

# Multiview Tensor Spectral Clustering via Co-regularization

Hongmin Cai<sup>1</sup>, *Senior Member, IEEE*, Yu Wang<sup>1,2</sup>, Fei Qi<sup>3,4</sup>,  
Zhuoyao Wang<sup>2</sup> and Yiu-Ming Cheung<sup>5\*</sup>, *Fellow, IEEE*



**Abstract**—Multi-view clustering seeks to partition objects by utilizing cross-views to provide a complementary description of the same object. It remains challenging to accurately cluster the samples with high-dimension features and low sample size. The key difficulty lies in ineptness of fusing the multi-view samples under various order affinities into a highly informative form which fuse cross-view and inner relationships, thus facilitating adequate assignment of the objects. To tackle this issue, we propose to achieve multi-view uniform clustering via latent representation co-regularization. Firstly, the sample affinities are encoded by both popular dyadic affinity and recent high-order affinities to comprehensively characterize spatial distributions of the samples. Secondly, a fused latent representation is learned through aligning the different view latent representation by co-regularization to effectively grasp the cross-view information. The learning of the fused representation is modeled by a high-order eigenvalue problem within manifold space to preserve the intrinsic connections and nonlinear properties of original data. A numerical scheme via manifold minimization is designed to solve the high-order eigenvalue problem efficaciously. Experiments on eight synthetic datasets and real-world datasets show the superiority of our proposed method in comparison with the existing benchmark methods.

**Index Terms**—High-order affinity, Clustering, Fusing affinity, Manifold optimization, Tensor, Spectral graph.

## 1 INTRODUCTION

CLUSTERING is one of the crucial topics in unsupervised learning [?], [?], [?], [?]. The goal of clustering is to partition unlabeled data into different subgroups. Traditional clustering methods have been extensively studied and have found wide applications in bioinformatics, computer vision, and other fields. However, in real-world scenarios, data can be obtained from different sources or multiple feature extractors [?], [?]. As multi-view data becomes increasingly popular, effectively integrating heterogeneous data from different views has become a key focus in multi-view clustering. The heterogeneity of multi-view data poses a key challenge in multi-view clustering. One essential problem is to find consensus information across views by uniformly representing data from different views [?].

Recent studies have proposed different strategy to integrate complementary information from different views, with the aim of improving clustering performance [?], [?]. A naive approach to achieving this is to directly concatenate different view data as vectors and then applying conventional single-view methods on the multi-view data vector [?]. However, this manner ignores the heterogeneity and difference of scale among different view data. Graph-based multi-view clustering methods are proposed to align the affinity graphs for the uniform representation [?]. Intuitively, although different view data are heterogeneous, using affinity matrices with same scale can eliminate this kind of difference. Therefore, the task is to jointly analyze these graphs under different views and achieve multi-view clustering by extracting consensus and complementary information from multiple graphs. Accordingly, various approaches have been proposed to refine a consensus representation from multiple graphs [?]. For example, Kumar et al. [?] proposed to refine the affinity matrix via performing co-training of spectral results. However, there is a critical prerequisite in aforementioned methods: the data relationship can be accurately described by pairwise affinity. This can be challenging in real applications, especially for high-dimension  $m$  yet low-sample-size  $n$  (HDLSS) data when  $n \gg m$  [?], [?]. The clustering performance of HDLSS data is hindered by the concentration effect, also known as the “curse of dimensionality” [?]. This effect occurs when the pairwise Euclidean distance between samples collapses to a constant value, causing the sample-wise affinity to become indiscriminate in high-dimensional feature spaces [?]. This effect poses an insurmountable roadblock for most clustering methods that rely on pairwise affinity to achieve precise clustering. Hence, conventional dyadic affinity-based methods have struggled to accurately elucidate the relationship of HDLSS data. Recent researches have proposed to utilize high-order affinity to describe the spatial distribution of multiple samples. The use of high-order affinities is expected to overcome the concentration effects on dyadic affinity and reveal essential data relations. Along this line, Mei et al. [?] proposed a hidden multi-order affinity that incorporates both second-order and third-order similarities to better capture the high-order relationships. By leveraging these measures, their models aim to grasp more nuanced and complex relationships, leading to improved performance on clustering. Also, Ghoshdasti-

<sup>1</sup>School of Future Technology, South China University of Technology, Guangzhou, 511442, China. <sup>2</sup>Peng Cheng Laboratory, Shenzhen, 518055, China. <sup>3</sup>School of Computer Science and Engineering, South China University of Technology, Guangzhou, 510006, China. <sup>4</sup>School of Data Science and Information Engineering, Guizhou Minzu University, China. <sup>5</sup>Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China.  
\*Corresponding Author, E-mail: ymc@comp.hkbu.edu.hk.

dar [?] demonstrates a relationship between the relaxation of hypergraph clustering and the multilinear singular value decomposition with consideration of order- $k$  affinities. In practice, IPS2 proved and validated that the higher-order affinities can be a complementary description of pairwise data relations to enhance clustering performance [?]. Follow this vein, our preliminary work [?] unifies different order affinities into a framework to perform clustering in single view data. This model effectively overcomes the problem of classical metric invalidation caused by concentration effect and achieves splendid results on HDLSS data. However, practical applications pose limitations on the extraction of consensus representation for multi-view data, making the aforementioned methods unsuitable.

To solve the key problems of learning consensus information and accurately revealing relationships from multi-view data, especially for HDLSS data, we propose the **Co-Regularized multi-view clustering via Manifold Alignment on Tensor Spectral embedding (CRMATS)** method. Our method presents a unified multi-view clustering framework that accurately characterizes samples using both low-order and high-order similarities across all views. By aligning the latent representations of each view, we obtain a fused representation that effectively incorporates cross-view information through introducing multiple high-order affinities as complementary information to dyadic affinity. The consensus embeddings from each view are aligned via co-regularization in the manifold space, overcoming the instability of rigid alignment within the Euclidean space. To accelerate the solution of the unified model, an efficient iterative strategy is designed to solve this model efficaciously. In this way, the final embedding aligned from different views can guide the multi-order embedding to learn more discriminative representations in each view. This improves the quality of the aligned consensus embedding for better partition as feedback. Extensive experiments on both synthetic and real datasets validate nice performance of the proposed CRMATS method, especially for HDLSS data. The framework of CRMATS is shown in Fig. 1, and the main contributions of this paper are summarized as follows:

- To capture the complex interactions among samples and the key characteristics of each view, a method is proposed to encode sample affinities using multiple order affinities that supplement the classical dyadic affinity and intrinsic connections of the samples.
- For aligning the corresponding representation of each view to yield the consensus graph, we incorporate co-regularized learning into our framework. This approach enables us to simultaneously leverage cross-view complementary information and ensure consistent locality, resulting in a more robust and accurate consensus graph.
- To effectively capture the local geometrical structure of each view and exploit the nonlinear properties, we adopt the manifold constraint to encode the local structure for learning low-dimensional representations. Additionally, we adopt a singular value decomposition-based method to solve the quadratic problem on the Stiefel manifold space that significantly speeds up the solving efficiency.

## 2 BACKGROUND ON MULTIVIEW TENSOR SPECTRAL CLUSTERING

**Notations.** In this paper, we use bold calligraphy, uppercase letters, and lowercase letters to represent tensors, matrices, and vectors, respectively. For a matrix  $\mathbf{X} \in \mathbb{R}^{n \times m}$ , the  $j$ -th column is denoted as  $\mathbf{X}_{:,j}$ , and the trace of  $\mathbf{X}$  is represented as  $\text{Tr}(\mathbf{X})$ . The square of the Frobenius norm of  $\mathbf{X}$  is denoted as  $\|\mathbf{X}\|_F^2$ . The Khatri-Rao and Kronecker products are denoted by  $*$  and  $\otimes$ , respectively. For a third-order tensor  $\mathcal{X} \in \mathbb{R}^{n \times m \times l}$ , we denote the  $v$ -th frontal slices of  $\mathcal{X}$  as  $\mathbf{X}^{(v)} \in \mathbb{R}^{n \times m}$ , and  $\mathbf{I}$  represents the identity matrix.

### 2.1 Tensor Based Approaches for Multiview Clustering

Tensor-based clustering techniques harness the high-order representation of multi-view data through tensors, thereby elucidating the intricate inter-view relationships prior to performing cluster analysis [?]. In this process, most of the early works relied on tensor decomposition techniques. For example, Yu et al. [?] stacked the original data into a tensor and applied tensor-based factorization to obtain factor matrices that capture high-order relationships. Similarly, Nie et al. [?] proposed a co-clustering method via tensor factorization to learn a low-rank approximation for discovering high-order relationships. Fast Multi-view Bilateral K-means (FMVBKM) [?] employed a low-rank constraint to explore high-order relationships. Following this approach, Guo et al. [?] utilized tensor logarithmic Schatten- $p$  norm to obtain a more compact low-rank structure, which effectively explores complementary information and accurately characterizes the high-order correlation among multi-views. Meanwhile, Ji et al. [?] combined tensor decomposition for consistent tensor and complementary tensor reconstruction. Then enhanced tensor rank constraints were imposed on the consistent tensor to explore high-order consistency. However, the clustering performance of HDLSS data is hindered by the concentration effect. The aforementioned methods require the assurance that data relationships can be accurately described by pairwise affinities, which is difficult to guarantee in HDLSS data. Due to the fact that the high-order affinity can intuitively reveal the intrinsic correlations within the data, Li et al. [?] introduced high-order affinity to maintain the correlation between views and samples. They constructed a hypergraph by unifying affinity matrices to preserve the high-order geometric structure of the data. Wang et al. [?] enhanced tensor-based multi-view clustering using high-order statistical measures of the views, which to some extent alleviates the concentration effect in HDLSS data.

The aforementioned approaches still have two limitations when dealing with complex HDLSS scenarios. First, the tensor decomposition-based techniques suffer from high computational complexity in high-dimensional data. Second, performing affinity measurement and solving in Euclidean space leads to computation collapse when dealing with non-linear features and noise in HDLSS data.

### 2.2 The Revisiting of Classical Spectral Clustering

Spectral clustering is a classical method that employs dyadic affinity to learn an optimal low-dimensional embedding

from raw data for clustering purposes. Given a data matrix  $\mathbf{X} \in \mathbb{R}^{n \times d}$ , where  $n$  is the number of samples and  $d$  is the feature dimension, the objective of spectral clustering is to divide these  $n$  samples into  $c$  subgroups by reformulating the clustering problem as a minimum cost problem of Graph-cut. The crucial step of this method is to construct the similarity graph by calculating the dyadic affinity matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$ . Specifically, the  $(i, j)$ -th element of  $\mathbf{A}$  is calculated as  $A_{ij} = d(\mathbf{x}_i, \mathbf{x}_j)$ , where  $d(\cdot, \cdot)$  is a pairwise metric. The  $i$ -th diagonal element of the degree matrix  $\mathbf{D} \in \mathbb{R}^{n \times n}$  is denoted as  $D_{ii} = \sum_{j=1}^n A_{ij}$ . The Laplacian matrix is then defined as  $\mathbf{L} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$ . The spectral clustering seeks a low-dimensional embedding by minimizing the following objective model:

$$\min_{\mathbf{V}} \text{Tr}(\mathbf{V}^\top \mathbf{L} \mathbf{V}) \quad \text{s.t. } \mathbf{V}^\top \mathbf{V} = \mathbf{I} \quad (1)$$

This model is equivalent to solving the problem of maximum partition in graph-cut [?]. It can be equivalently solved by seeking the dominant eigenvector of the Laplacian matrix  $\mathbf{L}$ , and thus degenerates to a standard eigenvalue problem. Alternatively, one can define a normalized dyadic affinity matrix  $\hat{\mathbf{L}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$ , other than the early ones to impose a strong positive definite Laplacian graph. The spectral clustering can also be popularly expressed as a maximization problem:

$$\max_{\mathbf{V}} \text{Tr}(\mathbf{V}^\top \hat{\mathbf{L}} \mathbf{V}) \quad \text{s.t. } \mathbf{V}^\top \mathbf{V} = \mathbf{I} \quad (2)$$

Then, one can perform a clustering task like k-means on the obtained embedding.

### 2.3 Tensor Spectral Clustering

The core of model (2) is to maximize the intra-cluster affinity in order to preserve the volume of each subgraph after graph cut. In our previous work [?], we propose a normalized affinity entropy measure that can assess the volume of affinity using any number of samples.

**Definition 1.** *The total normalized similarity: Let  $\mathbf{C}$  be a group of samples belonging to the dataset  $\mathbf{X}$ , and  $\mathcal{S}$  be an order- $k$  similarity tensor with elements in  $\mathbb{R}^{n_1 \times n_2 \times \dots \times n_k}$ . The total normalized similarity of the samples in  $\mathbf{C}$  is defined as follows:*

$$\text{Sim}(\mathbf{C}) = \sum_{\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_k} \in \mathbf{C}} \mathcal{L}_{\mathbf{x}_{i_1}, \mathbf{x}_{i_2}, \dots, \mathbf{x}_{i_k}} \quad (3)$$

where  $\mathcal{L}$  represents the normalized affinity tensor, which is computed using  $\mathcal{S}$ . Furthermore, given a partition of  $\mathbf{X}$  into  $\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_k$ , the normalized associativity ( $N_{\text{Assoc}}$ ) of the resulting clustering can be defined as follows:

$$N_{\text{Assoc}}(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_k) = \sum_{i=1}^k \frac{\text{Sim}(\mathbf{C}_i)}{|\mathbf{C}_i|^k} \quad (4)$$

where  $|\mathbf{C}_i|$  denotes the cardinality of cluster  $\mathbf{C}_i$ .

Tensor spectral clustering leverages multiple-order affinities to learn a low-dimensional latent representation. Let  $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_k] \in \mathbb{R}^{n \times k}$  denote the latent representation, where  $\mathbf{H}_{ik} = |\mathbf{C}_k|^{-1}$  if  $\mathbf{x}_i \in \mathbf{C}_k$  and zero otherwise. To obtain an optimal sample assignment  $\mathbf{C}_1, \dots, \mathbf{C}_k$ , we maximize the normalized associativity in Eq. (4). This

can be achieved through algebraic manipulation, which allows us to reformulate the problem as follows:

$$\max_{\mathbf{C}_1, \dots, \mathbf{C}_k} \sum_{j=1}^k (\mathcal{L} \otimes_m \mathbf{h}_j \otimes_{m-1} \mathbf{h}_j \cdots \otimes_1 \mathbf{h}_j) \quad (5)$$

where  $\mathbf{h}_j$  denotes the  $j$ -th column of  $\mathbf{H}$ , and  $\otimes_m$  is mode- $m$  product. Due to the NP-hardness of the maximum normalized associativity problem, it is often necessary to use relaxation techniques to make the problem more tractable. One such technique involves relaxing the binary assignment matrix to an orthonormal matrix  $\mathbf{V} \in \mathbb{R}^{n \times k}$ , where  $\mathbf{V}^\top \mathbf{V} = \mathbf{I}$ . This reduces the expectation of binary assignment and simplifies the problem. The simplified problem can then be formulated as follows:

$$\max_{\mathbf{V}^\top \mathbf{V} = \mathbf{I}} \sum_{j=1}^k \mathcal{L} \otimes_1 \mathbf{v}_j \otimes_2 \mathbf{v}_j \cdots \otimes_m \mathbf{v}_j \quad (6)$$

where  $\mathbf{v}_j$  represents the  $j$ -th column of  $\mathbf{V}$ .

### 2.4 Co-Regularized Multiview Clustering

In multi-view clustering, the common question is to ensure the consistency of the learned representations across views and capture the local view-specific geometric structure of each view. To overcome this issue, an approach is to automatically learn a consensus latent representation that can determine the contribution of each individual latent representation [?]. However, the quality of affinity from different views may be affected by noise or errors, which can lead to suboptimal performance. To overcome this issue, the consensus latent representation should be less similar to the corresponding deteriorating views [?]. Based on this hypothesis, co-regularized learning is applied to align the low-dimensional latent representation of different views. This approach can effectively reduce the impact of noise or errors in the individual views and improve the overall quality of the consensus representations.

Suppose there is a multi-view dataset  $\mathcal{X} \in \mathbb{R}^{n \times m \times l}$  with  $l$  views, and the corresponding  $i$ -th low-dimensional latent representation  $\mathbf{V}^{(i)} \in \mathbb{R}^{n \times k}$  for  $i = 1, \dots, l$ , as well as a consensus latent representation  $\mathbf{W} \in \mathbb{R}^{n \times k}$ . One can seek to learn a consensus sample embedding  $\mathbf{W}$  by jointly learning sample embedding  $\mathbf{V}^{(i)}$  for each view while maximizing the geodesic distance between the view-dependent affinity  $\mathbf{V}^{(i)} \mathbf{V}^{(i)\top}$  and view-independent affinities  $\mathbf{W} \mathbf{W}^\top$  for each view. The model can be mathematically formulated as:

$$\max_{\mathbf{V}^{(i)}, \mathbf{W}} \sum_{i=1}^l \text{Tr}(\mathbf{V}^{(i)\top} \hat{\mathbf{L}}_2^{(i)} \mathbf{V}^{(i)} + \lambda^{(i)} \mathbf{V}^{(i)} \mathbf{V}^{(i)\top} \mathbf{W} \mathbf{W}^\top) \quad (7)$$

$$\text{s.t. } \mathbf{V}^{(i)\top} \mathbf{V}^{(i)} = \mathbf{I}; \mathbf{W}^\top \mathbf{W} = \mathbf{I}$$

where  $\hat{\mathbf{L}}_2^{(i)} \in \mathbb{R}^{n \times n}$  is the affinity matrix of the  $i$ -th view,  $\mathbf{W}$  is the suitable latent representation to align each view,  $\lambda^{(i)}$  is used to measure the weight of view  $i$ .

## 3 MULTIVIEWED TENSOR SPECTRAL CLUSTERING VIA CO-REGULAZATION

Existing multiviewed methods are mostly based on linear spaces and are not suitable for multi-dimensional data with

intrinsic nonlinear structures [?]. Fortunately, Khan et al. [?] introduce manifold representation methods to mitigate this dilemma and achieve significant improvements in clustering performance. These methods allow for the representation of nonlinear manifold structures in high-dimensional spaces, which can be particularly useful for HDLSS data analysis. Inspired by previous work, a unified tensor clustering model can be developed by combining high-order affinities with dyadic affinities, effectively mitigating the concentration effects in HDLSS data. To further enhance the discriminative power of the low-dimensional embedding and make it more suitable for HDLSS tasks, the embedding can be constrained to lie on the Stiefel manifold. To demo the effectiveness of our framework in both odd-order and even-order higher-order affinities, we introduce third and fourth-order affinities. As a result, a low-dimensional embedding learning model that incorporates multiple high-order affinities for each view is proposed:

$$\min_{\mathbf{V} \in \mathcal{M}_k} \sum_{j=1}^k -\hat{\mathbf{L}}_2 \otimes_2 \mathbf{v}_j \otimes_1 \mathbf{v}_j - \mathcal{L}_3 \otimes_3 \mathbf{v}_j \otimes_2 \mathbf{v}_j \otimes_1 \mathbf{v}_j - \mathcal{L}_4 \otimes_4 \mathbf{v}_j \otimes_3 \mathbf{v}_j \otimes_2 \mathbf{v}_j \otimes_1 \mathbf{v}_j \quad (8)$$

where  $\mathcal{M}_k = \{\mathbf{V} \in \mathbb{R}^{n \times k} | \mathbf{V}^\top \mathbf{V} = \mathbf{I}\}$  is a Stiefel manifold,  $\hat{\mathbf{L}}_3$  and  $\hat{\mathbf{L}}_4$  are unfold form of normalized triadic and tetradic affinity tensor, calculated by:

$$\hat{\mathbf{L}}_3 = \hat{\mathbf{D}}_{3_1}^{-\frac{1}{2}} \hat{\mathbf{T}}_3 \hat{\mathbf{D}}_{3_2}^{-\frac{1}{2}}, \hat{\mathbf{L}}_4 = \hat{\mathbf{D}}_4^{-\frac{1}{2}} \hat{\mathbf{T}}_4 \hat{\mathbf{D}}_4^{-\frac{1}{2}} \quad (9)$$

where  $\hat{\mathbf{D}}_3$  and  $\hat{\mathbf{D}}_4$  are the degree matrix corresponding to  $\hat{\mathbf{T}}_3$  and  $\hat{\mathbf{T}}_4$ . The symbols  $\hat{\mathbf{T}}_3$  and  $\hat{\mathbf{T}}_4$  are the unfold form of triadic and tetradic affinity tensors. Introducing the degree matrix  $\mathbf{D}_2$  of second-order affinity  $\hat{\mathbf{L}}_2$ , We can calculate the two diagonal matrices  $\hat{\mathbf{D}}_{3_1} = \text{diag}(\sqrt{d}) \otimes \text{diag}(\sqrt{d}) = \mathbf{D}_2 \otimes \mathbf{D}_2$ ,  $\hat{\mathbf{D}}_{3_2} = \text{diag}(\sqrt{d}) * \text{diag}(\sqrt{d}) = \mathbf{D}_2 \otimes \mathbf{D}_2$ . The symbol  $d_j = \sum_{i=1}^{n^2} \hat{\mathbf{T}}_{3_{ij}} = \mathbf{D}_{2jj} \mathbf{D}_{2jj}$ ,  $d \in \mathbb{R}^{n \times 1}$ . In this paper, tensor affinities  $\mathcal{T}_3$  and  $\mathcal{T}_4$  can be defined as follows:

$$\mathcal{T}_{3_{ijk}} = 1 - \frac{(\mathbf{x}_i - \mathbf{x}_j), (\mathbf{x}_k - \mathbf{x}_j)}{d_{ij} d_{jk}} \quad (10)$$

$$\mathcal{T}_{4_{ijkl}} = \exp(-\sigma \frac{d_{ij} + d_{kl}}{d_{ik} + d_{jl} + \varepsilon}) \quad (11)$$

for  $i, j, k, l \in n$ , where  $d_{ij}$  denotes the distance between samples  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , and  $\sigma$  is an scaling constant.

### 3.1 Co-Regularized Multi-View Clustering via Manifold Alignment on Tensor Spectral Embedding

Model (8) seeks the low-dimensional embedding in each view by integrating multi-order affinities. Fortunately, co-regularized learning can be adopted to align embeddings between different views with the manifold space [?]. The ultimate goal of the proposed method is to find a consensus embedding for clustering that can effectively leverage the cross-view complementary information in HDLSS data. Therefore, by combining model of Eq. (8) and co-regularized learning component of Eq. (7), a Co-Regularized multi-view clustering via Manifold Alignment on Tensor Spectral embedding (CRMATS) method is proposed as follows:

$$\begin{aligned} \min_{\substack{\mathbf{V}_1^{(i)} \in \mathcal{M}_k, \\ \mathbf{W}, \lambda^{(i)}}} & - \sum_{i=1}^l \text{Tr}(\mathbf{V}_1^{(i)\top} \hat{\mathbf{L}}_2^{(i)} \mathbf{V}_1^{(i)} + (\mathbf{V}_1^{(i)} * \mathbf{V}_1^{(i)})^\top \hat{\mathbf{L}}_3^{(i)} \mathbf{V}_1^{(i)} \\ & + (\mathbf{V}_1^{(i)} * \mathbf{V}_1^{(i)})^\top \hat{\mathbf{L}}_4^{(i)} (\mathbf{V}_1^{(i)} * \mathbf{V}_1^{(i)}) - \lambda^{(i)} (\mathbf{V}_1^{(i)} \mathbf{V}_1^{(i)\top} \mathbf{W} \mathbf{W}^\top)) \\ \text{s.t. } & \mathbf{W}^\top \mathbf{W} = \mathbf{I} \end{aligned} \quad (12)$$

We introduce multi-order affinities to accurately describe the relationships between samples under each view. This allows us to learn an embedding that accurately reflects the true spatial relationships within the manifold space. To further improve the alignment of the embeddings between different views, we adopt the co-regularized learning approach with manifold constraints. This involves jointly learning the embeddings for all views while minimizing the disagreement between the embeddings and the consensus representation  $\mathbf{W}$  on the Stiefel manifold. This approach effectively integrates the multi-order affinities and co-regularized learning components, allowing us to leverage the complementary information across all views and reveal the underlying structure of HDLSS data. This, in turn, makes the consensus embedding more suitable for clustering. However, the computation in Eq. (12) involves high-order polynomial function that may lead to numerical difficulties. To circumvent this issue, one can introduce a slack variable  $\mathbf{V}_2^{(i)}$  to approximate the term  $\mathbf{V}_1^{(i)} * \mathbf{V}_1^{(i)}$  in Eq. (12). This approach avoids computing higher-order polynomial operations and can be extended to multi-view scenarios. Furthermore, we introduce hyperparameters  $\gamma_1, \gamma_2$  to trade off the affinities term and co-regularization term. The total model can be represented as:

$$\begin{aligned} \min_{\substack{\mathbf{V}_1^{(i)} \in \mathcal{M}_k, \\ \mathbf{V}_2^{(i)}, \mathbf{W}, \lambda^{(i)}}} & -\gamma_1 \sum_{i=1}^l \text{Tr}(\mathbf{V}_1^{(i)\top} \hat{\mathbf{L}}_2^{(i)} \mathbf{V}_1^{(i)} + \mathbf{V}_2^{(i)\top} \hat{\mathbf{L}}_3^{(i)} \mathbf{V}_1^{(i)} \\ & + \mathbf{V}_2^{(i)\top} \hat{\mathbf{L}}_4^{(i)} \mathbf{V}_2^{(i)}) - \gamma_2 \sum_{i=1}^l \lambda^{(i)} \text{Tr}(\mathbf{V}_1^{(i)} \mathbf{V}_1^{(i)\top} \mathbf{W} \mathbf{W}^\top) \\ \text{s.t. } & \mathbf{V}_1^{(i)} * \mathbf{V}_1^{(i)} = \mathbf{V}_2^{(i)}, i = 1, \dots, N; \quad \mathbf{W}^\top \mathbf{W} = \mathbf{I} \end{aligned} \quad (13)$$

Eq. (13) aims to fuse various affinity matrices to yield a uniform low-dimensional latent representation. This approach is resistant to noise and the concentration effects caused by high dimensionality. The proposed co-regularized learning method aligns the low dimensional latent representations of different views. The clustering task is then accomplished by applying k-means clustering, leading to the final group assignments.

### 3.2 Numerical Scheme to Solve CRMATS

An efficient alternating direction minimization strategy is employed to solve CRMATS. Using the Augmented Lagrange formulation methodology, the corresponding function of Eq. (13) is obtained by:

$$\begin{aligned}
& \min_{\substack{\mathbf{V}_1^{(i)} \in \mathcal{M}_k, \\ \mathbf{V}_2^{(i)}, \mathbf{W}, \lambda^{(i)}}} -\gamma_1 \text{Tr} \sum_{i=1}^l (\mathbf{V}_1^{(i)\top} \hat{\mathbf{L}}_2^{(i)} \mathbf{V}_1^{(i)} + \mathbf{V}_2^{(i)\top} \hat{\mathbf{L}}_3^{(i)} \mathbf{V}_1^{(i)} \\
& + \mathbf{V}_2^{(i)\top} \hat{\mathbf{L}}_4^{(i)} \mathbf{V}_2^{(i)}) - \gamma_2 \sum_{i=1}^l \lambda^{(i)} \text{Tr}(\mathbf{V}_1^{(i)} \mathbf{V}_1^{(i)\top} \mathbf{W} \mathbf{W}^\top) \\
& + \langle \mathbf{Z}, \mathbf{W}^\top \mathbf{W} - \mathbf{I} \rangle + \sum_{i=1}^l \frac{\mu_1^{(i)}}{2} \|\mathbf{V}_1^{(i)} * \mathbf{V}_1^{(i)} - \mathbf{V}_2^{(i)}\|_F^2 \\
& + \sum_{i=1}^l \langle \mathbf{Y}_i, \mathbf{V}_1^{(i)} * \mathbf{V}_1^{(i)} - \mathbf{V}_2^{(i)} \rangle + \frac{\mu_2}{2} \|\mathbf{W}^\top \mathbf{W} - \mathbf{I}\|_F^2
\end{aligned} \tag{14}$$

where  $\mathbf{Y}_i$  and  $\mathbf{Z}$  are Lagrange multipliers. The constant  $\mu_1^{(i)}$  and  $\mu_2 > 0$  are penalty parameters. We decompose the problem into some subproblems with respect to the following variables alternatively. Specifically, each variable is solved while fixing other variables. The process is updated iteratively until convergence.

**Step 1): Solving the subproblem with respect to the variable  $\mathbf{V}_1^{(i)}$**

When keeping the related items, Eq. (14) can be:

$$\begin{aligned}
& \min_{\mathbf{V}_1^{(i)} \in \mathcal{M}_k} \gamma_1 \sum_{i=1}^l \text{Tr}(\mathbf{V}_1^{(i)\top} (-\hat{\mathbf{L}}_2^{(i)} + \frac{\gamma_2}{\gamma_1} \lambda^{(i)} \mathbf{W} \mathbf{W}^\top)) \mathbf{V}_1^{(i)} \\
& - 2\mathbf{V}_1^{(i)\top} (\frac{1}{2} \hat{\mathbf{L}}_3^{(i)\top} \mathbf{V}_2^{(i)}) + \frac{\mu_1^{(i)}}{2} \|\mathbf{V}_1^{(i)} * \mathbf{V}_1^{(i)} - \mathbf{V}_2^{(i)} + \frac{\mathbf{Y}_i}{\mu_1^{(i)}}\|_F^2
\end{aligned} \tag{15}$$

Specifically, there are quadratic term  $-\hat{\mathbf{L}}_2^{(i)} + \frac{\gamma_2}{\gamma_1} \lambda^{(i)} \mathbf{W} \mathbf{W}^\top$  and first order term  $\frac{1}{2} \hat{\mathbf{L}}_3^{(i)\top} \mathbf{V}_2^{(i)}$  in Eq. (15). Combining with the manifold constraint, Eq. (15) can be simplified as a quadratic optimization problem in each view. In order to solve the above problem, we introduce a universal Stiefel manifold  $\mathcal{M}_p = \{\mathbf{V} \in \mathbb{R}^{n \times k} | \mathbf{V}^\top \mathbf{V} = \mathbf{I}\}$ , and the quadratic and first-order terms of each view as  $\mathbf{A}$  and  $\mathbf{B}$ .

$$\min_{\mathbf{V} \in \mathcal{M}_p} \text{Tr}(\mathbf{V}^\top \mathbf{A} \mathbf{V} - 2\mathbf{V}^\top \mathbf{B}) \tag{16}$$

where  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is a symmetric matrix. In order to solve the problem of Eq. (16), the problem can be relaxed into:

$$\max_{\mathbf{V} \in \mathcal{M}_p} \text{Tr}(\mathbf{V}^\top \hat{\mathbf{A}} \mathbf{V} + 2\mathbf{V}^\top \mathbf{B}) \tag{17}$$

where  $\hat{\mathbf{A}} = \alpha \mathbf{I} - \mathbf{A} \in \mathbb{R}^{n \times n}$ . The parameter  $\alpha$  is an arbitrary constant such that  $\hat{\mathbf{A}}$  is a positive definite matrix. A closed-form solution to problem (17) can be achieved through the corresponding derivation  $M$  [?]. Suppose that  $\mathbf{M} \leftarrow \mathbf{F} \mathbf{\Sigma} \mathbf{R}^\top$  represents the details of the singular value decomposition, in which  $\mathbf{F} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{\Sigma} \in \mathbb{R}^{n \times k}$ ,  $\mathbf{R} \in \mathbb{R}^{k \times k}$ . Thus, the problem (17) can be solved equally:

$$\max_{\mathbf{V} \in \mathcal{M}_k} \text{Tr}(\mathbf{V}^\top \mathbf{M}) \tag{18}$$

Then we have the following equations:

$$\begin{aligned}
\text{Tr}(\mathbf{V}^\top \mathbf{M}) &= \text{Tr}(\mathbf{V}^\top \mathbf{F} \mathbf{\Sigma} \mathbf{R}^\top) = \text{Tr}(\mathbf{\Sigma} \mathbf{R}^\top \mathbf{V}^\top \mathbf{F}) \\
&= \text{Tr}(\mathbf{\Sigma} \mathbf{Z}) = \sum_{i=1}^k \sigma_{ii} z_{ii}
\end{aligned} \tag{19}$$

#### Algorithm 1 Generalized power iteration method (GPI)

**Input:** The matrix  $\mathbf{A} \in \mathbb{R}^{n \times n}$ , matrix  $\mathbf{B} \in \mathbb{R}^{n \times k}$  and  $\alpha$   
**Output:** The orthonormal base  $\mathbf{V}$  of Stiefel manifold  $\mathcal{M}_p = \{\mathbf{V} \in \mathbb{R}^{n \times k} | \mathbf{V}^\top \mathbf{V} = \mathbf{I}\}$ .

- 1: Initialize a random orthonormal base of Stiefel manifold, definite matrix  $\hat{\mathbf{A}} = \alpha \mathbf{I} - \mathbf{A}$  via  $\alpha$ .
- 2: **while** Not Converged **do**
- 3:   Compute the gradient,  $\mathbf{M} \in \mathbb{R}^{n \times k} \leftarrow 2\hat{\mathbf{A}}\mathbf{V} + 2\mathbf{B}$ .
- 4:   Calculate  $\mathbf{M} = \mathbf{F} \mathbf{\Sigma} \mathbf{R}^\top$  via the compact SVD of  $\mathbf{M}$ .
- 5:   Update  $\mathbf{V} \in \mathbb{R}^{n \times k} \leftarrow \mathbf{F} \mathbf{R}^\top$ .
- 6: **end while**
- 7: **return** The orthonormal base  $\mathbf{V}$ .

where  $\mathbf{Z} = \mathbf{R}^\top \mathbf{V}^\top \mathbf{F} \in \mathbb{R}^{k \times n}$ . Apparently, we conduct that  $z_{ii} \leq 1$  since  $\mathbf{Z} \mathbf{Z}^\top = \mathbf{I}$ . In this way,  $\text{Tr}(\mathbf{V}^\top \mathbf{M})$  reaches maximum when the matrix  $\mathbf{Z} = [\mathbf{I}, \mathbf{0}] \in \mathbb{R}^{k \times n}$ . Thus, the optimal solution of  $\mathbf{V}$  can be:

$$\mathbf{V} = \mathbf{F} \mathbf{Z}^\top \mathbf{R}^\top = \mathbf{F} [\mathbf{I}; \mathbf{0}] \mathbf{R}^\top \tag{20}$$

Algorithm 1 summarizes the manifold algorithm of quadratic optimization. The convergence of Algorithm 1 will be presented in Section A of the supplementary materials. After obtaining the objective function of the subproblem, we can use Algorithm 1 to find the solution for  $\mathbf{V}^{(i)}$ . As we are imposing the orthogonal constraint of manifold, the gradient of a differentiable function  $f: \mathcal{V}_{k,m} \rightarrow \mathbb{R}$  can be given as  $\nabla_{\mathbf{X}} f = f_{\mathbf{X}} - \mathbf{X} f_{\mathbf{X}}^\top \mathbf{X}$  with  $f_{\mathbf{X}} = \frac{\partial f}{\partial \mathbf{X}}$  [?]. Thus, the gradient of  $\mathbf{V}_1^{(i)}$  in Eq. (15) is:

$$\begin{aligned}
\frac{\partial \mathbb{L}}{\partial \mathbf{V}_1^{(i)}} &= \gamma_1 (-2(\hat{\mathbf{L}}_2^{(i)} + \frac{\gamma_2}{\gamma_1} \lambda^{(i)} \mathbf{W} \mathbf{W}^\top) \mathbf{V}_1^{(i)} \\
&+ \mathbf{V}_1^{(i)} \mathbf{V}_1^{(i)\top} (\hat{\mathbf{L}}_2^{(i)} + \frac{\gamma_2}{\gamma_1} \lambda^{(i)} \mathbf{W} \mathbf{W}^\top)^\top \mathbf{V}_1^{(i)} \\
&- \hat{\mathbf{L}}_3^{(i)\top} \mathbf{V}_2^{(i)} + \mathbf{V}_1^{(i)} \mathbf{V}_2^{(i)\top} \hat{\mathbf{L}}_3^{(i)} \mathbf{V}_1^{(i)}) \\
&+ \mu_1^{(i)} (\sum_{j=1}^k (\mathbf{V}_{1,j}^{(i)} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{V}_{1,j}^{(i)})^\top \\
&(\mathbf{V}_{1,j}^{(i)} * \mathbf{V}_{1,j}^{(i)} - \mathbf{V}_{2,j}^{(i)} + \mathbf{Y}_{i,j} / \mu_1^{(i)}) \\
&+ \mathbf{V}_1^{(i)} (\sum_{j=1}^k (\mathbf{V}_{1,j}^{(i)} * \mathbf{V}_{1,j}^{(i)} - \mathbf{V}_{2,j}^{(i)} + \mathbf{Y}_{i,j} / \mu_1^{(i)})^\top \\
&(\mathbf{V}_{1,j}^{(i)} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{V}_{1,j}^{(i)}) \mathbf{V}_1^{(i)})
\end{aligned} \tag{21}$$

where  $\mathbf{V}_1^{(i)}$  can be updated iteratively through Eq. (21) and Algorithm 1.

**Step 2): Solving the subproblem with respect to the variable  $\mathbf{V}_2^{(i)}$**

Discarding the items irrelevant to  $\mathbf{V}_2^{(i)}$ , the augmented Lagrange function can be simplified as:

$$\begin{aligned}
& \min_{\mathbf{V}_2^{(i)}} -\gamma_1 \sum_{i=1}^l \text{Tr}(\mathbf{V}_2^{(i)\top} \hat{\mathbf{L}}_3^{(i)} \mathbf{V}_1^{(i)} + \mathbf{V}_2^{(i)\top} \hat{\mathbf{L}}_4^{(i)} \mathbf{V}_2^{(i)}) \\
& + \frac{\mu_1^{(i)}}{2} \|\mathbf{V}_1^{(i)} * \mathbf{V}_1^{(i)} - \mathbf{V}_2^{(i)} + \frac{\mathbf{Y}_i}{\mu_1^{(i)}}\|_F^2
\end{aligned} \tag{22}$$

The gradient of the objective function is:



$$\begin{aligned} \frac{\partial \mathbb{L}}{\partial \mathbf{V}_2^{(i)}} = & -\gamma_1 (\hat{\mathbf{L}}_3^{(i)} \mathbf{V}_1^{(i)} + 2\hat{\mathbf{L}}_4^{(i)} \mathbf{V}_2^{(i)}) \\ & + \mu_1^{(i)} (\mathbf{V}_2^{(i)} - \mathbf{V}_1^{(i)} * \mathbf{V}_1^{(i)} - \frac{\mathbf{Y}_i}{\mu_1^{(i)}}) \end{aligned} \quad (23)$$

By setting the gradient to zero,  $\hat{\mathbf{L}}_4^{(i)}$  is symmetric and its diagonal elements are not yet zero, one can obtain the implicit solution as:

$$\mathbf{V}_2^{(i)*} = (\mu_1^{(i)} \mathbf{I} - 2\gamma_1 \hat{\mathbf{L}}_4^{(i)})^{-1} (\mu_1^{(i)} \mathbf{V}_1^{(i)} * \mathbf{V}_1^{(i)} + \gamma_1 \hat{\mathbf{L}}_3^{(i)} \mathbf{V}_1^{(i)} + \mathbf{Y}_i) \quad (24)$$

**Step 3): Solving the subproblem with respect to the latent representation  $\mathbf{W}$ .**

When the terms associated with  $\mathbf{W}$  are kept, the following subproblem is obtained:

$$\max_{\mathbf{W}} \gamma_2 \text{Tr}(\mathbf{W}^\top (\sum_{i=1}^l \lambda^{(i)} \mathbf{V}_1^{(i)} \mathbf{V}_1^{(i)\top}) \mathbf{W}) \quad s.t. \mathbf{W}^\top \mathbf{W} = \mathbf{I} \quad (25)$$

where Eq. (25) is actually a standard k-means problems with a specific kernel  $\sum_{i=1}^l \lambda^{(i)} \mathbf{V}_1^{(i)} \mathbf{V}_1^{(i)\top}$ . We can obtain  $\mathbf{W}$  via eigenvalue decomposition on this specific matrix.

**Step 4): Solving the subproblem with respect to  $\lambda^{(i)}$ .**

$$\max_{\lambda^{(i)}} \gamma_2 \text{Tr}(\sum_{i=1}^l \lambda^{(i)} \mathbf{V}_1^{(i)} \mathbf{V}_1^{(i)\top} \mathbf{W} \mathbf{W}^\top) \quad (26)$$

By defining  $\text{Tr}(\mathbf{V}_1^{(i)} \mathbf{V}_1^{(i)\top} \mathbf{W} \mathbf{W}^\top) = d^{(i)}$ , and combining the Cauchy-Schwarz inequality [?], the optimal solution for  $\lambda^{(i)}$  can be obtained as:

$$\lambda^{(i)} = \frac{d^{(i)}}{\sqrt{\sum_{i=1}^l d^{(i)2}}}, \quad (27)$$

**Step 5): Updating the multipliers  $\mathbf{Y}_i$  and  $\mathbf{Z}$ , their formulations are follows:**

$$\mathbf{Y}_i^{(t+1)} = \mathbf{Y}_i^{(t)} + \mu_1^{(i)} (\mathbf{V}_1^{(i)(t)} * \mathbf{V}_1^{(i)(t)} - \mathbf{V}_2^{(i)(t)}) \quad (28)$$

$$\mathbf{Z}^{(t+1)} = \mathbf{Z}^{(t)} + \mu_2 (\mathbf{W}^\top \mathbf{W} - \mathbf{I}) \quad (29)$$

where  $t$  is current number of iterations.

The five steps are iteratively updated until convergence or until a stopping criterion is met:  $\max(\|\mathbf{V}_1^{(i)(t+1)} - \mathbf{V}_1^{(i)(t)}\|_\infty, \|\mathbf{V}_2^{(i)(t+1)} - \mathbf{V}_2^{(i)(t)}\|_\infty, \|\mathbf{V}_1^{(i)(t+1)} \otimes \mathbf{V}_1^{(i)(t+1)} - \mathbf{V}_2^{(i)(t+1)}\|_\infty, \|\mathbf{Y}_1^{(i)(t+1)} - \mathbf{Y}_1^{(i)(t)}\|_\infty, \|\mathbf{Z}^{(t+1)} - \mathbf{Z}^{(t)}\|_\infty) < \epsilon$ . To summarize the solving process described above, please refer to Algorithm 2 for detailed information of CRMATS. The convergence proof of CRMATS is elaborated in Section A of the supplementary materials.

## 4 EXPERIMENTS

In this section, we conduct a comprehensive experimental study on both synthetic and real-world datasets to show the effectiveness of the proposed CRMATS method. All of the experiments are implemented in Matlab 2020a on 64-bit Windows OS PC with an Intel 2.30-GHz CPU.

### Algorithm 2 Co-Regularized multi-view clustering via Manifold Alignment on Tensor Spectral embedding (CRMATS)

**Input:** Multi-view dataset  $\mathcal{X} \in \mathbb{R}^{n \times m \times l}$ , Cluster number  $c$ .  
**Output:** The consensus matrix  $\mathbf{W}$ .

- 1: Construct the affinity  $\hat{\mathbf{L}}_2^{(i)} \in \mathbb{R}^{n \times n}$ ,  $\hat{\mathbf{L}}_3^{(i)} \in \mathbb{R}^{n \times n^2}$ ,  $\hat{\mathbf{L}}_4^{(i)} \in \mathbb{R}^{n^2 \times n^2}$ . Set  $\mathbf{Y}_i = \mathbf{0}$ ,  $\mathbf{W} = \mathbf{0}$ ,  $\mathbf{Z} = \mathbf{0}$ ,  $\mu_1^{(i)} = \mu_2 = 10^{-1}$ ,  $\mu^{max} = 10^6$ ,  $\rho = 1.1$  and  $\epsilon = 10^{-6}$ .
- 2: Initialize  $\mathbf{V}_1^{(i)}, \mathbf{V}_2^{(i)} \in \mathcal{M}_k$ .
- 3: **while** Not Converged **do**
- 4:   **for**  $i = 1, \dots, l$  **do**
- 4:     Update  $\mathbf{V}_1^{(i)}$  via Eq. (21) as well as Algorithm 1.
- 4:     Computer  $\mathbf{V}_2^{(i)}$  by Eq. (24).
- 4:     Update  $\mathbf{Y}_i$  and  $\lambda^{(i)}$  by Eq. (28) and Eq. (27).
- 4:      $\mu_1^{(i)} = \min(\rho \mu_1^{(i)}, \mu^{max})$ .
- 5:   **end for**
- 6:    $\mu_2 = \min(\rho \mu_2, \mu^{max})$ .
- 7:   Update  $\mathbf{Z}$  and  $\mathbf{W}$  via Eq. (29) as well as Eq. (25).
- 8:   Check the convergence conditions.
- 9: **end while**
- 10: **return** Latent representation  $\mathbf{W}$ .
- 11: Perform spectral clustering on  $\mathbf{W}$  to have sample assignment  $\mathbf{Y}_{pred}$ .

## 4.1 Comparative Datasets and Methods

Eight datasets are used, including two synthesized datasets with high dimensions and low sample sizes: Syndata1 and Syndata2. Syndata1 consists of 120 samples divided into two categories, with each category containing 60 samples. Each sample is described from three views. To verify the robustness of our methods on HDLSS data, we extend the dimensions and the number of views, and reduce the sample size on Syndata1 to obtain Syndata2. Specifically, Syndata2 comprises 90 samples divided into 3 categories, with each sample being described from four views. Each subcategory of synthetic data is generated from independent and identically distributed normal distributions with mean values of 2 and standard deviations of 0.5. In addition, we evaluate the effectiveness of CRMATS in six public benchmark datasets, including Coil-20 [?], MSRC\_v1 [?], Yale [?], BBCSport [?], 3Sources [?] and Reuters [?]. To demonstrate the effectiveness of our method on HDLSS datasets, we randomly select some samples from these datasets for experiments. More details about the datasets are provided in TABLE 1, and our method is compared with other multi-view clustering

TABLE 1  
Statistics on tested datasets

Dataset	Instances	Views	Features	Classes	Types
Syndata1	120	3	2352,5097,9998	2	Synthetic
Syndata2	90	4	2341,3988,7236,16996	3	Synthetic
Coil-20	100	3	1024, 3304, 6750	5	Image
Yale	165	3	4096, 3304, 6750	15	Image
MSRC_v1	90	5	24,576,512,256,254	2	Image
BBCSport	78	2	3183, 3203	3	Text
3Sources	169	3	3560, 3631, 3068	6	Text
Reuters	120	5	21526,24892,34121,15487,11539	6	Text

TABLE 2

Comparison results (%): the mean and p value measured by different clustering methods on all the corresponding datasets and the best scores are highlighted in bold.

Dataset	Method	ACC	NMI	Purity	Fscore	Dataset	ACC	NMI	Purity	Fscore
Syndata1	SMSC	83.30(0.00)	46.91(0.00)	83.32(0.00)	73.32(0.00)	Syndata2	56.72(0.00)	30.61(0.09)	62.21(0.00)	52.21(0.00)
	RMSC	51.72(0.00)	10.91(0.00)	51.74(0.00)	49.51(0.01)		56.11(0.08)	17.91(0.00)	55.60(0.00)	43.39(0.00)
	PGSC	50.82(0.00)	0.81(0.00)	59.17(0.01)	65.90(0.00)		2.31(0.01)	2.30(0.00)	67.81(0.00)	48.61(0.00)
	LTBPL	3.33(0.00)	41.10(0.00)	53.12(0.00)	41.11(0.01)		52.19(0.02)	73.30(0.01)	54.37(0.00)	49.21(0.00)
	LSGMC	70.00(0.00)	32.47(0.00)	68.88(0.00)	53.52(0.00)		53.33(0.00)	32.28(0.00)	54.44(0.00)	49.49(0.00)
	CDMGc	51.67(0.00)	61.69(0.00)	51.67(0.00)	65.57(0.00)		40.00(0.00)	4.25(0.00)	41.12(0.00)	37.73(0.00)
	Co-reg	58.31(0.00)	62.05(0.00)	58.28(0.01)	50.60(0.00)		38.91(0.00)	22.11(0.00)	38.94(0.00)	31.17(0.00)
	CoMSC	74.44(0.00)	58.66(0.00)	74.44(0.01)	58.66(0.00)		76.67(0.00)	28.64(0.00)	76.67(0.00)	28.64(0.00)
	EMGC2F	55.10(0.00)	0.71(0.00)	55.03(0.00)	49.47(0.00)		60.01(0.00)	25.21(0.01)	60.03(0.00)	45.91(0.00)
	MVCT	51.71(0.00)	76.63(0.03)	91.72(0.04)	62.53(0.03)		35.63(0.01)	76.90(0.01)	65.61(0.00)	38.61(0.00)
	T-UMC	53.83(0.00)	57.74(0.00)	53.83(0.00)	52.51(0.01)		77.33(0.02)	61.99(0.00)	77.33(0.00)	67.19(0.00)
	WTSNM	60.00(0.00)	52.96(0.00)	60.00(0.00)	51.62(0.00)		56.67(0.00)	43.44(0.00)	56.67(0.00)	51.81(0.00)
Coil-20	HCETR	78.33(0.01)	64.93(0.00)	78.33(0.01)	65.65(0.00)	Yale	80.00(0.00)	59.34(0.00)	80.00(0.00)	69.72(0.00)
	CRMATS	<b>100.00(0.00)</b>	<b>100.00(0.00)</b>	<b>100.00(0.00)</b>	<b>100.00(0.00)</b>		<b>98.92(0.01)</b>	<b>95.53(0.05)</b>	<b>97.71(0.02)</b>	<b>98.91(0.01)</b>
	SMSC	94.01(0.00)	89.10(0.02)	94.04(0.00)	88.41(0.00)		50.91(0.00)	55.73(0.09)	52.71(0.00)	36.41(0.00)
	RMSC	48.01(0.00)	69.42(0.00)	60.03(0.00)	55.22(0.01)		73.25(0.00)	86.54(0.01)	73.91(0.00)	50.11(0.00)
	PGSC	42.00(0.00)	4.15(0.00)	86.00(0.00)	41.53(0.00)		7.91(0.00)	10.21(0.00)	61.51(0.00)	12.11(0.00)
	LTBPL	68.51(0.21)	67.53(0.00)	78.51(0.00)	73.93(0.00)		60.04(0.00)	62.92(0.00)	60.02(0.00)	36.41(0.00)
	LSGMC	84.00(0.00)	75.38(0.00)	84.12(0.00)	74.86(0.00)		64.85(0.00)	68.16(0.00)	62.23(0.00)	45.84(0.00)
	CDMGc	80.00(0.00)	77.52(0.00)	80.00(0.00)	75.41(0.00)		66.67(0.00)	69.46(0.00)	67.27(0.00)	48.63(0.00)
	Co-reg	87.10(0.00)	87.71(0.02)	87.02(0.00)	80.72(0.00)		67.91(0.00)	72.13(0.00)	67.93(0.00)	53.92(0.00)
	CoMSC	36.33(0.00)	33.54(0.00)	44.33(0.00)	36.67(0.00)		76.07(0.00)	48.65(0.00)	76.07(0.00)	56.17(0.00)
	EMGC2F	24.21(0.00)	56.73(0.00)	31.03(0.00)	31.62(0.00)		66.69(0.00)	70.81(0.00)	67.32(0.00)	50.21(0.00)
	MVCT	25.03(0.00)	14.01(0.00)	26.01(0.00)	30.43(0.00)		54.53(0.00)	59.52(0.00)	65.41(0.01)	35.71(0.00)
MSRC_v1	T-UMC	65.51(0.00)	71.09(0.00)	65.51(0.00)	70.03(0.00)		74.62(0.01)	75.89(0.00)	74.67(0.00)	64.84(0.00)
	WTSNM	43.51(0.00)	37.89(0.00)	50.00(0.00)	39.72(0.00)	BBCSport	81.52(0.00)	85.82(0.00)	82.73(0.00)	80.88(0.00)
	HCETR	70.00(0.00)	57.06(0.00)	70.00(0.00)	59.05(0.00)		59.39(0.00)	68.01(0.00)	63.03(0.00)	51.27(0.00)
	CRMATS	<b>97.02(0.13)</b>	<b>96.65(0.11)</b>	<b>98.64(0.03)</b>	<b>97.13(0.21)</b>		<b>91.12(0.01)</b>	<b>88.12(0.01)</b>	<b>91.12(0.01)</b>	<b>83.33(0.00)</b>
	SMSC	78.90(0.00)	47.73(0.02)	78.90(0.00)	66.11(0.00)		66.67(0.00)	56.51(0.15)	66.67(0.00)	66.22(0.00)
	RMSC	54.41(0.00)	55.43(0.00)	66.70(0.00)	62.93(0.00)		88.51(0.00)	69.41(0.00)	88.54(0.01)	79.10(0.00)
	PGSC	33.44(0.00)	2.31(0.00)	67.82(0.00)	48.56(0.00)		64.64(0.00)	2.60(0.00)	67.44(0.00)	48.36(0.00)
	LTBPL	67.81(0.00)	64.83(0.06)	67.79(0.00)	71.61(0.07)		74.34(0.00)	52.71(0.00)	74.44(0.00)	60.62(0.00)
	LSGMC	95.55(0.00)	83.59(0.00)	92.22(0.00)	91.10(0.00)		35.90(0.00)	4.74(0.00)	36.96(0.00)	11.76(0.00)
	CDMGc	71.11(0.00)	50.88(0.00)	71.11(0.00)	66.73(0.00)		51.28(0.00)	23.60(0.00)	53.85(0.00)	50.88(0.00)
	Co-reg	94.43(0.00)	81.21(0.02)	94.39(0.00)	89.21(0.01)		73.32(0.00)	57.32(0.00)	74.44(0.00)	66.51(0.00)
	CoMSC	44.81(0.00)	9.25(0.00)	45.19(0.00)	10.26(0.00)		58.55(0.00)	20.76(0.00)	59.83(0.00)	21.20(0.00)
	EMGC2F	<b>99.44(0.05)</b>	<b>95.91(0.06)</b>	<b>99.43(0.03)</b>	<b>96.89(0.07)</b>		91.01(0.00)	81.81(0.00)	91.32(0.00)	84.23(0.00)
3Sources	MVCT	40.00(0.00)	7.77(0.00)	56.73(0.00)	39.29(0.00)		51.31(0.00)	29.81(0.00)	79.54(0.01)	50.81(0.00)
	T-UMC	78.11(0.00)	67.73(0.00)	78.11(0.00)	68.16(0.00)	Reuters	94.62(0.00)	82.81(0.00)	94.62(0.01)	90.13(0.00)
	WTSNM	79.05(0.00)	66.31(0.00)	79.05(0.00)	64.07(0.00)		74.36(0.00)	68.99(0.00)	74.36(0.00)	60.64(0.00)
	HCETR	84.29(0.00)	74.01(0.00)	84.29(0.00)	72.61(0.00)		82.05(0.00)	57.72(0.00)	82.05(0.00)	70.20(0.00)
	CRMATS	98.13(0.02)	93.90(0.01)	96.28(0.01)	95.12(0.02)		<b>100.00(0.00)</b>	<b>100.00(0.00)</b>	<b>100.00(0.00)</b>	<b>100.00(0.00)</b>
	SMSC	26.04(0.00)	7.88(0.15)	43.21(0.00)	21.77(0.00)		53.50(0.00)	43.76(0.00)	53.50(0.00)	50.83(0.00)
	RMSC	47.34(0.00)	38.65(0.00)	61.54(0.00)	40.17(0.00)		51.50(0.00)	51.06(0.00)	51.50(0.00)	64.88(0.00)
	PGSC	34.91(0.00)	3.13(0.00)	77.44(0.00)	37.46(0.00)		51.05(0.00)	50.15(0.00)	71.05(0.00)	66.23(0.00)
	LTBPL	63.31(0.00)	53.46(0.00)	68.86(0.00)	55.12(0.00)		51.00(0.00)	55.17(0.00)	51.00(0.00)	66.06(0.00)
	LSGMC	75.10(0.00)	65.17(0.00)	76.16(0.00)	69.96(0.00)		66.73(0.01)	61.09(0.00)	66.73(0.01)	74.12(0.00)
	CDMGc	69.11(0.00)	56.17(0.00)	68.26(0.00)	60.95(0.00)		64.63(0.00)	59.93(0.00)	64.63(0.00)	77.12(0.00)
	Co-reg	43.20(0.00)	43.52(0.00)	64.54(0.00)	38.64(0.00)		52.00(0.00)	58.49(0.00)	52.00(0.00)	56.48(0.00)
	CoMSC	76.10(0.00)	60.17(0.00)	79.83(0.00)	62.20(0.00)		71.11(0.00)	65.66(0.00)	71.11(0.00)	77.13(0.00)
	EMGC2F	63.31(0.00)	63.12(0.00)	76.92(0.00)	59.35(0.00)		70.31(0.01)	65.99(0.00)	70.31(0.01)	73.42(0.00)
3Sources	MVCT	60.36(0.00)	54.28(0.00)	83.43(0.00)	56.63(0.00)		51.00(0.00)	61.73(0.00)	51.15(0.00)	69.93(0.00)
	T-UMC	76.91(0.00)	71.28(0.00)	83.61(0.00)	72.35(0.00)		75.67(0.00)	64.35(0.00)	75.67(0.00)	66.73(0.00)
	WTSNM	68.46(0.00)	57.12(0.00)	63.85(0.00)	70.75(0.00)		60.61(0.00)	61.36(0.00)	62.17(0.00)	75.35(0.00)
	HCETR	66.86(0.00)	48.50(0.00)	66.86(0.00)	50.50(0.00)		85.00(0.00)	64.22(0.00)	85.00(0.00)	74.75(0.00)
	CRMATS	<b>95.19(0.00)</b>	<b>87.46(0.00)</b>	<b>93.95(0.00)</b>	<b>89.64(0.00)</b>		<b>87.68(0.00)</b>	<b>77.38(0.00)</b>	<b>87.68(0.00)</b>	<b>79.95(0.00)</b>

methods. The details of the comparative methods are shown as follows:

- I. **Scalable Multi-view Subspace Clustering (SMSC)** [?] constructs latent graph after anchor learning.
- II. **Pure graph-guided multi-view subspace clustering (PGSC)** [?] learns consensus graph by leveraging the sparsity and connectivity of each affinity graph.
- III. **Robust Multi-View Spectral Clustering (RMSC)** [?] considers low rankness and sparsity of matrix

to learn a common graph after decomposition.

- IV. **Low-rank Tensor Based Proximity Learning (LTBPL)** [?] performs probability affinity to recover the low rankness and high-order correlations.
- V. **Multiview Subspace Clustering via Low-Rank Symmetric Affinity Graph (LSGMC)** [?] pursues a consistent low-rank structure across views.
- VI. **Measuring Diversity in Graph Learning: A Unified Framework for Structured Multi-View Clustering**

TABLE 3

CHI results: the mean and p value measured by different clustering methods on benchmark datasets and the best scores are highlighted in bold.

Method	Dataset	CHI	Dataset	CHI	Dataset	CHI	Dataset	CHI
SMSC	Syndata1	1.38(0.00)	Syndata2	1.61(0.00)	Coil	335.85(0.00)	Yale	36.86(0.00)
RMSC		1.38(0.00)		1.13(0.00)		134.91(0.00)		48.81(0.00)
PGSC		0.86(0.00)		1.12(0.00)		108.65(0.00)		13.87(0.00)
LTBPL		1.34(0.00)		0.41(0.00)		224.46(0.00)		44.74(0.00)
LSGMC		1.35(0.00)		1.41(0.00)		292.13(0.00)		44.31(0.00)
CGMGC		1.22(0.00)		1.13(0.00)		274.65(0.00)		45.29(0.00)
Co-reg		1.21(0.00)		1.23(0.00)		305.67(0.00)		45.95(0.00)
CoMSC		1.67(0.00)		1.47(0.00)		83.89(0.00)		50.31(0.00)
EMGC2F		1.41(0.00)		1.18(0.00)		30.84(0.00)		45.30(0.00)
MVCT		1.17(0.00)		1.13(0.00)		34.52(0.00)		38.80(0.00)
T-UMC		1.68(0.00)		1.24(0.00)		211.35(0.00)		49.54(0.00)
WTSNM		1.38(0.00)		1.26(0.00)		128.18(0.00)		53.23(0.00)
HCETR		1.74(0.00)		1.53(0.00)		229.77(0.00)		41.40(0.00)
CRMATS		<b>2.13(0.00)</b>		<b>1.86(0.00)</b>		<b>349.00(0.00)</b>		<b>58.37(0.00)</b>
Truth		2.13		1.90		362.02		63.12
SMSC	MSRC	12.29(0.00)	BBCSport	56.47(0.00)	3Scources	8.82(0.00)	Reuters	4.27(0.00)
RMSC		11.57(0.00)		104.62(0.00)		9.04(0.00)		4.28(0.00)
PGSC		10.94(0.00)		51.99(0.00)		8.91(0.00)		4.27(0.00)
LTBPL		11.96(0.00)		73.38(0.00)		9.20(0.00)		4.27(0.00)
LSGMC		12.79(0.00)		11.39(0.00)		9.32(0.00)		4.32(0.00)
CGMGC		12.06(0.00)		22.53(0.00)		9.26(0.00)		4.31(0.00)
Co-reg		12.76(0.00)		71.13(0.00)		8.99(0.00)		4.27(0.00)
CoMSC		11.28(0.00)		38.56(0.00)		9.33(0.00)		4.34(0.00)
EMGC2F		<b>12.90(0.00)</b>		110.14(0.00)		9.20(0.00)		4.33(0.00)
MVCT		11.13(0.00)		22.59(0.00)		9.17(0.00)		4.27(0.00)
T-UMC		12.27(0.00)		118.10(0.00)		9.33(0.00)		4.35(0.00)
WTSNM		12.30(0.00)		73.42(0.00)		9.25(0.00)		4.30(0.00)
HCETR		12.45(0.00)		90.38(0.00)		9.25(0.00)		4.38(0.00)
CRMATS		12.87(0.00)		<b>129.97(0.00)</b>		<b>9.52(0.00)</b>		<b>4.40(0.00)</b>
Truth		12.93		129.97		9.58		4.64

(CDMGC) [?] leverages the multi-view consistency and the diversity in a unified framework.

- VII. **Co-regularized kernel k-means for multi-view clustering (Co-reg)** [?] combines similarities of different view and latent representation for clustering.
- VIII. **Multiview Clustering via Co-Training Robust Representation (CoMSC)** [?] finds a consensus matrix and complementary information.
- IX. **Efficient Multi-view Graph Clustering (EMGC)** [?] finds a consistent cluster indicator matrix with a Super Nodes Similarity Minimization module.
- X. **Multi-View Clustering on Topological Manifold (MVCT)** [?] integrates multiple affinity graphs into a consensus one with the topological relevance.
- XI. **A Tensor Approach for Uncoupled Multiview Clustering (T-UMC)** [?] couples the representation matrix to explore high-order relationship.
- XII. **Multiview Subspace Clustering by an Enhanced Tensor Nuclear Norm (WTSNM)** [?] studies the Schatten  $p$ -norm to solve the minimization problem.
- XIII. **High-order Complementarity Multi-View Clustering with Enhanced Tensor Rank (HCETR)** [?] adopts Tensor Rank to find high-order consistency.

## 4.2 Clustering Performance

We evaluate the clustering performance using several commonly used metrics, including **Accuracy (ACC)**, **Normalized mutual information (NMI)**, Purity, and Fscore. A larger value indicates better performance, and the best results are

indicated in bold [?]. Considering that the clustering problem does not include the number of groups present in the data, we further use the Calinski Harabasz index (CHI) to evaluate the similarity of the created groups. CHI measures the compactness and separation of clustering results by calculating the ratio of inter class variance to intra class variance [?]. A higher CHI value indicates better clustering results. These five metrics provide a comprehensive evaluation as each measure represents a specific property of the clustering results. TABLE 2 and TABLE 3 present the clustering performance of our method and several compared methods on eight benchmark datasets. Furthermore, we provide the CHI results between the ground truth labels and data features for each dataset. To ensure the robustness, each algorithm is repeated 20 times to obtain the mean value. We then use Student's t-test to test the statistical significance of the results, with the p-value represented in parentheses.

### 4.2.1 Experiments on Synthetic Datasets

To evaluate the stability of our method on Syndata1 and Syndata2, we conducted experiments, and the results are summarized in TABLE 2. Notably, our method demonstrates substantial improvements in terms of NMI compared to other methods on Syndata1. Specifically, we observe improvements of 37.95%, 41.34%, 58.90%, 67.53%, 23.37%, 42.26%, 47.04% and 21.67% when compared to Co-reg, CoMSC, LTBPL, LSGMC, MVCT, T-UMC, WTSNM, and HCETR, respectively. These results highlight the effectiveness of the co-regularization term in Eq. (13) in integrating cross-perspective information and leveraging complemen-



tary information from multiple views on Syndata1. Regarding Syndata2, our method outperforms Co-reg, CoMSC, LTBPL, LSGMC, MVCT, T-UMC, WTSNM, and HCETR by 73.42%, 66.89%, 25.83%, 67.25%, 18.63%, 33.54%, 52.09% and 36.19%, respectively. This indicates that fusing different order affinities is more effective in fully capturing the data structure than traditional pairwise affinity, low-rankness and tensor based method. Our method integrates multi-order affinity and achieves correct clustering and partitioning of all samples in the dataset. These experimental results demonstrate the robustness of our method on synthetic data, and CRMATS achieves the best score among all the compared methods.

In order to assess the discriminative power of the consensus representation, we employ  $t$ -SNE to visualize the differences by projecting the latent representation onto a two-dimensional space. We specifically select Syndata2, which possesses the highest dimensionality, to represent the synthetic scenarios. When visualizing the raw data from their respective views using  $t$ -SNE (Fig. 2 (a)-(d)), it is evident that most of the samples in Syndata2 appear intermingled and lack clear separation. In contrast, the consensus representation achieved by our method successfully separates subcategories without any overlap (Fig. 2 (e)). This outcome demonstrates that our method effectively learns a consensus representation that captures the underlying structure of Syndata2 and enables clear discrimination of the original data. To validate the effectiveness of CRMATS, we analyze the heatmaps of similarities to assess the differences between groups. In Fig. 2 (f)-(i), the affinity heatmap of the raw samples lacks clear boundaries and block structures. However, Fig. 2 (j) shows that the affinity obtained from the low-dimensional embedding after applying CRMATS exhibits distinct boundaries, indicating the method’s ability to mitigate potential biases. These results demonstrate that high-order affinities effectively compensate for the limitations of pairwise affinity, particularly in high-dimensional feature scenarios. By integrating multi-order affinities and co-regularization techniques, our method accurately captures the underlying structure of high-dimensional data, leading to improved clustering performance on Syndata2. In addition, CRMATS encompasses the advantages of conventional affinity methods (e.g., Co-reg, CoMSC), low-rankness methods (e.g., LTBPL, LSGMC), manifold methods (e.g., MVCT), and tensor-based methods (e.g., T-UMC, WTSNM, HCETR). By integrating these diverse techniques, CRMATS surpasses all other state-of-the-art methods in terms of performance on synthetic data, as demonstrated by the results presented in TABLE 2 and the visualizations.

#### 4.2.2 Experiments on Real Datasets

To further validate the efficacy of CRMATS in real-world scenarios, we evaluate its performance on several benchmark datasets, including Coil-20, Yale, MSRC\_v1, BBCSport, 3Sources, and Reuters. Detailed information and corresponding results can be found in TABLE 1, 2 and 3, respectively. On Coil-20, CRMATS demonstrates superior performance compared to LTBPL, LSGMC, MVCT, CoMSC, Co-reg, T-UMC, WTSNM, and HCETR, with NMI clustering results exceeding them by 29.12%, 21.27%, 82.64%, 63.11%, 11.94%, 25.56%, 58.76%, and 39.59%, respectively. Likewise,

on Yale, CRMATS outperforms Co-reg, CoMSC, LTBPL, LSGMC, MVCT, T-UMC, WTSNM, and HCETR with performance improvements of 15.99%, 39.47%, 25.20%, 19.96%, 28.60%, 12.23%, 2.30%, and 20.11%, respectively. Comparing against alternative methods such as SMSC, RMSC, PGSC, CDMGC, and EMGC2F, CRMATS achieves significantly higher performance, surpassing them by 32.39%, 1.58%, 77.91%, 18.66%, and 17.31%, respectively. Furthermore, CRMATS exhibits superior performance on BBCSport, MSRC\_v1, 3Sources, and Reuters. Notably, the NMI result of CRMATS on MSRC\_v1 reaches 93.90%, making it the second-best performer. Additionally, CRMATS achieves a perfect NMI result of 100.00% on BBCSport, surpassing the second-best performance of 81.00% achieved by EMGC2F. In the case of the 3Sources dataset, CRMATS achieves an NMI result of 87.46%, which is over 22.29% higher than the second-best method, LSGMC. Similarly, on the Reuters dataset, CRMATS achieves an NMI result of 77.38%, surpassing the second-best method, LSGMC, by 11.39%. TABLE 3 presents the CHI results of CRMATS and the comparison methods in the dataset. In Syndata1 and BBCSport, CRMATS achieved results identical to the ground truth labels, and it also demonstrated outstanding performance in the remaining comparison datasets, which aligns with the trend observed in TABLE 2.

To visually demonstrate the effectiveness of our method on Coil-20, we present the spatial distributions of the raw data and consensus representation obtained using  $t$ -SNE in Fig. 3. Additionally, Fig. 4 showcases the corresponding heatmaps and consensus representations learned by our method. These visualizations provide compelling evidence supporting the superiority of our approach. In the pairwise affinity-based visualization (Fig. 3 (a)-(c)), most samples appear mixed together, making it challenging to accurately distinguish between different subgroups. However, our method generates a consensus representation that exhibits improved separation and reduced overlap, leading to enhanced clustering performance. The effectiveness of our method is demonstrated through the comparison of affinity heatmaps. In Coil-20, the affinity heatmaps of each view on the raw samples (Fig. 4 (a)-(c)) exhibit blurred boundaries and lack a clear block structure. However, after applying our method and generating the affinity matrix from the consensus representation, the boundaries become well-defined. This is supported by the  $t$ -SNE visualization of the consensus representation, where most samples are clustered with their corresponding partners, and distinct blocks are observed on the diagonal of the heatmap (Fig. 3 (d) and Fig. 4 (d)). Overall, CRMATS yields superior clustering performance on all real-world datasets, as shown in TABLE 2 and 3. Our method not only performs well on synthetic datasets but also has excellent performance on real datasets. Furthermore, we conducted an analysis on the randomness of views and resource consumption in CRMATS, which can be found in Section B and Section C of the supplementary materials.

Based on the experimental results presented above, CRMATS demonstrates several advantages in HDLSS data. First, our method effectively utilizes the complementarity of high and low-order affinities to comprehensively capture the spatial structure of HDLSS data. Second, our co-

regularization based method avoids producing suboptimal clustering results by aligning different latent representations to seek the consensus graph and supplement cross-view complementary information. Finally, by learning a low-dimensional latent representation on the manifold space, we consider the intrinsic connections between different samples and the nonlinear properties of the original data, leading to a decisive effect on the clustering performance.

### 4.3 Convergence Analysis

An alternate minimizing algorithm is developed to solve the optimization problem. The proposed method CRMATS can be guaranteed to convergence eventually since the objective function Eq. (13) is quadratic differentiable with the continuous iterations. In this subsection, we compare objective value of the benchmark datasets with diverse backgrounds to eliminate randomness and then illustrate clearly in Fig. 5. We show the objective value within 50 epochs.

The corresponding objective value of each benchmark dataset decreases sharply within the first 5 iterations and then stays steady with more iterations, implying that CRMATS converges steadily after just a few iterations. Combining the experimental performance and algorithm analysis, it has been significantly demonstrated that the proposed method considers efficient performance, especially in dealing with the HDLSS data. The fast convergence of CRMATS is due to the alternate minimizing algorithm, which updates each variable separately and guarantees the convergence of the optimization problem. Moreover, the incorporation of co-regularization and high-order affinity effectively captures the underlying structure of the data, leading to improved clustering performance in HDLSS scenarios. Therefore, the proposed method not only achieves superior clustering performance but also has a fast convergence speed.

### 4.4 Hyperparameter Sensitivity Analysis

We investigate the influence of hyperparameter sets on the clustering performance in terms of NMI, which are represented in Fig. 6. Notice that Eq. (13) consists of two parts, the affinities term and co-regularization term. We set hyperparameters for each part, denoted as  $\gamma_1$  and  $\gamma_2$ , respectively. We use a combination of hyperparameters [1e-6, 1e-5, 1e-4, 1e-3, 0.01, 0.1, 1, 10, 100] for both  $\gamma_1$  and  $\gamma_2$ , and evaluate the performance on four benchmark datasets. According to Fig. 6, which shows the clustering performance in terms of NMI, the clustering performance is influenced by  $\gamma_1$  in some cases. For example, the NMI of Syndata1 and Yale (Fig. 6 (a) and (c)) decreases with lower  $\gamma_1$ . In addition, the NMI changes little while  $\gamma_2$  increases. Although the performance of CRMATS changes with different combinations of  $\gamma_1$  and  $\gamma_2$ , as shown in Fig. 6 (b) and (d), CRMATS still outperforms its comparative methods on some datasets, demonstrating the stability of the proposed model.

Furthermore, the results in Fig. 6 demonstrate that CRMATS is relatively robust to the choice of hyperparameters, as it consistently outperforms the comparative methods on some datasets across different settings. This indicates that our method is not sensitive to the choice of hyperparameters and can achieve good performance with a wide

range of hyperparameter values. Overall, the results in Fig. 6 demonstrate the effectiveness and robustness of our CRMATS in multi-view clustering tasks, as well as its ability to handle different hyperparameter settings. This highlights the practicality and versatility of our proposed method. By considering the complementarity of high and low-order affinities and incorporating co-regularization, our proposed method can effectively capture the underlying structure of HDLSS data and achieve superior clustering performance.

### 4.5 Ablation Analysis

In this subsection, we conduct an ablation study of our proposed CRMATS method to investigate the roles of different orders of affinities and their combinations. To investigate the importance of multi-order affinities, we perform the ablation study as follows. First, we discard  $\hat{L}_3$  and  $\hat{L}_4$ , and the resulting method deteriorates into the traditional Co-reg method. This method is denoted as CRMATS-L2. Next, we apply each high-order affinity separately in the singular situation. We denote the resulting methods as CRMATS-L3 and CRMATS-L4, respectively. In the multiple affinities situation, we combine  $\hat{L}_2$ ,  $\hat{L}_3$ , and  $\hat{L}_4$  in pairs. For instance, CRMATS-L23 is the combination of  $\hat{L}_2$  and  $\hat{L}_3$ . We evaluate the clustering performance of each ablation method using the same benchmark datasets and metrics as in the previous experiments.

The experimental results presented in Fig. 7 demonstrate the importance and effectiveness of incorporating multi-order affinities in multi-view clustering tasks. First, our proposed CRMATS method achieves the best performance on all five benchmark datasets, indicating its effectiveness in capturing the underlying structure of the data in multi-view clustering tasks. Second, singular affinity experiments demonstrate that high-order affinities (CRMATS-L3, CRMATS-L4) supplement the inherent information of traditional second-order affinity. For example, the performance of CRMATS-L3 is better than CRMATS-L2 on Syndata1 and BBC. Finally, the fusion of second, third, and fourth order affinities exhibits superior performance in comparison to any two-order fusion (e.g., CRMATS-L23, CRMATS-L24, CRMATS-L34). This observation suggests that the inclusion of each order of fusion contributes to the enhancement of internal information, resulting in incremental improvements in the overall results. Overall, the experimental results confirm that high-order affinities and the combination strategies ( $\hat{L}_2$ ,  $\hat{L}_3$ , and  $\hat{L}_4$ ) are indispensable to ensure the effectiveness of our method in multi-view clustering tasks.

## 5 CONCLUSION

In this paper, we introduce CRMATS, a multi-view clustering approach aimed at mitigating concentration effects in HDLSS data. By jointly modeling affinities of different orders and incorporating co-regularization learning, our method can effectively capture the underlying structure of multi-view HDLSS data. The incorporation of pairwise matrix affinity, triadic tensor affinities, and tetradic tensor affinities enables the integration of diverse affinity information types, facilitating a more comprehensive representation of the intricate relationships among samples. The

co-regularization learning helps align the low-dimensional representations and extract the consensus graph, leading to improved clustering performance and robustness. The inclusion of a manifold constraint is a vital aspect of CRMATS, as it enables the accurate capturing of the local geometric structure within each view and leverages the nonlinear characteristics inherent in the original data. By encoding the local structure of the intrinsic subspace, we ensure that the consensus representations are meaningful and aligned with the underlying structure of the HDLSS data. We also use alternating minimization strategy and singular value decomposition to improve the effectiveness and efficiency of CRMATS. The alternate minimizing strategy improves the convergence rate and stability of the algorithm, while the singular value decomposition method speeds up the solving efficiency and allows for more efficient implementation of the algorithm. Experimental results on synthetic and real-world datasets have demonstrated the superiority and effectiveness of our proposed method compared to other popular approaches. Overall, our proposed method has provided a robust and effective framework for multi-view clustering of HDLSS data, which can be used to extract meaningful information and insights from complex and heterogeneous datasets. This highlights the potential applications of our method in various domains such as medical imaging, bioinformatics, and social network analysis.

Although our proposed method effectively addresses the concentration effect in high-dimensional data clustering and outperforms several baseline methods on the benchmark dataset, there are still potential directions for improvement. First, our strategy of improving the HDLSS clustering results through high-order affinity requires more time and memory costs for computation. Additionally, existing high-order affinity methods have limitations when dealing with complex graph data. To address these limitations, we consider incorporating deep graph neural network for high-dimensional learning. The deep neural network can leverage GPU computing units to reduce memory costs within each mini-batch. Furthermore, the graph neural network enables us to extract more potential high-order information from high-dimensional data.

## 6 ACKNOWLEDGMENTS

This work was supported in part by the National Key Research and Development Program of China (2022YFE0112200), the Key-Area Research and Development of Guangdong Province (2022A0505050014, 2022B1111050002), the Key-Area Research and Development Program of Guangzhou City (202206030009, 2023B01J0002), the National Natural Science Foundation of China (U21A20520, 62172112), Guangdong Key Laboratory of Human Digital Twin Technology (2022B1212010004).

## REFERENCES

fig/HM1.png

fig/wangyu.png

fig/QiFei.jpg

fig/zhuoyaowang.jpg

fig/CHEUNG-Yiumin

**Hongmin Cai** (Senior Member, IEEE) received the BS and MS degrees in mathematics from the Harbin Institute of Technology, Harbin, China, in 2001 and 2003, respectively, and the PhD degree in applied mathematics from Hong Kong University, in 2007. He is a professor with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. From 2005 to 2006, he was a research assistant with the Center of Bioinformatics, Harvard University, and the Section for Biomedical Image Analysis, University of Pennsylvania. His current research interests include bioinformatics, machine learning, and medical image analysis.

**Yu Wang** is currently pursuing the PH.D. degree in School of Future Technology, South China University of Technology, Guangzhou, Guangdong, China. He received the B.S and M.S degrees in the Hunan University of Technology and Business, Changsha, Hunan, China, in 2019 and 2022, respectively. His research interests include machine learning and image processing.

**Fei Qi** is currently pursuing the PH.D. degree in Computer Science and Engineering from South China University of Technology, Guangzhou, Guangdong, China. He received the B.S and M.S degrees in the Xiamen University, Xiamen, Fujian, China, in 2013 and 2016, respectively. His research interests include machine learning and image processing.

**Zhuoyao Wang** received the Ph.D. degree in electrical and computer engineering with a minor in mathematics from the University of New Mexico, Albuquerque, NM, USA, in 2016. He is currently a research scientist with Peng Cheng Laboratory, Shenzhen, China. He also spent several years working in industry at large Blue Chip organizations (Tencent, ZTE) and hi-tech startups. His research interests include risk analysis of cyber-physical systems, cloud computing and industrial internet. He is currently a co-chair of the CCSA TC13 WG4.

**Yiu-Ming Cheung** (Fellow, IEEE) received the PhD degree from the Department of Computer Science and Engineering, Chinese University of Hong Kong, Hong Kong. He is currently a chair professor with the Department of Computer Science, Hong Kong Baptist University, Hong Kong. His current research interests include machine learning, pattern recognition, and visual computing. He is the founding chairman of the Computational Intelligence Chapter of the IEEE Hong Kong Section. He is the Editor-in-Chief of IEEE Transactions on Emerging Topics in Computational Intelligence. Also, he serves as an associate editor for IEEE Transactions on Cybernetics, IEEE Transactions on Cognitive and Developmental Systems, IEEE Transactions on Neural Networks and Learning Systems (2014-2020), Pattern Recognition, Knowledge and Information Systems, and Neurocomputing, as well as the Guest Editor in several international journals. He is a fellow of the IEEE, AAAS, IET and BCS.

fig/framework-r6.pdf

**Fig. 1.** Description of the CRMATS method. The CRMATS method is adept at accurately learning latent representations from multiple views in a manifold space. Specifically, given a multi-view dataset  $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(N)}$ , the corresponding pairwise affinities, third-order and fourth-order affinities are computed. Next, we use the manifold constraint to encode the local structure of the intrinsic subspace and obtain low-dimensional orthogonal representations  $\mathbf{V}_1^{(1)}, \dots, \mathbf{V}_1^{(N)}$ . To align the representations of each view, we incorporate co-regularized learning, resulting in a consensus graph  $\mathbf{W}$ . Additionally, a self-weighting fusion module is adopted to compute the corresponding weights  $\lambda^{(1)}, \dots, \lambda^{(N)}$  during the alignment process. Finally, we use spectral clustering on the consensus representation to obtain results.

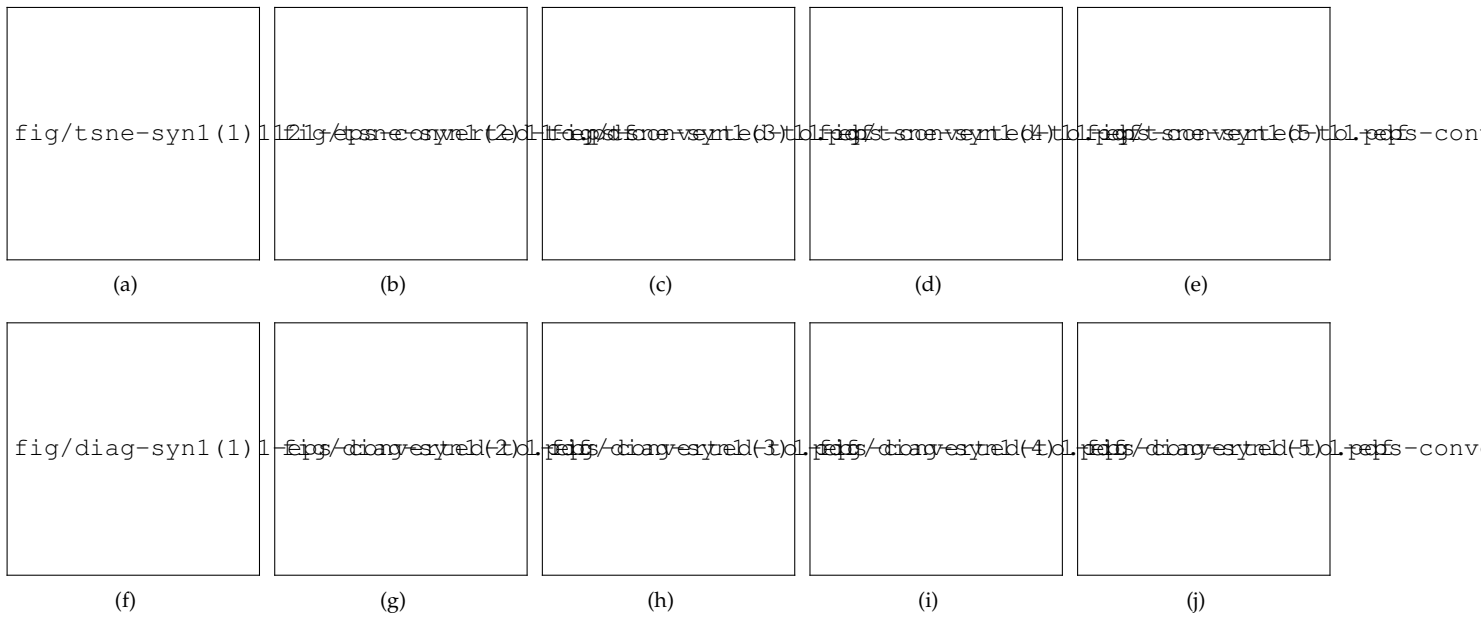


Fig. 2. Visualizations of the consensus representation, raw data with  $t$ -SNE, and the heatmap on Syndata2. Subfigure (a)-(d) depict the spatial distribution of Syndata2 from different views using  $t$ -SNE. Subfigure (e) is the visualization of the obtained consensus representations  $\mathbf{W}$  using  $t$ -SNE. Subfigure (f)-(i) show the heatmap of Syndata2 from different views. The affinity heatmap on raw samples has blurred boundaries and no apparent block structures. In contrast, subfigure (j) shows the affinity from consensus representations has clear-cut boundaries.



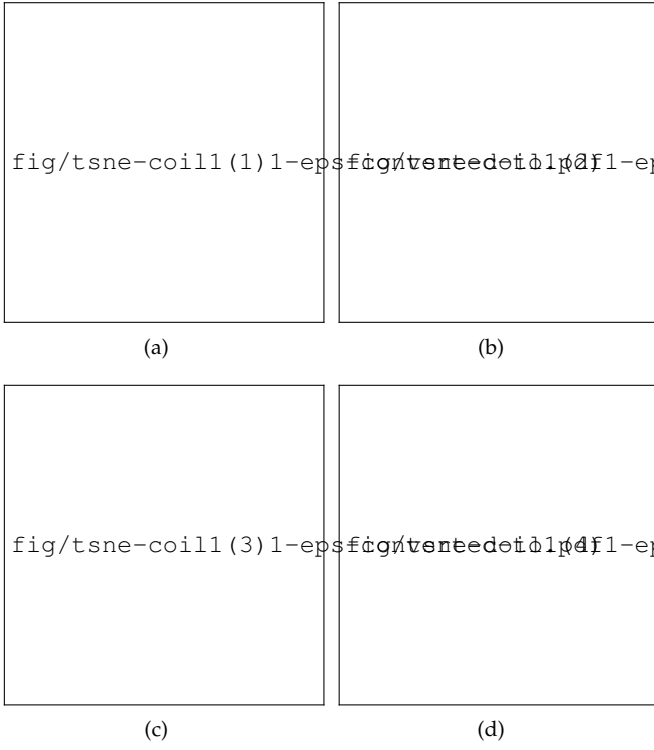


Fig. 3. Visualization of the consensus representations and raw data. Subfigure (a)-(c) depict the visualization of Coil-20 using  $t$ -SNE. Subfigure (d) is visualization of the consensus representations  $W$ .

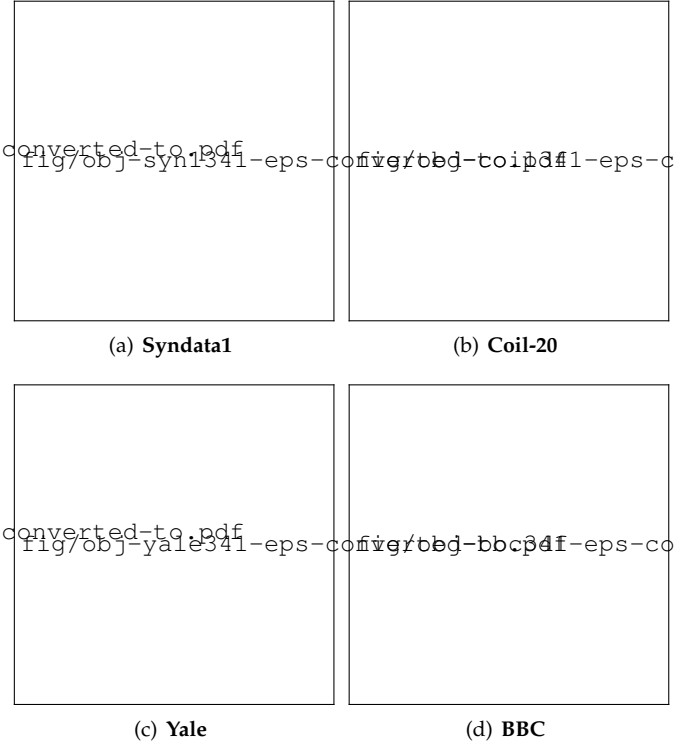


Fig. 5. The convergence curve of the proposed CRMATS method, on (a) Syndata1; (b) Coil-20; (c) Yale; (d) BBC. The objective value decreases consistently with respect to the iteration number.

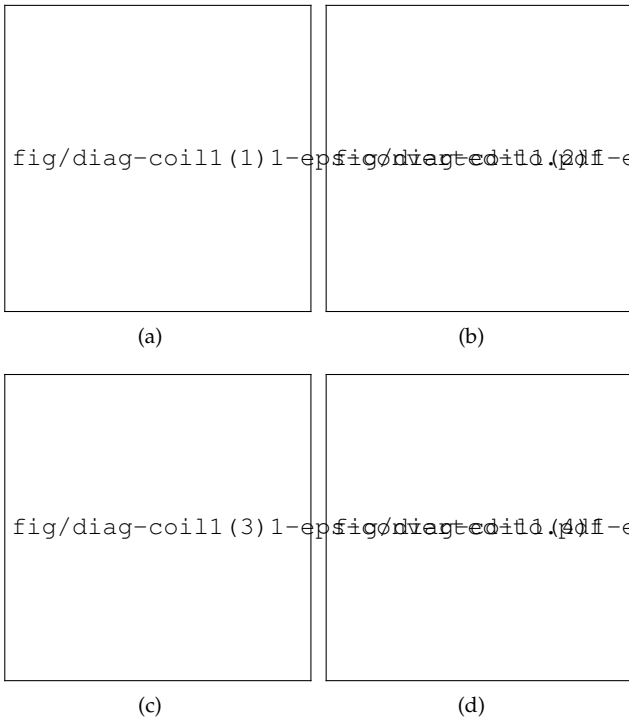


Fig. 4. Visualization of heatmap on Coil-20. Subfigure (a)-(c) are the heatmap from different views. The affinity heatmap on raw samples has blurred boundaries. In contrast, subfigure (d) shows the affinity from consensus representation has clear-cut boundaries.

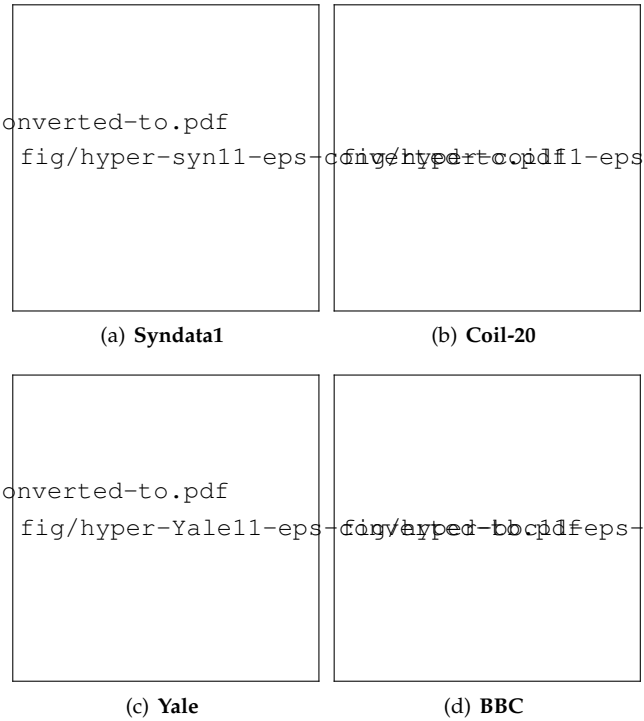
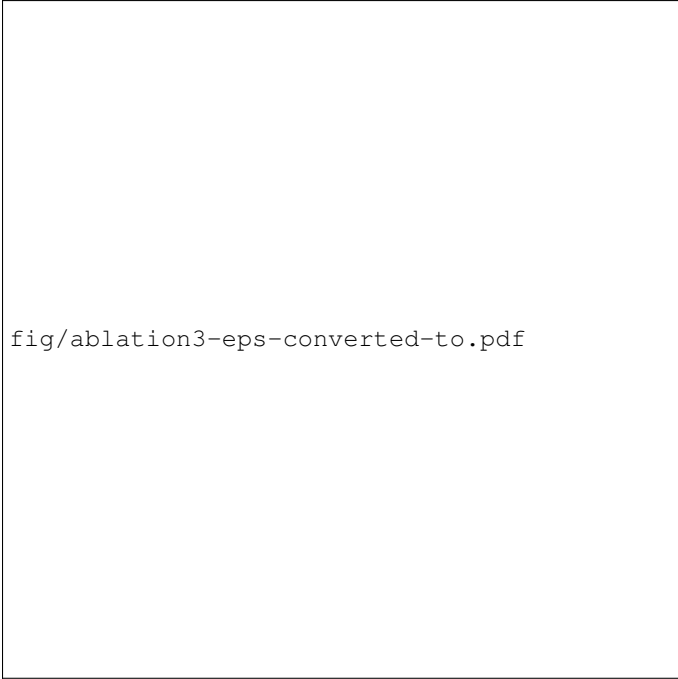


Fig. 6. Hyperparameter tuning ( $\gamma_1$  and  $\gamma_2$ ) in terms of NMI, on (a) Syndata1; (b) Coil-20; (c) Yale; (d) BBC.



fig/ablation3-eps-converted-to.pdf

Fig. 7. Ablation analysis experiments of the CRMATS: 1) CRMATS-L2 which only uses the pairwise affinity matrix  $\hat{L}_2$ ; 2) CRMATS-L3, which only uses the third-order affinity matrix  $\hat{L}_3$ ; 3) CRMATS-L4, which only uses the fourth-order affinity matrix  $\hat{L}_4$ ; 4) CRMATS-L23, which combines  $\hat{L}_2$  and  $\hat{L}_3$ ; 5) CRMATS-L24 combines  $\hat{L}_2$  and  $\hat{L}_4$ ; 6) CRMATS-L34 combines  $\hat{L}_3$  and  $\hat{L}_4$ ; 7) CRMATS is our model.