EDISON: Enhanced Dictionary-Induced Tensorized Incomplete Multi-View Clustering with Gaussian Error Rank Minimization

Zhibin Gu [1 2] **Zhendong Li** [1 2] **Songhe Feng** [1 2]

Abstract

This paper presents an efficient and scalable incomplete multi-view clustering method, referred to as **E**nhanced **D**ictionary-**I**nduced ten**S**orized incomplete multi-view clustering with Gaussian err**O**r ra**N**k minimization (**EDISON**). Specifically, EDISON employs an enhanced dictionary representation strategy as the foundation for inferring missing data and constructing anchor graphs, ensuring robustness to less-than-ideal data and maintaining high computational efficiency. Additionally, we introduce Gaussian error rank as a concise approximation of the true tensor rank, facilitating a comprehensive exploration of the diverse information encapsulated by various singular values in tensor data. Furthermore, we integrate a hyper-anchor graph Laplacian manifold regularization into the tensor representation, allowing for the simultaneous utilization of inter-view high-order correlations and intra-view local correlations. Extensive experiments demonstrate the superiority of the EDISON model in both effectiveness and efficiency compared to SOTA methods.

1. Introduction

Multi-View Clustering (MVC), focusing on effectively partitioning data points into distinct subclusters by leveraging the diverse information encapsulated within multiple views, has gained considerable acclaim and active research (Peng et al., 2019; Zhang et al., 2020; Trosten et al., 2021; 2023; Chen et al., 2022; Tang & Liu, 2022a;b; Xu et al., 2022b; Liang et al., 2023b; Zeng et al., 2023; Wen et al., 2023b; 2021a; Dong et al., 2023a;b). However, most existing MVC methods typically operate under the assumption that each instance universally exists in all views, contrasting with real-world scenarios where certain instances may be absent from specific views, significantly compromising their practical utility. Consequently, Incomplete Multi-View Clustering (IMVC) has emerged as a pivotal subdivision in multi-view clustering, aiming to identify underlying data structures and patterns while addressing missing information through effective fusion of multiple views. In general, existing IMVC methods can be broadly classified into four categories: matrix factorization-based methods (Zhao et al., 2016; Tao et al., 2021; Hu & Chen, 2019), kernel-based methods (Liu et al., 2019; 2021b; Liu, 2023), graph-based methods (Wang et al., 2022; Wen et al., 2023a; Liang et al., 2023a), and deep learning-based methods (Wang et al., 2021; Yang et al., 2023; Jin et al., 2023). Among them, graph-based approaches have drawn considerable interest for their ability to uncover inherent data structures and their solid mathematical underpinnings.

Graph-based incomplete multi-view clustering aims to derive complete graphs from the observed incomplete multiple features, facilitating the partitioning of all instances. For example, Liu et al. (2021a) integrated missing data imputation and graph learning into a unified framework to learn a consensus graph for clustering. Wen et al. (2023a) proposed a confidence neighbor-driven consensus graph learning framework that leverages similar-nearest-neighbor hypothesis to explore the group-wise structure information among samples. Wang et al. (2022) proposed leveraging bipartite graphs to capture the similarity between examples and anchors, thereby improving clustering performance and efficiency. Furthermore, Liang et al. (2023a) incorporated contributions from different views and accounted for the impact of missing samples within a unified GIMC framework to learn a high-quality graph for clustering. However, this matrix-constrained approach only captures the within-view linear relationships among data pairs, neglecting the higher-order correlations between samples across different views, leading to an underutilization of valuable priors in multi-view data. Consequently, recent methods have endeavored to integrate multiple view-specific graphs into a low-rank 3D tensor, allowing for effective exploration of higher-order correlations among data points. For instance, Wen et al.

[1]Key Laboratory of Big Data & Artificial Intelligence in Transportation (Beijing Jiaotong University), Ministry of Education, Beijing, China, [2]School of Computer Science and Technology, Beijing Jiaotong University, Beijing 100044, China. Correspondence to: Songhe Feng <shfeng@bjtu.edu.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

(2021c) seamlessly integrated missing-view inference and low-rank tensor learning into a unified framework, facilitating the simultaneous exploration of hidden information within missing views and high-order correlations among multiple views. Similarly, Zhao et al. (2023) introduced tensor low-rank regularization for missing data recovery and improved inter-graph similarity through inter-view constraint. Furthermore, Zhang et al. (2023) decomposed the complete 3D tensor into low-rank and sparse structures, enabling the simultaneous capture of valuable information and noise within multi-view data.

Notwithstanding the impressive clustering results of the Tensorized Graph-based Incomplete Multi-View Clustering (TGIMC) model, there are still four aspects that merit additional improvement. Firstly, preceding methods relied on the observed data itself to serve as a foundation for inferring missing information and constructing similarity graphs, necessitating observed feature vectors to contain sufficient noise-free sampled points. Otherwise, it may lead to inaccuracies in the recovered features and the constructed similarity graph. Secondly, previous methods often entail extensive computational complexities, posing significant challenges when dealing with large-scale datasets. Thirdly, the mainstream method adopts the Tensor Nuclear Norm (TNN) as the convex surrogate for low-rank tensor, which involves recovering the low-rank tensor by uniformly scaling all singular values. However, large and small singular values are respectively perceived as carrying valuable information and marginal noise. Consequently, TNN fails to exploit the inherent variability among different singular values. Fourthly, the traditional method prioritizes pursuing the global low-rank structure within the 3D tensor, yet disregards the local manifold information between data points within the view, consequently leading to a suboptimal tensor representation.

Building upon the aforementioned insights and motivations, this paper introduces an **E**nhanced **D**ictionary-**I**nduced TGIMC model with Gau**S**sian err**O**r ra**N**k minimization (**EDISON**), which efficiently partitions incomplete multi-view data by exploring rich priors across dictionary, subspace, and tensor representations. Specifically, EDISON incorporates an Enhanced Dictionary Representation (EDR) strategy to recover anchor subspace representations, addressing simultaneously the challenges of incomplete data, insufficient sampling, and high computational complexity encountered in real-world scenarios. Furthermore, we introduce Gaussian Error Rank (GER) as a concise approximation of the tensor rank, providing nuanced penalties on tensor singular values to comprehensively explore valuable prior information within tensor data. Finally, a Hyper-anchor graph Laplacian Regularization (HLR) is incorporated into the anchor subspace learning, ensuring that local geometric manifold information enriches anchor subspace representations. In comparison to the existing TGIMC algorithms, the contributions and innovations of this paper can be summarized as follows:

- EDISON employs an Enhanced Dictionary Representation (EDR) strategy as the foundation for inferring missing data and constructing anchor graphs, effectively addressing the challenges of data incompleteness, insufficient sampling, and high computational complexity encountered in real-world scenarios.
- EDISON adopts the non-convex Gaussian Error Rank (GER) as a concise surrogate of tensor rank, enabling a comprehensive exploration of distinctive information conveyed by various singular values within tensor data.
- EDISON integrates the Hyper-anchor graph Laplacian Regularization (HLR) into the anchor graph learning, enabling the derived tensor representations to retain a low-rank structure while adeptly harnessing rich local geometric manifold information.
- An optimization algorithm with theoretically guaranteed convergence is introduced. Extensive experiments demonstrate the superiority of the EDISON model in both effectiveness and efficiency compared to SOTA methods.

2. Preliminary

Within dataset $\mathbf{X} \in \mathbb{R}^{d\times n}$ containing n samples and d feature dimensions, the Low-Rank Representation (LRR) (Liu et al., 2010; 2013) technique leverages the dataset itself as a dictionary to reconstruct a coefficient matrix of minimal rank as the similarity graph, which can be formulated as:

$$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{Z}\|_* + \alpha\|\mathbf{E}\|_{2,1}, \text{s.t.}\, \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \tag{1}$$

where $\mathbf{Z} \in \mathbb{R}^{n\times n}$ and $\mathbf{E} \in \mathbb{R}^{d\times n}$ represent the graph representation and the reconstruction error, respectively. $\|\cdot\|_*$ and $\|\cdot\|_{2,1}$ denote the nuclear norm and the $\ell_{2,1}$-norm, with α being utilized as the balancing parameter. However, LRR is limited by the requirement that the observed data vectors must include sufficient noise-free sampling points. Otherwise, the resulting subspace representation may degenerate into the identity matrix, leading to the ineffectiveness of LRR. To this end, Liu and Yan (2011) proposed LatLRR, a strategy that collectively incorporates observed and hidden data into a unified dictionary representation for subspace recovery:

$$\min_{\mathbf{Z},\mathbf{H},\mathbf{E}} \|\mathbf{Z}\|_* + \alpha\|\mathbf{E}\|_{2,1}, \text{s.t.}\, \mathbf{X} = [\mathbf{X};\mathbf{H}]\mathbf{Z} + \mathbf{E}, \tag{2}$$

where $[;]$ signifies the concatenation of two matrices along the column direction, and $\mathbf{H}$ represents the hidden data.

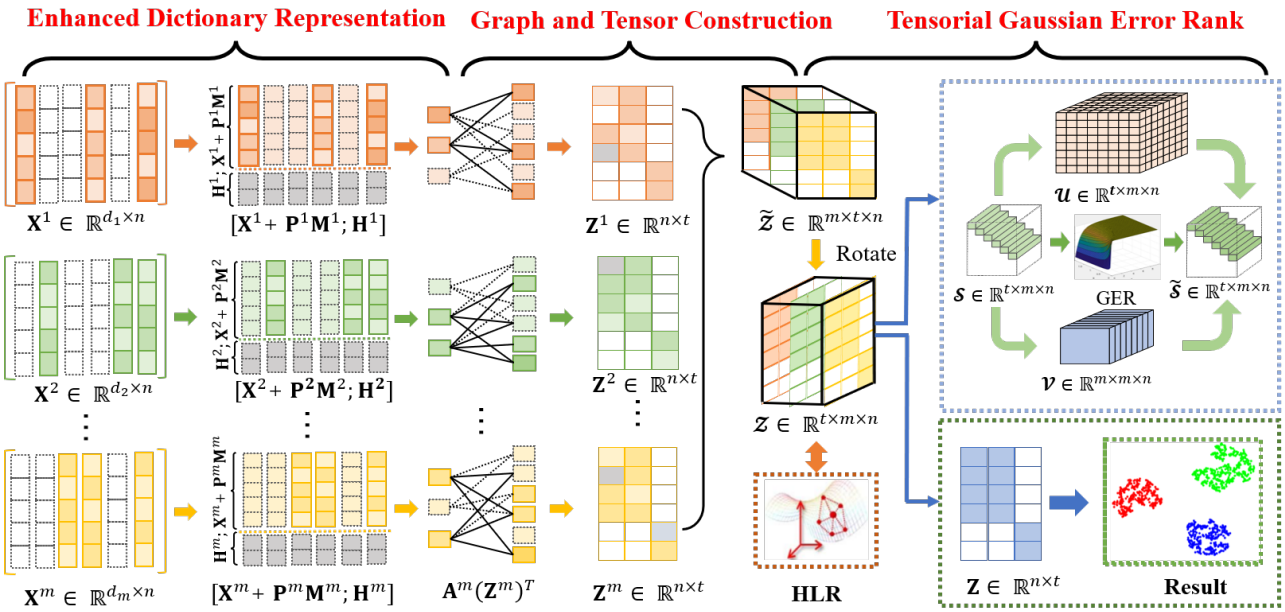


Figure 1. The framework of EDISON. EDISON first employs Enhanced Dictionary Representation (EDR) strategy as the foundation for constructing anchor graphs, ensuring robustness to less-than-ideal data and maintaining high computational efficiency. Then, the 3D tensor organized from multiple graphs undergoes Gaussian Error Rank (GER) minimization, facilitating a comprehensive exploration of the diverse information encapsulated by various singular values. Additionally, a Hyper-anchor graph Laplacian Regularization (HLR) is introduced, allowing for the simultaneous utilization of inter-view high-order correlations and intra-view local correlations.

The combination $[\mathbf{X}; \mathbf{H}]$ forms a comprehensive dictionary, facilitating the recovery of a robust subspace representation. However, due to the completely unknown nature of $\mathbf{H}$, it cannot be acquired directly. Interestingly, LatLRR (Liu & Yan, 2011) indicates that the effects of $\mathbf{H}$ can be formulated as the following dual nuclear norm optimization problem:

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{W},\mathbf{E}} \|\mathbf{Z}\|_* + \|\mathbf{W}\|_* + \alpha\|\mathbf{E}\|_{2,1}, \\ &\text{s.t. } \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{W}\mathbf{X} + \mathbf{E}, \end{aligned} \tag{3}$$

where $\mathbf{W} \in \mathbb{R}^{d \times d}$ is an auxiliary variable derived from the skinny SVD theory, introduced to capture the impact of hidden data (Fu et al., 2021; Li et al., 2023).

For a tensor $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its tensor singular value decomposition (t-SVD) is defined as $\mathcal{Z} = \mathcal{U} * \mathcal{S} * \mathcal{V}$, where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a f-diagonal tensor with diagonal matrices as frontal slices. Based on t-SVD, the definition of the Tensor Nuclear Torm (Lu et al., 2016; Zhou et al., 2021) can be expressed as follows:

Definition 2.1. For the given tensor $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, then the Tensor Nuclear Norm (TNN) can be defined as follows:

$$\|\mathcal{Z}\|_{\text{TNN}} = \frac{1}{n_3}\sum_{k=1}^{n_3} \mathcal{Z}_f^k = \frac{1}{n_3}\sum_{k=1}^{n_3}\sum_{i=1}^{h} \mathcal{S}_f^k(i,i) \tag{4}$$

where $h = min(n_1, n_2)$, and $\mathcal{Z}^k$ is the k-th frontal slice of the tensor $\mathcal{Z}$, and $\mathcal{S}_f^k(i,i)$ is the i-th singular value of $\mathcal{Z}^k$.

3. Proposed Method

In this section, we provide a detailed exposition of the EDISON model, with its framework outlined in Figure 1.

3.1. Tensorized Graph-Based Multi-View Clustering

The tensorized graph-based multi-view clustering typically encompasses the aggregation of multiple view-specific graphs into a 3D low-rank tensor, thereby exploring higher-order correlations among views. Formally, considering a multi-view dataset $\{\mathbf{X}^v \in \mathbb{R}^{d^v \times n}\}_{v=1}^m$ with m views, where d^v is the feature dimension of the v-th view, and n is the number of instances, the self-representation-induced generalized TGIMC model can be articulated as follows:

$$\begin{aligned} &\min_{\{\mathbf{Z}^v,\mathbf{E}^v\}} \|\mathcal{Z}\|_{\text{TNN}} + \alpha\|\mathbf{E}^v\|_{2,1} \\ &\text{s.t. } \forall v,\ \mathbf{X}^v = \mathbf{X}^v\mathbf{Z}^v + \mathbf{E}^v,\ \mathcal{Z} = \psi(\mathbf{Z}^1, ..., \mathbf{Z}^m), \end{aligned} \tag{5}$$

where $\mathbf{Z}^v \in \mathbb{R}^{n \times n}$ and $\mathbf{E}^v \in \mathbb{R}^{d \times n}$ respectively denote view-specific graphs and reconstruction errors. $\psi(\cdot)$ serves is tensorization operator, reconstructing multiple graphs into a 3D tensor $\mathcal{Z} \in \mathbb{R}^{n \times m \times n}$. α is a balancing parameter.

3.2. Enhanced Dictionary Representation

Despite effectively leveraging higher-order correlations within multi-view data, the dependence of Model (5) on observed data $\{\mathbf{X}^v\}_{v=1}^m$ as dictionary for subspace reconstruction overlooks three challenges prevalent in real-world scenarios: 1) **Incomplete observed data**, 2) **Insufficient feature sampling**, and 3) **Excessive computational complexity**. Specifically, the incompleteness of observed multi-view data results in distinct dimensions for view-specific graph structures, impeding the effective integration of multiple features. Sufficiency of feature sampling demands a sufficient number of noise-free sampled points in the feature vectors; otherwise, the graph structure may manifest as an identity matrix, leading to modeling failure in capturing the similarity. The high computational complexity stems from the cubic increase in computational demands during tensor operations with a growing number of samples, presenting challenges for large-scale datasets.

To address the challenges mentioned above in real-world scenarios, three solutions are proposed: 1) Missing Data Imputation, 2) Hidden Effects Consideration, and 3) Anchor Graph Construction.

[1. Missing Data Imputation:] Assuming $\{\mathbf{X}^v \in \mathbb{R}^{d_v\times n}\}_{v=1}^m$ is a data matrix containing n_o^v observed and n_u^v unobserved instances (i.e., $n = n_o^v + n_u^v$), where unobserved samples are filled with zeros. To recover missing data in a more informative manner rather than simple zero imputation, we incorporate the imputation of missing data as a learnable term into the graph learning process of Model (5). This integration establishes a synergy between missing data imputation and the overall model, leading to the inference of meaningful missing data. Specifically, let $\mathbf{P}^v \in \mathbb{R}^{d^v\times n_u^v}$ represents n_u missing instances intended for recovery and $\mathbf{M}^v \in \mathbb{R}^{n_u^v\times n}$ serves as an indicator matrix describing the indices of the missing instances within the v-th view. The term $\{\mathbf{X}^v + \mathbf{P}^v\mathbf{M}^v\}_{v=1}^m$ is then seamlessly integrated into the learning process of Model (5), facilitating the inference of missing information, where $\mathbf{M}_{ij}^v = 1$ when the j-th missing instance in $\mathbf{P}^v$ aligns with the i-th instance in $\mathbf{X}^v$; otherwise, it assumes a value of 0 (Wen et al., 2021c).

[2. Hidden Effects Consideration:] Model (5) employs the observed data $\mathbf{X}^v$ as a dictionary for recovering a similarity graph, requiring that the feature vectors of $\mathbf{X}^v$ contain sufficient noise-free sampled points. Otherwise, the learned graph fails to effectively capture similarity relationships between data pairs. To this end, inspired by LatLRR (Liu & Yan, 2011), we assume the existence of hidden data $\mathbf{H}^v$ as an ideal complement to $\mathbf{X}^v$, serving as the foundation for graph recovery. However, deriving $\mathbf{H}^v$ from $\mathbf{X}^v$ is not practical. Inspired by the theory of shrink SVD, we translate the effects of $\mathbf{H}^v$ into a convex optimization term. Considering the combined dictionary $[\mathbf{X}^v;\mathbf{H}^v]$, we

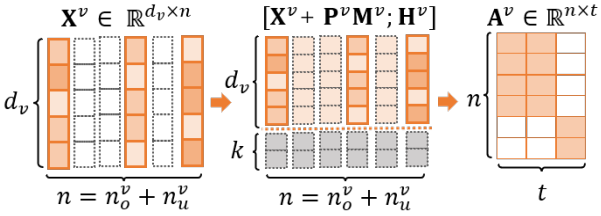


Figure 2. Illustration of Enhanced Dictionary Representation.

obtain the dictionary representation $\mathbf{X}^v = [\mathbf{X}^v;\mathbf{H}^v]\mathbf{Z}_*^v + \mathbf{E}^v$, where $\mathbf{Z}_*^v \in \mathbb{R}^{n\times n}$ is the optimal solution (with respect to $\mathbf{Z}^v$). Following skinny SVD theory (i.e., $\mathbf{X}=\mathbf{U}\Sigma\mathbf{V}_\mathbf{X}^\top$), we have $[\mathbf{X}^v;\mathbf{H}^v]\mathbf{Z}_*^v = \mathbf{X}^v\mathbf{Z}^v + \mathbf{U}^v\Sigma^v\mathbf{V}_\mathbf{Y}^{v\top}\mathbf{V}_\mathbf{H}^v(\Sigma^v)^{-1}\mathbf{U}^{v\top}\mathbf{X}^v$. Denote $\mathbf{P}^v = \mathbf{U}^v\Sigma^v\mathbf{V}_\mathbf{Y}^{v\top}\mathbf{V}_\mathbf{H}^v(\Sigma^v)^{-1}\mathbf{U}^{v\top}$, the dictionary learning, accounting for $\mathbf{H}^v$, can be represented as $\mathbf{X}^v = \mathbf{X}^v\mathbf{Z}^v + \mathbf{W}^v\mathbf{X}^v + \mathbf{E}^v$. Thus, even without obtaining $\mathbf{H}^v$, the effects of $\mathbf{H}^v$ are recovered, leading to an enhancement in the performance of the subspace representation $\mathbf{Z}^v$.

[3. Anchor Graph Construction:] To adapt Model (5) for large-scale datasets, we construct anchor graphs to model similarity, defined as $\mathbf{X}^v = \mathbf{A}^v(\mathbf{Z}^v)^\top$, $\mathbf{A}^v(\mathbf{A}^v)^\top = \mathbf{I}$, where $\mathbf{A}^v \in \mathbb{R}^{d_v\times t}$ represents anchor points with an orthogonality constraint for enhanced discriminability, and $\mathbf{Z}^v \in \mathbb{R}^{n\times t}$ represents the learned anchor graph structure, encoding the similarity between data points and anchor points.

By incorporating the three aforementioned improved dictionary strategies into Model (5), we obtain an enhanced dictionary-induced TGIMC model:

$$\begin{aligned}&\min_{\{\mathbf{Z}^v,\mathbf{P}^v,\mathbf{W}^v,\mathbf{A}^v\},\mathbf{E}} \|\mathcal{Z}\|_{\mathrm{TNN}} + \alpha\sum_{v=1}^m \|\mathbf{W}^v\|_{\mathrm{F}} + \beta\|\mathbf{E}^v\|_{2,1},\\ &\text{s.t.}\,\forall v,\ \mathbf{X}^v + \mathbf{P}^v\mathbf{M}^v = \mathbf{A}^v(\mathbf{Z}^v)^\top + \mathbf{W}^v(\mathbf{X}^v + \mathbf{P}^v\mathbf{M}^v) + \mathbf{E}^v,\\ &\mathbf{E} = [\mathbf{E}^1,\ldots,\mathbf{E}^m]^\top, \mathbf{A}^v(\mathbf{A}^v)^\top = \mathbf{I}, \mathcal{Z} = \psi(\mathbf{Z}^1,\ldots,\mathbf{Z}^m),\end{aligned} \tag{6}$$

where $\mathbf{E} = [\mathbf{E}^1;\ldots;\mathbf{E}^m]^\top$ is formed by horizontally concatenating elements. α and β are trade-off parameters.

3.3. Tensorial Gaussian Error Rank Minimization

Model (6) approximates the tensor rank using the TNN, which reconstructs a low-rank tensor by uniformly scaling all singular values. Nevertheless, this uniform scaling overlooks differentiated information between large and small singular values in the tensor data, resulting in suboptimal tensor representations. This occurs because large singular values signify valuable information, while small singular values represent noise. Treating them equally fails to adequately penalize noise-related features,

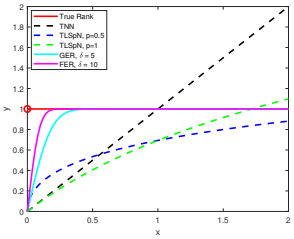


Figure 3. Comparing Tensor Rank Estimation Methods.

resulting in suboptimal tensor representations. To tackle this issue, we propose a novel tensor rank, referred to as the Gaussian Error Rank (GER), as a concise non-convex approximation for the true rank function. Its definition is as follows:

Definition 3.1. For the given tensor $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its Gaussian Error Rank (GER) is defined as follows:

$$\begin{aligned} \|\mathcal{Z}\|_{\mathrm{GER}} &:= \frac{1}{n_3}\sum_{k=1}^{n_3}\|\mathcal{Z}_f^k\|_{\mathrm{GER}} \\ &= \frac{1}{n_3}\sum_{k=1}^{n_3}\sum_{i=1}^{h}\frac{2}{\sqrt{\pi}}\int_0^{\delta \mathcal{S}_f^k(i,i)} e^{-l^2}dl, \end{aligned} \tag{7}$$

where $\delta > 0$, which can be used to adjust the penalty strength for different singular values within tensor data.

3.4. Hyper-Anchor Graph Laplacian Regularization

While model (6) effectively addresses the issue of incomplete and under-sampled observational data in real-world scenarios, its primary focus on uncovering the global low-rank structure of tensor representations neglects the local geometric information within views. This results in the underutilization of rich priors in multi-view data. To address this issue, we introduce Hyper-anchor graph Laplacian Regularization (HLR) to explore the intrinsic geometric structures of the data. To be specific, given a hyper-anchor graph, where vertices serve as anchors. Here, $\mathcal{V}^V$ denotes the set of vertices, $\mathcal{E}^V$ represents the set of hyperedges, and each hyperedge is empirically set to connect to three neighbors. $\mathbf{P}_e^V$ is the weight function, with $p(e)$ indicating the importance of connections within the hypergraph. The vertex-hyperedge relationship is captured by the incidence matrix $\mathbf{O}^v$, with dimensions $|\mathcal{V}^v| \times |\mathcal{E}^v|$, where entries are defined as:

$$\mathbf{O}_{ij} = \begin{cases} 1, & \text{if } \boldsymbol{v}_i \in \boldsymbol{e}_j \\ 0, & \text{otherwise} \end{cases} \tag{8}$$

Subsequently, the degrees of each vertex and hyperedge can be separately computed as $d(\boldsymbol{v}_i) = \sum_{\boldsymbol{e}_j \in \mathcal{E}} p(\boldsymbol{e}_j) o(\boldsymbol{v}_i, \boldsymbol{e}_j)$ and $d(\boldsymbol{e}_j) = \sum_{\boldsymbol{v}_i \in \mathcal{V}} o(\boldsymbol{v}_i, \boldsymbol{e}_j)$. Let $\mathbf{D}_h^v$ and $\mathbf{D}_e^v$ be diagonal matrices representing vertex and hyperedge degrees. The unnormalized hyper-anchor Laplacian matrix is defined as $\mathbf{L}_h^v = \mathbf{D}_h^v - \mathbf{H}^v\mathbf{W}_e^v(\mathbf{D}_e^v)^{-1}\mathbf{H}^v$ (Zhou et al., 2006; Gao et al., 2022). Furthermore, the hyper-anchor graph Laplacian manifold regularization can be formulated as follows:

$$\|\mathbf{Z}^v\|_{\mathrm{HLR}} := \mathrm{Tr}(\mathbf{Z}^v\mathbf{L}_h^v\mathbf{Z}^v), \tag{9}$$

3.5. The Overall EDISON Model

Through the consolidation of Eqs. (6), (7), and (9), the overall EDISON model can be formulated as follows:

$$\begin{aligned} \min_{\{\mathbf{Z}^v,\mathbf{P}^v,\mathbf{W}^v,\mathbf{A}^v\},\mathbf{E}} & \|\mathcal{Z}\|_{\mathrm{GER}} + \alpha\sum_{v=1}^{m}\|\mathbf{W}^v\|_{\mathrm{F}} \\ & + \beta\|\mathbf{E}\|_{2,1} + \gamma\sum_{v=1}^{m}\mathrm{Tr}(\mathbf{Z}^v\mathbf{L}_h^v\mathbf{Z}^v) \\ \text{s.t.} \quad & \forall v,\ \mathbf{X}^v + \mathbf{P}^v\mathbf{M}^v = \mathbf{A}^v(\mathbf{Z}^v)^\top \\ & + \mathbf{W}^v(\mathbf{X}^v + \mathbf{P}^v\mathbf{M}^v) + \mathbf{E}^v, \\ & \mathbf{E} = [\mathbf{E}^1,\ldots,\mathbf{E}^m]^\top, \quad \mathbf{A}^v(\mathbf{A}^v)^\top = \mathbf{I}, \\ & \mathcal{Z} = \psi(\mathbf{Z}^1,\ldots,\mathbf{Z}^m), \end{aligned} \tag{10}$$

where γ is a trade-off parameter. Finally, actively obtain clustering labels by applying the k-means to the left singular vector of the concatenated matrix $\bar{\mathbf{Z}} = \frac{1}{\sqrt{m}}[\mathbf{Z}^1, ..., \mathbf{Z}^m] \in \mathbb{R}^{n \times tm}$, as explained in (Kang et al., 2020).

Remark 1. [Advantages of EDR] Diverging from the previous TGIMC method (Zhang et al., 2023), the EDISON model employs an enhanced dictionary representation (EDR) strategy for inferring missing information and reconstructing subspace representations, ensuring both reliability and efficiency. Specifically, Figure 2 illustrates how EDR achieves reliability by accounting for latent effects in the dictionary representation, ensuring accurate subspace recovery. Furthermore, by selecting a highly discriminative subset of samples from the recovered dictionary representation as anchor dictionaries for subspace recovery, the EDISON model maintains relatively low computational complexity, thus enhancing its efficiency in processing large-scale datasets.

Remark 2. [Benefits of GER] We employ the Gaussian error function (i.e., $f(x) = \frac{2}{\sqrt{\pi}}\int_0^{\delta x} e^{-l^2}dl$) as a non-convex approximation for tensor rank, facilitating the exploration of differentiated information among singular values within tensor data. Figure 3 intuitively compares GER with representative tensor rank approximation methods, namely, TNN (Lu et al., 2016) and TLSpN (Guo et al., 2023). Notably, when $x = 0$, $f(x) = 0$, aligns closely with the true tensor rank. Additionally, as $x \to 0$, $f_{\text{GER}}(x) \gg f_{\text{TNN}}(x)$ and $f_{\text{GER}}(x) \gg f_{\text{TLS}p\text{N}}(x)$, while approaching larger values of x, $f_{\text{GER}}(x) \to 1$. This signifies that GER can robustly penalize small singular values associated with noise, preserving valuable large singular values effectively, thus leveraging valuable priors in tensor data exploration.

Remark 3. [Merits of HLR] Hyper-anchor graph Laplacian Regularization is introduced into the EDISON model to leverage the local geometric information present in multi-view data. Compared to previous Laplacian regularization approaches (Xie et al., 2020; Ji & Feng, 2023), HLR offers two clear advantages: 1) The use of hypergraphs captures

not only linear relationships between anchor points but also effectively captures high-order relationships among them; 2) The anchor-induced subspace representation maintains compactness in scale, reducing computational complexity and enhancing scalability for large datasets.

4. Theoretical Analysis

In this section, we propose an algorithm to optimize the objective function of EDISON, floowed by its convergence and complexity analysis.

4.1. Optimization

We employ the ADMM algorithm (Lin et al., 2011) to optimize the problem (10). This entails introducing auxiliary variables $\mathcal{G}$ and $\{\mathbf{R}^v\}_{k=1}^m$, thereby transforming the constrained problem Eq.(10) into an unconstrained problem. Subsequently, it is decomposed into seven sub-problems for further optimization (i.e., $\mathbf{Z}^v, \mathbf{E}^v, \mathbf{W}^v, \mathbf{A}^v, \mathbf{H}^v, \mathbf{P}^v, \mathcal{G}$). Due to space constraints, detailed optimization steps and algorithm pseudocode can be found in Appendix A.

4.2. Convergence Analysis

The convergence of Algorithm 1 is guaranteed through the validation presented in Theorem 4.1, with comprehensive and rigorous details available in the Appendix B.

Theorem 4.1. *The sequence stemming from the Appendix A (i.e., $\mathcal{S}_t = \{\mathbf{Z}_t^v, \mathbf{E}_t^v, \mathbf{P}_t^v, \mathbf{W}_t^v, \mathbf{A}_t^v, \mathbf{Y}_{1t}^v, \mathbf{Y}_{2t}^v, \mathcal{G}_t\}_{t=1}^{\infty}$) conforms to the following two essential principles:*

- *$\{\mathcal{S}_t\}_{t=1}^{\infty}$ remains bounded;*
- *Each accumulation point of the sequence $\{\mathcal{S}_t\}_{t=1}^{\infty}$ converges to a stationary Karush-Kuhn-Tucker (KKT) point.*

4.3. Complexity Analysis

The computational load of EDISON involves two components: 1) variable optimization and 2) clustering. In the initial phase, updating seven variables ($\mathbf{L}_h^v, \mathbf{Z}^v, \mathbf{E}^v, \mathbf{W}^v, \mathbf{A}^v, \mathbf{H}^v, \mathbf{P}^v, \mathcal{G}$) has specific time complexities: $\mathcal{O}(t^2 m \log(t)$, $\mathcal{O}(nt^2 + ntd)$, $\mathcal{O}(nd)$, $\mathcal{O}(nd^2 + ntd)$, $\mathcal{O}(ntd^v + t^2 d^v)$, $\mathcal{O}(nt^2)$, $\mathcal{O}(ntd + n^2 m n_v)$, $\mathcal{O}(mnt \log(mn) + nm^2 t)$. The second stage has a time complexity of $\mathcal{O}(nt^2 m^2)$, where $d = \sum_{v=1}^m d^v$. Since $t \ll n$, the overall time complexity is $O(ntd + n^2 m n_v + mnt \log(mn))$, which exhibits a quadratic relationship with the number of samples. The space complexity, $O(ntm + nd_{\max})$, with $d_{\max} = \max(d^v)$, demonstrates linear scaling with n.

5. Experiments

In this section, extensive experiments are conducted to verify the effectiveness and superiority of our EDISON. Notably, we performed clustering experiments on all datasets with missing rates of 10%, 30%, and 50%, and recorded the corresponding clustering results for comparison with other methods. For the remaining experimental parts, such as runtime comparison, model analysis, and ablation studies, we maintained a consistent missing rate of 10%. Due to space limitations, we present partial experimental results; for more detailed findings, please refer to the Appendix C.

5.1. Experimental Setup

Datasets: Six multi-view datasets spanning diverse types and scales (including NGs, BBCSport, HW, Scene15, MSRCV1, and ALOI-100) are utilized for clustering purposes. Detailed statistical information about these datasets is documented in Table 1.

Table 1. Statistical Characteristics of Six Datasets.

Datasets	Type	Samples	Clusters	Views
NGs	Text	500	5	3
BBCSport	Text	544	5	2
HW	Digit	2000	10	2
Scene15	Scene	4485	15	3
MSRCV1	Object	210	7	5
ALOI-100	Object	10800	100	4

Comparison Methods: We utilized eight representative methods as baselines to highlight the effectiveness of the EDISON model, encompassing six shallow-based methods (BSV (Ng et al., 2001), Concat (Ng et al., 2001), DAIMC (Hu & Chen, 2018) OPIMC (Hu & Chen, 2019), IMSR (Liu et al., 2021a), HCLS-CGL (Wen et al., 2023a)) and two deep-based methods (CDIMC (Wen et al., 2021b) and DSIMVC (Xu et al., 2022a)).

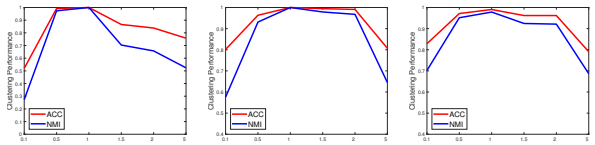
Evaluation Metrics: Three metrics, including ACC, NMI, and PUR, are employed for a comprehensive evaluation of clustering performance, where higher metric values indicate improved clustering quality.

Incomplete Data Construction: Under the condition that each sample retains at least one view, we randomly remove p% of instances from each view (p=10%, 20%, 50%), thereby creating incomplete multi-view datasets.

Implementation Details: Parameters for the comparison methods undergo fine-tuning by guidelines provided in the respective literature, with the best results documented. Regarding EDISON, empirical settings configure the parameter δ to 1 and the number of anchors to c, followed by fine-tuning as explained in the subsection 5.3. Trade-off

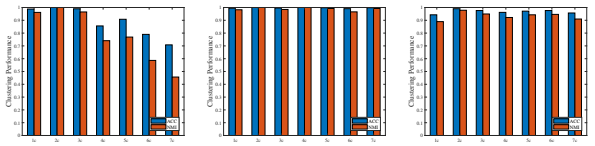
Table 2. Performance Comparison (Mean±Standard Deviation) of Multiple Algorithms Across Six Datasets.

Dataset	NGs									BBCSport								
Metric	ACC(%)			NMI(%)			PUR(%)			ACC(%)			NMI(%)			PUR(%)		
Method\Rate	10%	30%	50%	10%	30%	50%	10%	30%	50%	10%	30%	50%	10%	30%	50%	10%	30%	50%
BSV	37.60±0.00	36.60±0.00	20.80±0.00	19.84±0.00	16.03±0.00	1.00±0.00	39.80±0.00	39.80±0.00	21.00±0.00	58.82±0.00	36.03±0.00	35.85±0.00	43.27±0.00	0.83±0.00	0.87±0.00	68.20±0.00	36.21±0.00	36.21±0.00
Concat	69.40±0.00	21.80±0.00	24.40±0.00	58.24±0.00	2.05±0.00	3.55±0.00	69.40±0.00	22.00±0.00	25.60±0.00	36.03±0.00	46.51±0.00	34.19±0.00	1.34±0.00	28.32±0.00	11.94±0.00	36.58±0.00	49.45±0.00	43.38±0.00
DAIMC	86.20±0.00	63.80±0.00	44.40±0.00	69.83±0.00	40.21±0.00	20.58±0.00	86.20±0.00	63.80±0.00	44.80±0.00	74.45±0.00	63.60±0.00	41.73±0.00	63.06±0.00	45.54±0.00	26.12±0.00	78.86±0.00	69.12±0.00	54.04±0.00
OPIMC	30.86±0.44	45.60±6.96	44.28±0.89	12.84±1.92	25.94±4.31	20.00±0.68	31.06±0.44	51.18±5.00	44.46±0.82	47.59±1.69	52.30±3.20	42.61±3.02	26.82±1.77	30.89±1.82	8.20±1.10	51.14±1.28	55.33±2.33	43.49±2.56
CDIMC	20.00±0.00	20.00±0.00	21.20±0.00	0.00±0.00	0.00±0.00	0.90±0.00	20.00±0.00	20.00±0.00	21.20±0.00	41.00±0.00	34.90±0.40	44.60±0.60	12.40±0.00	3.30±0.60	14.70±0.50	41.50±0.00	36.50±0.30	44.60±0.60
IMSR	97.80±0.00	94.00±0.00	77.80±0.00	92.93±0.00	82.32±0.00	50.85±0.00	97.80±0.00	94.00±0.00	77.80±0.00	92.46±0.00	87.32±0.00	42.65±0.00	79.43±0.00	74.77±0.00	28.84±0.00	92.46±0.00	87.32±0.00	57.17±0.00
DSIMVC	63.00±4.80	57.00±4.10	49.70±4.30	48.10±4.60	39.60±4.90	27.60±4.50	63.70±3.90	57.90±3.80	50.70±3.70	56.60±6.00	54.80±7.30	55.50±4.20	36.40±6.40	33.80±7.90	34.30±5.50	62.70±6.10	61.80±5.70	63.40±4.40
HCLS_CGL	90.87±0.23	52.90±0.14	51.70±1.27	77.85±0.52	42.14±0.76	34.08±1.63	90.87±0.23	53.70±0.14	52.40±0.85	77.21±0.00	77.94±0.00	43.57±0.00	66.08±0.00	63.83±0.00	25.58±0.00	81.07±0.00	81.62±0.00	51.47±0.00
ours	**100.00±0.00**	**99.20±0.00**	**99.00±0.00**	**100.00±0.00**	**97.21±0.00**	**96.69±0.00**	**100.00±0.00**	**99.20±0.00**	**99.00±0.00**	**100.00±0.00**	**99.08±0.00**	**93.93±0.00**	**100.00±0.00**	**96.70±0.00**	**85.23±0.00**	**100.00±0.00**	**99.08±0.00**	**93.93±0.00**
Dataset	MSRCV1									HW								
Metric	ACC(%)			NMI(%)			PUR(%)			ACC(%)			NMI(%)			PUR(%)		
Method\Rate	10%	30%	50%	10%	30%	50%	10%	30%	50%	10%	30%	50%	10%	30%	50%	10%	30%	50%
BSV	58.10±0.00	42.38±0.00	36.67±0.00	46.12±0.00	28.66±0.00	26.52±0.00	58.57±0.00	42.86±0.00	39.05±0.00	67.65±0.00	60.30±0.00	45.15±0.00	63.11±0.00	50.11±0.00	36.18±0.00	71.45±0.00	60.30±0.00	45.15±0.00
Concat	80.00±0.00	67.14±0.00	42.86±0.00	69.02±0.00	57.60±0.00	31.06±0.00	80.00±0.00	69.05±0.00	45.24±0.00	56.70±0.00	62.60±0.00	37.25±0.00	64.14±0.00	58.47±0.00	33.13±0.00	64.20±0.00	62.80±0.00	38.40±0.00
DAIMC	66.19±0.00	60.48±0.00	60.48±0.00	55.41±0.00	48.59±0.00	49.60±0.00	68.57±0.00	60.48±0.00	64.29±0.00	84.65±0.00	74.65±0.00	67.40±0.00	76.70±0.00	71.49±0.00	55.17±0.00	84.65±0.00	78.25±0.00	67.75±0.00
OPIMC	48.71±0.45	24.57±0.90	20.95±0.00	37.61±1.35	9.59±1.69	4.49±0.11	48.76±0.60	24.86±1.81	21.05±0.30	62.52±0.89	27.50±1.42	15.64±0.30	62.84±1.80	21.01±3.75	2.96±0.07	62.66±0.76	28.15±1.88	16.09±0.27
CDIMC	75.40±0.80	61.80±3.30	45.00±1.40	71.20±0.40	54.10±3.20	40.40±1.30	77.40±0.80	64.60±3.20	47.00±0.90	80.40±3.50	95.50±0.10	70.10±1.90	86.10±2.20	90.50±0.20	69.90±0.70	84.70±3.30	95.50±0.10	71.30±1.50
IMSR	87.14±0.00	74.76±0.00	72.86±0.00	77.34±0.00	63.91±0.00	60.04±0.00	87.14±0.00	76.67±0.00	72.86±0.00	92.10±0.07	90.23±0.11	82.10±0.00	84.73±0.14	81.93±0.19	70.20±0.00	92.10±0.07	90.23±0.11	82.10±0.00
DSIMVC	54.40±3.40	55.10±4.80	54.00±5.50	47.90±4.20	48.20±4.60	46.30±5.10	55.60±3.30	56.40±4.60	55.80±5.30	81.90±5.30	81.80±3.00	78.60±3.20	80.00±3.40	78.90±1.40	75.30±1.80	83.00±4.10	82.50±1.80	78.90±2.70
HCLS_CGL	75.24±0.00	71.43±0.00	71.19±0.34	66.22±0.00	61.92±0.00	57.29±0.30	75.24±0.00	71.43±0.00	71.19±0.34	85.10±0.00	86.75±0.00	84.30±0.00	87.57±0.00	86.21±0.00	81.21±0.00	87.35±0.00	86.80±0.00	84.35±0.00
ours	**99.05±0.00**	**92.38±0.00**	**74.29±0.00**	**97.84±0.00**	**85.02±0.00**	**67.13±0.00**	**99.05±0.00**	**92.38±0.00**	**75.24±0.00**	**99.60±0.00**	**99.50±0.00**	**84.45±0.00**	**98.91±0.00**	**98.66±0.00**	**88.89±0.00**	**99.60±0.00**	**99.50±0.00**	**87.25±0.00**
Dataset	scene15									Aloi-100								
Metric	ACC(%)			NMI(%)			PUR(%)			ACC(%)			NMI(%)			PUR(%)		
Method\Rate	10%	30%	50%	10%	30%	50%	10%	30%	50%	10%	30%	50%	10%	30%	50%	10%	30%	50%
BSV	31.68±0.00	25.66±0.00	22.65±0.00	31.77±0.00	24.62±0.00	18.03±0.00	36.37±0.00	29.19±0.00	23.99±0.00	37.53±0.00	29.61±0.00	21.27±0.00	55.14±0.00	44.76±0.00	32.49±0.00	40.38±0.00	32.74±0.00	24.29±0.00
Concat	32.29±0.00	25.80±0.00	22.52±0.00	29.56±0.00	23.84±0.00	18.82±0.00	34.65±0.00	29.28±0.00	24.79±0.00	34.56±0.00	27.47±0.00	20.91±0.00	55.29±0.00	45.76±0.00	38.34±0.00	38.77±0.00	31.15±0.00	24.02±0.00
DAIMC	37.37±0.00	31.22±0.00	25.17±0.00	34.62±0.00	27.55±0.00	18.46±0.00	41.98±0.00	34.76±0.00	27.40±0.00	32.78±0.00	14.31±0.00	10.83±0.00	51.90±0.00	29.63±0.00	25.41±0.00	35.59±0.00	16.32±0.00	12.34±0.00
OPIMC	23.55±0.49	15.19±0.54	14.66±0.02	17.71±1.08	9.81±0.63	9.80±0.07	23.79±0.57	16.81±0.58	16.97±0.07	9.26±1.60	2.29±0.08	2.99±0.28	19.84±3.02	3.34±0.10	5.42±0.59	9.49±1.66	2.30±0.09	3.01±0.28
CDIMC	29.60±1.40	22.10±1.80	13.00±3.20	33.40±2.60	28.00±3.70	10.10±8.20	30.50±1.10	22.60±1.80	13.40±3.50	58.10±1.80	17.50±1.10	6.70±0.50	77.70±1.50	31.30±2.00	17.90±1.30	58.70±1.90	18.30±1.10	7.80±0.60
IMSR	37.28±0.44	28.85±0.09	25.52±0.02	33.84±0.36	25.03±0.49	22.48±0.14	40.47±0.88	31.18±0.08	28.84±0.02	48.57±0.98	37.38±0.25	27.68±0.42	65.79±0.10	57.25±0.13	50.77±0.08	50.69±1.19	39.70±0.04	29.95±0.26
DSIMVC	27.20±1.00	25.90±0.90	26.20±0.80	28.60±1.00	27.30±1.10	26.90±0.80	31.00±0.70	30.10±1.10	29.80±0.70	38.50±1.70	37.00±1.80	34.10±2.00	66.30±0.80	63.80±1.00	60.40±0.90	40.30±1.70	38.90±1.60	36.00±2.00
HCLS_CGL	37.07±0.01	37.78±0.02	31.64±0.00	34.68±0.00	33.49±0.03	26.32±0.00	39.18±0.00	39.99±0.02	33.15±0.00	76.60±1.11	72.97±0.38	**68.45±0.67**	84.59±0.28	78.62±0.28	**71.70±0.09**	78.63±0.69	74.27±0.50	**69.38±0.12**
ours	**87.78±0.00**	**50.50±0.00**	**40.02±0.00**	**86.57±0.00**	**54.94±0.00**	**40.18±0.00**	**88.43±0.00**	**53.56±0.00**	**44.46±0.00**	**84.95±0.00**	**73.65±0.00**	46.75±0.00	**93.18±0.00**	**84.04±0.00**	62.60±0.00	**87.38±0.00**	**76.63±0.00**	50.14±0.00



(a) NGs (b) BBCSport (c) MSRCV1

Figure 4. The Performance Variation of EDISON with Different δ.



(a) NGs (b) BBCSport (c) MSRCV1

Figure 5. EDISON Performance with Varying Anchor Numbers.

parameters α, β, and γ are fine-tuned within the range of [1e-5, 1e+1]. Each experiment is iterated 10 times, and final results, along with standard deviations, are recorded. MATLAB 2018a is used for experiments with the shallow model on a computer equipped with a 3.70GHz i9-10900k CPU and 64GB RAM, while experiments with the deep model are conducted using PyTorch 1.12 on an RTX 2080Ti GPU.

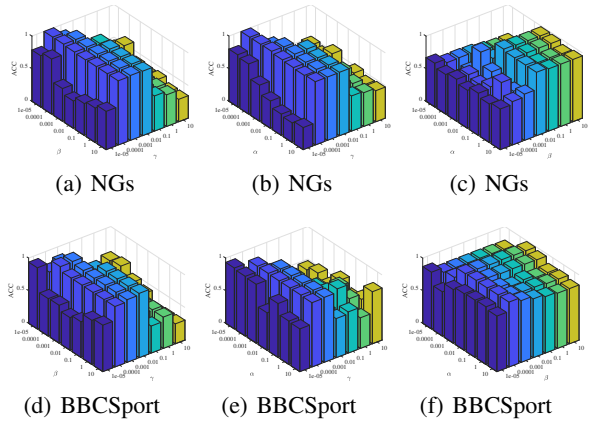
5.2. Performance and Runtime Comparison

Performance Comparison: Table 2 presents a comprehensive comparison between the EDISON model and eight SOTA models across six datasets. The best and second-best performances are highlighted in **red** and blue, respectively. Three significant observations can be inferred from the analysis of Table 2: (1) In most scenarios, our EDISON model demonstrates superior clustering performance. For example, on the BBCSport dataset with a 30% missing rate, the EDISON model surpasses the second-best IMSR model by 12.76%, 21.93%, and 12.76% in terms of ACC, NMI, and PUR, respectively. Furthermore, on both NGs and BBCSport datasets, it achieves a remarkable 100% clustering performance across different metrics. This notable achievement can be attributed to the clever integration of EDR, GER, and HLR into a unified framework within the EDISON model, ensuring a thorough exploration of prior information in dictionaries, subspaces, and tensors and resulting in a notable enhancement in clustering performance. (2) Compared to representative deep-based methods CDIMC (Wen et al., 2021b) and DSIMVC (Xu et al., 2022a), our EDISON model outperforms them across all datasets. This indicates that by judiciously exploiting valuable priors from multiple views, shallow models can still achieve excellent clustering performance.

Runtime Comparison: Table 4 provides a comparison of the runtime of our method with others on a dataset with over 2000 samples, highlighting the fastest and slowest times in **red** and blue, respectively. While our model may not be the absolute fastest, it still yields competitive results. For instance, on the ALOI-100 dataset, the runtime of EDISON is quicker than the SOTA HCLS-CGL 974 seconds. This highlights our method's effectiveness in addressing high computational complexity by constructing anchor graphs, enabling efficient handling of large-scale datasets.

5.3. Model Analysis

Influence of δ in GER: To examine the properties of Gaussian Error Rank (GER), we investigate the impact of the built-in parameter δ on the performance of EDISON. Specifically, we vary the values of δ within the range [0.1, 0.5, 1,



(a) NGs (b) NGs (c) NGs

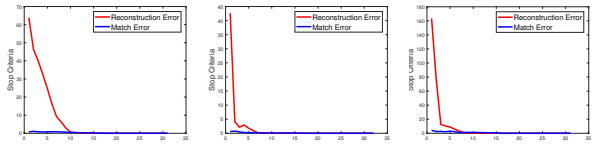
(d) BBCSport (e) BBCSport (f) BBCSport

Figure 6. Parameters Sensitivity Analysis of EDISON Model.

Table 3. Runtime Comparison (in Seconds) of Multi-view Clustering Methods on Datasets with over 2,000 Samples.

Datasets	BSV	Concat	OPIMC	DAIMC	DSIMVC	CDIMC	IMSR	HCLS_CGL	EDISON
HW	0.21	**0.11**	0.46	43.87	896.24	529.41	22.01	62.42	13.53
Scene15	0.49	**0.24**	0.38	206.59	576.86	503.84	66.68	181.98	35.34
ALOI-100	4.25	**1.95**	4.84	3084.00	3526.60	3472.82	2112.20	1169.70	193.28

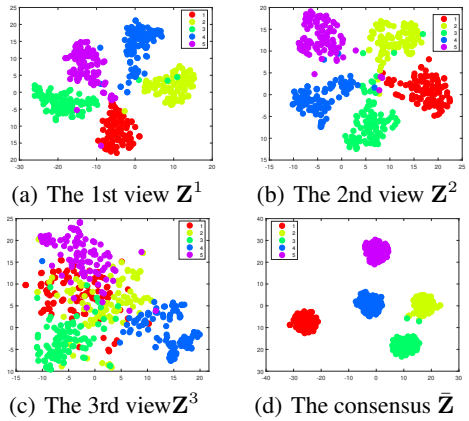
1.5, 2, 5] and record the clustering performance in Figure 5. Analyzing Figure 4 reveals that, globally, EDISON's clustering performance fluctuates to some extent with the changing parameter values. However, within certain local ranges, the EDISON model consistently demonstrates stable and efficient performance. This phenomenon arises from δ determining the penalty strength for tensor singular values, leading to different performances in tensor representation and clustering efficiency under varying penalty strengths.



(a) NGs (b) BBCSport (c) MSRCV1

Figure 7. Convergence Curves of EDISON on Three Datasets.

Impact of Anchor Quantity: The efficiency of the EDISON model is influenced to a certain extent by the chosen number of anchor points. Here, we investigate the impact of varying numbers of anchor points on the performance of the EDISON model. Figure 5 depicts the trend in clustering performance as the number of anchor points changes, with the range of anchor point values spanning $[c, 7c]$, where c represents the true number of clusters. It can be observed that, when the number of anchors is either c or $2c$, the EDISON model achieves excellent performance across different datasets. However, as the number of anchors increases, its performance shows a declining trend on the NGs dataset. This suggests that our model can learn a small number of highly discriminative anchors to effectively cover all samples, ensuring computational efficiency.



(a) The 1st view $\mathbf{Z}^1$ (b) The 2nd view $\mathbf{Z}^2$

(c) The 3rd view$\mathbf{Z}^3$ (d) The consensus $\bar{\mathbf{Z}}$

Figure 8. t-SNE Visualization of Graph Obtained by EDISON.

Parameters Analysis: Three trade-off parameters, denoted as α, β, and γ, control the importance of different regularization terms in the EDISON model and need to be fine-tuned. We fix one parameter and employ a gradient search strategy to fine-tune the remaining two within the range [1e-5, 10], as illustrated in Figure 6. It can be observed that, globally, the EDISON model exhibits some fluctuation with different parameter values. However, within considerable local ranges, the EDISON model consistently demonstrates stable performance. This to some extent indicates the robustness of the EDISON model.

Convergence Behavior: Figure 7 depicts the convergence curve of the EDISON model, with stopping criteria set as reconstruction error (RE $= \sum_{v=1}^{m} \|\mathbf{X}^v + \mathbf{P}^v\mathbf{M}^v - \mathbf{A}^v(\mathbf{Z}^v)^\top - \mathbf{W}^v(\mathbf{X}^v + \mathbf{P}^v\mathbf{M}^v) - \mathbf{E}^v\|_\infty$) and matching error (ME $= \|\mathcal{Z} - \mathcal{J}\|_\infty$). It can be observed that, during the initial iterations, the stopping criteria experience a sharp decline, approaching 0 after approximately 10 iterations and eventually stabilizing. This emphasizes the effective convergence achieved by the EDISON model.

Graph Visualization: To visually illustrate the graph structures learned by the EDISON model, we present the t-SNE visualization of each view anchor graph and the consensus anchor graph in Figure 8. The clustering structure of the consensus graph is clearer compared to the view-specific graphs, indicating the successful integration of rich information from multiple views by the EDISON model.

5.4. Ablation Study

Finally, we perform ablation studies to evaluate the influence of different modules within the EDISON model on the overall performance. We systematically set parameters α, β, and γ to 0, effectively excluding the corresponding

Table 4. Ablation Experiments for EDISON Model.

Datasets			NGs			BBCSport		
EAD	RE	HLR	ACC	NMI	PUR	ACC	NMI	PUR
✓			0.632	0.463	0.652	0.632	0.612	0.739
	✓		0.500	0.325	0.530	0.805	0.664	0.805
		✓	0.964	0.890	0.964	0.991	0.970	0.991
	✓	✓	0.978	0.935	0.978	0.619	0.601	0.735
✓		✓	0.978	0.935	0.978	0.619	0.601	0.735
✓	✓	‘	0.526	0.304	0.552	0.881	0.739	0.881
✓	✓	✓	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

modules EDR, Reconstruction Error (RE), and HLR. The experiment results are displayed in Table 4, where '✓' signifies the removal of the respective module, and the best results are highlighted in **red**. Eliminating one or two modules from the EDISON model leads to a noticeable decline in performance compared to the complete EDISON model. This underscores the adept integration of multiple modules within the EDISON model into a unified framework, effectively leveraging the rich priors inherent in multi-view data and thereby enhancing clustering performance.

6. Conclusion

This paper presents EDISON, a scalable incomplete multi-view clustering model that unifies EDR, GER, and HLR. This integrated framework aims to exploit valuable information in less-than-ideal multi-view data, thereby enhancing clustering performance while maintaining efficient computational complexity. Extensive experimental results demonstrate the superiority of the EDISON model over SOTA methods in terms of clustering effectiveness and efficiency.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

Acknowledgements

This work was supported by the Beijing Natural Science Foundation (No. 4242046) and the Fundamental Research Funds for the Central Universities (No. 2022JBZY019).

References

Chen, M.-S., Wang, C.-D., Huang, D., Lai, J.-H., and Yu, P. S. Efficient orthogonal multi-view subspace clustering. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 127–135, 2022.

Dinh, T. P. and Thi, H. A. L. Convex analysis approach to d.c. programming: Theory, algorithm and applications. *Acta mathematica vietnamica*, 50(4):289–356, 1997.

Dong, Z., Jin, J., Xiao, Y., Wang, S., Zhu, X., Liu, X., and Zhu, E. Iterative deep structural graph contrast clustering for multiview raw data. *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2023a.

Dong, Z., Wang, S., Jin, J., Liu, X., and Zhu, E. Cross-view topology based consistent and complementary information for deep multi-view clustering. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 19383–19394, 2023b.

Fu, Z., Zhao, Y., Chang, D., Zhang, X., and Wang, Y. Double low-rank representation with projection distance penalty for clustering. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5316–5325, 2021.

Gao, Y., Zhang, Z., Lin, H., Zhao, X., Du, S., and Zou, C. Hypergraph learning: Methods and practices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2548–2566, 2022.

Guo, J., Sun, Y., Gao, J., Hu, Y., and Yin, B. Logarithmic schatten-pp norm minimization for tensorial multi-view subspace clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3396–3410, 2023.

Hu, M. and Chen, S. Doubly aligned incomplete multi-view clustering. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pp. 2262–2268, 2018.

Hu, M. and Chen, S. One-pass incomplete multi-view clustering. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*, pp. 3838–3845, 2019.

Ji, J. and Feng, S. Anchor structure regularization induced multi-view subspace clustering via enhanced tensor rank minimization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 19343–19352, October 2023.

Jin, J., Wang, S., Dong, Z., Liu, X., and Zhu, E. Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11600–11609, 2023.

Kang, Z., Zhou, W., Zhao, Z., Shao, J., Han, M., and Xu, Z. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 4412–4419, 04 2020.

Lemaréchal, C. and Sagastizábal, C. Practical aspects of the moreau–yosida regularization: Theoretical preliminaries. *SIAM Journal on Optimization*, 7(2):367–385, 1997. doi: 10.1137/S1052623494267127.

Li, H., Xu, T., Wu, X.-J., Lu, J., and Kittler, J. LRRNet: A novel representation learning guided fusion network for infrared and visible images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):11040–11052, 2023.

Liang, N., Yang, Z., and Xie, S. Incomplete multi-view clustering with sample-level auto-weighted graph fusion. *IEEE Transactions on Knowledge and Data Engineering*, 35(6):6504–6511, 2023a.

Liang, W., Liu, X., Liu, Y., Ma, C., Zhao, Y., Liu, Z., and Zhu, E. Consistency of multiple kernel clustering. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 1–10, 2023b.

Lin, Z., Liu, R., and Su, Z. Linearized alternating direction method with adaptive penalty for low-rank representation. In *Proceedings of the Advances in Neural Information Processing Systems*, volume 24, pp. 1–9, 2011.

Liu, G. and Yan, S. Latent low-rank representation for subspace segmentation and feature extraction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1615–1622, 2011.

Liu, G., Lin, Z., and Yu, Y. Robust subspace segmentation by low-rank representation. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, pp. 663–670, 2010.

Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., and Ma, Y. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):171–184, 2013.

Liu, J., Liu, X., Zhang, Y., Zhang, P., Wenxuan, T., Wang, S., Liang, W., Wang, S., and Yang, Y. Self-representation subspace clustering for incomplete multi-view data. pp. 2726–2734, 10 2021a.

Liu, X. Hyperparameter-free localized simple multiple kernel k-means with global optimum. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):8566–8576, 2023.

Liu, X., Zhu, X., Li, M., Wang, L., Tang, C., Yin, J., Shen, D., Wang, H., and Gao, W. Late fusion incomplete multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(10):2410–2423, 2019.

Liu, X., Li, M., Tang, C., Xia, J., Xiong, J., Liu, L., Kloft, M., and Zhu, E. Efficient and effective regularized incomplete multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8):2634–2646, 2021b.

Lu, C., Feng, J., Chen, Y., Liu, W., Lin, Z., and Yan, S. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5249–5257, 2016.

Ng, A. Y., Jordan, M. I., and Weiss, Y. On spectral clustering: analysis and an algorithm. In *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, pp. 849–856, 2001.

Peng, X., Huang, Z., Lv, J., Zhu, H., and Zhou, J. T. COMIC: Multi-view clustering without parameter selection. In *Proceedings of International Conference on Machine Learning*, volume 97, pp. 5092–5101, 2019.

Tang, H. and Liu, Y. Deep safe incomplete multi-view clustering: Theorem and algorithm. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162, pp. 21090–21110, 2022a.

Tang, H. and Liu, Y. Deep safe multi-view clustering: Reducing the risk of clustering performance degradation caused by view increase. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 202–211, 2022b.

Tao, H., Hou, C., Yi, D., Zhu, J., and Hu, D. Joint embedding learning and low-rank approximation: A framework for incomplete multiview learning. *IEEE Transactions on Cybernetics*, 51(3):1690–1703, 2021.

Trosten, D. J., Løkse, S., Jenssen, R., and Kampffmeyer, M. Reconsidering representation alignment for multi-view clustering. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1255–1265, 2021.

Trosten, D. J., Løkse, S., Jenssen, R., and Kampffmeyer, M. C. On the effects of self-supervision and contrastive alignment in deep multi-view clustering. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23976–23985, 2023.

Wang, Q., Lian, H., Sun, G., Gao, Q., and Jiao, L. iCmSC: Incomplete cross-modal subspace clustering. *IEEE Transactions on Image Processing*, 30:305–317, 2021.

Wang, S., Liu, X., Liu, L., Tu, W., Zhu, X., Liu, J., Zhou, S., and Zhu, E. Highly-efficient incomplete largescale multiview clustering with consensus bipartite graph. In

2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9766–9775, 2022.

Wen, J., Yan, K., Zhang, Z., Xu, Y., Wang, J., Fei, L., and Zhang, B. Adaptive graph completion based incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 23:2493–2504, 2021a.

Wen, J., Zhang, Z., Xu, Y., Zhang, B., Fei, L., and Xie, G.-S. Cdimc-net: cognitive deep incomplete multi-view clustering network. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, pp. 3230–3236, 2021b.

Wen, J., Zhang, Z., Zhang, Z., Zhu, L., Fei, L., Zhang, B., and Xu, Y. Unified tensor framework for incomplete multi-view clustering and missing-view inferring. In *AAAI Conference on Artificial Intelligence*, pp. 10273–10281, 2021c.

Wen, J., Liu, C., Xu, G., Wu, Z., Huang, C., Fei, L., and Xu, Y. Highly confident local structure based consensus graph learning for incomplete multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15712–15721, June 2023a.

Wen, J., Zhang, Z., Fei, L., Zhang, B., Xu, Y., Zhang, Z., and Li, J. A survey on incomplete multiview clustering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2):1136–1149, 2023b.

Xie, Y., Zhang, W., Qu, Y., Dai, L., and Tao, D. Hyper-laplacian regularized multilinear multiview self-representations for clustering and semisupervised learning. *IEEE Transactions on Cybernetics*, 50(2):572–586, 2020.

Xu, J., Li, C., Ren, Y., Peng, L., Mo, Y., Shi, X., and lan Zhu, X. Deep incomplete multi-view clustering via mining cluster complementarity. In *AAAI Conference on Artificial Intelligence*, pp. 8761–8769, 2022a.

Xu, J., Tang, H., Ren, Y., Peng, L., Zhu, X., and He, L. Multi-level feature learning for contrastive multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16030–16039, 2022b.

Yang, M., Li, Y., Hu, P., Bai, J., Lv, J., and Peng, X. Robust multi-view clustering with incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):1055–1069, 2023.

Zeng, P., Yang, M., Lu, Y., Zhang, C., Hu, P., and Peng, X. Semantic invariant multi-view clustering with fully incomplete information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–12, 2023.

Zhang, C., Fu, H., Hu, Q., Cao, X., Xie, Y., Tao, D., and Xu, D. Generalized latent multi-view subspace clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(1):86–99, 2020.

Zhang, C., Li, H., Lv, W., Huang, Z., Gao, Y., and Chen, C. Enhanced tensor low-rank and sparse representation recovery for incomplete multi-view clustering. In *AAAI Conference on Artificial Intelligence*, pp. 11174–11182, 2023.

Zhao, H., Liu, H., and Fu, Y. R. Incomplete multi-modal visual data grouping. In *International Joint Conference on Artificial Intelligence*, pp. 2392–2398, 2016.

Zhao, S., Wen, J., Fei, L., and Zhang, B. Tensorized incomplete multi-view clustering with intrinsic graph completion. In *AAAI Conference on Artificial Intelligence*, pp. 11327–11335, 2023.

Zhou, D., Huang, J., and Schölkopf, B. Learning with hypergraphs: Clustering, classification, and embedding. In *Proceedings of the 19th International Conference on Neural Information Processing Systems*, pp. 1601–1608, 2006.

Zhou, P., Lu, C., Feng, J., Lin, Z., and Yan, S. Tensor low-rank representation for data recovery and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(5):1718–1732, 2021.

In this appendix, we provide supplementary explanations for the theoretical analysis and proofs presented in the main manuscript, along with additional experimental results. The document is organized as follows: Appendix A offers detailed derivations for model optimization; Appendix B presents the convergence proof for Algorithm 1. Lastly, Appendix C provides a comprehensive overview of the experimental results.

A. Optimization

We employ the ADMM algorithm (Lin et al., 2011) to optimize the problem (10). This entails introducing auxiliary variables $\mathcal{G}$ and $\mathbf{K}^v$, thereby transforming Eq. (10) into an unconstrained optimization problem.

$$\begin{aligned}\min_{\{\mathbf{Z}^v,\mathbf{P}^v,\mathbf{W}^v,\mathbf{H}^v,\mathbf{A}^v\},\mathbf{E},\mathcal{G}} &\|\mathcal{G}\|_{\text{GER}} + \frac{\alpha}{2}\sum_{v=1}^{m}\|\mathbf{W}^v\|_F^2 + \beta\|\mathbf{E}\|_{2,1} + \gamma\sum_{v=1}^{m}\text{Tr}(\mathbf{H}^v\mathbf{L}_h^v(\mathbf{H}^v)^\top) + \frac{\rho}{2}\|\mathcal{Z} - \mathcal{G} + \frac{\mathcal{J}}{\rho}\|_F^2 \\ &+ \frac{\mu_1}{2}\sum_{v=1}^{m}\|\mathbf{Y}^v - \mathbf{A}^v(\mathbf{Z}^v)^\top - \mathbf{W}^v\mathbf{Y}^v - \mathbf{E}^v + \frac{\mathbf{B}_1^v}{\mu_1}\|_F^2 + \frac{\mu_2}{2}\sum_{v=1}^{m}\|\mathbf{Z}^v - \mathbf{H}^v + \frac{\mathbf{B}_2^v}{\mu_2}\|_F^2,\end{aligned} \tag{11}$$

where $\mathbf{Y}^v = \mathbf{X}^v + \mathbf{P}^v\mathbf{M}^v$, $\mathcal{J}$, $\mathbf{B}_1^v$, and $\mathbf{B}_2^v$ function as Lagrange multipliers, while ρ, μ_1, and μ_2 play roles as penalty factors. This leads to the further reduction of the problem (11) into seven sub-problems outlined as follows:

• **Solving $\{\mathbf{Z}^v\}$ with fixed $\{\mathbf{A}^v\}$, $\{\mathbf{P}^v\}$, $\{\mathbf{W}^v\}$, $\{\mathbf{H}^v\}$, $\mathbf{E}$ and $\mathcal{G}$.** In this setting, attaining the optimal solution for $\mathbf{Z}^v$ involves equating the partial derivative of Eq. (11) concerning $\mathbf{Z}^v$ to zero, leading to the derivation of the following solution:

$$\begin{aligned}\mathbf{Z}^v = &(\rho\mathcal{G}^v - \mathcal{J}^v + \mu_1(\mathbf{Y}^v)^\top\mathbf{A}^v - \mu_1(\mathbf{Y}^v)^\top(\mathbf{W}^v)^\top\mathbf{A}^v \\ &- \mu_1(\mathbf{E}^v)^\top\mathbf{A}^v + (\mathbf{B}_1^v)^\top\mathbf{A}^v + \mu_2\mathbf{H}^v - \mathbf{B}_2^v) \times [(\rho+\mu_2)\mathbf{I} + \mu_1(\mathbf{A}^v)^\top\mathbf{A}^v]^{-1}.\end{aligned} \tag{12}$$

• **Solving $\{\mathbf{W}^v\}$ with fixed $\{\mathbf{Z}^v\}$, $\{\mathbf{A}^v\}$, $\{\mathbf{P}^v\}$, $\{\mathbf{H}^v\}$, $\mathbf{E}$ and $\mathcal{G}$.** In the current context, setting the first derivative of the problem (11) concerning $\mathbf{W}^v$ to zero results in the following expression for the optimal solution of $\mathbf{W}^v$:

$$\mathbf{W}^v = (\mu_1\mathbf{Y}^v - \mu_1\mathbf{A}^v(\mathbf{Z}^v)^\top - \mu_1\mathbf{E}^v + \mathbf{B}_1^v) \times (\mathbf{Y}^v)^\top \times [\alpha\mathbf{I} + \mu_1\mathbf{Y}^v(\mathbf{Y}^v)^\top]^{-1}. \tag{13}$$

• **Solving $\mathbf{E}$ with fixed $\{\mathbf{Z}^v\}$, $\{\mathbf{A}^v\}$, $\{\mathbf{P}^v\}$, $\{\mathbf{W}^v\}$, $\{\mathbf{H}^v\}$ and $\mathcal{G}$.** In such circumstances, the optimal solution for variable $\mathbf{E}$ can be derived by optimizing the following problem:

$$\min_{\mathbf{E}} \frac{\beta}{\mu_1}\|\mathbf{E}\|_{2,1} + \frac{1}{2}\|\mathbf{E} - \mathbf{D}\|_F^2. \tag{14}$$

where $\mathbf{D}$ is formed by horizontally stacking matrix $\mathbf{Y}^v - \mathbf{A}^v(\mathbf{Z}^v)^\top - \mathbf{W}^v\mathbf{Y}^v + \frac{\mathbf{B}_1^v}{\mu_1}$ along the rows. The optimal solution can be attained through the utilization of the minimization thresholding operator as outlined below:

$$\mathbf{E}^*_{:,i} = \begin{cases} \frac{\|\mathbf{D}_{:,i}\|_2 - \frac{\beta}{\mu_1}}{\|\mathbf{D}_{:,i}\|_2}\mathbf{D}_{:,i}, & \|\mathbf{D}_{:,i}\|_2 > \frac{\beta}{\mu_1} \\ 0, & otherwise \end{cases} \tag{15}$$

• **Solving $\{\mathbf{H}^v\}$ with fixed $\{\mathbf{Z}^v\}$, $\{\mathbf{A}^v\}$, $\{\mathbf{P}^v\}$, $\{\mathbf{W}^v\}$, $\mathbf{E}$ and $\mathcal{G}$.** In this scenario, optimizing problem (11) involves taking the partial derivative concerning $\mathbf{H}^v$, setting it to zero, and obtaining the optimal $\mathbf{H}^v$ as follows:

$$\mathbf{H}^v = (\mu_2\mathbf{Z}^v - \mathbf{B}_2^v)(2\gamma\mathbf{L}_h^v + \mu_2\mathbf{I})^{-1}, \tag{16}$$

• **Solving $\{\mathbf{P}^v\}$ with fixed $\{\mathbf{Z}^v\}$, $\{\mathbf{A}^v\}$, $\{\mathbf{W}^v\}$, $\{\mathbf{H}^v\}$, $\mathbf{E}$ and $\mathcal{G}$.** In this scenario, optimizing problem (11) involves taking the partial derivative concerning $\mathbf{Q}^v$, setting it to zero, and obtaining the optimal $\mathbf{Q}^v$ as follows:

$$\mathbf{P}^v = [(\mathbf{I} - \mathbf{W}^v)^{-1}(\mathbf{A}^v(\mathbf{Z}^v)^\top + \mathbf{E}^v - \frac{\mathbf{B}_1^v}{\mu_1}) - \mathbf{X}^v] * Pinv(\mathbf{M}^v) \tag{17}$$

• **Solving $\{\mathbf{A}^v\}$ with fixed $\{\mathbf{Z}^v\}$, $\{\mathbf{P}^v\}$, $\{\mathbf{W}^v\}$, $\{\mathbf{H}^v\}$, $\mathbf{E}$ and $\mathcal{G}$.** Under such conditions, the optimization of variable $\mathbf{A}^v$ can be streamlined into the following problem:

$$\max_{\mathbf{A}^v} \text{Tr}((\mathbf{A}^v)^\top\mathbf{B}), \quad \text{s.t.}\, (\mathbf{A}^v)^\top\mathbf{A}^v = \mathbf{I}, \tag{18}$$

where $\mathbf{B}=\sum_{v=1}^{m}\frac{\mu_1}{2}(\mathbf{Y}^v-\mathbf{W}^v\mathbf{Y}^v-\mathbf{E}^v+\frac{\mathbf{B}_1^v}{\mu_1})\mathbf{Z}^v$. Given the SVD of $\mathbf{B}$ as $\mathbf{B}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, the optimal $\mathbf{A}=\mathbf{U}\mathbf{V}^\top$.

• **Solving $\mathcal{G}$ with fixed $\{\mathbf{Z}^v\}$, $\{\mathbf{A}^v\}$, $\{\mathbf{P}^v\}$, $\{\mathbf{W}^v\}$, $\{\mathbf{H}^v\}$ and E.** In this scenario, the optimal value of $\mathcal{G}$ is determined by solving the following problem:

$$\min_{\mathcal{G}} \|\mathcal{G}\|_{\mathrm{GER}}+\frac{\rho}{2}\left\|\mathcal{Z}-\mathcal{G}+\frac{\mathcal{J}}{\rho}\right\|_F^2 \tag{19}$$

The Eq. (19) can be regarded as a tensorial hyperbolic tangent rank minimization (HTRM) problem, and it can be effectively resolved by leveraging the following theorem:

Theorem A.1. *Given tensor $\mathcal{G}\in\mathbb{R}^{n_1\times n_2\times n_3}$ and its t-SVD $\mathcal{G}=\mathcal{U}*\mathcal{V}*\mathcal{W}^\top$. The problem of tensorial hyperbolic tangent rank minimization can be formulated as follows:*

$$\min_{\mathcal{J}} \tau\|\mathcal{J}\|_{\mathrm{GER}}+\frac{1}{2}\|\mathcal{J}-\mathcal{G}\|_F^2, \tag{20}$$

whose optimal solution is $\mathcal{J}^*=\mathcal{U}*ifft(\Theta_{f,\tau}(\mathcal{V}_f),[],3)*\mathcal{W}^\top$, where $ifft(\Theta_{f,\tau}(\mathcal{V}_f^{(k)}),[],3)$ is a f-diagonal tensor, and satisfies $\Theta_{f,\tau}(\mathcal{V}_f^{(k)}(ii))=\min_{x\geq 0}\frac{1}{2}(x-\mathcal{V}_f^k(ii)^2)+\tau f(x)$, where $f(x)=\frac{2}{\sqrt{\pi}}\int_0^{\delta x}e^{-l^2}dl$.

Here we provide the proof of Theorem A.1. To establish the validity of Theorem A.1, we shall initially present the subsequent lemma.

Lemma A.2. *Given two matrices $\mathbf{J}$, $\mathbf{G}\in\mathbb{R}^{m\times n}$, and $\mathbf{G}=\mathbf{U}\mathbf{V}_{\mathbf{G}}\mathbf{W}^\top$ represents the singular value decomposition (SVD) of $\mathbf{G}$, and $\tau\geq 0$, we seek an optimal solution to the following problem:*

$$\min_{\mathbf{J}} \tau\|\mathbf{J}\|_{\mathrm{TRS}}+\frac{1}{2}\|\mathbf{J}-\mathbf{G}\|_F^2, \tag{21}$$

is $\mathbf{J}^*=\mathbf{U}\mathbf{V}_{\mathbf{J}}\mathbf{W}^\top$, where $\mathbf{P}_{\mathbf{J}}=diag(\delta^*)$, $\delta^*=\Theta_{f,\tau}(\delta_{\mathbf{G}})$ and $\Theta_{f,\tau}(\delta_{\mathbf{G}})$ refers to the Moreau-Yosida operator (Lemaréchal & Sagastizábal, 1997):

$$\Theta_{f,\tau}(\delta_{\mathbf{G}}):=\min_{\delta\geq 0}\tau f(\delta)+\frac{1}{2}\|\delta-\delta_G\|_2^2. \tag{22}$$

where $f(x)=\frac{2}{\sqrt{\pi}}\int_0^{\delta x}e^{-l^2}dl$.

Proof. In the Fourier domain, given the linearity of the FFT and the relationship $\|\mathcal{X}\|_F^2=\frac{1}{n_3}\|\mathcal{X}_f\|_F^2$, then Eq. (20) can be rephrased as:

$$\begin{aligned}&\frac{1}{2}\|\mathcal{J}-\mathcal{G}\|_F^2+\tau\|\mathcal{J}\|_{\mathrm{HTR}},\\ =&\frac{1}{2n_3}\|\mathcal{J}_f-\mathcal{G}_f\|_F^2+\frac{\tau}{n_3}\sum_{k=1}^{n_3}\|\mathcal{J}_f^k\|_{\mathrm{HTR}}\\ =&\frac{1}{n_3}\sum_{k=1}^{n_3}\left(\frac{1}{2}\left\|\mathcal{J}_f^k-\mathcal{G}_f^k\right\|_F^2+\tau\left\|\mathcal{J}_f^k\right\|\right)\end{aligned} \tag{23}$$

Consequently, the initial tensor optimization problem can be reformulated into n_3 independent matrix optimization problems.

$$\arg\min_{\mathcal{J}_f^k}\left(\frac{1}{2}\left\|\mathcal{J}_f^k-\mathcal{G}_f^k\right\|_F^2+\tau\left\|\mathcal{J}_f^k\right\|\right) \tag{24}$$

for $1\leq k\leq n_3$.

The SVD of $\mathcal{G}_f^k$ is $\mathcal{G}_f^k=\mathcal{U}_f^k\mathcal{V}_f^k(\mathcal{W}_f^k)^H$. Subsequently, leveraging Lemma 1, the optimal solution for Eq. (24) can be derived.

$$\mathcal{J}_f^{k^*}=\mathcal{U}_f^k\Theta_{f,\tau}\mathcal{V}_f^k(\mathcal{W}_f^k)^H. \tag{25}$$

Algorithm 1 Algorithm for solving Eq. (10)

Input: Incomplete multi-view data $\chi = \{X^v \in \mathbb{R}^{d_v \times n}\}_{v=1}^m$, index matrix $\{M^v \in \mathbb{R}^{n_v \times n}\}_{v=1}^m$, anchor points matrix$\{A^v \in \mathbb{R}^{d_v \times t}\}_{v=1}^m$ trade-off parameters $\alpha, \beta, \gamma, \delta > 0$ and cluster number c .
Output: Clustering results
Initialize $\mathbf{Z}^v, \mathbf{W}^v, \mathbf{P}^v, \mathbf{H}^v, \mathbf{E}^v, \mathbf{B}_1^v, \mathbf{B}_2^v$ with zero matrix,$\mathbf{Y}^v = \mathbf{X}^v + \mathbf{P}^v\mathbf{M}^v$, $\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{J}} = 0$, $\mu_1 = \mu_2 = \rho = 10^{-5}$, $\eta = 2$, $\mu_{max} = \rho_{max} = 10^{10}$, $\epsilon = 10^{-7}$
repeat
  Compute hyper-Laplacian matrices $\{L_h^v\}_{v=1}^m$ from $\{A^v\}_{v=1}^m$
  Update $\mathbf{Z}^v$ by solving Eq. (14)
  Update $\mathbf{W}^v$ by solving Eq. (15)
  Update $\mathbf{E}^v$ by solving Eq. (17)
  Update $\mathbf{H}^v$ by solving Eq. (18)
  Update $\mathbf{P}^v$ by solving Eq. (19)
  Update $\mathbf{Y}^v$ by $\mathbf{Y}^v = \mathbf{X}^v + \mathbf{P}^v\mathbf{M}^v$
  Update $\mathbf{A}^v$ by solving Eq. (20)
  Update $\boldsymbol{\mathcal{G}}$ by using Eq. (23)
  Update Lagrange multiplier and corresponding penalty parameters by using Eq. (24);
until $\|\mathbf{Y}^v - \mathbf{A}^v(\mathbf{Z}^v)^\top - \mathbf{W}^v\mathbf{Y}^v - \mathbf{E}^v\|_\infty < \epsilon$ and $\|\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{G}}\|_\infty < \epsilon$

where $\Theta_{f,\tau}$ is determined through the resolution of the following problem:

$$\arg\min_{\delta \geq 0} \frac{1}{2}(x - \boldsymbol{\mathcal{V}}_f^k(i,i))^2 + \tau f(x) \tag{26}$$

where $f(x) = \frac{2}{\sqrt{\pi}} \int_0^{\delta x} e^{-l^2} dl$.

To begin, it's essential to acknowledge that both $f(x)$ and $\frac{1}{2}(x - \boldsymbol{\mathcal{V}}_f^k(i,i))^2$ exhibit non-decreasing behavior when $\delta > \boldsymbol{\mathcal{V}}_f^k(i,i))$. Consequently, the value $\Theta_{f,\tau}\boldsymbol{\mathcal{V}}_f^k(i,i)$ remains within the range [0, $\boldsymbol{\mathcal{V}}_f^k(i,i)$]. Moreover, when $\boldsymbol{\mathcal{V}}_f^k(i,i) = 0$, it's evident that the optimal solution $\Theta_{f,\tau}\boldsymbol{\mathcal{V}}_f^k(i,i) = 0$.

Therefore, once we acquire the optimal solution $\boldsymbol{\mathcal{J}}_f^*$ through rule (25), the subsequent step involves utilizing the inverse Fast Fourier Transform (FFT) operator: ifft$[\boldsymbol{\mathcal{J}}_f^*, [], 3] = \boldsymbol{\mathcal{U}}$ifft$(\Theta_{f,\tau}\boldsymbol{\mathcal{V}})\boldsymbol{\mathcal{W}}^\top$. This process enables us to derive the value $\boldsymbol{\mathcal{J}}_f^*$. □

Eq. (26) presents a combination of both concave and convex functions, making it amenable to the application of the difference of convex (DC) programming (Dinh & Thi, 1997), which, in turn, enables the derivation of a closed-form solution.

$$\phi^{iter+1} = \left(\boldsymbol{\mathcal{V}}_f^{(k)}(ii) - \frac{\beta \partial f(\phi^{iter})}{\rho}\right)_+ \tag{27}$$

where $\phi = \Theta_{f,\frac{\beta}{\rho}}(\boldsymbol{\mathcal{V}}_f^{(k)}(ii))$, $f(x) = \frac{2}{\sqrt{\pi}} \int_0^{\delta x} e^{-l^2} dl$ and $iter$ represents the number of iterations.

Finally, the Lagrange multiplier and associated penalty parameters are updated as follows:

$$\begin{cases} \boldsymbol{\mathcal{J}} = \boldsymbol{\mathcal{J}} + \rho(\boldsymbol{\mathcal{Z}} - \boldsymbol{\mathcal{G}}), \\ \mathbf{B}_1^v = \mathbf{B}_1^v + \mu_1(\mathbf{Y}^v - \mathbf{A}^v(\mathbf{Z}^v)^\top - \mathbf{W}^v\mathbf{Y}^v - \mathbf{E}^v), \\ \mathbf{B}_2^v = \mathbf{B}_2^v + \mu_2(\mathbf{Z}^v - \mathbf{H}^v), \\ \mu_i = \min(\eta\mu_i, \mu_{\max}), i = 1, 2, \\ \rho = \min(\eta\rho, \rho_{\max}). \end{cases} \tag{28}$$

At this juncture, the closed-form solution for each variable in problem (11) has been optimized, and the optimization process is summarized in Algorithm 1.

B. Convergence Proof

Theorem B.1. *The sequence stemming from the Algorithm 1 (i.e.,* $\mathcal{S}_t = \{\mathbf{Z}_t^v, \mathbf{E}_t^v, \mathbf{P}_t^v, \mathbf{A}_t^v, \boldsymbol{\mathcal{J}}_t, \mathbf{Y}_{1\,t}^v, \mathbf{Y}_{2\,t}^v, \boldsymbol{\mathcal{K}}_t\}_{t=1}^\infty$*) conforms to the following two essential principles:*

- $\{\mathcal{S}_t\}_{t=1}^{\infty}$ *remains bounded;*
- *Each accumulation point of the sequence* $\{\mathcal{S}_t\}_{t=1}^{\infty}$ *converges to a stationary Karush-Kuhn-Tucker (KKT) point.*

To validate Theorem B.1, we start by introducing two crucial lemmas.

Lemma B.2. *In the setting of a real Hilbert space* $\mathcal{H}$ *endowed with an inner product* $\langle\cdot,\cdot\rangle$ *and a norm* $|\cdot|$ *with its dual norm* $\|\cdot\|^{dual}$*. Consider* $\mathbf{y}\in\partial|\mathbf{x}|$*, where* $\partial f(\cdot)$ *denotes the subdifferential of the function* $f(\cdot)$*. For* $\mathbf{x}\neq\mathbf{0}$*, it holds that* $|\mathbf{y}|^{dual}=1$*, and for* $\mathbf{x}=\mathbf{0}$*, it holds that* $|\mathbf{y}|^{dual}\leq 1$*.*

Lemma B.3. *Let* $\mathbf{F}:\mathbb{R}^{m\times n}\to\mathbb{R}$ *be a function defined as* $\mathbf{F}(\mathbf{X})=f\circ\delta(\mathbf{X})$*, where* $\mathbf{X}\in\mathbb{R}^{m\times n}$ *and* $\delta(\mathbf{X})=(\sigma_1(\mathbf{X}),\ldots,\sigma_r(\mathbf{X}))$ *represents the singular value vector obtained from the SVD of* $\mathbf{X}$*. Here,* $r=\min(m,n)$*, and* $f(\cdot):\mathbb{R}^r\to\mathbb{R}$ *is a differentiable and symmetric function with respect to its arguments in* $\delta(\mathbf{X})$*. The subdifferential of* $F(\mathbf{X})$ *at* $\mathbf{X}$ *is given by:*

$$\frac{\partial F(\mathbf{X})}{\partial\mathbf{X}}=\mathbf{U}Diag(\partial f(\delta(\mathbf{X})))\mathbf{W}^\top,$$

where $\partial f(\delta(\mathbf{X}))=\left(\frac{\partial f(\sigma_1(x))}{\partial\mathbf{X}},\ldots,\frac{\partial f(\sigma_r(x))}{\partial\mathbf{X}}\right)$.

Proof of the 1st principle: In the $(t+1)$-th iteration, the updating rule of $\mathbf{E}_{t+1}^v$ guarantees the fulfillment of the first-order optimal condition. Therefore, we can infer that:

$$\begin{aligned}0\in&\ \beta\partial\|\mathbf{E}_{t+1}^v\|_{2,1}\\&+\mu_{1t}\|\mathbf{E}_{t+1}^v-(\mathbf{X}_{t+1}^v-\mathbf{A}^v(\mathbf{Z}_{t+1}^v)^\top-\mathbf{P}_{t+1}^v\mathbf{X}^v+\frac{\mathbf{Y}_{1t}^v}{\mu_1})\|_F^2\\&=\beta\partial\|\mathbf{E}_{t+1}^v\|_{2,1}-\mathbf{Y}_{1,t+1}^v,\end{aligned}\tag{29}$$

Thus, we have:

$$\frac{1}{\beta}[\mathbf{Y}_{1,t+1}^v]_{:,j}=\partial\|[\mathbf{E}_{t+1}^v]_{:,j}\|_2,\tag{30}$$

where $[\mathbf{Y}_{1,t+1}^v]_{:,j}$ and $[\mathbf{E}_{t+1}^v]_{:,j}$ represent the j-th column of $\mathbf{Y}_{1,t+1}^v$ and $\mathbf{E}_{t+1}^v$, respectively. Additionally, considering the self-duality of the ℓ_2 norm and relying on B.2, we can establish that $\frac{1}{\beta}[\mathbf{Y}_{1,t+1}^v]_{:,j}\leq 1$. Therefore, the sequence $[\mathbf{Y}_{1,t+1}^v]$ is bounded.

In a similar vein, in accordance with the update regulation of $\mathbf{Q}_{t+1}^v$, $\mathbf{Y}_{2,t+1}$ attains optimality, meeting the requirements of the first-order optimality condition. Consequently,

$$\begin{aligned}0\in&\ \gamma\partial\mathrm{Tr}(\mathbf{Q}_{t+1}^v\mathbf{L}_{t+1}^v(\mathbf{Q}_{t+1}^v)^\top)+\mu_2(\mathbf{Z}_{t+1}^v-\mathbf{Q}_{t+1}^v+\frac{\mathbf{Y}_{2t}^v}{\mu_2})\\&=\gamma\partial\mathrm{Tr}(\mathbf{Q}_{t+1}^v\mathbf{L}_{t+1}^v(\mathbf{Q}_{t+1}^v)^\top)-\mathbf{Y}_{2,t+1}^v\end{aligned}\tag{31}$$

where the equation holds true due to the assertion that $\mathbf{Y}_{2,t+1}^v=\mathbf{Y}_{2,t}^v+\mu_2(\mathbf{Z}_{t+1}^v-\mathbf{Q}_{t+1}^v)$. Hence, it follows that $\{\mathbf{Y}_{2,t+1}^v\}$ is bounded.

Next, we confirm the boundedness of $\{\mathcal{K}_{t+1}\}$. Given the update rule for $\mathcal{J}$, the optimality of $\mathcal{J}_{t+1}$ ensures it meets the first-order optimality condition. This indicates that:

$$\partial\|\mathcal{J}_{t+1}\|_{\mathrm{HTR}}=\mathcal{K}_{t+1}.\tag{32}$$

Given the t-SVD of tensor $\mathcal{J}$ as $\mathcal{U}*\mathcal{V}*\mathcal{W}^T$, the utilization of B.3 leads us to the following outcome:

$$\begin{aligned}&\|\partial\|\mathcal{J}_{t+1}\|_{\mathrm{HTR}}\|_F^2=\left\|\frac{1}{n}\mathcal{U}*ifft(\partial f(\mathcal{V}_f),[],3)*\mathcal{W}^T\right\|\\&=\left\|\frac{1}{n^2}(ifft(\partial f(\mathcal{V}_f),[],3))\right\|_F^2=\left\|\frac{1}{n^3}(\partial f(\mathcal{V}_f))\right\|_F^2\\&\leq\frac{1}{n^3}\sum_{k=1}^{n}\sum_{j=1}^{min(n,m)}[(\partial f(\mathcal{V}_f^k(jj))]^2.\end{aligned}\tag{33}$$

This implies that $\partial\|\mathcal{J}_{f,t+1}\|_{\mathrm{TRS}}$ is bounded. Moreover, the boundedness of $\partial\|\mathcal{J}_{t+1}\|_{\mathrm{HTR}}$ is also ensured. Thus, utilizing Eq. (33) reveals the bounded nature of $\{\mathcal{J}_{t+1}\}$.

Throughout the iterative steps outlined in Algorithm 1, we can deduce the following:

$$\begin{aligned}
&\mathcal{L}(\mathbf{Z}_{t+1}^v, \mathbf{E}_{t+1}^v, \mathbf{P}_{t+1}^v, \mathbf{A}_{t+1}^v, \mathcal{J}_{t+1}, \mathbf{Y}_{1,t}^v, \mathbf{Y}_{2,t}^v, \mathcal{K}_t, \rho_t, \mu_{1,t}, \mu_{2,t}) \\
&\leq \mathcal{L}(\mathbf{Z}_t^v, \mathbf{E}_t^v, \mathbf{P}_t^v, \mathbf{A}_t^v, \mathcal{J}_t, \mathbf{Y}_{1,t}^v, \mathbf{Y}_{2,t}^v, \mathcal{K}_t, \rho_t, \mu_{1t}, \mu_{2,t}) \\
&= \mathcal{L}(\mathbf{Z}_t^v, \mathbf{E}_t^v, \mathbf{P}_t^v, \mathbf{A}_t^v, \mathcal{J}_t, \mathbf{Y}_{1,t-1}^v, \mathbf{Y}_{2,t-1}^v, \mathcal{K}_{t-1}, \rho_{t-1}, \mu_{1,t-1}, \mu_{2,t-1}) \\
&+ \frac{\rho_t - \rho_{t-1}}{2\rho_{t-1}^2} \|\mathcal{K}_t - \mathcal{K}_{t-1}\|_F^2 + \frac{\mu_{1,t} - \mu_{1,t-1}}{2\mu_{1,t-1}^2} \left\|\mathbf{Y}_{1,t}^v - \mathbf{Y}_{1,t-1}^v\right\|_F^2 \\
&+ \frac{\mu_{2,t} - \mu_{2,t-1}}{2\mu_{2,t-1}^2} \left\|\mathbf{Y}_{2,t}^v - \mathbf{Y}_{2,t-1}^v\right\|_F^2
\end{aligned} \tag{34}$$

Hence, by summing both sides of Eq. (34) over t from 1 to n, we obtain the following outcome:

$$\begin{aligned}
&\mathcal{L}(\mathbf{Z}_{t+1}^v, \mathbf{E}_{t+1}^v, \mathbf{P}_{t+1}^v, \mathbf{A}_{t+1}^v, \mathcal{J}_{t+1}, \mathbf{Y}_{1,t}^v, \mathbf{Y}_{2,t}^v, \mathcal{K}_t, \rho_t, \mu_{1,t}, \mu_{2,t}) \\
&\leq \mathcal{L}(\mathbf{Z}_1^v, \mathbf{E}_1^v, \mathbf{P}_1^v, \mathbf{A}_1^v, \mathcal{J}_1, \mathbf{Y}_{1,0}^v, \mathbf{Y}_{2,0}^v, \mathcal{K}_0, \rho_0, \mu_{1,0}, \mu_{2,0}) \\
&+ \sum_{t=1}^{n} \frac{\rho_t - \rho_{t-1}}{2\rho_{t-1}^2} \|\mathcal{K}_t - \mathcal{K}_{t-1}\|_F^2 \\
&+ \sum_{t=1}^{n} \frac{\mu_{1,t} - \mu_{1,t-1}}{2\mu_{1,t-1}^2} \left\|\mathbf{Y}_{1,t}^v - \mathbf{Y}_{1,t-1}^v\right\|_F^2 \\
&+ \sum_{t=1}^{n} \frac{\mu_{2,t} - \mu_{2,t-1}}{2\mu_{2,t-1}^2} \left\|\mathbf{Y}_{2,t}^v - \mathbf{Y}_{2,t-1}^v\right\|_F^2
\end{aligned} \tag{35}$$

Since $\mathcal{L}(\mathbf{Z}_1^v, \mathbf{E}_1^v, \mathbf{P}_1^v, \mathbf{A}_1^v, \mathcal{J}_1, \mathbf{Y}_{1,0}^v, \mathbf{Y}_{2,0}^v, \mathcal{K}_0, \rho_0, \mu_{1,0}, \mu_{2,0})$ is finite. Furthermore, sequences $\{\mathcal{K}_t\}$, $\{\mathbf{Y}_{1,t}\}$, $\{\mathbf{Y}_{2,t}\}$, $\sum_{t=1}^{n} \frac{\rho_t - \rho_{t-1}}{2\rho_{t-1}^2}$, $\sum_{t=1}^{n} \frac{\mu_{1,t} - \mu_{1,t-1}}{2\mu_{1,t-1}^2}$ and $\sum_{t=1}^{n} \frac{\mu_{2,t} - \mu_{2,t-1}}{2\mu_{2,t-1}^2}$ are bounded. Thus, we can confidently state that $\mathcal{L}(\mathbf{Z}_{t+1}^v, \mathbf{E}_{t+1}^v, \mathbf{P}_{t+1}^v, \mathbf{A}_{t+1}^v, \mathcal{J}_{t+1}, \mathbf{Y}_{1,t}^v, \mathbf{Y}_{2,t}^v, \mathcal{K}_t, \rho_t, \mu_{1,t}, \mu_{2,t})$ remains bounded as well. Moreover, considering the boundedness of $\|\mathcal{J}_{t+1}\|_{\mathrm{HTR}}$, it is apparent that all singular values of $\mathcal{J}_{t+1}$ are also bounded. Additionally, building upon the subsequent equation:

$$\begin{aligned}
\|\mathcal{J}_{t+1}\|_F^2 &= \frac{1}{n_3} \|\mathcal{J}_{t+1}\|_F^2 \\
&= \frac{1}{n_3} \sum_{i=1}^{n_3} \sum_{j=1}^{\min(n_1,n_2)} [((\boldsymbol{\mathcal{V}}_f^{(i)}(jj))]^2,
\end{aligned} \tag{36}$$

Therefore, we can affirm that the sequence $\{\mathcal{J}_{t+1}\}$ being bounded implies the boundedness of the sequences $\{\mathbf{A}_{t+1}^v\}$, $\{\mathbf{Z}_{t+1}^v\}$, and $\{\mathbf{P}_{t+1}^v\}$ are also bounded.

Drawing from the preceding analysis, we can confidently conclude that the sequence $\mathcal{S}_t = \{\mathbf{Z}_t^v, \mathbf{E}_t^v, \mathbf{P}_t^v, \mathbf{A}_t^v, \mathcal{J}_t, \mathbf{Y}_{1t}^v, \mathbf{Y}_{2t}^v, \mathcal{K}_t\}_{t=1}^{\infty}$ is bounded.

Proof of the 2nd principle: By appealing to the Weierstrass-Bolzano theorem, we establish the existence of at least one accumulation point, denoted as $\mathcal{S}_t^* = \{\mathbf{Z}_t^v, \mathbf{E}_t^v, \mathbf{P}_t^v, \mathbf{A}_t^v, \mathcal{K}_t, \mathbf{Y}_{1t}^v, \mathbf{Y}_{2t}^v, \mathcal{J}_t\}_{t=1}^{\infty}$, in the sequence $\{\mathcal{S}_t\}_{t=1}^{\infty}$. As a result, we can assert:

$$\begin{aligned}
&\lim_{t\to\infty} (\mathbf{Z}_t^v, \mathbf{E}_t^v, \mathbf{P}_t^v, \mathbf{A}_t^v, \mathcal{J}_t, \mathbf{Y}_{1t}^v, \mathbf{Y}_{2t}^v, \mathcal{K}_t) \\
&\quad = (\mathbf{Z}_*^v, \mathbf{E}_*^v, \mathbf{P}_t^*, \mathbf{A}_*^v, \mathcal{J}_*, \mathbf{Y}_{1*}^v, \mathbf{Y}_{2*}^v, \mathcal{K}_*).
\end{aligned} \tag{37}$$

Referring to the update procedure of $\mathcal{K}$, $\mathbf{Y}_1$ and $\mathbf{Y}_2$, we can derive the following equations:

$$\begin{aligned}
&\mathcal{Z}_{t+1} - \mathcal{J}_{t+1} = (\mathcal{K}_{t+1} - \mathcal{K}_t)/\rho_t, \\
&\mathbf{X}_{t+1}^v - \mathbf{A}_{t+1}^v (\mathbf{Z}_{t+1}^v)^\top - \mathbf{P}_{t+1}^v \mathbf{X}^v - \mathbf{E}_{t+1}^v = (\mathbf{Y}_{1,t+1}^v - \mathbf{Y}_{1,t}^v)/\mu_{1,t}, \\
&\mathbf{Z}_{t+1}^v - \mathbf{Q}_{t+1}^v = (\mathbf{Y}_{2,t+1}^v - \mathbf{Y}_{2,t}^v)/\mu_{2,t}.
\end{aligned} \tag{38}$$

Given the boundedness of the sequences $\{\mathcal{K}_t\}$, $\{\mathbf{Y}^v_{1,t}\}$ and $\{\mathbf{Y}^v_{2,t}\}$, the the fact that $\lim_{t\to\infty}, \rho_t = \infty$, $\mu_{1,t} = \infty$, and $\mu_{2,t} = \infty$, we can draw the following conclusion:

$$\begin{aligned}
&\lim_{t\to\infty} \mathcal{Z}_{t+1} - \mathcal{J}_{t+1} = \lim_{t\to\infty} (\mathcal{K}_{t+1} - \mathcal{K}_t)/\rho_t = 0 \\
&\lim_{t\to\infty} \mathbf{X}^v_{t+1} - \mathbf{A}^v_{t+1}(\mathbf{Z}^v_{t+1})^\top - \mathbf{P}^v_{t+1}\mathbf{X}^v - \mathbf{E}^v_{t+1} = \lim_{t\to\infty} (\mathbf{Y}^v_{1,t+1}\mathbf{Y}^v_{1,t})/\mu_{1,t} = 0, \\
&\lim_{t\to\infty} \mathbf{Z}^v_{t+1} - \mathbf{Q}^v_{t+1} = \lim_{t\to\infty} (\mathbf{Y}^v_{2,t+1}\mathbf{Y}^v_{2,t})/\mu_{2,t} = 0.
\end{aligned} \tag{39}$$

Then, we have $\mathcal{Z}_* - \mathcal{J}_* = 0$, $\mathbf{X}^v - \mathbf{A}^v_*(\mathbf{Z}^v_*)^\top - \mathbf{P}^v_*\mathbf{X}^v - \mathbf{E}^v_* = 0$, $\mathbf{Z}^v_* - \mathbf{Q}^v_* = 0$. Moreover, owing to the fulfillment of the first-order optimality conditions for $\mathcal{J}_{t+1}$, $\mathbf{E}^v_{t+1}$ and $\mathbf{Q}^v_{t+1}$ we can deduce:

$$\begin{aligned}
&0 \in \partial\|\mathcal{J}_{t+1}\|_{\text{GER}} - \mathcal{K}_{t+1} \Rightarrow \mathcal{K}_* = \partial\|\mathcal{J}_*\|_{\text{GER}} \\
&0 \in \beta\partial\|\mathbf{E}^v_{t+1}\|_{2,1} - \mathbf{Y}^v_{1,t+1} \Rightarrow \mathbf{Y}^v_{1,*} = \beta\partial\|\mathbf{E}^v_*\|_{2,1} \\
&0 \in \gamma\partial\text{Tr}(\mathbf{Q}^v_{t+1}\mathbf{L}^v_{t+1}(\mathbf{Q}^v_{t+1})^\top) - \mathbf{Y}^v_{2,t+1} \\
&\quad \Rightarrow \mathbf{Y}^v_{2,*} = \gamma\partial\text{Tr}(\mathbf{Q}^v_*\mathbf{L}^v_*(\mathbf{Q}^v_*)^\top)
\end{aligned} \tag{40}$$

Hence, the KKT conditions are satisfied by the accumulation point of the sequencethe $\mathcal{S}_t = \{\mathbf{Z}^v_t, \mathbf{E}^v_t, \mathbf{P}^v_t, \mathbf{A}^v_t, \mathcal{J}_t, \mathbf{Y}^v_{1\,t}, \mathbf{Y}^v_{2\,t}, \mathcal{K}_t\}_{t=1}^\infty$, obtained through solving the objective function of the EDISON model.

C. More Experimental Results

This section presents the experimental results for all six challenging datasets, encompassing the impact of parameter δ, the influence of anchor points on clustering performance, parameter sensitivity analysis, convergence curves, and ablation experiments.

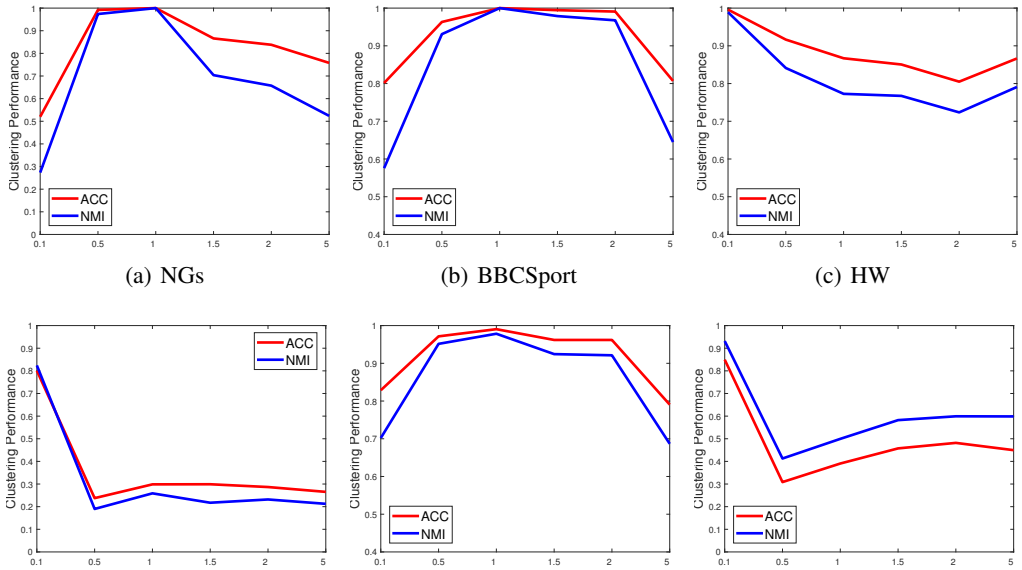
Influence of δ in GER: Figure 9 depicts the variation in clustering performance of the EDISON model concerning the internal parameter δ within the HTR across eight datasets. As anticipated, different values of δ exhibit an impact on clustering performance. Optimal clustering performance is achieved with a δ value of 1 for datasets NGs, BBCSpor and MSRCV1, while datasets HW, Scene15,and ALOI-100 demonstrate improved performance with δ values of 0.5 and 0.1, respectively.

Impact of Anchor Quantity: Figure 10 displays the analysis regarding the number of anchor points across all six datasets. It is apparent that greater quantity of selected anchors does not necessarily result in better performance. This indicates that a small, discriminative set of anchor points can significantly enhance clustering performance compared to a larger set of redundant ones. Moreover, it is evident that for all datasets, the number of anchor points is set equal to the number of clusters (i.e., $t = 2c$), resulting in optimal clustering outcomes. These experimental findings highlight our method's efficiency in effectively covering the entire dataset by employing a minimal number of anchor points, thereby enhancing both effectiveness and efficiency.

Parameters Analysis: The sensitivity analysis of the EDISON model concerning three trade-off parameters, α, β and γ across eight datasets is illustrated in Figure 11. It can be observed that varying combinations of these parameters lead to fluctuations in the clustering performance of the EDISON model. However, when these three parameters fluctuate within a certain range, the EDISON model demonstrates relatively stable clustering performance. This indicates the model's adeptness in amalgamating different regularization terms within a unified framework, effectively extracting rich information from multi-view data.

Convergence Behavior: Figure 12 presents the convergence curves for the reconstruction error (RE) and matching error (ME) of the EDISON model across eight datasets (RE $= \sum_{v=1}^m \|\mathbf{X}^v - \mathbf{A}^v(\mathbf{Z}^v)^\top - \mathbf{P}^v\mathbf{X}^v - \mathbf{E}^v\|_\infty$ and ME $= \|\mathcal{Z} - \mathcal{J}\|_\infty$ and ME $= \|\mathcal{Z} - \mathcal{J}\|_\infty$). It is evident that both RE and ME converge rapidly to 0 within 15 iterations across all datasets, affirming the robust convergence of the EDISON model.

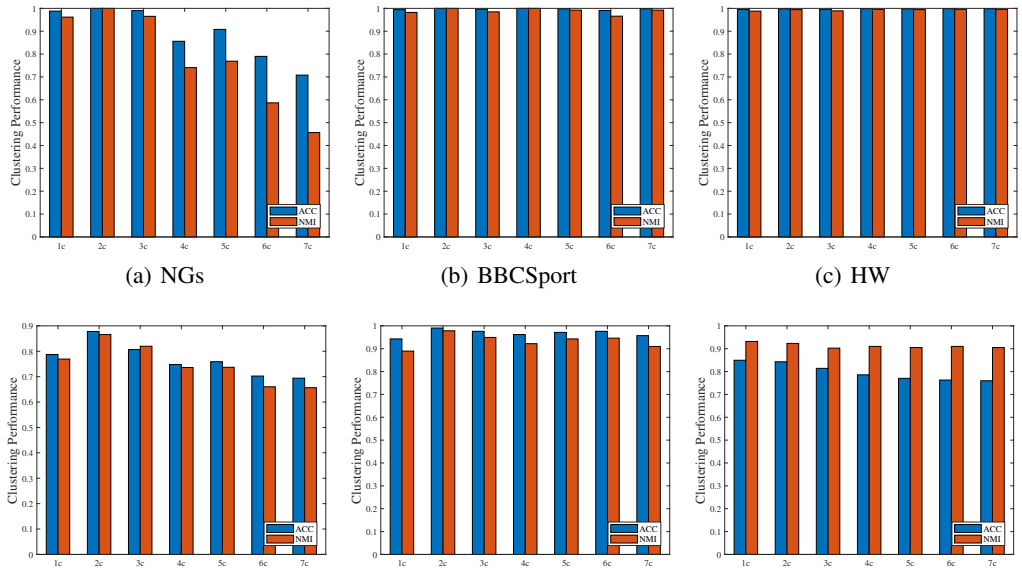
Ablation Study: Table 5-Table 7 report the results of ablation experiments conducted across six datasets, evaluating three metrics. These experiments involved selectively removing the EDR, GER, and HLR modules from the EDISON model in different combinations, indicated by checkmarks denoting their inclusion. The experimental outcomes distinctly reveal that the clustering performance of the downgraded models—resulting from the exclusion of one or two sub-modules from the



(a) NGs (b) BBCSport (c) HW

(d) Scene15 (e) MSRCv1 (f) ALOI-100

Figure 9. The Performance Variation of EDISON with Different δ.

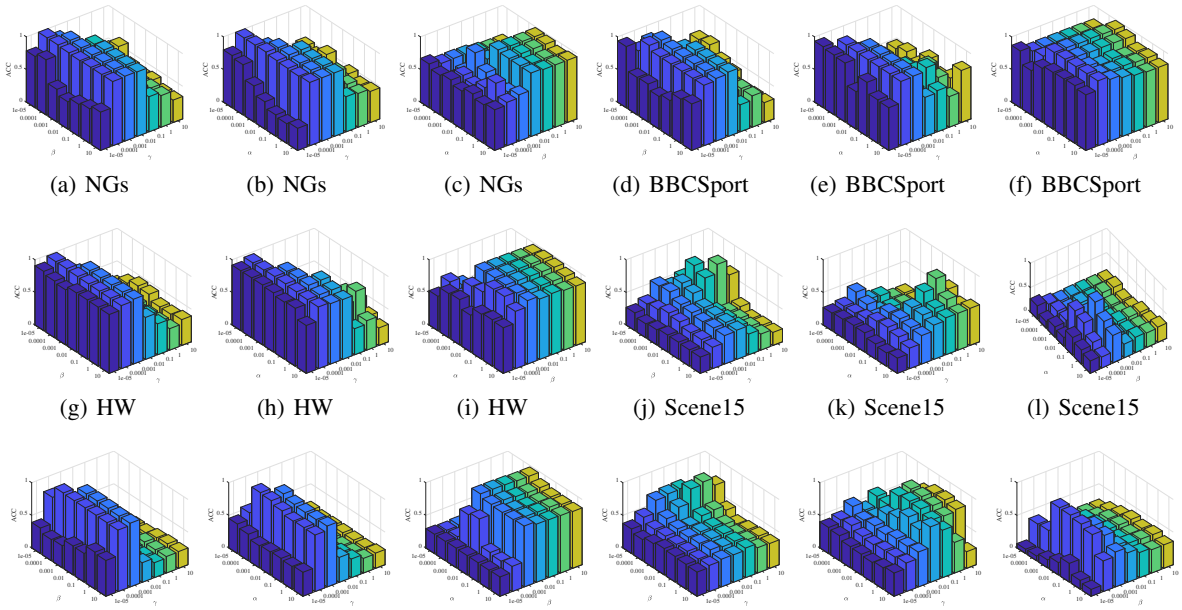


(a) NGs (b) BBCSport (c) HW

(d) Scene15 (e) MSRCv1 (f) ALOI-100

Figure 10. Parameters Sensitivity Analysis of EDISON Model.

EDISON model—significantly lags behind that of the complete EDISON model. This emphasizes the successful synergy among EDR, GER, and HLR within the EDISON framework, enabling them to collaboratively leverage the rich information inherent in multi-view data and achieve commendable clustering performance.

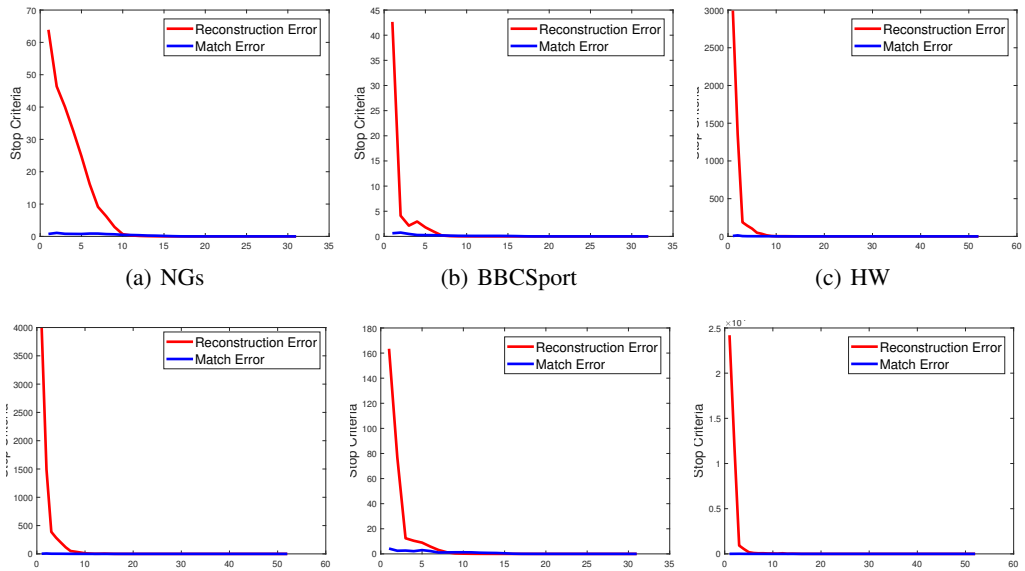


(a) NGs (b) NGs (c) NGs (d) BBCSport (e) BBCSport (f) BBCSport

(g) HW (h) HW (i) HW (j) Scene15 (k) Scene15 (l) Scene15

(m) MSRCv1 (n) MSRCv1 (o) MSRCv1 (p) ALOI-100 (q) ALOI-100 (r) ALOI-100

Figure 11. Parameters Sensitivity Analysis of EDISON Model.



(a) NGs (b) BBCSport (c) HW

(d) Scene15 (e) MSRCv1 (f) ALOI-100

Figure 12. Convergence Curves of EDISON on Three Datasets.

Table 5. Ablation Experiments for EDISON Model.

Datasets			NGs			BBCSport		
EAD	RE	HLR	ACC	NMI	PUR	ACC	NMI	PUR
✓			0.632	0.463	0.652	0.632	0.612	0.739
	✓		0.500	0.325	0.530	0.805	0.664	0.805
		✓	0.964	0.890	0.964	0.991	0.970	0.991
	✓	✓	0.978	0.935	0.978	0.619	0.601	0.735
✓		✓	0.978	0.935	0.978	0.619	0.601	0.735
✓	✓		0.526	0.304	0.552	0.881	0.739	0.881
✓	✓	✓	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

Table 6. Ablation Experiments for EDISON Model.

Datasets			MSRCV1			HW		
EAD	RE	HLR	ACC	NMI	PUR	ACC	NMI	PUR
✓			0.290	0.224	0.305	0.669	0.683	0.670
	✓		0.295	0.239	0.310	0.637	0.650	0.642
		✓	0.610	0.575	0.624	0.997	0.991	0.997
	✓	✓	0.405	0.353	0.414	0.852	0.912	0.891
✓		✓	0.367	0.319	0.367	0.857	0.924	0.896
✓	✓		0.295	0.262	0.305	0.557	0.589	0.560
✓	✓	✓	**0.991**	**0.978**	**0.991**	**0.996**	**0.989**	**0.996**

Table 7. Ablation Experiments for EDISON Model.

Datasets			Scene15			ALOI-100		
EAD	RE	HLR	ACC	NMI	PUR	ACC	NMI	PUR
✓			0.157	0.065	0.165	0.064	0.144	0.067
	✓		0.124	0.045	0.134	0.081	0.180	0.085
		✓	0.790	0.821	0.839	0.670	0.834	0.720
	✓	✓	0.766	0.835	0.806	0.714	0.870	0.755
✓		✓	0.643	0.633	0.659	0.706	0.865	0.740
✓	✓		0.127	0.067	0.144	0.073	0.173	0.076
✓	✓	✓	**0.878**	**0.866**	**0.884**	**0.850**	**0.932**	**0.874**