Self-supervised Graph Contrastive Learning for Incomplete Multi-view Clustering

Anni Chen, Shan Jin, Zhikui Chen*, Yanfan Li, Shuo Yu
University of Wollongong, Australia; Dalian University of Technology, China
Email: ac851@uowmail.edu.au, shan.jinjin@outlook.com, zkchen@dlut.edu.cn, liyanfan@dlut.edu.cn, shuo.yu@ieee.org

***Abstract*—Improving clustering performance on incomplete multi-view data is crucial for intelligent systems. However, most existing incomplete multi-view clustering methods assume consistent missing patterns across views, which rarely holds in real-world applications with heterogeneous and view-specific missingness. To address this issue, we propose SIGMA (Self-supervised Graph Contrastive Learning for Incomplete Multi-view Data Alignment), a method that enables effective multi-view representations under inconsistent missing patterns. Compared with existing studies, SIGMA extracts view-specific structures while independently inferring missing information for each view. Specifically, SIGMA initializes view-specific graphs and a global consistent graph using k-nearest neighbors, and dynamically propagates reliable structural information from the global graph to incomplete views. Furthermore, a weighted graph contrastive learning strategy is introduced to preserve both local discriminability and global structural consistency across recovered views. Extensive experiments on multiple benchmark datasets demonstrate that SIGMA consistently outperforms state-of-the-art incomplete multi-view clustering methods.**

***Index Terms*—graph learning, graph neural networks, multi-view learning, clustering**

I. Introduction

The diverse social behaviors of users on social platforms give rise to multi-view networks in web applications. Clustering analysis of multi-view social networks plays a crucial role in supporting downstream web services, including community detection, social recommendation, and online public opinion monitoring [1]. However, due to pervasive privacy constraints on web platforms, social information often suffers from various types of data missing, leading to incomplete social networks [2]. Therefore, improving clustering performance in incomplete social network scenarios is essential. This facilitates providing comprehensive data representations for downstream web services, ensuring service quality under privacy constraints.

To address the aforementioned issues, numerous studies on incomplete multi-view clustering (IMC) have been proposed [3]. Some studies have employed multi-kernel learning approaches to generate multi-view clustering representations [4]. By defining multiple kernel functions, these methods are capable of extracting salient information from diverse views [5]. However, these methods typically process each view independently, overlooking the consistent and interactive information shared across different views. Furthermore, some studies have incorporated graph learning approaches to align the consistent relationships across views [6], [7]. By representing each sample as a node and the relationships between samples as edges, these methods leverage information propagation mechanisms to infer missing information from both intra-view and inter-view structural relationships, significantly enhancing the capability of representation learning [5], [8].

This work was supported by the National Key R&D Program of China under Grant 2025YFF0514800, and in part by the Natural Science Foundation of Liaoning under Grant No. 2024-MSLH-073. * Corresponding author: Zhikui Chen (zkchen@dlut.edu.cn).

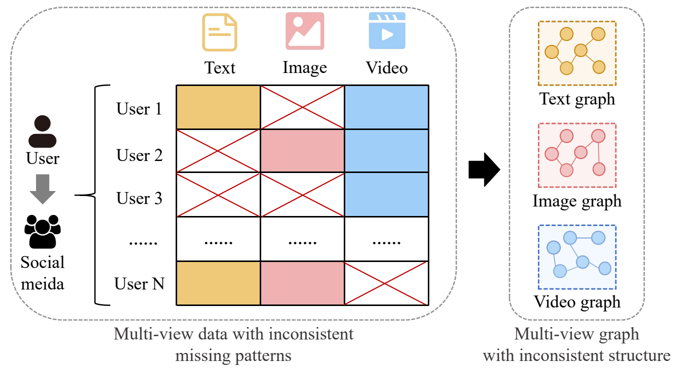

Fig. 1. Illustration of the incomplete multi-view data with inconsistent missing patterns. The inconsistent pattern of missing data directly leads to structural inconsistencies among the constructed multi-view graphs.

Although effective, existing methods typically assume that missing patterns are consistent across views. Unfortunately, in real-world web scenarios, influenced by variations in user behavior and platform generation mechanisms, missing patterns across multiple views are generally inconsistent [9]. For example, as shown in Figure 1, on social media platforms, some users provide only text and image information while lacking videos, others provide text and videos but no images, and yet others upload only images. This inconsistency in missing patterns limits the comparability of samples across views, induces disparities in the data structures among views, and constrains the ability of existing methods to infer missing information from inter-view structural relationships. This limitation leads to a substantial drop in the effectiveness of clustering analysis on incomplete social network data.

To overcome these limitations, we propose **SIGMA**, a novel method that effectively learns multi-view representations from multi-view data with inconsistent missing patterns. SIGMA first construct view-specific graphs and multi-view consistent graph using k-nearest neighbors method. Through this process, we align the topological structures of different views. Then, we introduce an information propagation mechanism that leverages non-missing neighbors in the multi-view consistent graph to infer missing representations in each view. Finally, we design an adaptive weighted graph contrastive learning strategy to ensure local and global structural consistency across different views, facilitating effective learning of multi-view representations.

The main contributions of this work are summarized as follows:

- We present the study to effectively learn multi-view clustering representations under highly diverse and view-specific missing patterns, paving the way for promising advances in practical applications.
- We propose SIGMA, a novel method that leverages a structure-aware multi-view graph completion strategy to recover missing data by learning the global consistent structure, and a weighted graph contrastive learning strategy to learn structure-enhanced multi-view representations.
- Extensive experiments on four real-world benchmark datasets demonstrate that SIGMA consistently outperforms state-of-the-art incomplete multi-view clustering methods.

II. Methodology

We consider a multi-view dataset denoted as $\mathcal{D} = \{D_1, D_2, \ldots, D_M\}$, where each view D_m corresponds to a distinct feature representation of the same set of N instances. The multi-view data are given as $\mathcal{X} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(M)}\}$, where $\mathbf{X}^{(m)} \in \mathbb{R}^{N\times d_m}$ is the feature matrix of the m-th view with feature dimension d_m. Since some instances may be missing in certain views, a binary mask matrix $\mathbf{M} \in \mathbb{R}^{N\times V}$ is introduced for each view m, where $\mathbf{M}_{im} = 1$ if the i-th instance is observed in view m, and $\mathbf{M}_{im} = 0$ otherwise.

In this paper, we aim to learn effective latent representations $\mathbf{Z}$ from incomplete multi-view data $\mathcal{V}$ and derive a consensus clustering assignment C that effectively groups instances despite view missingness. We present the overall framework of SIGMA in Figure 2.

A. Multi-View Graph Initialization

In real-world scenarios, multi-view data often exhibit missing views for certain samples. Directly learning representations from each view independently may result in misaligned information, thereby degrading the quality of representation learning. Therefore, we propose a k-nearest neighbors based multi-view graph initialization module that constructs view-specific similarity graphs and a unified global similarity graph, providing structured initial representations for incomplete samples via neighborhood and cross-view associations.

a) View-specific Graph Initialization

Given a multi-view dataset comprising $\mathcal{D}$ views and a binary mask matrix $\mathbf{M}$, each view $\mathbf{X}^{(v)}$ is partitioned into a complete subset $\mathbf{X}_C^{(v)}$ and an incomplete subset $\mathbf{X}_I^{(v)}$. To promote sparsity and computational efficiency, we construct view-specific graphs using a k-nearest neighbor (k-NN) strategy.

Specifically, for any pair of samples $\mathbf{X}_i^{(v)}$ and $\mathbf{X}_j^{(v)}$ in view v, the sample-specific similarity weight $\mathbf{W}_{ij}^{(v)}$ is defined based on the normalized distance gap, as shown in Equation (1).

$$\mathbf{W}_{ij}^{(v)} = \begin{cases} \frac{d_{i,k+1}-d_{ij}}{k d_{i,k+1}-\sum_{h=1}^{k} d_{ih}}, & j \in \mathcal{N}_i^{(v)}, \\ 0, & \text{otherwise}, \end{cases} \tag{1}$$

where $d_{ij} = \|\mathbf{X}_i^{(v)} - \mathbf{X}_j^{(v)}\|_2$ denotes the Euclidean distance, and $\mathcal{N}_i^{(v)}$ denotes the index set of the k nearest neighbors of sample i in view v.

b) Unified Global Graph Initialization

To integrate cross-view information under missing observations, we construct a unified global similarity graph by aggregating available view-specific similarities while ignoring missing entries, providing an effective structural initialization for incomplete multi-view data.

Formally, for each sample i, the global similarity vector $\mathbf{S}_i$ is defined as a masked average over the observed views, as shown in Equation (2).

$$\mathbf{S}_i = \frac{\sum_{v=1}^{V} \mathbf{M}_{iv} \otimes \mathbf{W}_i^{(v)}}{\sum_{v=1}^{V} \mathbf{M}_{iv}}, \tag{2}$$

where $\mathbf{M}_{iv} \in \{0, 1\}$ indicates the availability of sample i in view v, and $\otimes$ denotes the element-wise product.

To enforce cross-view structural consistency, the consistency mask $\mathbf{W}^*$ is defined as shown in Equation (3).

$$\mathbf{W}_{ij}^{*} = \prod_{v=1}^{V} \mathbb{I}\big(j \in \mathcal{N}_i^{(v)}\big), \tag{3}$$

where $\mathbb{I}(\cdot)$ is the indicator function, and $\mathcal{N}_i^{(v)}$ denotes the k-nearest neighbors of node i in view v.

B. Incomplete Multi-View Data Completion

To complete missing features and recover structural information, we employ a GNN-based propagation mechanism guided by the global similarity matrix $\mathbf{S}$.

For a sample $\mathbf{X}_p^{(v)}$ missing in view v, its representation $\mathbf{H}_p^{(v)}$ is inferred by aggregating neighborhood information in a similarity-weighted manner, as shown in Equation (4).

$$\mathbf{H}_p^{(v)} = \sigma\left(b + \sum_{q\in\mathcal{N}_p} \frac{\mathbf{S}_{pq}}{\sum_{q'\in\mathcal{N}_p} \mathbf{S}_{pq'}} \mathbf{M}_{qv}\mathbf{X}_q^{(v)}\right), \tag{4}$$

where $\sigma(\cdot)$ is the nonlinear activation function, b is the bias term.

For an observed sample, a self-loop term is added to preserve the original features, as shown in Equation (5).

$$\mathbf{H}_p^{(v)} = \sigma\left(b + \sum_{q\in\mathcal{N}_p} \frac{\mathbf{S}_{pq}}{\sum_{q'\in\mathcal{N}_p} \mathbf{S}_{pq'}} \mathbf{M}_{qv}\mathbf{X}_q^{(v)} + \mathbf{X}_p^{(v)}\right). \tag{5}$$

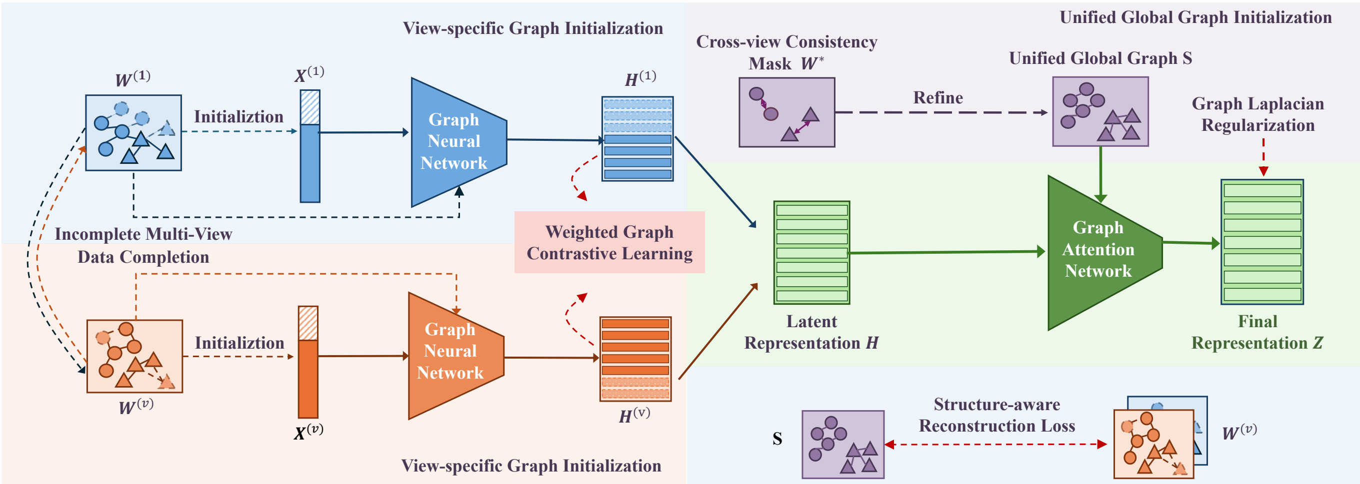

Fig. 2. Overview of the proposed SIGMA framework. SIGMA integrates view-specific graph initialization, unified global graph initialization, incomplete multi-view graph completion, and weighted graph contrastive learning to learn unified representations for clustering.

Both cases are unified into a single expression, as shown in Equation (6).

$$\mathbf{H}^{(v)} = \sigma\left(b + \operatorname{diag}(\mathbf{M}_{\cdot v})(\hat{\mathbf{S}} + \mathbf{I})\mathbf{X}^{(v)}\right), \tag{6}$$

where $\hat{\mathbf{S}}$ denotes the row-normalized similarity matrix.

The representations from all views are concatenated, as shown in Equation (7).

$$\mathbf{H} = [\mathbf{H}^{(1)}|\mathbf{H}^{(2)}|\cdots|\mathbf{H}^{(V)}], \tag{7}$$

and fed into a graph attention network to obtain the final embedding $\mathbf{Z}$.

C. Weighted Graph Contrastive Learning

To align view-specific representations with the global structure under heterogeneous missing patterns, we propose a weighted graph contrastive learning framework, where view-wise reliability and sample uncertainty are incorporated via a weight term $\omega_i^{(v)}$.

a) Weighted Intra-view Contrastive Loss

Given the latent representation $\mathbf{H}^{(v)}$ of view v, positive pairs for each sample are defined by its high-similarity neighbors, while dissimilar samples are treated as negatives. By incorporating the weight term $\omega_i^{(v)}$, the loss assigns different importance to samples according to their observation reliability, thereby preserving intra-view local structures while mitigating the influence of noisy or partially observed samples.

The sample-specific weight is defined as shown in Equation (8).

$$\omega_i^{(v)} = \frac{1}{2\|\mathbf{S}_i - \mathbf{W}_i^{(v)}\|_2}, \tag{8}$$

where $\mathbf{S}_i$ and $\mathbf{W}_i^{(v)}$ denote the global and view-specific similarity vectors of sample i.

The weighted intra-view contrastive loss is formulated as shown in Equation (9).

$$\begin{aligned}\mathcal{L}_{\text{view-contrast}} &= -\frac{1}{V}\sum_{v=1}^{V}\frac{1}{N}\sum_{i=1}^{N}\omega_i^{(v)}\log\left(\frac{P_i^{(v)}}{Q_i^{(v)}}\right),\\ P_i^{(v)} &= \sum_{j\in\mathcal{N}_i^{(v)}}\mathbf{S}_{ij}\, d\left(\mathbf{H}_i^{(v)},\mathbf{H}_j^{(v)}\right),\\ Q_i^{(v)} &= \sum_{j=1}^{N}\mathbf{W}_{ij}^{(v)}(1-\mathbf{S}_{ij})\, d\left(\mathbf{H}_i^{(v)},\mathbf{H}_j^{(v)}\right),\end{aligned} \tag{9}$$

where $d(\cdot,\cdot) = \exp(\operatorname{sim}(\cdot,\cdot)/\tau)$ and $\operatorname{sim}(\cdot,\cdot)$ denotes cosine similarity, and τ controls the sharpness of similarity scores.

b) Structure-aware Reconstruction Loss

To align view-wise similarity matrices $\mathbf{W}^{(v)}$ with the global structure $\mathbf{S}$, we impose a weighted reconstruction constraint, as shown in Equation (10).

$$\begin{aligned}\mathcal{L}_{\text{rec}} &= \sum_{v=1}^{V}\sum_{i=1}^{N}\omega_i^{(v)}\left\|\mathbf{S}_i - \mathbf{W}_i^{(v)}\right\|_2^2\\ &= \sum_{v=1}^{V}\operatorname{tr}\left((\mathbf{S}-\mathbf{W}^{(v)})^T\operatorname{diag}(\boldsymbol{\omega}^{(v)})(\mathbf{S}-\mathbf{W}^{(v)})\right).\end{aligned} \tag{10}$$

This loss encourages $\mathbf{S}$ to maintain a balanced representation across multiple views, thereby avoiding overfitting to any single view and mitigating the influence of noise or missing features.

c) Graph Laplacian Regularization

To encourage structural smoothness, we add a Laplacian regularization term based on the global similarity graph and the cross-view consistency graph. The graph Laplacian regularization loss is formulated, as shown in Equation (11).

$$\mathcal{L}_{\text{reg}} = \operatorname{tr}(\mathbf{Z}^T\mathbf{L}_S\mathbf{Z}) + \operatorname{tr}(\mathbf{Z}^T\mathbf{L}_{W^*}\mathbf{Z}), \tag{11}$$

where $\mathbf{L}_S$ is the Laplacian of the global similarity graph $\mathbf{S}$, and $\mathbf{L}_{W^*}$ corresponds to the Laplacian of the cross-view consistency graph.

Finally, the overall objective function of SIGMA is defined, as shown in Equation (12).

$$\mathcal{L} = \mathcal{L}_{\text{graph-contrast}} + \alpha\mathcal{L}_{\text{rec}} + \beta\mathcal{L}_{\text{reg}}, \tag{12}$$

where α and β are trade-off hyperparameters that control the relative importance of the reconstruction and regularization losses.

III. Experiments

A. Experimental Settings

1) Datasets

To evaluate SIGMA on four widely used benchmark datasets for incomplete multi-view clustering, including BBC4 [10], ORL [11], YALE [12], and BDGP [13]. To simulate incomplete multi-view settings, we randomly remove views from the original complete data according to a matching ratio $p \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, where p denotes the proportion of instances with all views observed. For example, $p = 0.1$ indicates that only 10% of the instances are fully observed, while the remaining samples contain randomly missing views.

2) Baselines

We compare SIGMA with seven state-of-the-art incomplete multi-view clustering methods, including CONCAT [14], UEAF [15], CBG [16], BSV [17], BGIMVC [18], SAGF_IMC [19], and GIMVC [20]. Clustering performance is evaluated using Accuracy (ACC), Normalized Mutual Information (NMI), and Adjusted Rand Index (ARI), averaged over five runs. All methods are trained using Adam with an initial learning rate of 10^{-4}. A ReduceLROnPlateau scheduler is applied with a decay factor of 0.5 and patience of 10 epochs, and the minimum learning rate is set to 10^{-6}. The maximum number of training epochs is fixed to 1,000 for all datasets.

B. Results

We evaluate SIGMA against seven baseline methods on four benchmark datasets. Tables I and II report the clustering performance in terms of ACC, NMI, and ARI under different matching ratios p. Overall, SIGMA achieves the best or highly competitive results across most datasets and missing-view settings. In particular, SIGMA consistently outperforms all baselines on ORL and YALE, demonstrating its effectiveness on image-centric multi-view data. On BBC4 and BDGP, SIGMA remains competitive under low matching ratios and achieves superior performance as p increases, indicating strong robustness to severe view incompleteness.

C. Ablation Study

We conduct ablation studies on the four selected datasets under different matching ratios p, and report ACC under various configurations.

1) Effectiveness of Update Strategies

We evaluate four global similarity update strategies:

- **Static S (SIGMA-1):** **S** is fixed during training.
- **Full Update (SIGMA-2):** **S**, GNN, and GAT are updated jointly.
- **Two-Phase With S Later (SIGMA-3):** **S** is updated after training GNN and GAT.
- **Two-Phase Only S Later (SIGMA-4):** Only **S** is optimized in the second phase.

As shown in Table III, the two-phase strategy achieves the best performance, especially under high matching ratios. SIGMA-1 performs worst across all p, confirming the necessity of dynamic graph alignment. SIGMA-2 and SIGMA-3 show similar results, indicating that updating the global graph during missing-view imputation introduces bias.

2) Effectiveness of Loss Functions

We conduct loss ablation experiments to evaluate the contribution of each term in SIGMA, as shown in Table IV. We compare three variants using only the weighted contrastive loss ($\mathcal{L}_{\text{graph-contrast}}$), the reconstruction loss ($\mathcal{L}_{\text{rec}}$), or the regularization loss ($\mathcal{L}_{\text{reg}}$).

Removing any loss term results in performance degradation, while the full objective consistently achieves the best results across all matching ratios p. When p is high, the weighted contrastive loss alone performs competitively, indicating its effectiveness under reliable cross-view observations. However, under severe incompleteness ($p = 0.1$), contrastive learning alone becomes insufficient, highlighting the necessity of complementary loss terms.

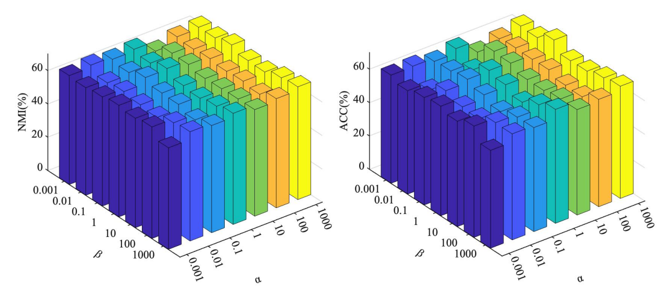

Fig. 3. The performance of NMI (left) and ACC (right) under different α and β on the YALE dataset.

D. Sensitivity Analysis

We analyze the sensitivity of SIGMA to the hyperparameters α and β, which control the reconstruction and regularization terms. Figure 3 shows a 2D grid of results with $\alpha, \beta \in \{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$, enabling a systematic evaluation of their influence on clustering performance.

As α increases from small values, ACC and NMI improve, indicating that emphasizing reconstruction helps preserve discriminative information from incomplete views. However, overly large α values lead to performance degradation, as excessive reconstruction may amplify noise introduced by missing or unreliable observations. Similarly, extreme β values cause instability by either weakening or over-constraining the structural consistency regularization, resulting in fluctuating performance across different parameter combinations. In contrast, moderate β values provide more stable and robust clustering results.

Overall, $\alpha, \beta \in \{0.1, 1, 10\}$ offer a favorable trade-off between reconstruction accuracy and structural consistency, and are therefore adopted in our experiments.

TABLE I
COMPARISON EXPERIMENTS ON FOUR DATASETS UNDER DIFFERENT MATCHING RATIOS p IN TERMS OF NMI AND ACC.

Dataset	**Method**	**NMI**					**ACC**				
		p=0.1	p=0.3	p=0.5	p=0.7	p=0.9	p=0.1	p=0.3	p=0.5	p=0.7	p=0.9
BBC4	BSV	0.0073	0.0084	0.0372	0.0592	0.0984	0.3342	0.3345	0.3514	0.3743	0.4126
	CONCAT	0.0086	0.0068	0.0074	0.0368	0.0510	0.3350	0.3302	0.3315	0.3537	0.3642
	UEAF	0.0704	0.1717	0.3639	0.4872	0.5470	0.3168	0.3804	0.5559	0.6639	0.7226
	CBG	0.0191	0.0154	0.0379	0.0181	0.0799	0.3336	0.3331	0.3447	0.3343	0.3781
	SAGF_IMC	0.2941	0.3624	0.4629	0.5050	0.4719	0.5661	0.6234	**0.7022**	0.7066	0.6964
	BGIMVC	0.2244	**0.4535**	0.4603	0.4333	0.4639	0.4035	**0.6528**	0.6993	0.6336	0.6219
	GIMVC	0.0302	0.0347	0.0693	0.1109	0.1584	0.2971	0.2901	0.3472	0.4060	0.4635
	SIGMA	**0.3086**	0.3734	**0.6144**	**0.6080**	**0.6219**	**0.5735**	0.6313	0.6964	**0.7067**	**0.7480**
ORL	BSV	0.3389	0.4191	0.4816	0.5618	0.6682	0.2628	0.3283	0.3700	0.4370	0.5045
	CONCAT	0.2115	0.3145	0.3989	0.5355	0.6587	0.1800	0.2508	0.2950	0.3910	0.4875
	UEAF	0.2306	0.3588	0.4628	0.5725	0.6998	0.1825	0.2825	0.3540	0.4290	0.5145
	CBG	0.3638	0.4967	0.5595	0.6303	0.7474	0.2063	0.3360	0.3780	0.4190	0.5510
	SAGF_IMC	0.4337	0.6358	0.6621	0.6812	0.6582	0.2855	0.4630	0.5030	0.4850	0.4825
	BGIMVC	0.5273	0.5613	0.5580	0.5774	0.5830	0.3268	0.3628	0.3515	0.3563	0.3538
	GIMVC	0.6647	0.7845	0.8314	0.8502	0.8501	**0.4975**	0.6317	0.6913	0.7163	0.7087
	SIGMA	**0.6694**	**0.7882**	**0.8529**	**0.8507**	**0.8611**	0.4605	**0.6472**	**0.7430**	**0.7265**	**0.7417**
YALE	BSV	0.2353	0.2643	0.3055	0.3909	0.5277	0.2527	0.2752	0.2933	0.3612	0.4733
	CONCAT	0.1407	0.1882	0.2318	0.2947	0.4650	0.1648	0.2073	0.2455	0.2709	0.4091
	UEAF	0.1966	0.2876	0.3120	0.5449	0.5567	0.2073	0.2945	0.2848	0.4958	0.5079
	CBG	0.2874	0.2846	0.3893	0.5243	0.4243	0.2491	0.2533	0.3339	0.4691	0.3703
	SAGF_IMC	0.2945	0.4653	0.5027	0.5534	0.5807	0.2267	0.4291	0.4521	0.5055	0.5285
	BGIMVC	0.3194	0.3015	0.4094	0.3522	0.4218	0.2673	0.2612	0.3521	0.2873	0.3539
	GIMVC	0.4087	0.5403	0.6520	0.6576	0.6435	0.3600	0.4988	0.6115	0.6127	0.6048
	SIGMA	**0.6150**	**0.6392**	**0.6684**	**0.6647**	**0.6548**	**0.6133**	**0.6321**	**0.6521**	**0.6130**	**0.6370**
BDGP	BSV	0.2212	0.2642	0.2868	0.2951	0.3656	0.3871	0.4189	0.4336	0.4440	0.5115
	CONCAT	0.2110	0.2679	0.2805	0.3378	0.3868	0.3634	0.4054	0.4096	0.4672	0.5322
	UEAF	0.1052	0.5151	0.5992	0.6842	0.7668	0.3128	0.7475	0.8038	0.8612	0.9122
	CBG	0.3631	0.4519	0.5155	0.5292	0.5668	0.5553	0.6448	0.7108	0.7416	0.7948
	SAGF_IMC	**0.5606**	0.6140	0.7016	0.7290	0.8107	**0.7072**	0.7256	0.8700	0.8836	0.9332
	BGIMVC	0.0201	0.0279	0.1739	0.0882	0.2281	0.2200	0.2264	0.3428	0.2816	0.3720
	GIMVC	0.5405	0.5934	0.5871	0.5877	0.6712	0.6858	0.6813	0.6787	0.6903	0.7517
	SIGMA	0.3895	**0.6913**	**0.8124**	**0.8690**	**0.9172**	0.5178	**0.8279**	**0.8952**	**0.9204**	**0.9677**

E. Visualization

We perform t-SNE visualization on the BDGP dataset with a matching ratio of $p = 0.5$. Figure 4 shows three stages: raw features, intermediate embeddings after 50 epochs, and final representations. Initially, samples are highly mixed due to incomplete and heterogeneous observations. As training progresses, SIGMA gradually complements missing information and enhances alignment across views, leading to improved intra-cluster cohesion. In the final stage, clusters become compact and well separated, indicating that the learned representations are more discriminative. These results qualitatively demonstrate SIGMA's effectiveness under incomplete multi-view settings, which is consistent with the quantitative clustering results reported above.

IV. Conclusion and Discussion

In this paper, we proposed SIGMA to address incomplete multi-view clustering under inconsistent missing patterns. By learning a globally consistent structure across heterogeneous views, SIGMA effectively infers missing information and integrates partial observations. Experiments on multiple public benchmark datasets demonstrate that SIGMA consistently outperforms state-of-the-art IMC methods across diverse missing-view settings. Through structure-aware graph optimization, SIGMA balances local discriminability and global structural consistency. Future work will extend SIGMA to dynamic and large-scale settings with evolving view availability.

References

[1] G. Chao, Y. Jiang, and D. Chu, "Incomplete contrastive multi-view clustering with high-confidence guiding," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, pp. 11 221–11 229.

[2] J. Pu, C. Cui, X. Chen, Y. Ren, X. Pu, Z. Hao, P. S. Yu, and L. He, "Adaptive feature imputation with latent graph for deep incomplete multi-view clustering," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, pp. 14 633–14 641.

[3] C. Liu, Z. Wu, J. Wen, Y. Xu, and C. Huang, "Localized sparse incomplete multi-view clustering," *IEEE Transactions on Multimedia*, vol. 25, pp. 5539–5551, 2023.

TABLE II
COMPARISON EXPERIMENTS ON FOUR DATASETS UNDER DIFFERENT MATCHING RATIOS p IN TERMS OF ARI.

Dataset	Method	ARI				
		p=0.1	p=0.3	p=0.5	p=0.7	p=0.9
BBC4	BSV	0.0012	0.0015	0.0030	0.0204	0.0386
	CONCAT	0.0009	0.0004	0.0002	0.0141	0.0253
	UEAF	0.0040	0.0474	0.2847	0.4882	0.5382
	CBG	0.0031	0.0002	0.0047	0.0033	0.0626
	SAGF_IMC	0.2720	0.3099	0.4629	0.5051	0.4766
	BGIMVC	0.1563	0.4531	0.4640	0.4362	0.4587
	GIMVC	0.0258	0.0217	0.0703	0.1171	0.1765
	SIGMA	**0.2954**	**0.3685**	**0.5603**	**0.6186**	**0.6219**
ORL	BSV	0.0177	0.0338	0.0610	0.1115	0.2495
	CONCAT	0.0026	0.0093	0.0251	0.0902	0.2507
	UEAF	0.0028	0.0144	0.0403	0.1191	0.3352
	CBG	0.0122	0.0346	0.0670	0.1396	0.3341
	SAGF_IMC	0.0497	0.2071	0.2371	0.2596	0.2381
	BGIMVC	0.1203	0.1552	0.1571	0.1618	0.1553
	GIMVC	0.2417	0.4579	0.5500	0.5869	0.5743
	SIGMA	**0.2907**	**0.4839**	**0.6282**	**0.5992**	**0.6226**
YALE	BSV	0.0202	0.0291	0.0532	0.1126	0.2561
	CONCAT	0.0020	0.0066	0.0160	0.0502	0.1745
	UEAF	0.0062	0.0236	0.0362	0.2826	0.3105
	CBG	0.0206	0.0209	0.0848	0.1913	0.1119
	SAGF_IMC	0.0447	0.1527	0.2237	0.3027	0.3464
	BGIMVC	0.0626	0.0692	0.1460	0.1229	0.1778
	GIMVC	0.0987	0.2496	0.4048	0.4286	**0.4052**
	SIGMA	**0.3581**	**0.4029**	**0.4261**	**0.4347**	0.3825
BDGP	BSV	0.0571	0.0801	0.0988	0.1054	0.1632
	CONCAT	0.0571	0.0855	0.1019	0.1394	0.1782
	UEAF	0.0207	0.4387	0.5471	0.6721	0.7920
	CBG	0.1498	0.2536	0.3617	0.4378	0.5579
	SAGF_IMC	**0.5448**	0.5970	0.7115	0.7385	0.8416
	BGIMVC	0.0005	0.0017	0.0893	0.0239	0.1410
	GIMVC	0.5019	0.5387	0.5330	0.5533	0.6339
	SIGMA	0.2594	**0.6553**	**0.7867**	**0.8334**	**0.9228**

TABLE III
ABLATION STUDY OF GLOBAL SIMILARITY UPDATE STRATEGIES ON THE ORL DATASET.

Method	p=0.1	p=0.3	p=0.5	p=0.7	p=0.9
SIGMA-1	0.3305	0.4282	0.4825	0.5045	0.5538
SIGMA-2	0.3505	0.4605	0.5092	0.5160	0.5678
SIGMA-3	0.3505	0.4605	0.5092	0.5160	0.5678
SIGMA-4	**0.4605**	**0.6472**	**0.7430**	**0.7265**	**0.7417**

TABLE IV
ABLATION STUDY OF LOSS FUNCTION ON THE ORL DATASET.

Loss Function	p=0.1	p=0.3	p=0.5	p=0.7	p=0.9
$\mathcal{L}_{\text{graph-contrast}}$	0.3345	0.4010	0.4977	0.4990	0.5195
$\mathcal{L}_{\text{Reg}}$	0.3278	0.4018	0.5418	0.5112	0.5068
$\mathcal{L}_{\text{Rec}}$	0.3125	0.4018	0.5052	0.4990	0.5125
$\mathcal{L}$ (Full Loss)	**0.4605**	**0.6460**	**0.7358**	**0.7245**	**0.7432**

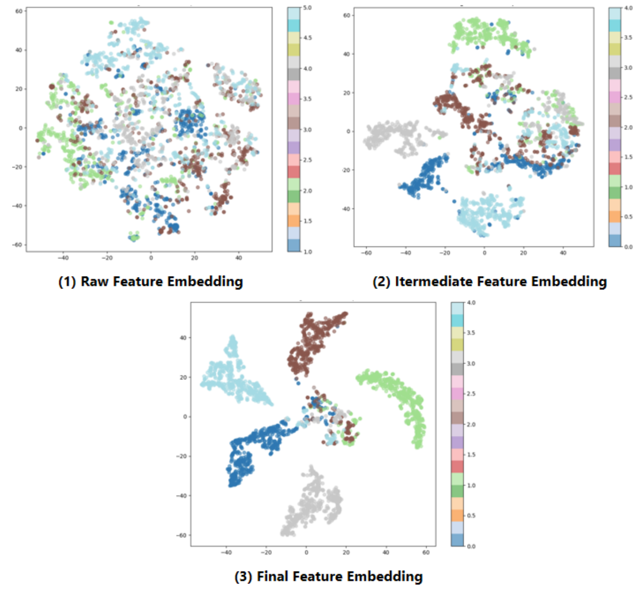

Fig. 4. Visualization of the raw features, intermediate features and final results on the BDGP.

[4] H. Wang, M. Yao, Y. Chen, Y. Xu, H. Liu, W. Jia, X. Fu, and Y. Wang, "Manifold-based incomplete multi-view clustering via bi-consistency guidance," *IEEE Transactions on Multimedia*, vol. 26, pp. 10 001–10 014, 2024.

[5] R. Wang, L. Li, X. Tao, P. Wang, and P. Liu, "Contrastive and attentive graph learning for multi-view clustering," *Information Processing & Management*, vol. 59, no. 4, p. 102967, 2022.

[6] J. Xu, H. Tang, Y. Ren, L. Peng, X. Zhu, and L. He, "Multi-level feature learning for contrastive multi-view clustering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16 051–16 060.

[7] Y. Yu, J. Wang, W. Yu, Z. Zhao, Z. Miao, and F. Nie, "Bidirectional fusion for deep contrastive multi-view clustering," *Expert Systems with Applications*, vol. 287, p. 128193, 2025.

[8] Z. Xu, C. Lang, L. Wei, L. Liang, T. Wang, Y. Li, and M. C. Kampffmeyer, "Unagi: Unified neighbor-aware graph neural network for multi-view clustering," *Neural Networks*, vol. 185, p. 107193, 2025.

[9] L. Wang and C. Zhu, "Global semantic space feature fusion for multi-view clustering," *Multimedia Systems*, vol. 31, no. 3, pp. 1–13, 2025.

[10] S. Huang, Y. Liu, Y. Ren, I. W. Tsang, Z. Xu, and J. Lv, "Learning smooth representation for multi-view subspace clustering," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, p. 3421–3429.

[11] Y. Tan, Y. Liu, S. Huang, W. Feng, and J. Lv, "Sample-level multi-view graph clustering," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 23 966–23 975.

[12] Y. Jia, G. Lu, H. Liu, and J. Hou, "Semi-supervised subspace clustering via tensor low-rank representation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 7, pp. 3455–3461, 2023.

[13] W. Feng, G. Sheng, Q. Wang, Q. Gao, Z. Tao, and B. Dong, "Partial multi-view clustering via self-supervised network," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, pp. 11 988–11 995.

[14] Q. Zheng, J. Zhu, Z. Li, S. Pang, J. Wang, and Y. Li, "Feature concatenation multi-view subspace clustering," *Neurocomputing*, vol. 379, pp. 89–102, 2020.

[15] J. Wen, Z. Zhang, Y. Xu, B. Zhang, L. Fei, and H. Liu, "Unified embedding alignment with missing views inferring for incomplete multi-view clustering," in *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*, 2019, pp. 5393–5400.

[16] S. Wang, X. Liu, L. Liu, W. Tu, X. Zhu, J. Liu, S. Zhou, and E. Zhu, "Highly-efficient incomplete large-scale multiview clustering with consensus bipartite graph," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 9766–9775.

[17] Q. Shen, X. Zhang, S. Wang, Y. Li, Y. Liang, and Y. Chen, "Dual completion learning for incomplete multi-view clustering," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 9, no. 1, pp. 455–467, 2025.

[18] Z. Chen, H. Zhang, Z. Liu, H. Zheng, and L. Zhao, "Beyond missing data: A multi-scale graph fusion framework for sustainable development insights," *Sustainability*, vol. 17, no. 3, p. 1136, 2025.

[19] N. Liang, Z. Yang, and S. Xie, "Incomplete multi-view clustering with sample-level auto-weighted graph fusion," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 6, pp. 6504–6511, 2023.

[20] S. Bai, Q. Zheng, X. Ren, and J. Zhu, "Graph-guided imputation-free incomplete multi-view clustering," *Expert Systems with Applications*, vol. 258, p. 125165, 2024.