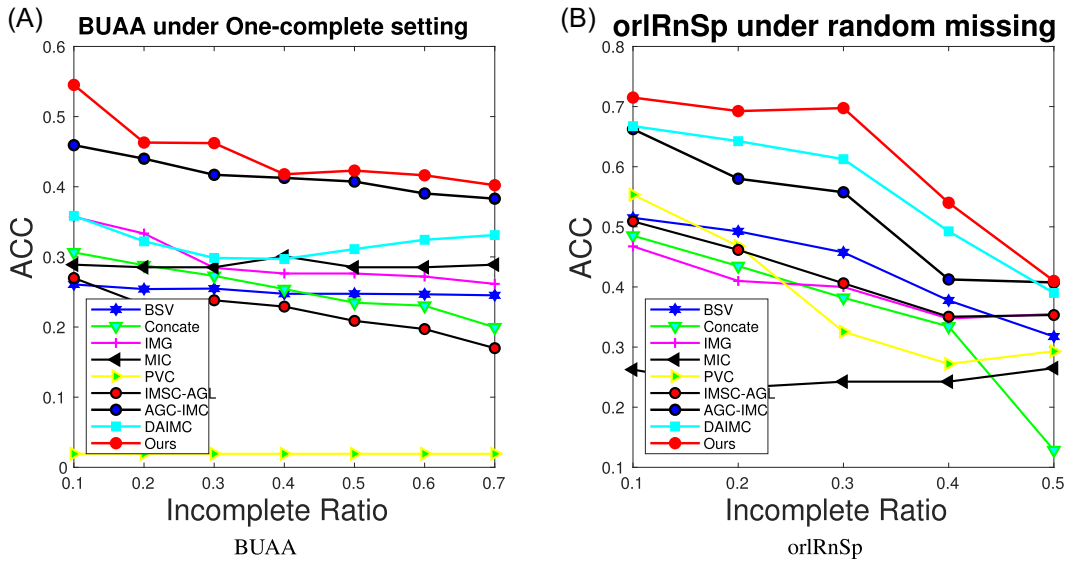


**FIGURE 1** The ACC performance variation with various incomplete ratio increasing on BUAA and orlRnSp under One-complete setting. ACC, accuracy [Color figure can be viewed at wileyonlinelibrary.com]

In summary, the above experimental results have well demonstrated the effectiveness of our proposed method comparing to other state-of-the-art methods. We attribute the superiority of proposed algorithm as two aspects: (i) We induce the $\ell_{2,1}$ regularization term into traditional NMF-based IMVC approaches. This regularization not only ensures sparsity in various base matrices but also performs feature selection in the original model. (ii) The constructed graph is utilized for preserving the locality in the consensus representation while also encourages the view-consistency and degrades the intra-view disagreements. These factors greatly promote the clustering performance of our proposed method.

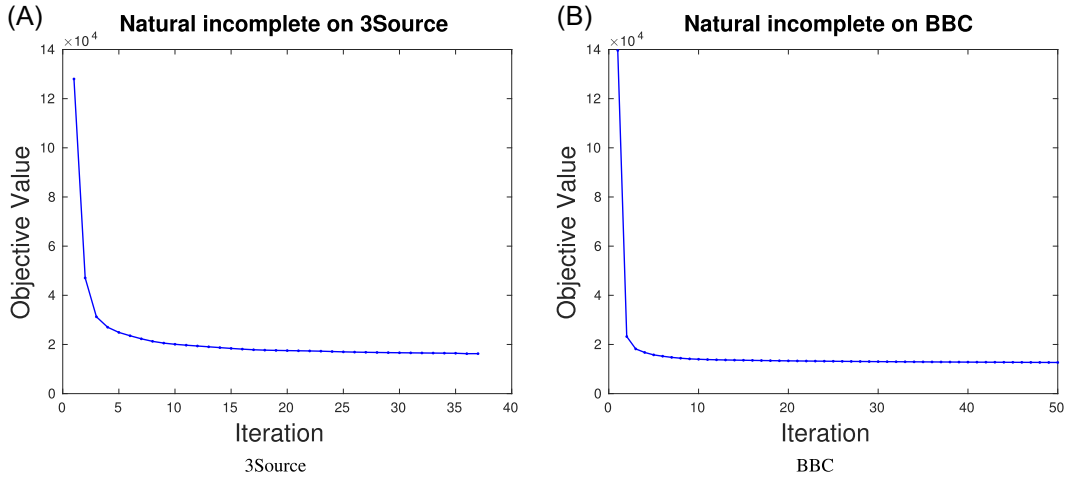


**FIGURE 2** The convergence curves of the objective values on data set orlRnSp under random incomplete setting [Color figure can be viewed at wileyonlinelibrary.com]

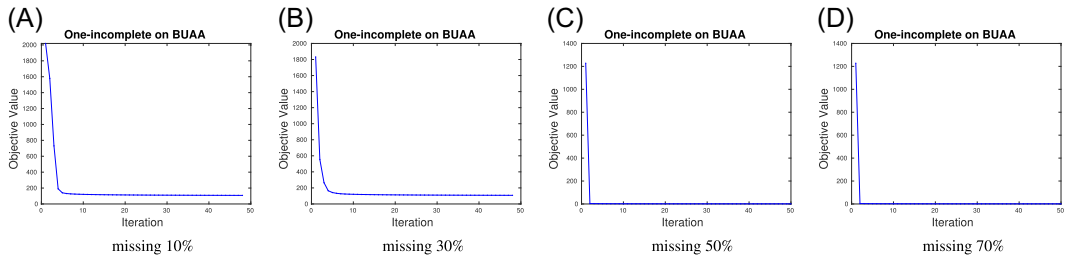


**FIGURE 3** The convergence curves of the objective values on data set BUAA under One-incomplete setting [Color figure can be viewed at wileyonlinelibrary.com]

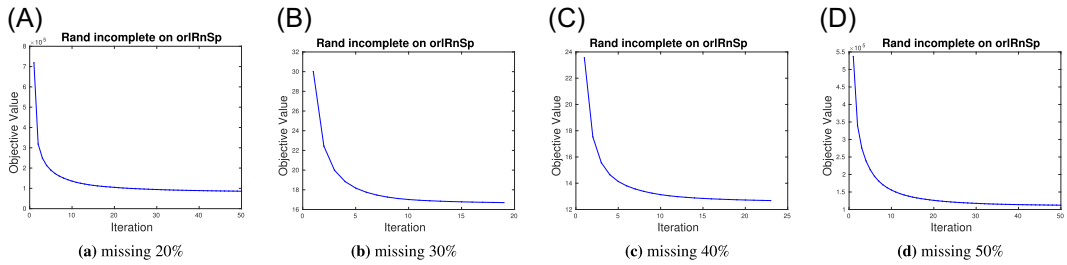


**FIGURE 4** The convergence curves of the objective values on data set orlRnSp under random incomplete setting [Color figure can be viewed at wileyonlinelibrary.com]

## 5.5 | Convergence and parameter sensitivity

*Convergence*. Our algorithm is theoretically guaranteed to converge to a local minimum. The examples of the evolution of the objective value on the experimental results are shown in Figures 2, 3, and 4. Figure 2 shows the convergence visualization of the natural incomplete data sets 3Source and BBC. Figures 3 and 4 are the curves of objective value under one-complete setting and random incomplete setting. In the above experiments, we observe that the objective

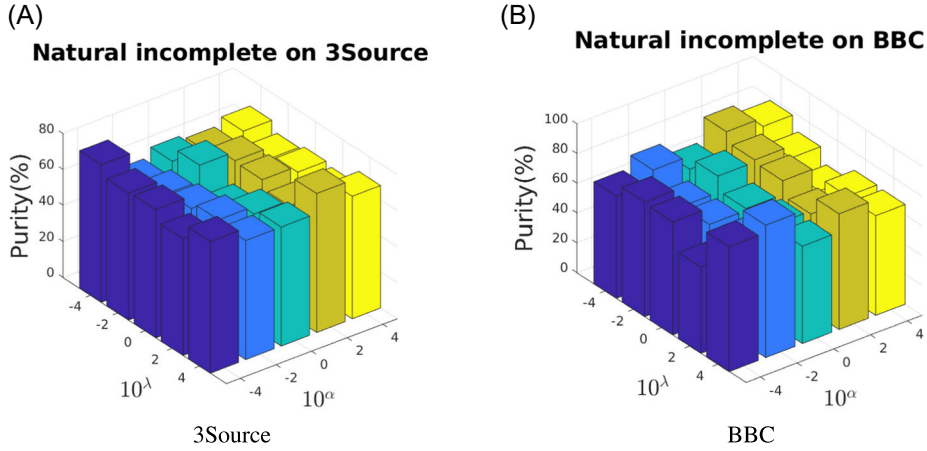


**FIGURE 5** The sensitivity of purity on natural incomplete data set 3Source and BBC [Color figure can be viewed at wileyonlinelibrary.com]

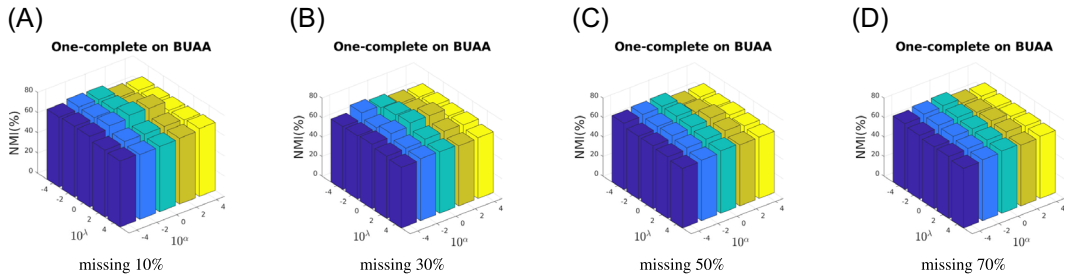


**FIGURE 6** The NMI sensitivity on data set BUAA under One-complete setting. NMI, normalized mutual information [Color figure can be viewed at wileyonlinelibrary.com]

value of our algorithm does monotonically decrease at each iteration and that it usually converges in less than 50 iterations. These results clearly verifies our proposed algorithm's convergence.

*Sensitivity*. In addition, we analyze the sensitivity of the clustering performance to the hyper parameters $\alpha$ and $\lambda$. Parameters are chosen in the range of $10^{-4}$ to $10^{4}$ in steps of $10^{2}$. We use three different evaluation metrics, ACC NMI and purity, in three different settings to demonstrate the effect of parameter variation on clustering performance. Figure 5 is the sensitivity of hyper-parameters on natural incomplete data set 3Source and BBC in terms of purity.

Figures 6 and 7 show the variety of clustering performance of the data set BUAA and orlRnSp under one-complete and random incomplete settings for all parameter combinations.

From these observations, we can find that the performance of BUAA and orlRnSp can maintain relatively stable with the change of parameters at some missing ratio. And it shows a trend of obtaining the optimal performance between the chosen range of parameters $\alpha$ and $\lambda$. However, the natural incomplete data sets 3Source and BBC do not display a clear pattern in their sensitivity bar chart.

## 6 | CONCLUSION

In this article, we propose a novel partial no-negative matrix factorization to handle IMVC. We incorporate NMF and local graph regularisation into a unified framework, whereas two processes can negotiate with each other serving for learning task. Moreover, we add the base coefficient regression $\ell_{2,1}$-norm regularization term to constraint the various base matrices among views. This

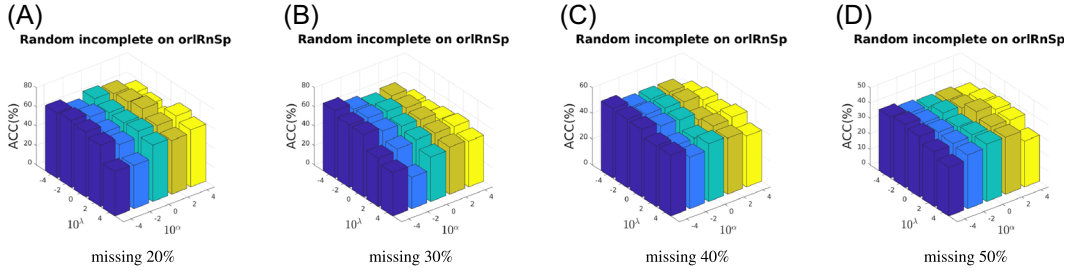


**FIGURE 7** The ACC sensitivity on data set orlRnSp under random incomplete setting. ACC, accuracy [Color figure can be viewed at wileyonlinelibrary.com]

term is here introduced to constraint sparseness in rows. The final learned graph not only considers the incomplete data features for the clustering tasks but also ensures locality among all views. In the future, we will explore to IMVC with more advanced consideration.

## ACKNOWLEDGEMENTS

This study was supported by the National Key R & D Program of China 2020AAA0107100 and the Natural Science Foundation of China (project no. 61976196).

## AUTHOR CONTRIBUTIONS

All authors contributed to the manuscript preparation of the paper. *Conceptualization*: Huiqiang Lian and Huiying Xu; *Methodology*: Huiqiang Lian and Siwei Wang; *Validation*: Miaomiao Li and Siwei Wang; *Formal analysis*: Huiqiang Lian, Huiying Xu, Xinwang Liu, and Xinzhong Zhu; *Writing–original draft preparation*: Huiqiang Lian; *Writing–review and editing*: Huiying Xu and Xinwang Liu; *Funding acquisition*: Xinwang Liu and Xinzhong Zhu. All authors reviewed and approved the final manuscipt.

## ENDNOTES

*https://github.com/hdzhao/IMG.

†https://github.com/GPMVCDummy/GPMVC/blob/master/partialMV/PVC/recreateResults/data/orlRnSp.mat.

‡http://erdos.ucd.ie/datasets/3sources.html.

## ORCID

*Huiying Xu* https://orcid.org/0000-0002-6704-0301
*Siwei Wang* https://orcid.org/0000-0001-9517-262X
*Miaomiao Li* https://orcid.org/0000-0001-7678-687X
*Xinzhong Zhu* https://orcid.org/0000-0002-0033-5260
*Xinwang Liu* https://orcid.org/0000-0001-9066-1475

## REFERENCES

1. Gao J, Han J, Liu J, Wang C. Multi-view clustering via joint nonnegative matrix factorization. In: *Proceedings of the 2013 SIAM International Conference on Data Mining*; 2013. https://epubs.siam.org/doi/10.1137/1.9781611972832.28
2. Liu G, Lin Z, Yan S, Sun J, Yu Y, Ma Y. Robust recovery of subspace structures by low-rank representation. *IEEE Trans Pattern Anal Mach Intell.* 2013;35(1):171-184.
3. Wang S, Zhu E, Hu J, et al. Efficient multiple kernel k-means clustering with late fusion. *IEEE Access*. 2019;7:61109-61120.
4. Wang S, Liu X, Zhu E, et al. Multi-view clustering via late fusion alignment maximization. In: *IJCAI*; 2019:3778-3784.
5. Liang W, Zhou S, Xiong J, et al. Multi-view spectral clustering with high-order optimal neighborhood laplacian matrix. *IEEE Trans Knowl Data Eng*. 2020;1-1. http://doi.org/10.1109/tkde.2020.3025100
6. Ou Q, Wang S, Zhou S, Li M, Guo X, Zhu E. Anchor-based multiview subspace clustering with diversity regularization. *IEEE MultiMedia*. 2020;27(4):91-101.
7. Liu X, Zhu X, Li M, et al. Late fusion incomplete multi-view clustering. *IEEE Trans Pattern Anal Mach Intell*. 2018;41(10):2410-2423.
8. Liu X, Zhu X, Li M, et al. Efficient and effective incomplete multi-view clustering. *Proc AAAI Conf Artif Intell*. 2019;33:4392-4399.

9. Liu X, Zhu X, Li M, et al. Multiple kernel *k* k-means with incomplete kernels. *IEEE Trans Pattern Anal Mach Intell*. 2019;42(5):1191-1204.
10. Liu X, Li M, Tang C, et al. Efficient and effective regularized incomplete multi-view clustering. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*; 2020.
11. Zhang P, Liu X, Xiong J, et al. Consensus one-step multi-view subspace clustering. In: *IEEE Transactions on Knowledge and Data Engineering*; 2020:1. https://doi.org/10.1109/TKDE.2020.3045770
12. Meng F, Tang J, Xu Z. Deriving priority weights from intuitionistic fuzzy multiplicative preference relations. *Int J Intell Syst*. 2019;34(11):2937-2969.
13. Fei L, Feng Y, Liu L. Evidence combination using OWA-based soft likelihood functions. *Int J Intell Syst*. 2019;34(9):2269-2290.
14. Merigó JM, Blanco-Mesa F, Gil-Lafuente AM, Yager RR. Thirty years of the International journal of intelligent systems: a bibliometric review. *Int J Intell Syst*. 2017;32(5):526-554.
15. Garg H, Rani D. Some results on information measures for complex intuitionistic fuzzy sets. *Int J Intell Syst*. 2019;34(10):2319-2363.
16. Torra V. Hesitant fuzzy sets. *Int J Intell Syst*. 2010;25(6):529-539.
17. Li SY, Jiang Y, Zhou ZH. Partial multi-view clustering. *Proc AAAI Conf Artif Intell*. 2014;28:1.
18. Zhao H, Liu H, Fu Y. Incomplete multi-modal visual data grouping. In: *IJCAI*; 2016:2392-2398.
19. Hu M, Chen S. Doubly aligned incomplete multi-view clustering. In: *Proceedings of the 27th International Joint Conference on Artificial Intelligence*; 2018:2262-2268.
20. Hu J, Zhu E, Wang S, Wang S, Liu X, Yin J. Two-stage unsupervised video anomaly detection using low-rank based unsupervised one-class learning with ridge regression. In: *2019 International Joint Conference on Neural Networks (IJCNN)*; 2019:1-8.
21. Wang S, Li M, Hu N, et al. K-means clustering with incomplete data. *IEEE Access*. 2019;7:69162-69171.
22. Zhuge W, Hou C, Liu X, Tao H, Yi D. Simultaneous representation learning and clustering for incomplete multi-view data. In: *IJCAI*; 2019:4482-4488.
23. Hu M, Chen S. One-pass incomplete multi-view clustering. *Proc AAAI Conf Artif Intell*. 2019;33:3838-3845.
24. Xu C, Guan Z, Zhao W, Wu H, Niu Y, Ling B. Adversarial incomplete multi-view clustering. In: *IJCAI*; 2019:3933-3939.
25. Zhuge W, Luo T, Tao H, Hou C, Yi D. Multi-view spectral clustering with incomplete graphs. *IEEE Access*. 2020;8:99820-99831.
26. Deng W, Liu L, Li J, Lin Y. Auto-weighted incomplete multi-view clustering. *IEEE Access*. 2020;8:138752-138762. http://doi.org/10.1109/access.2020.2997755
27. Ji M, Rao H, Li Z, Zhu J, Wang N. Partial multi-view clustering based on sparse embedding framework. *IEEE Access*. 2019;7:29332-29343.
28. Wang Q, Ding Z, Tao Z, Gao Q, Fu Y. Partial multi-view clustering via consistent GAN. In: *2018 IEEE International Conference on Data Mining (ICDM)*; 2018:1290-1295.
29. Shao W, He L, Philip SY. Multiple incomplete views clustering via weighted nonnegative matrix factorization with L2, 1 regularization. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*; 2015:318-334.
30. Gao H, Peng Y, Jian S. Incomplete multi-view clustering. In: *International Conference on Intelligent Information Processing*; 2016:245-255.
31. Zhang P, Wang S, Hu J, et al. Adaptive weighted graph fusion incomplete multi-view subspace clustering. *Sensors*. 2020;20(20):5755.
32. Ding CH, Li T, Jordan MI. Convex and semi-nonnegative matrix factorizations. *IEEE Trans Pattern Anal Mach Intell*. 2008;32(1):45-55.
33. Ren Z, Sun Q. Simultaneous global and local graph structure preserving for multiple kernel clustering. *IEEE Trans Neural Netw Learn Syst*. 2020;1-13.
34. Wen J, Fang X, Xu Y, Tian C, Fei L. Low-rank representation with adaptive graph regularization. *Neural Netw*. 2018;108:83-96.
35. Greene D, Cunningham P. Practical solutions to the problem of diagonal dominance in kernel document clustering. In: *Proceedings of the 23rd international conference on Machine learning*; 2006:377-384.
36. Wen J, Xu Y, Liu H. Incomplete multiview spectral clustering with adaptive graph learning. *IEEE Trans Cybernet*. 2020;50(4):1418-1429.

37. Elhamifar E, Vidal R. Sparse subspace clustering: algorithm, theory, and applications. *IEEE Trans Pattern Anal Mach Intell*. 2013;35(11):2765-2781.
38. Parsons L, Haque E, Liu H. Subspace clustering for high dimensional data: a review. *ACM Sigkdd Explorations Newsletter*. 2004;6(1):90-105.
39. Liu X, Wang L, Huang GB, Zhang J, Yin J. Multiple kernel extreme learning machine. *Neurocomputing*. 2015;149:253-264.
40. Liu X, Dou Y, Yin J, Wang L, Zhu E. Multiple kernel k-means clustering with matrix-induced regularization. In: *Thirtieth AAAI Conference on Artificial Intelligence*; 2016.
41. Wang H, Yang Y, Liu B. GMC: graph-based multi-view clustering. *IEEE Trans Knowl Data Eng*. 2019;32(6):1116-1129.
42. Cai X, Nie F, Huang H. Multi-view k-means clustering on big data. In: *Twenty-Third International Joint Conference on Artificial Intelligence*; 2013.
43. Nie F, Li J, Li X, et al. Parameter-free auto-weighted multiple graph learning: a framework for multiview clustering and semi-supervised classification. In: *IJCAI*; 2016:1881-1887.
44. Guo Y. Convex subspace representation learning from multi-view data. In: *Twenty-Seventh AAAI Conference on Artificial Intelligence*; 2013.
45. Greene D, Cunningham P. A matrix factorization approach for integrating multiple data views. In: *Joint European conference on machine learning and knowledge discovery in databases*; 2009:423-438.
46. Wang Y, Wu L, Lin X, Gao J. Multiview spectral clustering via structured low-rank matrix factorization. *IEEE Trans Neural Netw Learn Syst*. 2018;29(10):4833-4843.
47. Zhang Z, Liu L, Shen F, Shen HT, Shao L. Binary multi-view clustering. *IEEE Trans Pattern Anal Mach Intell*. 2018;41(7):1774-1782.
48. Chang W, Wei C, Wang YF. Multi-view nonnegative matrix factorization for clothing image characterization. In: *2014 22nd International Conference on Pattern Recognition*; 2014:1272-1277.
49. Yu S, Tranchevent L, Liu X, et al. Optimized data fusion for kernel k-means clustering. *IEEE Trans Pattern Anal Mach Intell*. 2012;34(5):1031-1039.
50. Chen J, Zhao Z, Ye J, Liu H. Nonlinear adaptive distance metric learning for clustering. In: *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; 2007:123-132.
51. Chaudhuri K, Kakade SM, Livescu K, Sridharan K. Multi-view clustering via canonical correlation analysis. In: *Proceedings of the 26th Annual International Conference on Machine Learning*; 2009:129-136.
52. Huang HC, Chuang YY, Chen CS. Multiple kernel fuzzy clustering. *IEEE Trans Fuzzy Syst*. 2012;20(1):120-134.
53. Liu X, Wang L, Yin J, Zhu E, Zhang J. An efficient approach to integrating radius information into multiple kernel learning. *IEEE Trans Cybernet*. 2013;43(2):557-569.
54. Mallah C, Cope J, Orwell J. Plant leaf classification using probabilistic integration of shape, texture and margin features. *Signal Process Pattern Recogn Appl*. 2013;5(1):45-54.
55. Zhu X, Liu X, Li M, et al. Localized incomplete multiple Kernel k-means. In: *IJCAI*; 2018:3271-3277.


**How to cite this article:** Lian H, Xu H, Wang S, Li M, Zhu X, Liu X. Partial multiview clustering with locality graph regularization. *Int J Intell Syst*. 2021;36:2991-3010. https://doi.org/10.1002/int.22409