Attribute-Missing Multi-view Graph Clustering

Bowen Zhao[1], Qianqian Wang[1,*], Zhengming Ding[2], Quanxue Gao[1]
[1]Xidian University, [2]Tulane University
bw_zhao@stu.xidian.edu.cn, qqwang@xidian.edu.cn, zding1@tulane.edu, qxgao@xidian.edu.cn

Abstract

*The success of existing deep multi-view graph clustering methods is based on the assumption that node attributes are fully available across all views. However, in practical scenarios, node attributes are frequently missing due to factors such as data privacy concerns or failures in data collection devices. Although some methods have been proposed to address the issue of missing node attributes, they come with the following limitations: i) Existing methods are often not tailored specifically for clustering tasks and struggle to address missing attributes effectively. ii) They tend to ignore the relational dependencies between nodes and their neighboring nodes. This oversight results in unreliable imputations, thereby degrading clustering performance. To address the above issues, we propose an **A**ttribute-**M**issing **M**ulti-view **G**raph **C**lustering (**AMMGC**). Specifically, we first impute missing node attributes by leveraging neighborhood information through an adjacency matrix. Then, to improve the consistency, we integrate a dual structure consistency module that aligns graph structures across multiple views, reducing redundancy and retaining key information. Furthermore, we introduce a high-confidence guidance module to improve the reliability of clustering. Extensive experiment results showcase the effectiveness and superiority of our proposed method on multiple benchmark datasets.*

1. Introduction

As a fundamental data type, graphs play a crucial role in representing complex relational data across various domains. With the rapid advancement of information technology, graph data is evolving in a more diversified manner. Taking the social network as an example, one graph view could capture the friendship connections between users, while another view could represent the frequency of interactions between users, such as comments, likes, and shares. Compared to single-view graphs, multi-view graphs can better represent the real graph data. As an important application of multi-view graphs, multi-view graph clustering (MVGC) [16–19, 34, 42] partitions data into several disjoint groups by exploiting consistent and complementary information across multiple views.

The existing MVGC methods can be divided into the following three categories: (1) GCNs-based methods [3, 19]. These methods employ GCNs to extract node representations from each view, which are then integrated to obtain a unified representation. (2) Contrastive-based methods [1, 6, 23]. They utilize contrastive learning to bring the representations of similar nodes closer while pushing apart those of dissimilar nodes, further enhancing the consistency of representations across different views. (3) Distribution alignment-based methods [5, 19]. They use KL divergence to align the representations across different views. Although these methods show promising results, existing MVGC methods typically assume complete data availability across all views. However, in real-world scenarios, this assumption often fails due to issues like data privacy concerns or data loss caused by damaged storage media, resulting in some node views being only partially accessible. These uncontrollable factors can readily lead to data sparsity and data absence issues, which in turn negatively impact the effectiveness of clustering performance. Graphs without node attributes can be broadly categorized into two types: (1) attribute-incomplete graphs, where a portion of attributes is missing for every node. (2) attribute-missing graphs, where certain nodes have lack all attribute information. In this work, we focus on the second type, as it presents greater challenges and more accurately reflects real-world situations. Current MVGC methods don't incorporate a specific attribute completion mechanism to address nodes with missing attributes. This limitation presents a major challenge in learning robust graph embeddings for accurate clustering on graphs with missing attributes.

To learn effective graph embeddings, many methods have been proposed to address the issue of missing attributes [2, 7, 8, 26, 41]. These methods can generally be divided into two categories. The first category is based on node embedding alignment, where an encoder is used

*Corresponding authors.

to separately encode node attributes and structural information, and then performs distribution matching in the latent space to reconstruct the missing attributes [2, 13]. The second category is based on data imputation methods, which tackle the challenge by filling in missing attributes through data imputation techniques [27, 40]. Though these methods have achieved promising results, they have the following drawbacks: 1) They are typically designed for single-view settings and are not tailored for clustering tasks. 2) These methods overlook the relationships between nodes and their neighbors, leading to unreliable imputation.

The above observations indicate that attribute-missing MVGC has become an urgent need in practical applications while remaining under-explored. This finding motivates us to propose an effective MVGC framework called Attribute-Missing Multi-view Graph Clustering (AMMGC). Specifically, we first design the context-aware imputation module to iteratively refine missing node attributes by incorporating neighborhood information. We then apply a Laplacian filter to the node attributes to reduce the high-frequency components. To further enhance both consistency and discriminative capacity, we propose a dual structure consistency module that aligns the node representation from multiple views with graph structure. This alignment reduces redundant information while preserving critical semantic information. Finally, we introduce the high-confidence guidance module to supervise the learning process, thereby improving clustering reliability. The main contributions of this paper are summarized as follows:

- We introduce a novel framework named AMMGC, specifically designed to tackle the challenge of attribute-missing MVGC. This approach iteratively refines the missing node features by effectively incorporating neighborhood information, ensuring enhanced attribute recovery and improved clustering performance.
- To enhance clustering consistency and boost discriminative capacity, we develop a dual-structure module that aims to reduce redundancy while retaining critical information.
- Comprehensive experiments on various benchmark datasets reveal that the proposed AMMGC framework consistently surpasses existing state-of-the-art clustering methods, especially in cases where a substantial amount of node attributes are missing.

2. Related Work

This section briefly describes the latest research progress of Deep Multi-view Graph Clustering and Clustering with Missing Attribute Data.

2.1. Deep Multi-view Graph Clustering

With the rapid development of GNNs, many deep multi-view graph clustering methods have been proposed [3, 17, 20, 23]. These methods leverage GNNs to explore attribute information and graph structural information simultaneously. O2MGC [5] introduces a one2multi graph autoencoder to learn node embeddings. MAGCN [3] designs a two-pathway encoder structure, effectively capturing graph embedding features while learning view-consistency information. MVGC [36] leverages the learned clustering labels by a subspace clustering module to self-supervise the learning process for node representations and the view-consensus coefficient matrix. Due to its outstanding representation learning capabilities, many graph contrastive methods are proposed to extract information from graphs. MCGC [23] learns a consensus graph regularized by graph contrastive loss. SCAGC [35] utilizes pseudo labels to guide contrastive learning. CCGC [38] employs high-confidence clustering labels to construct reliable positive and negative sample pairs. Although these methods achieve strong performance, they rely on the assumption that node attributes in multi-view graphs are complete. Therefore, their performance might decrease when node attributes are missing.

2.2. Clustering with Missing Attribute Data

Recently, the challenge of clustering with missing attribute data has attracted growing interest [14, 15, 31, 37]. For multi-view data with missing attributes, existing deep incomplete multi-view clustering methods can be broadly divided into three main categories: (1) GANs-based methods. These methods utilize generative adversarial networks (GANs) to explore shared representations across multiple views, directly generating the missing data [31, 32]. (2) GCNs-based methods. GCNs are employed to extract common representations by leveraging the structural information of different views [25, 33]. (3) Prototype-based methods. These methods learn prototypes from the missing view and utilize sample-prototype relationships in the observed views to reconstruct the data [9, 12].

For graph data with missing attributes, SAT [2] restores missing attributes through distribution matching and adversarial learning strategies. T2-GNN [7] proposes using teacher-student to enhance the performance of GNNs on incomplete graphs. ITR [27] utilizes structural information for the initial imputation phase, followed by a refinement process that incorporates both observed attributes and structural data to enhance the imputed latent variables. AMGC [28] integrates clustering and data imputation within a unified framework, where the two processes support each other through iterative optimization. RITR [29] initializes missing and incomplete data with noise or structure embeddings, and then refines these embeddings through a consistency-preserving mechanism. MATE [24] performs imputation directly in the input space using graph diffusion and parameter initialization. However, these methods are either de-

signed for single-view settings, not specifically tailored for clustering tasks, or tend to overlook the relationships between nodes and their neighbors.

3. The Proposed Method

In this section, we propose a novel Attribute-Missing Multi-view Graph Clustering (AMMGC). The overall framework is illustrated in Fig. 1. AMMGC mainly consists of three key modules: context-aware imputation, dual structure consistency, and high-confidence guidance. Detailed explanations of each module will be elaborated upon in the subsequent sections.

3.1. Notations and Problem Definition

Notations: Given an undirected multi-view graph data $\mathcal{G} = \{\mathbf{X}^v \in \mathbb{R}^{N\times D}, \mathbf{A} \in \mathbb{R}^{N\times N}\}_{v=1}^2$. Let $\mathcal{V} = \{v_1, v_2, ..., v_N\}$ be a set of N nodes with K classes and $\mathcal{E}$ be a set of edges. $\mathbf{X}^v$ denote the attribute matrix of all nodes for the v-th view, where D is the dimension of attributes. $\mathbf{A} \in \{0,1\}^{N\times N}$ denote the adjacency matrix, where $a_{ij} = 1$ if $(v_i, v_j) \in \mathcal{E}$, otherwise $a_{ij} = 0$. With the renormalization trick $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ in GCN [11], where $\mathbf{I}_N$ is the N-order identity matrix. The normalized adjacency matrix is denoted as $\tilde{\mathbf{A}} = \hat{\mathbf{D}}^{-1/2}\hat{\mathbf{A}}\hat{\mathbf{D}}^{-1/2}$. $\hat{\mathbf{D}} = diag(\hat{d}_1, \hat{d}_2, ..., \hat{d}_N) \in \mathbb{R}^{N\times N}$ is degree matrix and $\hat{d}_{ii} = \sum_{j=1}^{N} \hat{a}_{ij}$. The symmetric normalized graph Laplacian matrix is defined as $\tilde{\mathbf{L}} = \mathbf{I}_N - \tilde{\mathbf{A}}$.
Problem Statement: Clustering on an attribute-missing multi-view graph refers to the task of grouping nodes in multi-view graphs where some node attributes are missing. Here, we concentrate on cases where node attributes are either completely observed or entirely missing. The objective of attribute-missing multi-view graph clustering is to partition a multi-view graph with N unlabeled nodes into K disjoint clusters $\{\mathcal{C}_1, ..., \mathcal{C}_K\}$.

3.2. Context-Aware Imputation

An intuitive imputation method involves one-time imputation strategies, such as zero-filling, where the missing data is imputed once and then directly utilized for subsequent clustering tasks. These strategies introduce extra noises and thus degrade the clustering performance. In our work, we adopt a novel context-aware imputation strategy to solve this problem. We first construct different views by adding the random Gaussian noise to $\mathbf{X}^1$ as formulated:

$$\mathbf{X}^2 = \mathbf{X}^1 + \mathbf{N}, \tag{1}$$

where $\mathbf{N} \in \mathbb{R}^{N\times D}$ is sampled from the Gaussian distribution $\mathcal{N}(0, 0.01)$. Then, we initialize the attribute matrix as follows:

$$\mathbf{X}^v \leftarrow \mathbf{M}^v \odot \mathbf{X}^v, \tag{2}$$

where $\odot$ denotes the element-wise multiplication. $\mathbf{M}^v \in \{0,1\}^{N\times D}$ is missing feature mask matrices. When $\mathbf{M}_{ij}^v = 0$, it indicates that the j-th feature of the i-th node is observed in the v-th view. This operation preserves only the observed features in the attribute matrix $\mathbf{X}^v$, while setting the missing features to an initial value of zero. Next, we define the Dirichlet energy as

$$\ell(\mathbf{x}^v) = \frac{1}{2}\sum_{ij} \tilde{a}_{ij}(\mathbf{x}_i^v - \mathbf{x}_j^v)^2, \tag{3}$$

where $\tilde{a}_{ij}$ are the individual entries of the normalized adjacency matrix $\tilde{\mathbf{A}}$. The Dirichlet energy is widely used as a smoothness criterion, measuring the squared differences between the features of neighboring nodes and promoting feature similarity among adjacent nodes. We impute the missing node features by minimizing the energy function based on the known features.

Inspired by [26], we apply the Euler scheme to decrease the Dirichlet energy. Specifically, we perform iterative imputation of missing node features using the graph's adjacency matrix. The update rule for each iteration is defined as follows:

$$(\mathbf{X}^v)^{(t+1)} \leftarrow \mathbf{M}^v \odot \mathbf{X}^v + (1 - \mathbf{M}^v) \odot \tilde{\mathbf{A}} \cdot (\mathbf{X}^v)^{(t)}, \tag{4}$$

where the first term $\mathbf{M}^v \odot \mathbf{X}^v$ ensures that known features remain unchanged throughout the iterative process, while the second term imputes the missing features by propagating information from neighboring nodes via the normalized adjacency matrix $\tilde{\mathbf{A}}$.

The process is repeated for T iterations or until the energy function converges. After the final iteration, the reconstructed attribute matrix is denoted as:

$$\hat{\mathbf{X}}^v = (\mathbf{X}^v)^{(T)}. \tag{5}$$

At this point, the missing features have been inferred using the graph's structural information, while the original features are preserved. This iterative approach effectively utilizes the graph topology to reconstruct missing attributes, resulting in more reliable node representations.

3.3. Dual Structure Consistency

To effectively learn node representations in graph-based data, the majority of current methods heavily rely on graph convolutional network (GCN) encoders, which integrate node attributes with graph structure. However, the interaction between graph convolutional filters and weight matrices can degrade both the performance and robustness during the representation learning process [4].

Inspired by [38], we apply a widely-used Laplacian filter to reduce high-frequency noise in the node attributes, while integrating attribute information with graph structure. This process is formulated as follows:

$$\tilde{\mathbf{X}}^v = (\mathbf{I}_N - k\tilde{\mathbf{L}})^t \hat{\mathbf{X}}^v, \tag{6}$$

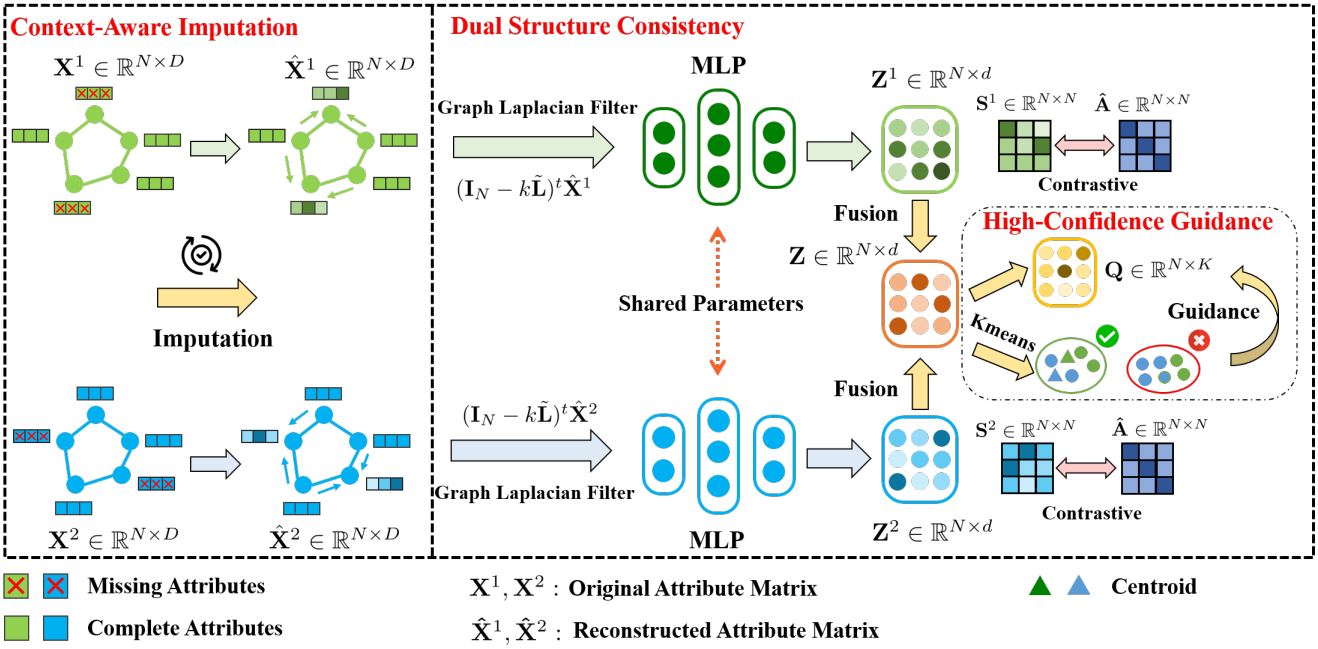

Figure 1. The framework of our proposed AMMGC. AMMGC consists of three modules: context-aware imputation module (CAI), dual structure consistency module (DSL), and high-confidence guidance module (HCG). Specifically, CAI iteratively imputes missing node features by utilizing information from neighboring nodes to ensure accurate estimations. DSL aligns the graph structures across multiple views to enhance both consistency and discriminative power. HCG enhances the reliability of clustering information by guiding the model's learning with high-confidence clustering data. Ultimately, we fuse the embeddings from different views and perform K-means to obtain final clustering results.

where k denotes a real-value parameter, and t specifies the number of filtering layers. $\tilde{\mathbf{L}}$ represents the symmetrically normalized graph Laplacian matrix. In our method, we set $k = 1$, transforming the filter into a GCN filter [4]. $\tilde{\mathbf{X}}^v$ indicates the resulting smoothed attribute matrix. Subsequently, node representations are derived by employing multi-layer perceptrons (MLPs) with shared parameters, which is expressed as:

$$\mathbf{Z}^v = \mathrm{MLP}(\tilde{\mathbf{X}}^v). \tag{7}$$

To maintain structural consistency across views, we design a dual structure consistency module. Concretely, we begin by integrating the node representations through an adaptive weight fusion mechanism, defined as follows:

$$\mathbf{Z} = \frac{\sum_{v=1}^{2} a_v \mathbf{Z}^v}{\sum_{v=1}^{2} a_v}, \tag{8}$$

where a_v is learnable parameters, and $\mathbf{Z}$ denotes the unified node representations. This mechanism dynamically learns the optimal weight for each view and assigns a higher weight to the view containing more informative features. Then, cross-view sample similarity matrices $\mathbf{S}^v$ by evaluating the similarity between $\mathbf{Z}$ and $\mathbf{Z}^v$, as formularized below:

$$\mathbf{S}^v = \frac{\mathbf{Z} \cdot (\mathbf{Z}^v)^\top}{\|\mathbf{Z}\| \, \|\mathbf{Z}^v\|}, \tag{9}$$

where $\mathbf{S}^v \in \mathbb{R}^{N \times N}$ is cross-view sample similarity matrix, and $\mathbf{s}_{ij}^v$ is the similarity between $\mathbf{z}_i$ and $\mathbf{z}_j^v$. We further enforce $\mathbf{S}^v$ to align with the adjacency matrix $\hat{\mathbf{A}}$ formulated as:

$$\mathcal{L}_s = \frac{1}{N^2} \sum_{v=1}^{2} (\mathbf{S}^v - \hat{\mathbf{A}})^2. \tag{10}$$

By minimizing $\mathcal{L}_s$, we ensure consistency between various views while preserving their unique semantic information.

3.4. High-Confidence Guidance

In this section, we introduce a high-confidence guidance module to obtain reliable clustering performance. Specifically, we begin by applying K-Means on fused $\mathbf{Z}$ to generate pseudo labels. Formally,

$$\hat{\mathbf{y}} = \mathrm{Kmeans}(\mathbf{Z}), \tag{11}$$

where $\hat{\mathbf{y}}$ represents the pseudo labels. We then select the most reliable pseudo-labels from $\hat{\mathbf{y}}$. Specifically, we begin by computing the distance of each sample to its corresponding cluster centroid. These distances are subsequently sorted in increasing order, and a threshold is applied to retain only those samples that are closer to their centroids,

while discarding those that are more distant. This process can be formally expressed as follows:

$$\mathbf{p} = \text{top}(\hat{\mathbf{y}}), \tag{12}$$

where $\mathbf{p}$ demotes reliable pseudo labels. Next, we define soft clustering assignment $\mathbf{q}$ as follows:

$$\mathbf{q}_{ik} = \frac{(1 + ||\mathbf{z}_i - \mathbf{c}_k||/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_{k' \in [K]}(1 + ||\mathbf{z}_i - \mathbf{c}_{k'}||/\alpha)^{-\frac{\alpha+1}{2}}}, \tag{13}$$

where $\mathbf{c}_k$ represents the cluster centroid, α is the degree of freedom of Student's t-distribution. In our experiment, we set α to 1. $\mathbf{q}_{ik}$ can be interpreted as the probability the i-th sample is assigned to k-th cluster. We align the derived soft clustering assignments with the reliable pseudo-labels to steer the network's learning process, as described below:

$$\mathcal{L}_p = \mathcal{H}(\mathbf{q}, \mathbf{p}) = -\sum_{i=1}^{N} \mathbf{q}_i \log(\mathbf{p}_i), \tag{14}$$

where $\mathcal{H}(\cdot)$ is cross-entropy loss. This loss function allows us to refine the quality of the soft clustering assignments, thereby bolstering the reliability of the clustering results.

3.5. The Objective Function

The object function of the proposed AMMGC contains the dual structure consistency loss $\mathcal{L}_s$ and the high-confidence guidance loss $\mathcal{L}_p$. In summary, the object function of AMMGC is formulated as follows:

$$\mathcal{L} = \mathcal{L}_s + \alpha \mathcal{L}_p, \tag{15}$$

where α is the trade-off parameters between $\mathcal{L}_s$ and $\mathcal{L}_p$.

To obtain the final clustering results, we directly perform the K-Means algorithm over $\mathbf{Z}$. Specifically, the fused representation $\mathbf{Z}$ is factorized as follows:

$$\begin{aligned} &\min_{\mathbf{U},\mathbf{V}} \ \|\mathbf{Z} - \mathbf{U}\mathbf{V}\|_{\mathrm{F}}^2 \\ &\text{s.t. } \mathbf{U}\mathbf{1} = \mathbf{1}, \ \mathbf{U} \geq \mathbf{0}, \end{aligned} \tag{16}$$

where $\mathbf{U} \in \mathbb{R}^{n \times k}$ is cluster indicator matrix, $\mathbf{V} \in \mathbb{R}^{k \times d}$ is the center matrix of clustering.

4. Experiments

In this section, we assess the performance of the proposed AMMGC across four commonly utilized graph datasets to validate its effectiveness and superiority. Additionally, we conduct an ablation study and a parameter sensitivity analysis to explore the characteristics and robustness of AMMGC.

Table 1. Statistical characteristics of four datasets.

Dataset	Type	Sample	Edge	Dimension	Clusters
CORA	Graph	2708	5429	1433	7
CITESEER	Graph	3327	4732	3703	6
AMAC	Graph	13752	245861	767	10
WIKI	Graph	2405	8261	4973	17

4.1. Experimental Setting

Datasets and Evaluation Metrics. We conduct experiments on four benchmark graph datasets, namely CORA [4], CITESEER [4], AMAC [23], WIKI [4]. The detailed characteristics are illustrated in Table 1. In this experiment, we set the missing rate of each dataset to [0.3, 0.6, 0.9]. To assess clustering performance, we utilize metrics such as Accuracy (ACC), Normalized Mutual Information (NMI), and Average Rand Index (ARI), with higher values indicating superior clustering results. To mitigate the effects of randomness, we independently run each method times and present the mean and standard deviation of these three metrics for each method.

Comparison Methods. To demonstrate the effectiveness of the proposed AMMGC, we conduct comparisons against state-of-the-art baseline methods, which include four deep graph clustering methods and three incomplete graph methods. **SAT** [2] restores missing attributes and improves graph tasks like link prediction and node completion through distribution matching and adversarial learning strategies. **FPGM** [26] utilizes feature propagation to recover the missing attributes of nodes. **T2-GNN** [7] proposes using teacher-student distillation to enhance the performance of GNNs on incomplete graphs. **CONVERT** [39] utilizes a reversible perturb-recover network to enhance the reliability of data augmentations. **CCGC** [38] enhances clustering performance by utilizing high-confidence clustering results to construct reliable positive and negative sample pairs. **SCGC** [21] employs a novel cross-view structural consistency to improve efficiency in deep graph clustering. **HSAN** [22] introduces a dynamic sample weighting strategy to effectively mine both hard negative and hard positive samples, thereby improving overall discriminative capability.

Implementation Details. The experiments are executed using the following hardware configuration: Intel Core i9-13900K CPU, NVIDIA GeForce RTX 4090 GPU, and 64GB RAM. Additionally, all experiments are implemented using the Pytorch framework, with a maximum training epoch limit of 400. The Adam optimizer [10] is employed to minimize the total loss, and the K-means algorithm was applied to the fused embeddings to derive the final clustering results.

Table 2. The clustering performance, averaged over ten runs on four benchmark datasets, is evaluated using three metrics, with results displayed as mean values and standard deviations. The highest and second-highest results are highlighted in **bold** and underlined, respectively. 'OOM' denotes the out-of-memory failure.

	Missing rate	0.3			0.6			0.9		
	Method	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
CORA	SAT	68.72±1.13	50.60±0.88	45.03±1.14	60.11±1.12	44.16±1.47	36.31±1.46	33.23±1.65	10.04±0.82	04.14±0.26
	FPGM	49.45±1.43	29.74±1.54	18.55±1.62	46.45±1.24	27.17±0.95	15.35±1.24	40.29±1.21	19.59±0.84	07.09±0.96
	T2-GNN	65.59±2.42	48.31±1.65	40.38±1.24	63.70±1.78	43.83±1.56	35.84±1.64	41.03±1.25	19.02±1.26	13.07±1.34
	CONVERT	72.40±1.04	54.15±0.97	48.17±0.78	71.04±0.92	50.27±1.16	46.94±1.53	50.46±1.07	32.71±1.88	20.49±1.89
	CCGC	71.95±1.81	54.20±2.30	49.26±2.79	53.15±2.77	37.36±4.40	37.36±4.35	33.21±1.35	11.27±1.47	04.35±1.05
	SCGC	72.88±0.84	54.97±1.43	50.39±1.49	68.15±1.19	49.33±1.41	46.30±1.57	31.45±1.37	07.71±3.30	02.40±0.97
	HSAN	73.49±0.95	**56.23±0.94**	**51.25±1.36**	67.45±1.81	46.75±2.07	41.68±2.90	36.15±3.46	14.84±4.22	04.36±2.22
	Ours	**73.64±1.37**	54.43±1.02	50.16±2.82	**72.05±0.82**	**53.14±1.36**	**48.29±0.80**	**68.94±0.46**	**51.20±0.36**	**45.87±0.92**
CITESEER	SAT	60.83±1.58	32.15±2.01	31.84±1.92	38.84±1.21	16.21±0.89	12.06±1.42	21.89±1.12	01.56±0.63	0.13±0.24
	FPGM	31.05±1.26	12.33±1.12	02.04±0.67	29.67±1.54	10.72±0.86	01.68±0.45	27.62±1.31	08.48±0.93	0.87±0.21
	T2-GNN	50.56±2.15	27.09±1.82	24.07±1.26	47.19±1.56	20.85±1.73	18.86±1.63	34.75±1.52	08.02±0.93	8.32±1.15
	CONVERT	65.93±0.93	38.56±1.06	38.85±1.48	61.63±0.80	33.04±0.99	34.11±1.32	47.07±0.97	23.00±0.59	16.62±0.81
	CCGC	66.51±2.20	38.73±2.56	39.47±3.17	45.72±3.95	21.02±3.21	18.75±3.87	23.65±3.12	04.01±1.56	0.87±0.64
	SCGC	68.60±0.67	41.38±0.78	42.42±1.10	59.28±1.40	31.66±1.69	31.67±1.77	24.54±0.52	03.97±1.35	0.95±0.47
	HSAN	66.47±1.90	38.87±1.57	38.07±3.57	48.41±3.34	22.28±2.88	13.70±3.26	22.02±2.01	01.46±1.07	0.07±0.16
	Ours	**69.08±0.75**	**41.82±0.78**	**43.15±0.74**	**66.87±1.13**	**38.91±0.84**	**40.18±0.96**	**55.25±0.29**	**27.47±0.62**	**25.71±0.30**
AMAC	SAT	48.63±1.76	29.54±1.46	25.35±1.86	36.84±2.45	19.32±2.34	14.23±1.54	22.02±1.43	03.21±1.45	01.46±0.87
	FPGM	28.90±1.45	08.52±0.87	-03.04±0.23	28.47±1.85	07.15±0.74	-03.93±0.43	29.10±1.64	05.19±0.83	-04.24±0.63
	T2-GNN	43.24±2.16	33.63±1.53	27.36±1.64	39.70±1.64	29.08±1.78	17.66±1.25	28.73±1.54	14.09±0.86	05.75±0.53
	CONVERT	53.76±3.27	39.72±2.14	29.69±3.93	50.89±1.64	36.13±1.21	28.43±0.94	37.50±0.00	11.37±0.00	-00.01±0.00
	CCGC	53.98±0.75	33.54±0.99	29.29±1.44	43.38±1.33	25.28±2.28	20.40±1.48	22.10±1.05	04.33±3.34	01.95±1.51
	SCGC	55.23±1.22	35.93±0.81	31.48±1.64	47.60±1.40	28.76±0.97	24.55±0.92	23.43±1.25	02.49±1.39	00.99±0.43
	HSAN		OOM			OOM			OOM	
	Ours	**56.04±0.46**	**40.92±0.76**	**34.12±0.95**	**55.26±0.89**	**38.25±1.23**	**31.84±1.00**	**53.38±0.95**	**33.91±1.23**	**28.65±1.48**
WIKI	SAT	44.85±1.46	41.23±1.53	18.21±1.67	38.30±2.54	34.88±1.17	11.64±1.17	27.34±2.13	21.68±2.44	05.67±2.15
	FPGM	29.44±1.56	27.26±1.84	03.77±0.45	29.23±1.31	28.02±1.25	03.30±0.25	29.06±1.16	26.23±1.24	04.11±0.85
	T2-GNN	50.10±1.94	45.28±1.51	30.97±1.35	46.78±1.25	40.01±1.36	24.06±1.56	34.97±1.67	32.25±1.68	10.46±1.28
	CONVERT	45.02±3.66	42.47±1.50	25.62±3.80	16.71±0.00	01.42±0.00	00.05±0.00	16.71±0.00	01.42±0.00	00.05±0.00
	CCGC	49.01±1.18	42.10±0.80	22.18±2.12	43.75±1.43	38.87±0.92	17.13±1.31	29.95±3.73	22.82±3.02	06.46±4.23
	SCGC	52.73±0.54	46.97±0.51	28.21±2.33	46.36±0.66	42.32±0.73	20.69±0.98	36.98±1.93	32.87±1.04	09.94±0.73
	HSAN	51.13±1.90	44.92±1.13	30.85±2.25	46.54±1.32	39.15±1.09	25.57±1.92	32.94±2.62	26.97±1.87	**12.71±1.74**
	Ours	**55.32±1.18**	**49.35±0.53**	**35.20±1.35**	**47.99±0.65**	**42.98±1.18**	**25.72±0.65**	**38.08±0.85**	**34.05±0.80**	12.55±0.49

4.2. Comparsion Results with State-of-The-Arts

In Table 2, we compare our proposed method with seven state-of-the-art methods on three metrics. The performance of AMMGC is evaluated on four benchmark datasets, taking into account three levels of missing data: [0.3, 0.6, 0.9]. From the above chart, we have the following observations:

- Our method consistently outperforms all compared approaches across most datasets, especially when the proportion of missing node attributes is high. AMMGC exhibits remarkable stability with varying missing rates. For instance, on the AMAC dataset, the accuracy experiences only decreases by 2.66% as the missing rate increases from 0.3 to 0.9.
- It can be seen that the AMMGC can still achieve the competitive effect with a high missing rate in the dataset. For instance, in the CORA dataset, AMMGC significantly surpasses the second-best method, achieving improvements of 18.48% in ACC, 18.49% in NMI, and 25.38% in ARI. These findings robustly support the effectiveness and resilience of AMMGC in addressing high-level missing data, highlighting its capability to sustain strong clustering performance under challenging attribute-missing conditions.

In summary, the experiments conducted have verified the effectiveness and stability of our proposed AMMGC. The method's superior performance can be attributed to three critical factors. First, the context-aware imputation effectively utilizes neighborhood information to iteratively refine the imputation of missing features, resulting in more robust and reliable node representations. Additionally, the dual structure consistency ensures information consistency

Table 3. Ablation study on four datasets. ✔ denotes AMMGC with the component (The missing rate is 0.6).

Datasets	Components			Metrics		
	CAI	DSC	HCG	ACC	NMI	ARI
CORA	✔	✔		71.65±0.80	52.26±2.06	47.13±1.53
	✔		✔	70.55±1.79	51.34±1.73	45.78±3.08
		✔	✔	55.74±0.72	40.40±0.84	29.38±1.01
	✔	✔	✔	**72.05±0.82**	**53.14±1.36**	**48.29±0.80**
CITESEER	✔	✔		66.18±1.13	37.82±1.06	38.67±1.58
	✔		✔	55.72±3.69	30.70±2.35	27.98±4.11
		✔	✔	63.54±1.12	36.06±1.09	34.38±1.08
	✔	✔	✔	**66.87±1.13**	**38.91±0.84**	**40.18±0.96**
AMAC	✔	✔		54.63±0.76	38.01±0.92	31.25±1.25
	✔		✔	53.83±0.72	37.96±1.36	30.93±1.55
		✔	✔	51.42±1.67	33.75±1.24	28.42±1.56
	✔	✔	✔	**55.26±0.89**	**38.25±1.23**	**31.84±1.00**
WIKI	✔	✔		47.60±0.49	42.77±0.66	25.60±0.58
	✔		✔	43.91±1.75	39.95±1.87	20.85±1.77
		✔	✔	41.87±0.64	40.53±0.79	20.89±0.85
	✔	✔	✔	**47.99±0.65**	**42.98±1.18**	**25.72±0.65**

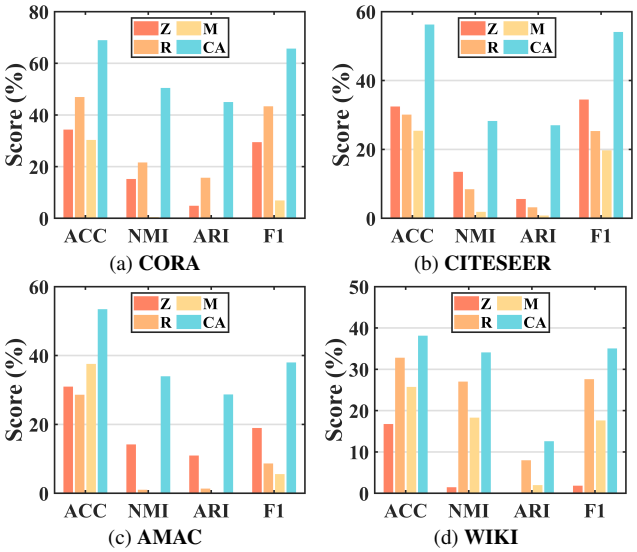

Figure 2. The experimental results of different Imputation strategies across four datasets (The missing rate is 0.9).

between views, minimizing redundancy and noise. Finally, high-confidence guidance allows the model to capture high-confidence clustering information, thereby improving the reliability of the clustering results.

4.3. Ablation Studies

In this subsection, we conduct ablation experiments on four benchmark datasets to verify the effectiveness of each component.

Ablation on Model Components. We begin by analyzing the influence of various components of our proposed method and explore how different imputation strategies impact the outcomes. For easier presentation, we refer to context-aware imputation, dual structure consistency, and high-confidence guidance as 'CAI', 'DSC', and 'HCG', respectively. As illustrated in Table 3, the optimal performance is achieved when all components are incorporated. Excluding any of the proposed components leads to a decrease in clustering performance, demonstrating the significance of each component in enhancing the overall effectiveness of the method.

Ablation on Imputation Strategies. We further investigated the effects of various imputation strategies on the results. Specifically, we assessed four distinct imputation methods: zero imputation, random imputation, mean imputation, and context-aware imputation. For clarity, we refer to these strategies as 'Z' for zero imputation, 'R' for random imputation, 'M' for mean imputation, and 'CA' for context-aware imputation. The detailed experimental results are presented in Fig. 2. The figure indicates that context-aware imputation outperforms the other strategies. This enhancement can be attributed to context-aware imputation's ability to effectively harness valuable information from neighboring nodes, resulting in more reliable node representations.

4.4. Parameter Sensitivity Analysis

In this subsection, we examine the sensitivity of hyper-parameters, specifically the number of graph Laplacian filter layer t and trade-off parameter α, on the clustering performance of the AMMGC across four datasets. we analyze the values of t and α in the range of $\{1, 2, 3, 4, 5\}$ and $\{10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}\}$, respectively. The experimental results are shown in Fig. 4, and we can observe that α has a greater impact on CITESEER, while both parameters exhibit minimal impact on the other three datasets.

4.5. Performance with Different Missing Rates

To further evaluate the robustness and effectiveness of our proposed method, we conducted experiments across four datasets (CORA, CITESEER, AMAC, and WIKI), systematically varying the missing rate r from 0.1 to 0.9 with increments of 0.1. From the results in Fig. 5, we could observe that: 1) AMMGC significantly outperforms all the tested baselines under most settings of missing rates; 2) with increasing the missing rate, the performance degradations of the compared methods are much larger than that of ours. For instance, on the AMAC dataset, both AMMGC and CONVERT initially achieve comparable accuracies (55.97 and 54.42 respectively) when $r = 0.1$. However, as the missing rate escalates, AMMGC demonstrates a marked advantage over CONVERT, maintaining a more stable accuracy and highlighting its capability to mitigate the adverse effects of high missing rates.

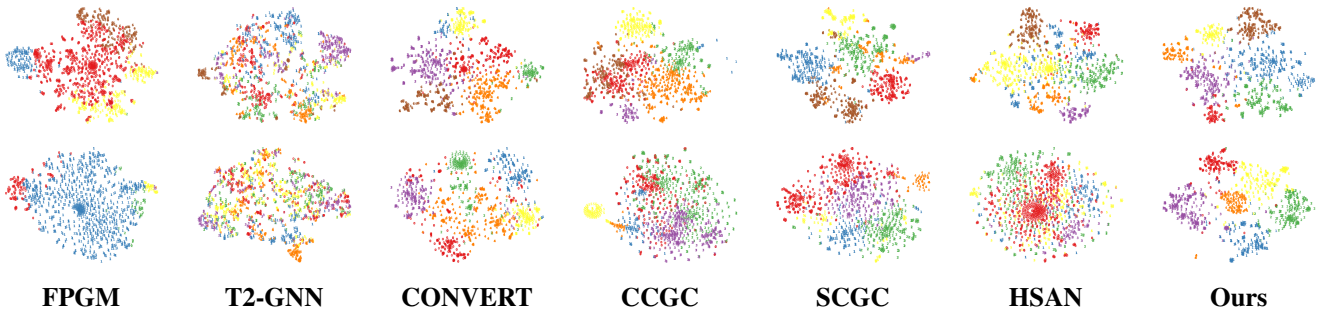

Figure 3. The Visualization of node representation using t-SNE algorithm. The first row and second row correspond to CORA and CITESEER, respectively (The missing rate is 0.6).

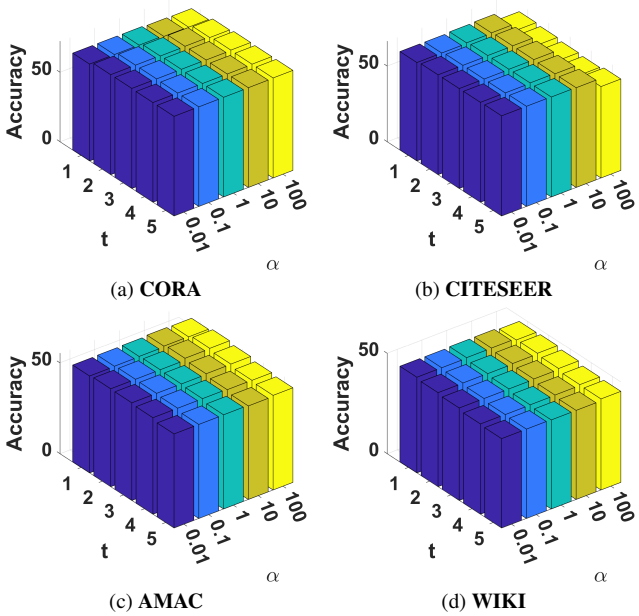

Figure 4. The sensitivity analysis of hyper-parameter t and α on four datasets (The missing rate is 0.6).

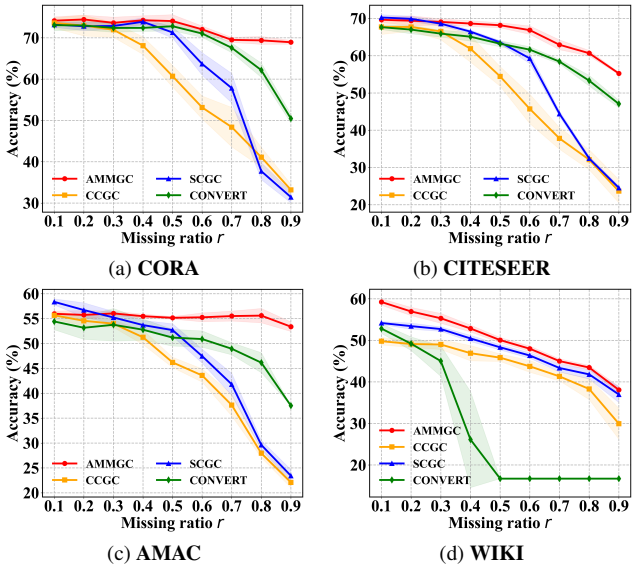

Figure 5. Error band plot of Accuracy on four datasets as missing rate increases, with the shaded area representing the standard deviation. The red curve denotes our method.

4.6. Visualization Analysis

To comprehensively validate the effectiveness of the proposed AMMGC method, we perform a visualization analysis using t-SNE [30] on the CORA and CITESEER dataset with a missing rate of $r = 0.6$, as shown in Fig. 3. This approach enables an intuitive comparison of clustering performance across various methods, including FPGM, T2-GNN, CONVERT, CCGC, SCGC, and HSAN, under severe missing data conditions. Our method shows a noticeably clearer separation of clusters, with well-defined boundaries between node embeddings, unlike the more dispersed or overlapping clusters produced by other methods. This enhanced clustering structure highlights AMMGC's strong ability to capture meaningful patterns and relationships among nodes. These visual results underscore the effectiveness of our approach in handling missing data, further validating its superiority in the attribute-missing MVGC setting.

5. Conclusion

In this paper, we propose an Attribute-Missing Multi-view Graph Clustering, termed AMMGC, which iteratively imputes missing node features through the use of neighborhood information. We first design a dual structure consistency module aimed at improving the alignment of graph structures across various views. Additionally, we implement high-confidence guidance to enhance the reliability of the clustering process. The experimental results demonstrate the effectiveness and superiority of our method, particularly in cases where a substantial number of node attributes are absent, underscoring its exceptional performance.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant 62176203, the Fundamental Research Funds for the Central Universities (ZYTS25267 and QTZX25004), and the Science and Technology Project of Xi'an (Grant 2022JH-JSYF-0009), Open Project of Anhui Provincial Key Laboratory of Multimodal Cognitive Computation, Anhui University (No. MMC202416), Selected Support Project for Scientific and Technological Activities of Returned Overseas Chinese Scholars in Shaanxi Province 2023-02, and the Xidian Innovation Fund (Project NoYJSJ25007)

References

[1] Man-Sheng Chen, Xi-Ran Zhu, Jia-Qi Lin, and Chang-Dong Wang. Contrastive multiview attribute graph clustering with adaptive encoders. *IEEE Transactions on Neural Networks and Learning Systems*, 2024. 1

[2] Xu Chen, Siheng Chen, Jiangchao Yao, Huangjie Zheng, Ya Zhang, and Ivor W Tsang. Learning on attribute-missing graphs. *IEEE transactions on pattern analysis and machine intelligence*, 44(2):740–757, 2020. 1, 2, 5

[3] Jiafeng Cheng, Qianqian Wang, Zhiqiang Tao, Deyan Xie, and Quanxue Gao. Multi-view attribute graph convolution networks for clustering. In *Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence*, pages 2973–2979, 2021. 1, 2

[4] Ganqu Cui, Jie Zhou, Cheng Yang, and Zhiyuan Liu. Adaptive graph encoder for attributed graph embedding. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 976–985, 2020. 3, 4, 5

[5] Shaohua Fan, Xiao Wang, Chuan Shi, Emiao Lu, Ken Lin, and Bai Wang. One2multi graph autoencoder for multi-view graph clustering. In *proceedings of the web conference 2020*, pages 3070–3076, 2020. 1, 2

[6] Kaveh Hassani and Amir Hosein Khasahmadi. Contrastive multi-view representation learning on graphs. In *International conference on machine learning*, pages 4116–4126. PMLR, 2020. 1

[7] Cuiying Huo, Di Jin, Yawen Li, Dongxiao He, Yu-Bin Yang, and Lingfei Wu. T2-gnn: Graph neural networks for graphs with incomplete features and structure via teacher-student distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4339–4346, 2023. 1, 2, 5

[8] Xinke Jiang, Zidi Qin, Jiarong Xu, and Xiang Ao. Incomplete graph learning via attribute-structure decoupled variational auto-encoder. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, pages 304–312, 2024. 1

[9] Jiaqi Jin, Siwei Wang, Zhibin Dong, Xinwang Liu, and En Zhu. Deep incomplete multi-view clustering with cross-view partial sample and prototype alignment. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11600–11609, 2023. 2

[10] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[11] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016. 3

[12] Haobin Li, Yunfan Li, Mouxing Yang, Peng Hu, Dezhong Peng, and Xi Peng. Incomplete multi-view clustering via prototype-based imputation. *arXiv preprint arXiv:2301.11045*, 2023. 2

[13] Mengran Li, Yong Zhang, Wei Zhang, Shiyu Zhao, Xinglin Piao, and Baocai Yin. Csat: Contrastive sampling-aggregating transformer for attribute-missing graph learning. *Available at SSRN 4278482*, 2022. 2

[14] Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering via contrastive prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11174–11183, 2021. 2

[15] Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. Dual contrastive prediction for incomplete multi-view representation learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4): 4447–4461, 2022. 2

[16] Zhiping Lin and Zhao Kang. Graph filter-based multi-view attributed graph clustering. In *IJCAI*, pages 2723–2729, 2021. 1

[17] Zhiping Lin, Zhao Kang, Lizong Zhang, and Ling Tian. Multi-view attributed graph clustering. *IEEE Transactions on knowledge and data engineering*, 35(2):1872–1880, 2021. 2

[18] Yawen Ling, Jianpeng Chen, Yazhou Ren, Xiaorong Pu, Jie Xu, Xiaofeng Zhu, and Lifang He. Dual label-guided graph refinement for multi-view graph clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8791–8798, 2023.

[19] Jing Liu, Fuyuan Cao, Xuechun Jing, and Jiye Liang. Deep multi-view graph clustering network with weighting mechanism and collaborative training. *Expert Systems with Applications*, 236:121298, 2024. 1

[20] Yue Liu, Wenxuan Tu, Sihang Zhou, Xinwang Liu, Linxuan Song, Xihong Yang, and En Zhu. Deep graph clustering via dual correlation reduction. In *Proceedings of the AAAI conference on artificial intelligence*, pages 7603–7611, 2022. 2

[21] Yue Liu, Xihong Yang, Sihang Zhou, Xinwang Liu, Siwei Wang, Ke Liang, Wenxuan Tu, and Liang Li. Simple contrastive graph clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2023. 5

[22] Yue Liu, Xihong Yang, Sihang Zhou, Xinwang Liu, Zhen Wang, Ke Liang, Wenxuan Tu, Liang Li, Jingcan Duan, and Cancan Chen. Hard sample aware network for contrastive deep graph clustering. In *Proceedings of the AAAI conference on artificial intelligence*, pages 8914–8922, 2023. 5

[23] Erlin Pan and Zhao Kang. Multi-view contrastive graph clustering. *Advances in neural information processing systems*, 34:2148–2159, 2021. 1, 2, 5

[24] Xin Peng, Jieren Cheng, Xiangyan Tang, Bin Zhang, and Wenxuan Tu. Multi-view graph imputation network. *Information Fusion*, 102:102024, 2024. 2

[25] Jingyu Pu, Chenhang Cui, Xinyue Chen, Yazhou Ren, Xiaorong Pu, Zhifeng Hao, S Yu Philip, and Lifang He. Adaptive feature imputation with latent graph for deep incomplete multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 14633–14641, 2024. 2

[26] Emanuele Rossi, Henry Kenlay, Maria I Gorinova, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M Bronstein. On the unreasonable effectiveness of feature propagation in learning on graphs with missing node features. In *Learning on graphs conference*, pages 11–1. PMLR, 2022. 1, 3, 5

[27] Wenxuan Tu, Sihang Zhou, Xinwang Liu, Yue Liu, Zhiping Cai, En Zhu, Changwang Zhang, and Jieren Cheng. Initializing then refining: A simple graph attribute imputation network. In *IJCAI*, pages 3494–3500, 2022. 2

[28] Wenxuan Tu, Renxiang Guan, Sihang Zhou, Chuan Ma, Xin Peng, Zhiping Cai, Zhe Liu, Jieren Cheng, and Xinwang Liu. Attribute-missing graph clustering network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 15392–15401, 2024. 2

[29] Wenxuan Tu, Bin Xiao, Xinwang Liu, Sihang Zhou, Zhiping Cai, and Jieren Cheng. Revisiting initializing then refining: an incomplete and missing graph imputation network. *IEEE Transactions on Neural Networks and Learning Systems*, 2024. 2

[30] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9 (11), 2008. 8

[31] Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao, and Yun Fu. Partial multi-view clustering via consistent gan. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 1290–1295. IEEE, 2018. 2

[32] Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao, and Yun Fu. Generative partial multi-view clustering. *arXiv preprint arXiv:2003.13088*, 2020. 2

[33] Yiming Wang, Dongxia Chang, Zhiqiang Fu, Jie Wen, and Yao Zhao. Incomplete multiview clustering via cross-view relation transfer. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(1):367–378, 2022. 2

[34] Yi Wen, Suyuan Liu, Xinhang Wan, Siwei Wang, Ke Liang, Xinwang Liu, Xihong Yang, and Pei Zhang. Efficient multi-view graph clustering with local and global structure preservation. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 3021–3030, 2023. 1

[35] Wei Xia, Qianqian Wang, Quanxue Gao, Ming Yang, and Xinbo Gao. Self-consistent contrastive attributed graph clustering with pseudo-label prompt. *IEEE Transactions on Multimedia*, 25:6665–6677, 2022. 2

[36] Wei Xia, Sen Wang, Ming Yang, Quanxue Gao, Jungong Han, and Xinbo Gao. Multi-view graph embedding clustering network: Joint self-supervision and block diagonal representation. *Neural Networks*, 145:1–9, 2022. 2

[37] Jie Xu, Chao Li, Liang Peng, Yazhou Ren, Xiaoshuang Shi, Heng Tao Shen, and Xiaofeng Zhu. Adaptive feature projection with distribution alignment for deep incomplete multi-view clustering. *IEEE Transactions on Image Processing*, 32:1354–1366, 2023. 2

[38] Xihong Yang, Yue Liu, Sihang Zhou, Siwei Wang, Wenxuan Tu, Qun Zheng, Xinwang Liu, Liming Fang, and En Zhu. Cluster-guided contrastive graph clustering network. In *Proceedings of the AAAI conference on artificial intelligence*, pages 10834–10842, 2023. 2, 3, 5

[39] Xihong Yang, Cheng Tan, Yue Liu, Ke Liang, Siwei Wang, Sihang Zhou, Jun Xia, Stan Z Li, Xinwang Liu, and En Zhu. Convert: Contrastive graph clustering with reliable augmentation. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 319–327, 2023. 5

[40] Jaemin Yoo, Hyunsik Jeon, Jinhong Jung, and U Kang. Accurate node feature estimation with structured variational graph autoencoder. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2336–2346, 2022. 2

[41] Jiaxuan You, Xiaobai Ma, Yi Ding, Mykel J Kochenderfer, and Jure Leskovec. Handling missing data with graph representation learning. *Advances in Neural Information Processing Systems*, 33:19075–19087, 2020. 1

[42] Zihao Zhang, Qianqian Wang, Zhiqiang Tao, Quanxue Gao, and Wei Feng. Dropping pathways towards deep multi-view graph subspace clustering networks. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 3259–3267, 2023. 1