Learn from Global Rather Than Local: Consistent Context-Aware Representation Learning for Multi-View Graph Clustering

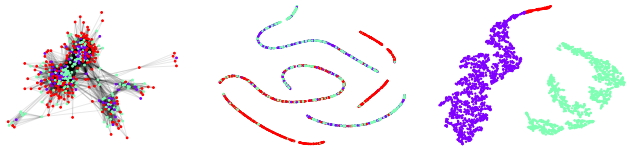
Lele Fu[1] , **Bowen Deng**[1] , **Sheng Huang**[1] , **Tianchi Liao**[1] , **Chuanfu Zhang**[1] and **Chuan Chen**[1*]
[1]Sun Yat-sen University
{fulle, dengbw3, huangsh253, liaotch}@mail2.sysu.edu.cn, {zhangchf9, chenchuan}@mail.sysu.edu.cn

Abstract

Multi-view graph clustering (MVGC) has been of widespread interest owing to the ability of capturing the complementary information among views, thereby enhancing the performance of node clustering. Despite the impressive achievements of existing methods, they are limited by a common deficiency, namely, the curse of local manifold while failing to perceive the global manifold structure. In light of this drawback, we propose a Consistent Context-Aware Representation Learning (CCARL) method for MVGC, aiming to learn node representations from global space rather than just local topology. Concretely, we define a set of anchors to establish the global coordinate, which are optimally mapped to multi-view graphs with minimal cost via fused Gromov-Wasserstein optimal transport. To fuse the complementary information in various views, the attention mechanism is employed to integrate multiple graph embeddings into a consistent representation. By transforming to the global coordinate connecting with anchors, the consistent representation captures the contextual information, and its clustering-friendliness is further enhanced through a self-training strategy. Finally, extensive experiments on four multi-view graph datasets demonstrate the effectiveness of the proposed CCARL over existing MVGC methods.

1 Introduction

Graph clustering [Xia *et al.*, 2022; Liu *et al.*, 2024c; Deng *et al.*, 2025] is an important data mining technique, aiming to couple node features and topology for grouping nodes into different clusters. Conventional graph clustering is frequently oriented towards single-view graph data, i.e., both node features and topology are unique. Nevertheless, with the diversification of information collection, the features or connections between nodes can be portrayed from various perspectives. For example, different products can be connected by the similarity or by sharing the same buyers. Such graph data integrating rich information from multiple sources is dubbed multi-view graph data. Compared to the single-view graph data, multi-view graph data can provide the complementary information to enhance the discriminability of latent embedding, thereby improving the performance of downstream tasks.



(a) YELP-View0 (b) GAE (c) CCARL

Figure 1: (a) illustrates extensive inter-class edges in YELP dataset. (b) and (c) show the learned consistent representations by GAE and the proposed CCARL. Notably, GAE averages multi-view graph embeddings into a consistent representation. It can be seen that the embedding of GAE fails to exhibit well intra-class aggregation and inter-class separation, whereas the proposed CCARL excels.

While multi-view data [Fu *et al.*, 2020; Yang *et al.*, 2022; Zhang *et al.*, 2023; Zhang *et al.*, 2025; Yu *et al.*, 2025] has more advanced information, it also poses a key challenge to the conventional graph clustering, namely how to effectively fuse information from multiple views. In response to the concern, extensive multi-view graph clustering (MVGC) methods surge to be encouraging solutions. [Fu *et al.*, 2022; Tang *et al.*, 2022; Wang *et al.*, 2024] learned a consistent label indicator matrix via spectral decomposition. [Liu *et al.*, 2024a; Guan *et al.*, 2024; Fu *et al.*, 2024] projected the multiple raw features into a common subspace with structural constraints. [Li *et al.*, 2023a; Li *et al.*, 2023b; Feng *et al.*, 2024] explored the high-order correlations between views by low-rank tensor optimization. The above methods lack a coupled treatment of features and topology for nodes, then the learned embedding might be suboptimal. In light of this drawback, [Chen *et al.*, 2023; Tu *et al.*, 2021; Tsitsulin *et al.*, 2023; Tu *et al.*, 2024; Chen *et al.*, 2024] introduced the graph convolutional neural networks (GCNs) to learn the latent embedding, which recursively handled the features and topology with a coupled manner. Generally, they adopted the GCNs to project the raw multi-view graph data into the latent space with some constraints, then the fusion strategies were developed for integrating heterogeneous information. Specifically, the fusion process is either prioritized or posteriorized [Lin *et al.*, 2024;

* Corresponding author.

Chen *et al.*, 2024], depending on whether the consistent representation is obtained during training or after training.

Thanks to the integration of the complementary information between views, the above MVGC methods achieve promising results. Nevertheless, they are inevitably trapped into an identical limitation, namely focusing on utilizing local manifold structure while ignoring the global manifold structure. Concretely, traditional MVGC methods frequently introduce the Laplacian regularization to enhance the latent representations. Likewise, GCN-based methods essentially use the Laplacian filters to fuse the neighboring node information. Laplacian-induced operations make close embeddings more similar, but also cause the model to fall into the curse of locality and fail to perceive the global context, which is dangerous for clustering tasks. Especially in graphs, there may be connection edges between two nodes of different categories as shown in Fig. 1(a), whose corresponding embeddings yielded by GAE are similar after graph convolution, resulting in a significant difficulty for correct clustering like in Fig. 1 (b). Hence, it is important not to refine the node embeddings only from a local perspective, but to consider the positions of them in the global context for making the most reasonable judgment. Just like in natural language processing, the same word expresses various meanings in different sentences, the specific meaning can only be accurately determined by connecting to the context. Furthermore, in addition to complementarity across views, consistency is another important principle for multi-view learning, which requires maximizing agreement among multiple heterogeneous views so that they can converge to a uniform clustering space. In a nutshell, *how to capture the consistent contextual information in global space is a challenging issue for achieving advanced MVGC*.

In this paper, we propose a Consistent Context-Aware Representation Learning (CCARL) method for MVGC in response to the above challenge. First, we attempt to use a set of anchors beyond multiple views to span the global space. Intuitively, these anchors are viewed to establish the new global coordinate. Thus, the global anchor graph is optimally mapped to multi-view graphs via fused Gromov-Wasserstein optimal transport (GW-OT), guaranteeing the global space being intact. To integrate the complementary information in diverse views, the attention mechanism is employed to combine various graph embeddings into a consistent representation. Importantly, the consistent representation is endowed with the context-awareness by connecting with the anchors for projecting it into the global coordinate. Last but not least, considering the peculiarity of the clustering task, samples in the same clusters have to be as close as possible to the cluster centroids. A self-training strategy, which tightens the prediction distribution to the target distribution, is adopted to drive the consistent context-aware representation to be clustering-friendly. Fig. 2 illustrates the overall framework for the proposed CCARL. In conclusion, the principal contributions of this paper are summarized from three-fold:

- We define a set of anchors beyond multiple views, which are used to span the global space and establish the global coordinate. Thus, the fused GW-OT is leveraged to find the optimal mapping between the anchor graph and multi-view graphs, ensuring that the anchors can cover the global space.
- The attention mechanism is employed to adaptively fuse the various graph embeddings into a consistent representation. By connecting with the global anchors, the consistent representation captures the context-aware information. Further, a self-training strategy is used to enhance its clustering-friendliness.
- A large number of experiments on four multi-view graph datasets are conducted to verify the effectiveness of the proposed CCARL. The comparative and ablation experimental results demonstrate the superiority of CCARL over the typical and SOTA methods.

2 Related Works

2.1 Multi-View Graph Clustering

Multi-view graph clustering improves the downstream clustering performance by fusing the complementary information among multi-source graph data. Current methods can be categorized into two types, one is based on conventional matrix factorization (MF) and the other is based on GCN. MF-based methods aim to decompose the raw matrix into several low-dimensional matrices, thereby discovering the underlying structure. For instance, [Huang *et al.*, 2022; Fu *et al.*, 2022; Liu *et al.*, 2024b] explored a uniform label indicator matrix from multiple similarity matrices, thus directly acquiring the clustering results. [Chen *et al.*, 2021; Guan *et al.*, 2024; Fu *et al.*, 2024] learned a self-expression representation from multi-view data, on which the low-rank constraint was often imposed for enhancing the clustering properties. Tensor is a high-order form of matrix that carries more complex data structures. [Li *et al.*, 2021; Pan *et al.*, 2024; Qin *et al.*, 2024] leveraged the low-rank tensor decomposition to capture the consistency and complementarity among various views. [Lin and Kang, 2021; Lin *et al.*, 2023; Qian *et al.*, 2024] introduced graph filters to smooth the node embedding, then performed the integration process. GCN is a simple yet powerful network for graph data, which aggregates the neighborhood information of nodes to enhance the embedding, many GCN-based MVGC approaches are witnessing a high moment in the spotlight. [Fan *et al.*, 2020; Lin *et al.*, 2024] jointly learned the fusion embedding and cluster assignment via drawing the Student's t-distribution to the target distribution. [Pan and Kang, 2021; Cai *et al.*, 2024; Liu *et al.*, 2025] performed the cross-view contrastive learning to promote the consensus between views. [Wang *et al.*, 2023; Zhuang *et al.*, 2024] developed novel multi-graph fusion mechanisms to aggregate multi-source information.

2.2 Optimal Transport

Optimal transport [Montesuma *et al.*, 2024] refers to the minimum cost for transferring one distribution to another distribution, which is often used to measure the distance between two distributions. Due to the well-established theory and delicate solution, OT has been introduced into substantial research works involving distribution alignment. For instance, [Courty *et al.*, 2016; Liu *et al.*, 2023; Groppe and

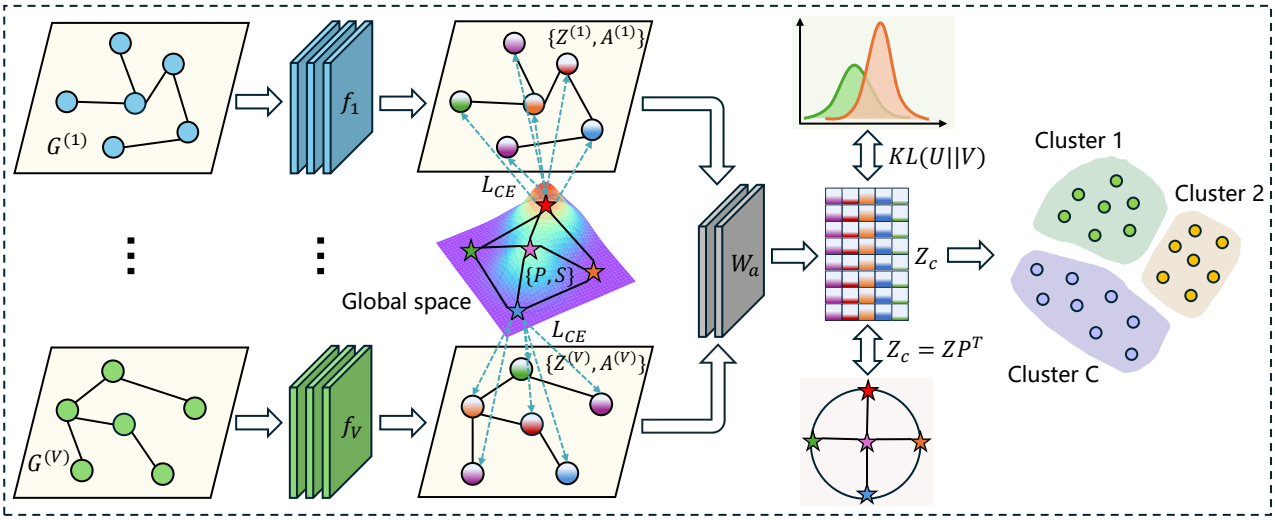


Figure 2: The framework of the proposed CCARL. The global anchor graph $\{\mathbf{P}, \mathbf{S}\}$ is optimally mapped to multi-view graphs $\{\mathbf{Z}^{(v)}, \mathbf{A}^{(v)}\}$ via GW-OT. The consistent context-aware representation $\mathbf{Z}_c$ is used to perform the clustering with Kmeans algorithm.

Hundrieser, 2024] achieved the alignment between different data domains via OT mapping. [Arjovsky *et al.*, 2017; Tolstikhin *et al.*, 2018] introduced the theory of OT into generative modeling methods for strengthening the data generation. Above OT methods can only measure the distance of distributions in the same space and fail when the distributions are in different spaces. In light of this drawback, GW-OT [Mémoli, 2011] was proposed and focused on the keeping minimal cost in terms of the structures during the alignment. Hence, GW-OT is naturally suitable to handle structured data such as graphs. [Titouan *et al.*, 2019; Ma *et al.*, 2024] took into account the peculiarities when aligning multiple graphs, and achieved the optimal mapping between varying topologies using GW-OT. Inspired by GW-OT, we leverage the fused GW-OT to align the anchor graph and multi-view graphs in this paper, exploring the consistent context-aware coordinate information. Then, then a unified representation incorporating context information can be further obtained.

3 Methodology

3.1 Notations

Given a multi-view graph data $\mathcal{G} = \{\mathcal{G}^{(v)}\}_{v=1}^{V}$, where V denotes the number of views. $\mathcal{G}^{(v)} = \{\mathcal{V}, \mathcal{E}^{(v)}, \mathbf{X}^{(v)}\}$ is the graph of the v-th view, where $\mathcal{V}$ is the node set, $\mathcal{E}^{(v)}$ is the edge set, $\mathbf{X}^{(v)} \in \mathbb{R}^{N \times d}$ is the feature matrix (N is the number of nodes, d is the dimension). According to the edge set $\mathcal{E}^{(v)}$, the adjacency matrix $\mathbf{A}^{(v)} \in \mathbb{R}^{N \times N}$ is constructed. For the v-th view, a L-layer GCN $f_v : \mathbb{R}^d \to \mathbb{R}^{d_e}$ is used to encode the raw data into latent embedding, where d_e denotes the dimension of embedding. The computation of the l-th layer is formulated as

$$\mathbf{Z}_l^{(v)} = \sigma\left(\mathbf{D}^{(v)^{-\frac{1}{2}}} \hat{\mathbf{A}}^{(v)} \mathbf{D}^{(v)^{-\frac{1}{2}}} \mathbf{Z}_{l-1}^{(v)} \mathbf{W}_l^{(v)}\right) \tag{1}$$

where $\sigma(\cdot)$ denotes the activation function. $\mathbf{Z}_l^{(v)}$ and $\mathbf{W}_l^{(v)}$ are the embedding and the parameters for the l-th layer, respectively. $\hat{\mathbf{A}}^{(v)} = (\mathbf{A}^{(v)} + \mathbf{I})$ is the adjacency matrix, $\mathbf{D}^{(v)} \in \mathbb{R}^{N \times N}$ is the degree matrix and defined as $\mathbf{D}_{ii}^{(v)} = \sum_j \hat{\mathbf{A}}_{ij}^{(v)}$, $\mathbf{I} \in \mathbb{R}^{N \times N}$ denotes an identity matrix. For a multi-view graph data $\mathcal{G} = \{\mathcal{G}^{(v)}\}_{v=1}^{V}$, a set of global anchors $\mathbf{P} = \{\mathbf{p}_1, ..., \mathbf{p}_M\} \in \mathbb{R}^{M \times d_e}$ are expected to be explored.

3.2 Graph Alignment via Fused GW-OT

To establish a new coordinate system in global space, we expect to explore M anchors $\mathbf{P} = \{\mathbf{p}_1, ..., \mathbf{p}_M\}$ beyond multi-view graphs. For good global anchors, they can be transferred to each view's graph with minimal cost. The alignment is naturally an OT problem. In terms of node features, the formulation of OT between the anchors and the i-th view's features is written as

$$\begin{aligned} &\min_{\pi^{(v)} \in \Pi} \operatorname{Tr}(-\mathbf{P}\mathbf{Z}^{(v)^T} \pi^{(v)}) - \epsilon \operatorname{H}(\pi^{(v)}) \\ &\text{s.t. } \Pi = \{\pi^{(v)} \in \mathbb{R}_+^{N \times M} | \pi^{(v)} \mathbf{1}_M = \nu, \pi^{(v)^T} \mathbf{1}_N = \mu\}, \end{aligned} \tag{2}$$

where $\mathbf{Z}^{(v)} = f_v(\mathbf{X}^{(v)}, \mathbf{A}^{(v)})$ is the embedding after encoding, $\epsilon > 0$ is a hyperparameter adjusting the smoothness of the unnormalized assignment, $\operatorname{H}(\cdot)$ denotes the entropy for a variable. $\pi^{(v)}$ denotes the transport matrix and its marginal distributions subject to node marginal distribution ν and anchor marginal distribution μ. $\mathbf{1}_M$ and $\mathbf{1}_N$ denote the M-dimensional and N-dimensional all-one vectors, respectively. Notably, the product $\mathbf{P}\mathbf{Z}^{(v)^T}$ measures the similarity between the v-th view's embedding $\mathbf{Z}^{(v)}$ and the anchors $\mathbf{P}$, then the negative product is viewed as the cost between them. The objective of Eq. (2) is to learn an optimal transport matrix $\pi^{(v)}$. Intuitively, each entry $\pi_{i,j}^{(v)}$ records the unnormalized transfer probability from the i-th node to the j-th anchor.

However, general OT like Eq. (2) can only find the best mapping at the feature level and ignore the mapping at the

topological level for graph data, resulting in the suboptimal solution. Considering this issue, we resort to the GW-OT, which emphasizes maintaining the structural variance in distribution transferring. Thus, it is used to perform the topology alignment between anchor topology and multi-view topologies. First, the anchor graph $\mathbf{S} \in \mathbb{R}^{M\times M}$ is constructed based on anchor relationships. The definition of GW-OT is expressed as follows.

Definition 1. *Given two metric measure spaces $\mathcal{M}_1 = (\mathcal{V}_1, \mathbf{D}_1, \mu_1)$ and $\mathcal{M}_2 = (\mathcal{V}_2, \mathbf{D}_2, \mu_2)$ with size N_1 and N_2, where $\mathcal{V}_1$ and $\mathcal{V}_2$ denote the node sets, $\mathbf{D}_1$ and $\mathbf{D}_2$ denote the distance matrices, μ_1 and μ_2 denote the data marginal distributions. Then, the GW-OT distance between the two metric measure spaces is defined as*

$$\begin{aligned} &\inf_{\pi\in\Pi} \sum_{i,j}^{N_1}\sum_{k,l}^{N_2} \left([\mathbf{D}_1]_{i,j} - [\mathbf{D}_2]_{k,l}\right)^p \pi_{i,k}\pi_{j,l} \\ &s.t.\ \Pi = \{\pi \in \mathbb{R}_+^{N_1\times N_2} | \pi \mathbf{1}_{N_2} = \mu_1, \pi^T \mathbf{1}_{N_1} = \mu_2\}, \end{aligned} \tag{3}$$

In terms of structures, the alignment between anchor topology and the i-th view's topology based on Definition 1 is formulated as

$$\begin{aligned} &\min_{\pi^{(v)}\in\Pi} \sum_{i,j}^{N}\sum_{k,l}^{M} \left|[\mathbf{A}^{(v)}]_{i,j} - [\mathbf{S}]_{k,l}\right| \pi_{i,k}^{(v)}\pi_{j,l}^{(v)}, \\ &\text{s.t. } \Pi = \{\pi^{(v)} \in \mathbb{R}_+^{N\times M} | \pi^{(v)}\mathbf{1}_M = \nu, {\pi^{(v)}}^T \mathbf{1}_N = \mu\}. \end{aligned} \tag{4}$$

Incorporating alignments of feature and topology, the fused GW-OT problem between anchor graph and the i-th view's graph is formulated as

$$\begin{aligned} &\min_{\pi^{(v)}\in\Pi} \sum_{i,j}^{N}\sum_{k,l}^{M} \Big((1-\alpha)(-[\mathbf{P}]_k[{\mathbf{Z}^{(v)}}^T]_i) \\ &+\alpha \left|[\mathbf{A}^{(v)}]_{i,j} - [\mathbf{S}]_{k,l}\right|\Big) \pi_{i,k}^{(v)}\pi_{j,l}^{(v)} - \epsilon \mathrm{H}(\pi) \\ &\text{s.t. } \Pi = \{\pi^{(v)} \in \mathbb{R}_+^{N\times M} | \pi^{(v)}\mathbf{1}_M = \nu, {\pi^{(v)}}^T \mathbf{1}_N = \mu\}, \end{aligned} \tag{5}$$

where α denotes the trade-off parameter that regulates the strength of topology alignment. The fused GW-OT simultaneously performs the distribution matching from feature and structure, more accurately locating the anchors in the global space. For the problem (5), we use the solution proposed by [Titouan *et al.*, 2019] for finding an optimal transport matrix $\pi^{(v)}$. Based on $\pi^{(v)}$, the normalized node-anchor assignment matrix $\mathbf{Q}^{(v)}$ can be obtained by

$$[\mathbf{Q}^{(v)}]_{n,m} = \frac{[\pi^{(v)}]_{n,m}}{\sum_{m'}[\pi^{(v)}]_{n,m'}}. \tag{6}$$

$[\mathbf{Q}^{(v)}]_{n,m}$ is viewed as the transfer probability from the n-th node to the m-th anchor, and larger indicates that the two are closer together. As mentioned above, $\mathbf{P}{\mathbf{Z}^{(v)}}^T$ depicts the similarity between graph embedding $\mathbf{Z}^{(v)}$ and anchors $\mathbf{P}$, the normalized similarity is also viewed as the transfer probability matrix, which is formulated as

$$[\mathbf{F}^{(v)}]_{n,m} = \frac{\exp\left([\mathbf{Z}^{(v)}]_n[\mathbf{P}^T]_m/\tau\right)}{\sum_{m'}\exp\left([\mathbf{Z}^{(v)}]_n[\mathbf{P}^T]_{m'}/\tau\right)}. \tag{7}$$

Then, we expect to draw $\mathbf{F}^{(v)}$ closer to $\mathbf{Q}^{(v)}$, since the latter is optimal. Notably, anchor graph $\mathbf{S}$ is conducted by $\mathbf{F}^{(v)}{\mathbf{F}^{(v)}}^T$ for each view. By constructing this optimization objective, the anchors can be iteratively optimized. In general, to achieve view alignment, $\mathbf{F}^{(v)}$ has to be drawn closer to $\{\mathbf{Q}^{(u)}\}_{u\neq v}$ in pairs, but the cross-computation imposes an excessive computational burden. To simplify the calculation, we obtain a uniform transfer probability matrix $\mathbf{F}$ via $\mathbf{F} = \frac{1}{V}\sum_v \mathbf{F}^{(v)}$, then the cross-entropy (CE) between $\mathbf{Q}^{(v)}$ and $\mathbf{F}$ is written as

$$\mathcal{L}_{ce} = -\frac{1}{V}\sum_{v=1}^{V}\sum_{n=1}^{N}\sum_{m=1}^{M}\left([\mathbf{Q}^{(v)}]_{n,m}\log[\mathbf{F}]_{n,m}\right). \tag{8}$$

3.3 Consistent Context-Aware Representation

How to effectively handle multi-view graph embeddings is consistently a principal concern in MVGC. To fuse the complementary information in multi-view graphs, we leverage the attention mechanism to adaptively integrate multi-view graph embeddings into a consistent representation. As well known, the attention mechanism essentially measures the contribution of various components. Different from the common weighting mechanism, which assigns a view-level weight for each view by assessing the similarities of view-specific representations to consistent representation, the adopted weighting manner based on attention mechanism is at the element-level and more flexible in assessing in the importance of different views. Specifically, the attention layer is introduced for the fusion, the calculation is formulated as

$$\mathbf{Z} = \sigma\left(\sum_{v=1}^{V}\mathbf{W}_a(\mathbf{D}^{(v)^{-\frac{1}{2}}}\hat{\mathbf{A}}^{(v)}\mathbf{D}^{(v)^{-\frac{1}{2}}}\mathbf{Z}_{L-1}^{(v)}\mathbf{W}_L^{(v)})\right), \tag{9}$$

where $\sigma(\cdot)$ denotes the activation function, L is the number of layers, $\mathbf{W}_a$ denotes the parameters of attention layer. $\mathbf{Z}$ is the consistent representation incorporating complementary information from multiple views with a balanced manner. Then, the consistent context-aware representation $\mathbf{Z}_c$ can be further obtained via $\mathbf{Z}_c = \mathbf{Z}\mathbf{P}^T$. Simply speaking, we transform the consistent representation $\mathbf{Z}$ into the global coordinate by connecting with the anchors $\mathbf{P}$. Since the anchors span the whole global space, the global manifold structure can be perceived. Therefore, $\mathbf{Z}_c$ is considered to be context aware. Furthermore, to enhance the clustering structure of embedding space, the self-training strategy is imposed on $\mathbf{Z}_c$ with following optimization objective:

$$\mathcal{L}_{clu} = \mathrm{KL}(\mathbf{U}||\mathbf{V}) = \sum_{i}^{N}\sum_{j}^{C}[\mathbf{U}]_{i,j}\log\frac{[\mathbf{U}]_{i,j}}{[\mathbf{V}]_{i,j}}, \tag{10}$$

where C denotes the number of categories, $\mathbf{U} \in \mathbb{R}^{N\times C}$ is the target distribution, $\mathbf{V} \in \mathbb{R}^{N\times C}$ is the prediction distribution. $[\mathbf{V}]_{i,j}$ describes the probability that the i-th sample belongs to the j-th cluster center, which is computed by

$$[\mathbf{V}]_{i,j} = \frac{\left(1 + ||[\mathbf{Z}_c]_i - \boldsymbol{\rho}_j||^2\right)^{-\frac{\xi+1}{2}}}{\sum_{j'}\left(1 + ||[\mathbf{Z}_c]_i - \boldsymbol{\rho}_{j'}||^2\right)^{-\frac{\xi+1}{2}}}, \tag{11}$$

where $\boldsymbol{\rho} = [\boldsymbol{\rho}_1, ..., \boldsymbol{\rho}_C]$ is the learnable cluster centroids, $[\mathbf{Z}_c]_i$ is the embedding for the i-th sample, ξ is the hyperparameter adjusting the tightness of prediction distribution $\mathbf{V}$. Based on $\mathbf{V}$, the target distribution $\mathbf{U}$ is calculated by

$$[\mathbf{U}]_{i,j} = \frac{[\mathbf{V}]_{i,j}^2 / \sum_{i'} [\mathbf{V}]_{i',j}}{\sum_{j'} \left([\mathbf{V}]_{i,j'}^2 / \sum_{j'} [\mathbf{V}]_{i,j'} \right)}. \tag{12}$$

$\mathcal{L}_{clu}$ draws the prediction distribution $\mathbf{V}$ closer to the target distribution $\mathbf{U}$, aiming to make the samples in same clusters more compact and in various clusters more distant. Finally, the overall loss function is formulated as

$$\mathcal{L} = \mathcal{L}_{ce} + \lambda \mathcal{L}_{clu}, \tag{13}$$

where λ denotes the trade-off parameter. The overall loss $\mathcal{L}$ guides the training of multi-view GCNs $\{f_v\}_{v=1}^V$ and global anchors $\mathbf{P}$. After the training process, the consistent context-aware representation $\mathbf{Z}_c$ is explored, then the Kmeans algorithm is used to obtain the final clustering results. Algorithm 1 lists the main steps of the proposed CCARL.

Algorithm 1 The main steps of CCARL

Input: Multi-view graph $\mathcal{G} = \{\mathcal{G}^{(v)}\}_{v=1}^V$, the number of anchors M, the trade-off parameters α and ϵ in fused GW-OT, the regularization parameter λ, training epoch E, learning rate η.
Output: The consistent context-aware representation $\mathbf{Z}_c$.

for $e = 1 : E$ **do**
  for view $v = 1 : V$ **do**
    Calculate the transport matrix $\pi^{(v)}$ via the fused GW-OT;
    // Normalize $\pi^{(v)}$ to the assignment matrix $\mathbf{Q}^{(v)}$ //
    $\mathbf{Q}^{(v)} \leftarrow \text{Normalize}(\pi^{(v)})$ via Eq. (6);
    // Calculate the transfer probability matrix $\mathbf{F}^{(v)}$ //
    $\mathbf{F}^{(v)} \leftarrow \text{Softmax}(\mathbf{Z}^{(v)}\mathbf{P}^T)$ via Eq. (7);
  end for
  Obtain the uniform transfer probability matrix $\mathbf{F}$ via $\mathbf{F} = \frac{1}{V}\sum_v \mathbf{F}^{(v)}$;
  // Calculate the CE loss //
  $\mathcal{L}_{ce} \leftarrow \frac{1}{V}\sum_v \text{CE}(\mathbf{Q}^{(v)}, \mathbf{F})$ via Eq. (8);
  // Embedding fusion via attention mechanism //
  $\mathbf{Z} \leftarrow \text{Attention}(\mathbf{W}_a, \{\mathbf{Z}^{(v)}\}_{v=1}^V)$ via Eq. (9);
  Obtain the consistent context-aware representation $\mathbf{Z}_c$ via $\mathbf{Z}_c = \mathbf{Z}\mathbf{P}^T$;
  Calculate the predicted distribution $\mathbf{V}$ and target distribution $\mathbf{U}$ via Eqs. (11) and (12), respectively;
  // Calculate the clustering loss //
  $\mathcal{L}_{clu} \leftarrow \text{KL}(\mathbf{U}||\mathbf{V})$;
  $\mathcal{L} \leftarrow \mathcal{L}_{ce} + \lambda\mathcal{L}_{clu}$;
  // Update the multi-view GCNs and global anchors //
  $f_v \leftarrow f_v - \eta\nabla\mathcal{L}$;
  $\mathbf{P} \leftarrow \mathbf{P} - \eta\nabla\mathcal{L}$;
end for
Perform Kmeans algorithm on the consistent context-aware representation $\mathbf{Z}_c$ to acquire the clustering results.

4 Experiments

4.1 Datasets

Four real-world multi-view graph datasets are selected to conduct the experiments, including **AMINER**, **BDGP**, **IMDB**, and **YELP**. Specifically, **AMINER** is an academic graph network, recording the relationships of 6,564 papers, four categories and two views are included. **BDGP** is a genetic graph network containing 2,500 samples covering five categories, it records two kinds of topologies. **IMDB** is a movie network containing 4,780 movies with three classes, two kinds of relationships are recorded. **YELP** is a business network with 2,614 nodes covering three classes of nodes, three views are contained.

4.2 Compared Methods

We compare the proposed CCARL with fifteen algorithms, which are classified into three categories. Specifically, **Kmeans** [Krishna and Murty, 1999], **LINE** [Tang *et al.*, 2015], **GAE** [Kipf and Welling, 2016] are three typical single-view clustering algorithm. When using the **Kmeans**, only the data feature is fed. The best performance is reported among all views for **LINE** and **GAE**. **SwMC** [Nie *et al.*, 2017], **GMC** [Wang *et al.*, 2020], **CGL** [Li *et al.*, 2022], **RCAGL** [Liu *et al.*, 2024b] are four conventional MVGC methods with decoupled processing features and topologies. **O2MA** [Fan *et al.*, 2020], **O2MAC** [Fan *et al.*, 2020], **MvAGC** [Lin *et al.*, 2023], **AHMcV** [Pan and Kang, 2023], **CMGEC** [Wang *et al.*, 2023], **LMGEC** [Fettal *et al.*, 2023], **CMAGC** [Chen *et al.*, 2024], **DIAGC** [Lin *et al.*, 2024] are eight GCN-based MVGC approaches.

4.3 Performance Comparison

The experimental results on four multi-view graph datasets are reported in Tables 1 and 2. Three salient observations are revealed. First, the single-view approaches achieve inferior performance, even when picking the best among all views, the results are less than satisfactory, which is because it is difficult to get a clear depiction of samples from the single-view data. Second, the conventional MVGC approaches improves the performance compared to single-view ones, but shallow variables and decoupled graph processing limit their abilities to capture the complicated data distributions. Third, the GCN-based approaches bring the performance to the forefront. Undoubtedly, GCN has better expression capability for graph data. However, all these methods only focus on the aggregation of local information and ignore the global contextual information, which may cause difficulties in clustering. Hence, the proposed CCARL aims to establish the new coordinate in global space, based on which the consistent context-aware representation is explored. This ability helps CCARL achieve the optimal results. In addition, the visualizations on BDGP dataset are presented in Fig. 3. It can be seen that CCARL achieves better intra-class aggregation and inter-class separation compared to other MVGC algorithms.

4.4 Ablation Study

Three important components GW-OT, Attention, ST (Self-training) are ablated to verify their roles, the ablation results are reported in Tables 3 and 4. Notably, when GW-OT is dis-

Method	AMINER						BDGP					
	ACC	NMI	Purity	ARI	Fscore	Precision	ACC	NMI	Purity	ARI	Fscore	Precision
Kmeans	38.68	23.60	69.40	-	53.68	51.69	71.36	58.55	66.20	35.60	63.94	63.73
LINE$_{best}$	26.68	0.10	62.72	0.07	30.12	37.65	21.96	0.17	40.00	-0.02	20.24	20.11
GAE$_{best}$	42.77	13.14	68.76	1.75	44.54	45.91	31.08	7.09	47.76	4.37	25.90	24.62
SwMC	61.64	33.95	73.77	22.21	55.18	48.91	72.44	73.36	77.96	64.39	72.48	63.73
GMC	49.77	15.28	62.16	1.82	48.47	38.32	56.60	51.07	58.60	42.21	57.87	42.64
CGL	51.01	0.93	54.71	2.23	51.50	36.81	36.40	17.17	37.04	13.57	34.05	28.70
RCAGL	50.42	14.11	61.56	0.19	48.01	37.48	78.64	55.84	78.64	55.52	64.60	63.18
O2MA	52.25	35.44	74.36	29.91	57.48	61.59	53.40	35.74	68.68	31.99	45.22	44.24
O2MAC	50.91	34.37	74.50	29.01	56.31	61.04	51.08	34.29	65.00	30.00	43.90	43.12
MvAGC	55.63	16.33	70.88	10.65	55.76	46.41	82.32	69.57	89.56	58.16	74.09	74.65
AHMcV	57.51	31.10	70.64	26.74	60.51	**63.77**	87.20	70.95	87.72	71.06	77.83	78.08
CMGEC	58.83	32.23	77.42	32.29	54.68	59.07	89.20	79.59	91.96	77.41	82.57	82.16
LMGEC	61.76	35.61	80.83	21.57	61.06	59.54	67.52	55.73	67.64	35.49	58.42	60.73
CMAGC	60.91	34.41	74.79	29.63	52.61	59.75	86.88	77.65	88.08	72.89	80.32	79.65
DIAGC	55.24	26.64	75.81	24.44	51.48	56.46	52.80	27.88	56.36	21.42	40.26	39.37
CCARL	**73.96**	**39.11**	**83.47**	**53.94**	**66.30**	63.76	**91.08**	**81.08**	**94.36**	**80.11**	**84.59**	**84.44**

Table 1: Comparison of experimental results (%) on AMINER and BDGP datasets for all compared methods, the optimal results are **bolded** and the suboptimal results are underlined.

Method	IMDB						YELP					
	ACC	NMI	Purity	ARI	Fscore	Precision	ACC	NMI	Purity	ARI	Fscore	Precision
Kmeans	53.43	0.41	55.73	1.79	51.84	40.61	49.08	15.13	81.34	5.25	57.48	43.41
LINE$_{best}$	36.55	0.10	54.58	0.20	36.74	40.47	34.39	0.02	82.02	-0.08	35.17	36.88
GAE$_{best}$	43.87	0.53	54.58	0.46	48.15	40.68	65.15	38.66	82.01	42.33	62.35	61.36
SwMC	54.75	0.46	54.98	0.39	**57.37**	40.51	46.79	9.10	54.59	7.04	41.73	41.20
GMC	54.37	0.29	54.71	0.00	57.37	40.39	43.69	1.54	44.30	0.52	52.22	37.02
CGL	37.68	0.00	54.58	0.00	38.66	40.40	35.34	0.00	42.19	0.00	35.16	36.84
RCAGL	48.09	0.10	54.58	0.25	49.57	40.49	61.17	25.80	66.56	27.81	56.74	52.18
O2MA	49.77	5.18	61.95	5.43	43.71	43.79	65.07	39.02	82.01	42.53	62.40	61.63
O2MAC	48.74	3.50	59.10	4.25	42.53	42.60	46.25	4.74	82.02	5.60	38.77	40.53
MvAGC	55.50	2.30	59.16	5.71	53.22	41.73	70.39	32.93	**83.58**	37.00	61.70	56.68
AHMcV	55.25	1.80	60.21	4.34	31.88	41.50	64.31	42.30	76.17	31.90	66.95	57.67
CMGEC	46.67	0.11	54.58	0.41	47.88	40.45	64.30	38.21	82.01	41.94	61.81	61.53
LMGEC	56.17	3.81	62.41	9.74	51.85	42.79	69.17	44.22	82.02	44.82	65.68	**64.75**
CMAGC	54.06	4.33	63.10	9.06	43.44	43.54	67.71	35.78	82.02	36.65	63.59	54.81
DIAGC	58.39	6.58	66.27	13.16	43.01	45.06	65.34	39.26	82.02	42.85	62.64	61.73
CCARL	**61.51**	**9.59**	**68.24**	**16.60**	53.45	**46.25**	**73.60**	**44.83**	82.02	**46.34**	**70.97**	61.57

Table 2: Comparison of experimental results (%) on IMDB and YELP datasets for all compared methods, the optimal results are **bolded** and the suboptimal results are underlined.

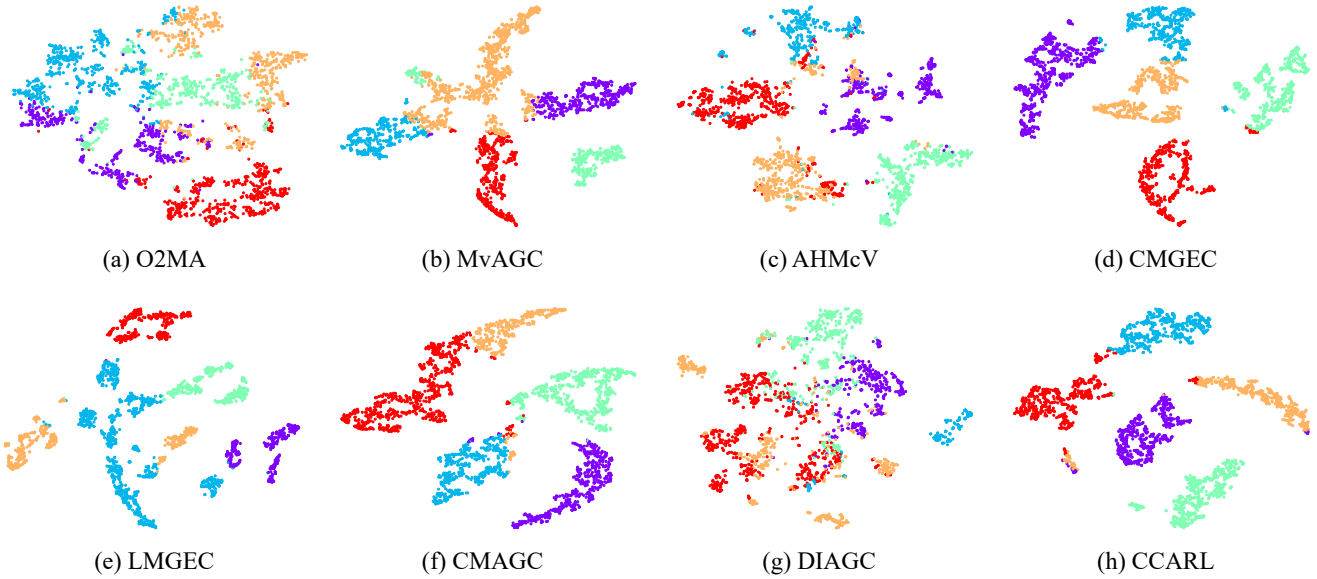
carded, only the general OT is used for distribution alignment at the node feature level. Overall, each component plays an influential role in achieving the best results, the absence of any one can result in the performance degradation. Interestingly, the performance drops the most when the self-training strategy is removed, which means it has a more crucial effect. This is because the clustering task puts more emphasis on intra-class aggregation.

GW-OT	Attention	ST	AMINER			
			ACC	NMI	ARI	Fscore
✗	✓	✓	63.54	36.47	38.82	57.98
✓	✗	✓	69.15	34.53	48.13	59.11
✓	✓	✗	62.07	23.42	31.58	50.96
✓	✓	✓	**73.96**	**39.11**	**53.94**	**66.30**

Table 3: Ablation results (%) with respect to three components on AMINER dataset, where ST denotes the self-training strategy.

4.5 Parameter Sensitivity Investigation

The proposed CCARL contains three important trade-off parameters for balancing the optimization objective, including α, ϵ, and λ. Fig. 4 shows the effect on the results with different values for α and ϵ, where α and ϵ are varied in $\{0.1, ..., 0.9\}$, $\{0.01, ..., 1\}$, respectively. It can be observed that different datasets are not equally tolerant to the value intervals, the proposed CCARL is more sensitive on AMINER dataset than YELP dataset, indicating that suitable values are



(a) O2MA (b) MvAGC (c) AHMcV (d) CMGEC

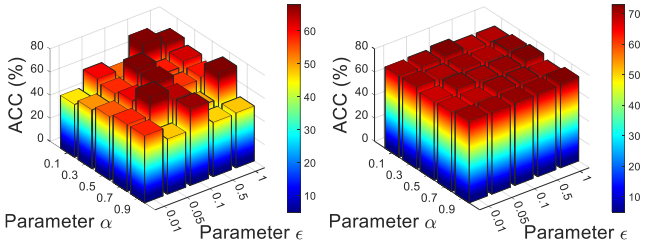
(e) LMGEC (f) CMAGC (g) DIAGC (h) CCARL

Figure 3: Scatter visualizations on BDGP dataset for eight MVGL methods, where the t-SNE is used for dimension reduction and different colors indicate different clusters.

necessary. Analogously, this situation is also shown for λ in Fig. 5, a larger λ induces a significant performance decrease on AMINER dataset. The reason for this is that the target distribution is inaccurate at the early training, then causing the model to converge to a undesirable solution.

GW-OT	Attention	ST	YELP			
			ACC	NMI	ARI	Fscore
✗	✓	✓	71.77	40.65	43.85	68.41
✓	✗	✓	72.49	41.13	44.84	68.96
✓	✓	✗	68.86	36.84	41.28	64.03
✓	✓	✓	**73.60**	**44.83**	**46.34**	**70.97**

Table 4: Ablation results (%) with respect to three components on YELP dataset, where ST denotes the self-training strategy.

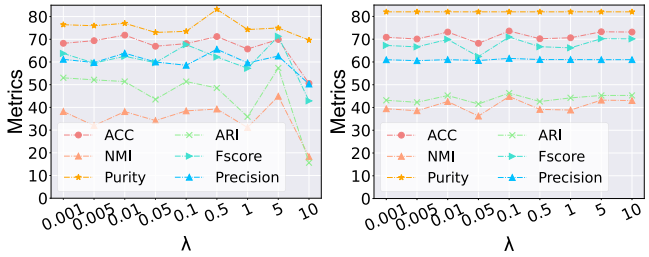


(a) AMINER (b) YELP

Figure 4: Sensitivity study with respect to α and ϵ on AMINER and YELP datasets, where α and ϵ are varied in $\{0.1, ..., 0.9\}$, $\{0.01, ..., 1\}$.

5 Conclusion

In this paper, we propose a novel consistent context-aware representation learning approach for MVGC, which learns from the global space and is not limited to the local structures. To outline the global space, a set of anchors are defined. Then, the anchor graph is optimally mapped to multi-view graphs via the fused GW-OT. The attention mechanism is employed for adaptively fusing the complementary information in multiple graph embeddings. After that, a consistent representation is obtained. Through the transformation of global coordinate spanned by anchors, the consistent representation is equipped with the contextual information, achieving the better clustering. However, the multi-view graph data is frequently incomplete in the practical applications, which imposes a tricky challenge for the distribution matching process via the fused GW-OT. Therefore, in the future work, we will further optimize the fused GW-OT procedure to adapt it to the incomplete scenario.



(a) AMINER (b) YELP

Figure 5: Sensitivity study with respect to λ on AMINER and YELP datasets, where λ is varied in $\{0.001, ..., 10\}$.

Acknowledgments

The research is supported by the National Key R&D Program of China (2023YFB2703700), the National Natural Science Foundation of China (62176269).

Contribution Statement

Lele Fu and Bowen Deng contributed equally as co-first authors.

References

[Arjovsky *et al.*, 2017] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *ICML*, pages 214–223, 2017.

[Cai *et al.*, 2024] Jinyu Cai, Yunhe Zhang, Jicong Fan, Yali Du, and Wenzhong Guo. Dual contrastive graph-level clustering with multiple cluster perspectives alignment. In *Proceedings of IJCAI*, pages 3770–3779, 2024.

[Chen *et al.*, 2021] Yongyong Chen, Xiaolin Xiao, Chong Peng, Guangming Lu, and Yicong Zhou. Low-rank tensor graph learning for multi-view subspace clustering. *IEEE TCSVT*, 32(1):92–104, 2021.

[Chen *et al.*, 2023] Zhaoliang Chen, Lele Fu, Jie Yao, Wenzhong Guo, Claudia Plant, and Shiping Wang. Learnable graph convolutional network and feature fusion for multi-view learning. *Information Fusion*, 95:109–119, 2023.

[Chen *et al.*, 2024] Man-Sheng Chen, Xi-Ran Zhu, Jia-Qi Lin, and Chang-Dong Wang. Contrastive multiview attribute graph clustering with adaptive encoders. *IEEE TNNLS*, 2024. doi=10.1109/TNNLS.2024.3391801.

[Courty *et al.*, 2016] Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE TPAMI*, 39(9):1853–1865, 2016.

[Deng *et al.*, 2025] Bowen Deng, Tong Wang, Lele Fu, Sheng Huang, Chuan Chen, and Tao Zhang. THESAURUS: contrastive graph clustering by swapping fused gromov-wasserstein couplings. In *Proceedings of AAAI*, pages 16199–16207, 2025.

[Fan *et al.*, 2020] Shaohua Fan, Xiao Wang, Chuan Shi, Emiao Lu, Ken Lin, and Bai Wang. One2multi graph autoencoder for multi-view graph clustering. In *WWW*, pages 3070–3076, 2020.

[Feng *et al.*, 2024] Wei Feng, Zhenwei Wu, Qianqian Wang, Bo Dong, Zhiqiang Tao, and Quanxue Gao. Federated multi-view clustering via tensor factorization. In *IJCAI*, pages 3962–3970, 2024.

[Fettal *et al.*, 2023] Chakib Fettal, Lazhar Labiod, and Mohamed Nadif. Simultaneous linear multi-view attributed graph representation learning and clustering. In *WSDM*, pages 303–311, 2023.

[Fu *et al.*, 2020] Lele Fu, Pengfei Lin, Athanasios V Vasilakos, and Shiping Wang. An overview of recent multi-view clustering. *Neurocomputing*, 402:148–161, 2020.

[Fu *et al.*, 2022] Lele Fu, Zhaoliang Chen, Yongyong Chen, and Shiping Wang. Unified low-rank tensor learning and spectral embedding for multi-view subspace clustering. *IEEE TMM*, 25:4972–4985, 2022.

[Fu *et al.*, 2024] Lele Fu, Sheng Huang, Lei Zhang, Jinghua Yang, Zibin Zheng, Chuanfu Zhang, and Chuan Chen. Subspace-contrastive multi-view clustering. *ACM TKDD*, 18(9), 2024.

[Groppe and Hundrieser, 2024] Michel Groppe and Shayan Hundrieser. Lower complexity adaptation for empirical entropic optimal transport. *JMLR*, 25(344):1–55, 2024.

[Guan *et al.*, 2024] Renxiang Guan, Zihao Li, Wenxuan Tu, Jun Wang, Yue Liu, Xianju Li, Chang Tang, and Ruyi Feng. Contrastive multiview subspace clustering of hyperspectral images based on graph convolutional networks. *IEEE TGRS*, 62:1–14, 2024.

[Huang *et al.*, 2022] Shudong Huang, Ivor W. Tsang, Zenglin Xu, and Jiancheng Lv. Measuring diversity in graph learning: A unified framework for structured multi-view clustering. *IEEE TKDE*, 34(12):5869–5883, 2022.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.

[Krishna and Murty, 1999] K Krishna and M Narasimha Murty. Genetic k-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 29(3):433–439, 1999.

[Li *et al.*, 2021] Zhenglai Li, Chang Tang, Xinwang Liu, Xiao Zheng, Wei Zhang, and En Zhu. Consensus graph learning for multi-view clustering. *IEEE TMM*, 24:2461–2472, 2021.

[Li *et al.*, 2022] Zhenglai Li, Chang Tang, Xinwang Liu, Xiao Zheng, Wei Zhang, and En Zhu. Consensus graph learning for multi-view clustering. *IEEE Transactions on Multimedia*, 24:2461–2472, 2022.

[Li *et al.*, 2023a] Jing Li, Quanxue Gao, Qianqian Wang, Ming Yang, and Wei Xia. Orthogonal non-negative tensor factorization based multi-view clustering. In *NIPS*, pages 18186–18202, 2023.

[Li *et al.*, 2023b] Xingfeng Li, Zhenwen Ren, Quansen Sun, and Zhi Xu. Auto-weighted tensor schatten p-norm for robust multi-view graph clustering. *Pattern Recognition*, 134:109083, 2023.

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

[Lin *et al.*, 2023] Zhiping Lin, Zhao Kang, Lizong Zhang, and Ling Tian. Multi-view attributed graph clustering. *IEEE TKDE*, 35(2):1872–1880, 2023.

[Lin *et al.*, 2024] Jia-Qi Lin, Man-Sheng Chen, Xi-Ran Zhu, Chang-Dong Wang, and Haizhang Zhang. Dual information enhanced multiview attributed graph clustering. *IEEE TNNLS*, 2024. doi=10.1109/TNNLS.2024.3401449.

[Liu *et al.*, 2023] Yang Liu, Zhipeng Zhou, and Baigui Sun. Cot: Unsupervised domain adaptation with clustering and optimal transport. In *CVPR*, pages 19998–20007, 2023.

[Liu *et al.*, 2024a] Chengliang Liu, Gehui Xu, Jie Wen, Yabo Liu, Chao Huang, and Yong Xu. Partial multi-view multi-label classification via semantic invariance learning and prototype modeling. In *Proceedings of ICML*, 2024.

[Liu *et al.*, 2024b] Suyuan Liu, Qing Liao, Siwei Wang, Xinwang Liu, and En Zhu. Robust and consistent anchor graph learning for multi-view clustering. *IEEE TKDE*, 36(8):4207–4219, 2024.

[Liu *et al.*, 2024c] Yue Liu, Xihong Yang, Sihang Zhou, Xinwang Liu, Siwei Wang, Ke Liang, Wenxuan Tu, and Liang Li. Simple contrastive graph clustering. *IEEE TNNLS*, 35(10):13789–13800, 2024.

[Liu *et al.*, 2025] Chengliang Liu, Jie Wen, Yong Xu, Bob Zhang, Liqiang Nie, and Min Zhang. Reliable representation learning for incomplete multi-view missing multi-label classification. *IEEE TPAMI*, pages 1–17, 2025.

[Ma *et al.*, 2024] Xinyu Ma, Xu Chu, Yasha Wang, Yang Lin, Junfeng Zhao, Liantao Ma, and Wenwu Zhu. Fused gromov-wasserstein graph mixup for graph-level classifications. *NIPS*, 36, 2024.

[Mémoli, 2011] Facundo Mémoli. Gromov–wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11:417–487, 2011.

[Montesuma *et al.*, 2024] Eduardo Fernandes Montesuma, Fred Maurice Ngolè Mboula, and Antoine Souloumiac. Recent advances in optimal transport for machine learning. *IEEE TPAMI*, 2024.

[Nie *et al.*, 2017] Feiping Nie, Jing Li, Xuelong Li, et al. Self-weighted multiview clustering with multiple graphs. In *IJCAI*, pages 2564–2570, 2017.

[Pan and Kang, 2021] Erlin Pan and Zhao Kang. Multi-view contrastive graph clustering. pages 2148–2159, 2021.

[Pan and Kang, 2023] Erlin Pan and Zhao Kang. High-order multi-view clustering for generic data. *Information Fusion*, 100:101947, 2023.

[Pan *et al.*, 2024] Baicheng Pan, Chuandong Li, Hangjun Che, Man-Fai Leung, and Keping Yu. Low-rank tensor regularized graph fuzzy learning for multi-view data processing. *IEEE TCE*, 70(1):2925–2938, 2024.

[Qian *et al.*, 2024] Xiaowei Qian, Bingheng Li, and Zhao Kang. Upper bounding barlow twins: A novel filter for multi-relational clustering. In *AAAI*, pages 14660–14668, 2024.

[Qin *et al.*, 2024] Yalan Qin, Zhenjun Tang, Hanzhou Wu, and Guorui Feng. Flexible tensor learning for multi-view clustering with markov chain. *IEEE TKDE*, 36(4):1552–1565, 2024.

[Tang *et al.*, 2015] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In *WWW*, pages 1067–1077, 2015.

[Tang *et al.*, 2022] Chang Tang, Zhenglai Li, Jun Wang, Xinwang Liu, Wei Zhang, and En Zhu. Unified one-step multi-view spectral clustering. *IEEE TKDE*, 35(6):6449–6460, 2022.

[Titouan *et al.*, 2019] Vayer Titouan, Nicolas Courty, Romain Tavenard, and Rémi Flamary. Optimal transport for structured data with application on graphs. In *ICML*, pages 6275–6284, 2019.

[Tolstikhin *et al.*, 2018] Ilya Tolstikhin, Olivier Bousquet, Sylvain Gelly, and Bernhard Schoelkopf. Wasserstein auto-encoders. In *ICLR*, 2018.

[Tsitsulin *et al.*, 2023] Anton Tsitsulin, John Palowitch, Bryan Perozzi, and Emmanuel Muller. Graph clustering with graph neural networks. *JMLR*, 24(127):1–21, 2023.

[Tu *et al.*, 2021] Wenxuan Tu, Sihang Zhou, Xinwang Liu, Xifeng Guo, Zhiping Cai, En Zhu, and Jieren Cheng. Deep fusion clustering network. In *AAAI*, pages 9978–9987, 2021.

[Tu *et al.*, 2024] 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 *AAAI*, pages 15392–15401, 2024.

[Wang *et al.*, 2020] Hao Wang, Yan Yang, and Bing Liu. Gmc: Graph-based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 32(6):1116–1129, 2020.

[Wang *et al.*, 2023] Yiming Wang, Dongxia Chang, Zhiqiang Fu, and Yao Zhao. Consistent multiple graph embedding for multi-view clustering. *IEEE TMM*, 25:1008–1018, 2023.

[Wang *et al.*, 2024] Siwei Wang, Xinwang Liu, Suyuan Liu, Wenxuan Tu, and En Zhu. Scalable and structural multi-view graph clustering with adaptive anchor fusion. *IEEE TIP*, 33:4627–4639, 2024.

[Xia *et al.*, 2022] Wei Xia, Quanxue Gao, Qianqian Wang, Xinbo Gao, Chris Ding, and Dacheng Tao. Tensorized bipartite graph learning for multi-view clustering. *IEEE TPAMI*, 45(4):5187–5202, 2022.

[Yang *et al.*, 2022] Jing-Hua Yang, Chuan Chen, Hong-Ning Dai, Meng Ding, Le-Le Fu, and Zibin Zheng. Hierarchical representation for multi-view clustering: From intra-sample to intra-view to inter-view. In *Proceedings of CIKM*, pages 2362–2371, 2022.

[Yu *et al.*, 2025] Zixiao Yu, Lele Fu, Yongyong Chen, Zhiling Cai, and Guoqing Chao. Hyper-laplacian regularized concept factorization in low-rank tensor space for multi-view clustering. *IEEE TETCI*, 9(2):1728–1742, 2025.

[Zhang *et al.*, 2023] Lei Zhang, Lele Fu, Tong Wang, Chuan Chen, and Chuanfu Zhang. Mutual information-driven multi-view clustering. In *Proceedings of CIKM*, pages 3268–3277, 2023.

[Zhang *et al.*, 2025] Yunhe Zhang, Jinyu Cai, Zhihao Wu, Pengyang Wang, and See-Kiong Ng. Mixture of experts as representation learner for deep multi-view clustering. In *Proceedings of AAAI*, pages 22704–22713, 2025.

[Zhuang *et al.*, 2024] Shuman Zhuang, Sujia Huang, Wei Huang, Yuhong Chen, Zhihao Wu, and Ximeng Liu. Enhancing multi-view graph neural network with cross-view confluent message passing. In *ACM MM*, pages 10065–10074, 2024.