FairGC: Fostering Individual and Group Fairness for Deep Graph Clustering

Haodong Zhang[1*], Xinyue Wang[1*], Tao Ren[1], Yifan Wang[2†], Siyu Yi[3], Fanchun Meng[1], Zeyu Ma[4], Qingqing Long[5], Wei Ju[6]

[1]Software College, Northeastern University, Shenyang, China
[2]School of Artificial Intelligence and Data Science, University of International Business and Economics, Beijing, China
[3]College of Mathematics, Sichuan University, Chengdu, China
[4]Jiangnan University, Wuxi, China
[5]Computer Information Center, Chinese Academy of Sciences, Beijing, China
[6]College of Computer Science, Sichuan University, Chengdu, China
2110496@stu.neu.edu.cn, chinaxywang@163.com, chinarentao@163.com, yifanwang@uibe.edu.cn, siyuyi@scu.edu.cn, chinafcmeng@163.com, zeyu.ma@stu.hit.edu.cn, qqlong@cnic.cn, juwei@scu.edu.cn

Abstract

The widespread adoption of graph neural networks (GNNs) has brought increased attention to fairness issues related to sensitive attributes, such as gender and race, in practical scenarios. However, this concern remains largely unexplored in the context of graph clustering. Conventional fair graph clustering methods primarily depend on spectral clustering approaches. Meanwhile, we argue that existing graph learning works mainly focus on a single type of fairness, whereas graph clustering should achieve group equality-informed individual fairness. In this paper, we introduce for the first time a fairness-aware framework termed FairGC for deep graph clustering, which integrates the dual objectives of individual and group fairness while maintaining accurate clustering results. Specifically, we construct two views with distinct semantics using Siamese encoders. Then, we apply multi-step random walks on view-specific affinity graphs to capture high-order affinities of node pairs, thereby reformulating the contrastive learning with a focus on individual similarity. Besides, we utilize adversarial learning by making node representations independent of the estimated sensitive attributes to further eliminate group biases of clustering results. Extensive experiments on four benchmarks demonstrate the effectiveness and superiority of our proposed framework FairGC.

Introduction

Graphs serve as a crucial data structure for representing complex relationships and interactions among objects and have been extensively studied over the years. In particular, graph learning algorithms, which leverage rich information encoded from graphs, have shown great potential in various tasks like link prediction (Cai and Ji 2020; Zhang and Chen 2018), node/graph classification (Tu et al. 2024b, 2025; Zhang et al. 2025), and anomaly detection (Liu et al. 2023b,a). Among these various directions of graph learning, one key and challenging task, namely graph clustering, has recently attracted intensive attention and has been extensively applied across various domains, including social network analysis (Newman 2006), community detection (Tu et al. 2018) and recommender system (Liu et al. 2024b,c).

The main idea of graph clustering is to divide the nodes into distinct groups, ensuring that nodes within the same cluster share higher similarity than those from different clusters (Ren et al. 2025). Thanks to the powerful capabilities of deep learning, particularly graph neural networks (GNNs), a number of deep graph clustering approaches have been

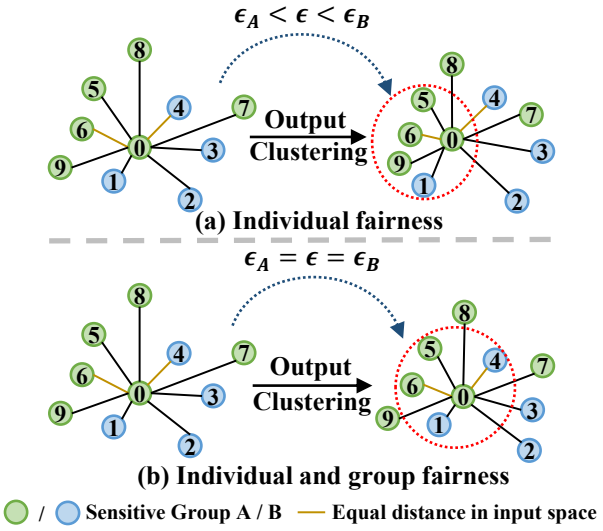


Figure 1: An illustration of the disparity in individual fairness among different sensitive groups. (a) Individual fairness fails to impose equal constraints on similar outputs between groups. For example, given node v_6 in group A and v_4 in group B, which share equal distances to v_0, are identified to be distinct clusters due to their unequal constraining scalars between the groups, namely $\epsilon_A < \epsilon_B$. (b) By incorporating group fairness, the constraint scalar for each group can be maintained at the same level, namely $\epsilon_A = \epsilon = \epsilon_B$, yielding a better fairness-aware clustering result.

*These authors contributed equally.
†Correspondence author.

proposed. Reconstructive and adversarial methods aim at learning cluster-oriented node representations by recovering graph information and using fake sample generation-recognition mechanisms, respectively (Liu et al. 2022b; Yi et al. 2023). More recently, contrastive methods reduce distortion relative to pre-learned clustering centers, and more discriminative contrastive loss functions are designed for network training (Yang et al. 2023; Liu et al. 2023d).

Despite their encouraging performance, fairness in graph clustering still remains largely underexplored. Intuitively, clustering results could be affected by sensitive attributes such as gender and race (Caton and Haas 2024), and unfair graph clustering can result in discriminatory outcomes. For example, a particular group is more likely to be clustered together, which can perpetuate stereotypes or result in unequal treatment. Although a few studies investigate graph partitions with fairness constraints (Wang et al. 2023; Li, Wang, and Merchant 2023; Ghodsi, Seyedi, and Ntoutsi 2024), they rely on classical spectral clustering methods, which struggle to handle high-dimensional node features effectively. Thus, it is essential to implement deep fair graph clustering methods that are unbiased by sensitive node attributes.

In the literature, existing fair graph learning works mainly focus on a single type of fairness. Individual fairness ignores sensitive features and ensures that individuals who are similar in non-sensitive features receive similar treatment or outcomes from the system (Kang et al. 2020), while group fairness strives to provide equal outcome rates across various demographic groups defined by sensitive features (Dai and Wang 2021). Nevertheless, as an example illustrated in Figure. 1, given a graph with two sensitive groups, since individual fairness approaches often fail to enforce equal constraints on similarity outputs between the groups, it is possible that nodes from the two groups, despite having the same similarity in the input space, may be assigned to different clusters. Thus, an ideal way for fairness-aware graph clustering is to achieve individual and group fairness simultaneously. In this way, similar nodes should yield similar representations, which are more likely to be distributed in the same cluster. Meanwhile, nodes in each sensitive group are approximately proportionally represented in each cluster.

Motivated by the observation, in this paper, we propose **FairGC**, a unified Fairness-aware framework fostering individual and group fairness for deep Graph Clustering. The core idea of our FairGC is to constrain group equality informed individual fairness and ensure that each cluster maintains the same proportion of groups as in the original dataset for the clustering results. Specifically, we take advantage of Siamese encoders to construct two views with different semantics. Then, given the calculated affinity graph of the two views in the embedding space, we apply the multi-step random walks for each node to obtain the high-order affinity of node pairs globally. A fairness-aware cross-view contrastive learning approach is proposed between the two views to ensure that similar nodes receive similar embeddings, thereby promoting individual fairness. Furthermore, we use adversarial learning to ensure group fairness by adding a discriminator to eliminate the sensitive information for clustering.

To summarize, we make the following contributions:

- *Conceptual*: We highlight the sensitive attributes and introduce the novel problem of fairness-aware deep graph clustering. To the best of our knowledge, we are the first to study fairness in deep graph clustering.
- *Methodological*: We propose a unified framework that not only leverages the fairness-aware cross-view contrastive learning but also introduces fairness-aware adversarial learning to foster individual and group fairness.
- *Experimental*: We conduct extensive experiments on 4 benchmark datasets to evaluate the performance of FairGC. The experimental results demonstrate the superiority of our proposed framework, while also maintaining fairness considerations for deep graph clustering.

Related Work

Deep Graph Clustering

Deep graph clustering has emerged as a prominent framework aimed at partitioning graph nodes into disjoint clusters using neural networks (Liu et al. 2023c; Tu et al. 2024a). From the perspective of the learning paradigm, deep graph clustering can be broadly classified into three types: 1) Reconstructive methods, which prioritize the use of intra-data information within the graph, learning meaningful representations by integrating both graph structure and node attributes (Liu et al. 2025b,a); 2) Adversarial methods, which enhance the quality of node representations through an adversarial training scheme between the generator and the discriminator (Pan et al. 2019; Gong et al. 2022); 3) Contrastive methods, which improves feature discriminativeness by drawing positive sample pairs closer while pushing negative sample pairs apart (Cui et al. 2020; Yang et al. 2023; Xia et al. 2022). However, prior efforts tend to propagate information among nodes with the same sensitive attributes due to the message-passing mechanism in GNNs, significantly heightening the risk of discrimination towards sensitive attributes. Despite their great success in clustering, they may fall into the trap of fairness issues. In this paper, we propose a scheme that integrates individual and group fairness to ensure effective clustering while preserving fairness.

Fairness-Aware Graph Learning

Fairness-aware graph learning has garnered considerable attention as the application of GNNs has widely expanded in the field of recommendation systems and social networks. And previous fairness-aware graph learning can be broadly divided into three categories: group fairness (Guo, Chu, and Li 2023; Dai and Wang 2021; Li et al. 2024), individual fairness (Song et al. 2022; Ghodsi, Seyedi, and Ntoutsi 2024; Zhan et al. 2024), and counterfactual fairness (Agarwal, Lakkaraju, and Zitnik 2021; Guo et al. 2023). For example, FatraGNN (Li et al. 2024) focuses on node classification by minimizing the representation distances for each sensitive group through adversarial learning. Guide (Song et al. 2022) equalizes the level of individual fairness across different groups in node classification. CAF (Guo et al. 2023) attains model fairness from a causal perspective by contrasting original scenarios with their counterfactual alternatives.

While these approaches show satisfactory performance across extensive benchmarks, they primarily focus on supervised learning tasks, such as node classification (Guo et al. 2023; Song et al. 2022) and link prediction (Cao et al. 2023), leaving graph clustering—an unsupervised learning scheme—largely unexplored. Some conventional methods, such as FNM (Li, Wang, and Merchant 2023) and iFairNMTF (Ghodsi, Seyedi, and Ntoutsi 2024), have made efforts to enhance fairness in clustering. However, they fail to handle high-dimensional node features, resulting in poor clustering performance. To the best of our knowledge, this is the first study addressing fairness-aware deep graph clustering.

Preliminary and Problem Definition

Notations. Let a tuple $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represent an undirected graph, where $\mathcal{E}$ denotes the set of edges and $\mathcal{V}$ denotes the set of N nodes. The adjacency matrix $\boldsymbol{A} = (a_{uv}) \in \mathbb{R}^{N \times N}$ characterizes the structure of the graph, with $a_{uv} = 1$ if $(u, v) \in \mathcal{E}$, otherwise, $a_{uv} = 0$. The degree matrix is given by $\boldsymbol{D} = \text{diag}(d_1, \ldots, d_N)$, where $d_i = \sum_{j=1}^{N} a_{ij}$. The Laplacian matrix can be defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ and is normalized to $\tilde{\boldsymbol{L}} = \boldsymbol{I} - \hat{\boldsymbol{D}}^{-\frac{1}{2}} \hat{\boldsymbol{A}} \hat{\boldsymbol{D}}^{-\frac{1}{2}}$, with $\hat{\boldsymbol{A}} = \boldsymbol{A} + \boldsymbol{I}$ with self-connections, and $\boldsymbol{I} \in \mathbb{R}^{N \times N}$ is the identity matrix. The node attribute set $\mathcal{X} = (\boldsymbol{X}, \boldsymbol{S})$ consists of non-sensitive attributes $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) \in \mathbb{R}^{N \times d'}$, where $\boldsymbol{x}_v$ is the feature vector of node v with dimension d', and sensitive attributes $\boldsymbol{S} = (s_1, \ldots, s_N) \in \mathbb{R}^N$, where $s_v \in \{1, \ldots, M\}$.

Deep Graph Clustering. Given an unlabeled graph with N nodes, deep graph clustering focuses on dividing these nodes into several disjoint groups without human annotations. Specifically, node embeddings $\boldsymbol{Z} \in \mathbb{R}^{N \times d}$ are first learned in an unsupervised manner by encoding the node and structure attributes with a deep neural network $\mathcal{F}(\cdot)$:

$$\boldsymbol{Z} = \mathcal{F}(\boldsymbol{A}, \boldsymbol{X}). \tag{1}$$

Then, a clustering algorithm (e.g., spectral clustering, K-means, or the clustering neural network layer (Bo et al. 2020)) can be employed to divide these nodes into K groups $\{\mathcal{C}_1, \ldots, \mathcal{C}_K\}$ based on the learned embedding.

Individual and Group Fairness. Individual fairness emphasizes that similar inputs should receive consistent and fair treatment. Let $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ be two nodes of $\mathcal{G}$, the output of model $\boldsymbol{z}_u$ and $\boldsymbol{z}_v$ are *individually fair* w.r.t. the input node similarity T_{in} and output distance measure T_{out} if the following Lipschitz constraint holds.

$$T_{out}(\boldsymbol{z}_u, \boldsymbol{z}_v) \leq \epsilon \cdot T_{in}(\boldsymbol{x}_u, \boldsymbol{x}_v) \; \forall u, v = 1, \ldots, N, \tag{2}$$

where $\epsilon > 0$ is a constant that rescales the level of the input distance. Moreover, group fairness is satisfied if ϵ across all groups is equal. Formally, we denote the node partition induced by the sensitive attribute as $\{\mathcal{V}_1, \ldots, \mathcal{V}_M\}$. The clustering result is *group fair* if each cluster maintains the same proportion of nodes from each group as observed in the original dataset, which can be defined as:

$$\frac{|\mathcal{V}_m \cap \mathcal{C}_k|}{|\mathcal{C}_k|} = \frac{|\mathcal{V}_m|}{|\mathcal{V}|}, \forall m = 1, \ldots, M, \tag{3}$$

where $\mathcal{C}_k$ denotes the k-th cluster, and M is the number of classes for sensitive attribute.

The Proposed Framework

Overview

The fundamental concept of our `FairGC` emphasizes the sensitive attribute of nodes, ensuring the encoding of both individual and group fairness-aware representations. As shown in Figure. 2, there are three components in our `FairGC` framework. Given the input undirected graph, we first employ two Siamese encoders to embed attribute and structure information of the node into the latent space as two views. Then, the affinity graphs are constructed, and multi-step random walks are performed for two views to identify the node pair similarities in a global manner. We treat the similarities of one view as input space similarities, and individual fairness-aware contrastive learning is performed to preserve the cross-view consistency. Finally, we seek to leverage adversarial learning to make the clustering results independent of the sensitive attributes for group fairness.

Attribute and Structure Encoding

Since both the attributes and structure of nodes are crucial for calculating node pair similarity (Liu et al. 2023d), we adopt two types of graph encoders in this section for node attribute and structure encoding.

Attribute Encoding. For attribute encoding, we adopt a widely used Laplacian filter (Cui et al. 2020) to conduct neighborhood attribute aggregation and filter out high-frequency noises in $\boldsymbol{X}$ as follows:

$$\tilde{\boldsymbol{X}} = (\prod_{i=1}^{l} (1 - \tilde{\boldsymbol{L}})) \boldsymbol{X} = (1 - \tilde{\boldsymbol{L}})^l \boldsymbol{X}, \tag{4}$$

where $(1 - \tilde{\boldsymbol{L}})^l$ is the graph Laplacian filter stacking up l times. $\tilde{\boldsymbol{X}}$ is the filtered feature matrix and is encoded with two encoders AE_1 and AE_2. Taking AE_1 as an example:

$$\begin{aligned} \boldsymbol{Z}_1 &= (\boldsymbol{z}_1^{v_1}, \ldots, \boldsymbol{z}_1^{v_N}) = \text{AE}_1(\tilde{\boldsymbol{X}}) \\ \boldsymbol{z}_1^{v_i} &= \frac{\boldsymbol{z}_1^{v_i}}{\|\boldsymbol{z}_1^{v_i}\|_2}, \; i = 1, \ldots, N, \end{aligned} \tag{5}$$

where $\boldsymbol{Z}_1$ is the attribute embedding of the first view. In practice, we simply utilize multi-layer perceptions (MLPs) for the implementation of AE_1 and AE_2.

Structure Encoding. For structure encoding, we further leverage two encoders SE_1 and SE_2 and take adjacency matrix $\boldsymbol{A}$ as input. For instance, consider SE_1 as:

$$\begin{aligned} \boldsymbol{E}_1 &= (\boldsymbol{e}_1^{v_1}, \ldots, \boldsymbol{e}_1^{v_N}) = \text{SE}_1(\boldsymbol{A}) \\ \boldsymbol{e}_1^{v_i} &= \frac{\boldsymbol{e}_1^{v_i}}{\|\boldsymbol{e}_1^{v_i}\|_2}, \; i = 1, \ldots, N, \end{aligned} \tag{6}$$

where $\boldsymbol{E}_1$ denotes the structure embedding of the first view. Similarly, SE_1 and SE_2 can be simple MLPs.

Individual Fairness-Aware Contrastive Learning

To promote the individual fairness of the learned representation, we construct affinity graphs via random walks (Lu et al. 2023) to capture high-order node similarity and conduct cross-view contrastive learning to satisfy the constraint.

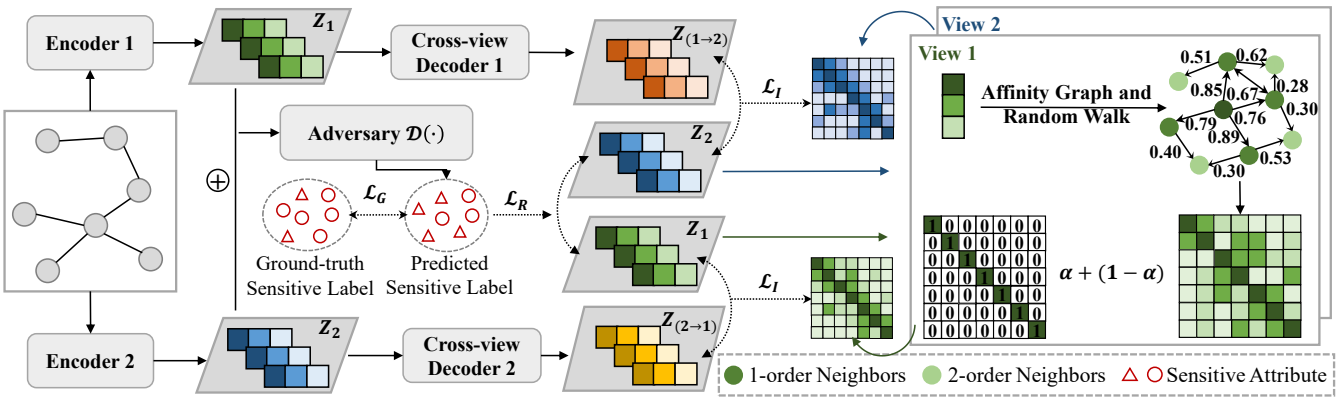


Figure 2: An overview of our FairGC. We first extract attribute and structure information of the node from two views. Then, the affinity graphs are constructed via a multi-step random walk on node pair similarities. Finally, we leverage adversarial learning to make the clustering results independent of the sensitive attributes.

Affinity Graph via Random Walks. Given that node pair similarity depends on both structure and attribute information, we calculate the attribute-structure similarity of the node pairs for each view, formulated as:

$$r_{uv}^{(1)} = \rho \cdot (\boldsymbol{z}_1^u)^{\mathrm{T}} \boldsymbol{z}_1^v + (1-\rho) \cdot (\boldsymbol{e}_1^u)^{\mathrm{T}} \boldsymbol{e}_1^v, \tag{7}$$

where ρ denotes the learnable parameter, and we could obtain the similarity matrix $\boldsymbol{R}_1 = (r_{uv}^{(1)}) \in \mathbb{R}^{N \times N}$ for the first view. In this way, the random walk transition matrix $\tilde{\boldsymbol{R}}_1$ is obtained by row-wise normalization of $\boldsymbol{R}_1$. The multi-step transition matrix, $\tilde{\boldsymbol{R}}_1^t$, is then derived by raising the transition matrix to the power of t, which represents the global similarity of node pairs. Finally, we use the $\tilde{\boldsymbol{R}}_1^t$ and identity matrix $\boldsymbol{I}_N$ to construct an affinity graph of the first view, i.e.,

$$\boldsymbol{T}_1 = \alpha \boldsymbol{I}_N + (1-\alpha)\tilde{\boldsymbol{R}}_1^t, \tag{8}$$

where α is a trade-off parameter between self and node pair similarities. In the implementation, we construct the affinity graph for one view and use it as the pseudo target similarity in the input space for the other view accordingly.

Cross-view Contrastive Learning. To preserve the node pair similarity under the individual fairness constraint, we propose cross-view contrastive learning by maximizing the consistency between the embedding similarity from one view and the pseudo target similarity:

$$\mathcal{L}_I = H(\boldsymbol{T}_1, p(\boldsymbol{Z}_{(2\to1)}, \boldsymbol{Z}_1)) + H(\boldsymbol{T}_2, p(\boldsymbol{Z}_{(1\to2)}, \boldsymbol{Z}_2)), \tag{9}$$

where $H(\cdot,\cdot)$ is the cross entropy, $p(\cdot,\cdot)$ denotes the pairwise similarity with the row-wise normalization operator,

$$[p(\boldsymbol{Z}_{(1\to2)}, \boldsymbol{Z}_1)]_{uv} = \frac{\exp(q(\boldsymbol{z}_{(1\to2)}^u, \boldsymbol{z}_2^v)/\tau)}{\sum_{v'=1}^{N} \exp(q(\boldsymbol{z}_{(1\to2)}^u, \boldsymbol{z}_2^{v'})/\tau)}, \tag{10}$$

where $q(\cdot,\cdot)$ is the similarity function, i.e., cosine similarity, τ is the temperature parameter. Notice that cross-view contrastive learning can be degraded into a common contrastive learning method when $\boldsymbol{T}$ is set as the identity matrix $\boldsymbol{I}_N$. Meanwhile, simply maximizing the consistency between cross-view embedding in a common space may overlook view-specific semantics. Instead, we introduce a cross-view decoder that maintains these semantics in different spaces while ensuring cross-view consistency, defined as:

$$\boldsymbol{Z}_{(1\to2)} = g_{(1\to2)}(\boldsymbol{Z}_1), \tag{11}$$

where $g_{(1\to2)}(\cdot)$ denotes the cross-view decoder and can be implemented via an MLP.

Group Fairness-Aware Adversarial Learning

To further ensure group fairness for graph clustering results, we deploy an adversarial learning approach to learn fair node representations and stabilize the learning process with the proposed covariance constraint.

Adversarial Learning. The general goal of adversarial learning is to make the learned node representations independent of the corresponding sensitive attributes. To be specific, we first fuse the two views of the extracted node embeddings as follows:

$$\boldsymbol{Z} = (\boldsymbol{z}_{v_1}, \ldots, \boldsymbol{z}_{v_N}) = \frac{1}{2}(\boldsymbol{Z}_1 + \boldsymbol{Z}_2). \tag{12}$$

Then, we introduce a discriminator $\mathcal{D}(\cdot)$ with parameter ψ and let θ denote the parameters in $\mathcal{F}(\cdot)$. Given the representation $\boldsymbol{z}_v$ of each node v, $\mathcal{D}(\cdot)$ tries to predict the corresponding sensitive attribute s_v, formulated as:

$$\mathcal{L}_G = -\sum_{v=1}^{N} [s_v \log \mathcal{D}(\boldsymbol{z}_v) + (1-s_v)\log(1-\mathcal{D}(\boldsymbol{z}_v))]. \tag{13}$$

We adopt the following min-max learning strategy:

$$\min_{\psi} \max_{\theta} \mathcal{L}_G(\theta, \psi). \tag{14}$$

During the training process, on the one hand, we minimize the adversarial objective w.r.t. ψ to enable the discriminator to recognize the sensitive attribute. On the other hand, we maximize the adversarial objective w.r.t. θ to confuse the discriminator. So, in this way, we can facilitate our model $\mathcal{F}(\cdot)$ to generate a fair representation of the nodes.

Covariance Constraint. The instability of adversarial learning during training is well-known in previous work (Arjovsky and Bottou 2017). Especially in adversarial debiasing, insufficient convergence can lead to discriminatory classifiers. As a remedy, we introduce a covariance constraint on the learned node representation $\boldsymbol{Z}$ for more stable group fairness-aware adversarial learning. The covariance constraint has also proven effective in achieving a fair classifier by minimizing the absolute covariance between the noisy sensitive attribute and the label prediction (Dai and Wang 2021). In our issue, we aim to adaptively decorrelate the sensitive attribute from the learned node representations to ensure group fairness in graph clustering. Thus, we regard the absolute covariance between the sensitive attribute s_v and each column of the extracted feature $\boldsymbol{z}_{v,i}, \forall i \in \{1, \ldots, d\}$ as the loss function, defined as:

$$\mathcal{L}_R = \sum_{v=1}^{N} \sum_{i=1}^{d} |Cov(s_v, \boldsymbol{z}_{v,i})|, \tag{15}$$

where $|\cdot|$ is the absolute value, $\mathbb{E}(\cdot)$ is the expectation operation and $Cov(\cdot)$ corresponds to covariance for each node v, which can be formulated as:

$$Cov(s, \boldsymbol{z}_{v,i}) = \mathbb{E}[(s - \mathbb{E}(s))(\boldsymbol{z}_{v,i} - \mathbb{E}(\boldsymbol{z}_{v,i}))]. \tag{16}$$

The covariance constraint ensures that the node representation $\boldsymbol{z}_v$ and sensitive attribute s_v are independent, thereby promoting effective adversarial learning.

Objective Function

To foster the individual and group fairness for deep graph clustering, we integrate each component, and the final loss objective of the proposed `FairGC` is given by:

$$\mathcal{L} = \mathcal{L}_I - \beta \mathcal{L}_G + \gamma \mathcal{L}_R, \tag{17}$$

where β and γ denote the hyperparameters to balance the weight of each component. The final optimizing algorithm of the `FairGC` can be formulated as:

$$\begin{cases} \min_\theta \mathcal{L}_I - \beta \mathcal{L}_G + \gamma \mathcal{L}_R, \\ \min_\psi \beta \mathcal{L}_G. \end{cases} \tag{18}$$

In a nutshell, all three loss objectives contribute to fairness-aware deep graph clustering. And we optimize our framework in an alternative manner. The final clustering results are obtained by performing K-means algorithm (Hartigan and Wong 1979) over $\boldsymbol{Z}$ directly.

Experiment

Experiment Setup

Datasets. We evaluate the performance of `FairGC` on four widely-used benchmark graph datasets, including NBA (Dai and Wang 2021), Credit (Yeh and Lien 2009; Luo et al. 2024), Income (Dong et al. 2023; Luo et al. 2024), and Bail (Jordan and Freiburger 2015; Luo et al. 2024). Specifically, we use Nationality as the sensitive attribute for the NBA dataset, Age for the Credit dataset, and Race for both the Income dataset and the Bail dataset. The clustering task is to forecast the player's position for NBA, credit card payments for Credit, individuals based on their occupations for Income, and the likelihood of committing violent or nonviolent crimes for Bail.

Baselines. We compare our proposed `FairGC` against nine approaches. Specifically, these methods can be categorized into three groups: classical deep graph clustering methods (*i.e.,* DAEGC (Wang et al. 2019), SDCN (Bo et al. 2020), DFCN (Tu et al. 2021)), contrastive deep graph clustering methods (*i.e.,* DCRN (Liu et al. 2022a), CCGC (Yang et al. 2023), HSAN (Liu et al. 2023d), MAGI (Liu et al. 2024a)), and fair graph clustering methods (*i.e.,* FNM (Li, Wang, and Merchant 2023), iFairNMTF (Ghodsi, Seyedi, and Ntoutsi 2024)).

Metric. Following prior work (Yang et al. 2023), we employ accuracy (ACC) to evaluate the clustering performance. Additionally, we evaluate the performance of individual fairness with metrics: overall individual fairness (IF) (Song et al. 2022) and group disparity of individual fairness (GDIF) (Song et al. 2022). To measure the group fairness of clustering, we adopt *Balance* following previous research (Li, Wang, and Merchant 2023).

Implementation. Our model is implemented using the PyTorch deep learning framework. In practice, all the encoders and decoders are implemented as unshared one-layer MLPs with 256 hidden units. The maximum training epoch is 400 with Adam optimizer (Kingma and Ba 2015) to update the model's parameters. More concretely, we fix the rectified weight α to 0.5, the filter times l to 8, and the influence factors of covariance loss and adversarial loss to 0.5 and 0.1, respectively. We search for the optimal value of the random walk step t in the range $\{2, 4, 6, 8\}$. For baseline methods, we adopt their original settings and reproduce the results.

Performance Comparison

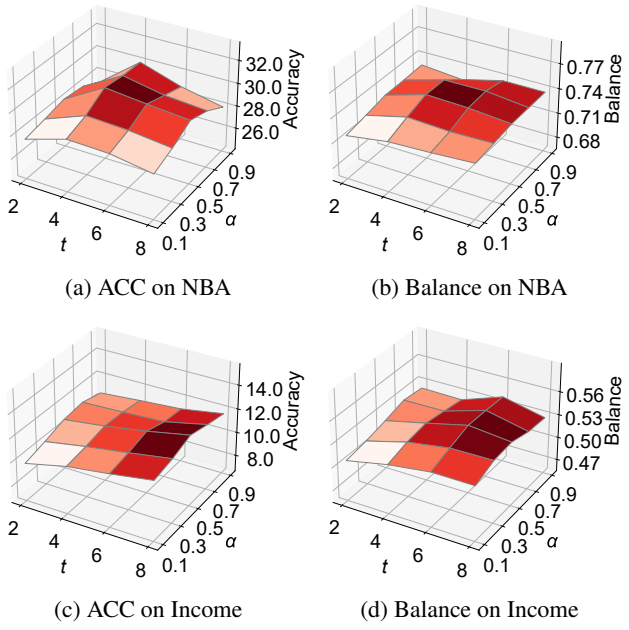
Table 1 shows the performance of our `FairGC` against all baselines. *Note that individual fairness metrics rely on encoded embeddings to compute node pair similarity, which traditional non-deep clustering methods (e.g., FNM and iFairNMTF) do not support; we do not report their results in Table 1.* We have the following conclusions. ❶ Compared with classical deep graph clustering methods, contrastive deep graph clustering methods achieve comprehensive improvement in both clustering performance and fairness. ❷ Fair graph clustering methods that consider sensitive attributes exhibit competitive performance regarding fairness while lagging behind other baselines in terms of clustering metrics due to their disregard for node features. ❸ Overall, our proposed `FairGC` consistently outperforms other baselines, enhancing both individual and group fairness while maintaining superior clustering performance.

Parameter Analysis

Effect of the Affinity Graph. For the random walk step t and trade-off parameter α, we search the optimal value in the range of $\{2, 4, 6, 8\}$ and $\{0.1, 0.3, 0.5, 0.7, 0.9\}$, respectively. As shown in Figure. 3, we can find that: ❶ The model achieves optimal performance while maintaining individual fairness when node pair similarity is appropriately integrated (i.e., $\alpha = 0.5$). ❷ However, the optimal value of t differs across datasets ($t = 4$ for the NBA dataset and $t = 6$ for the Income dataset). This can be attributed to the smaller scale

Dataset	Metric	Classical Deep Graph Clustering			Contrastive Deep Graph Clustering				Fair Graph Clustering		
		DAEGC IJCAI 19	SDCN WWW 20	DFCN AAAI 21	DCRN AAAI 22	HSAN AAAI 23	CCGC AAAI 23	MAGI SIGKDD 24	FNM ECAI 23	iFairNMTF PAKDD 24	FairGC Ours
NBA	ACC(↑)	26.55	28.78	24.42	24.37	29.88	24.76	24.62	26.50	24.52	**31.74**
	Balance(↑)	0.59	0.54	0.44	0.44	0.66	0.57	0.61	0.72	0.65	**0.74**
	IF(↓)	165.29	62.72	11.52	73.80	0.36	73.10	59.94	-	-	**0.09**
	GDIF(↓)	1.42	1.31	1.30	1.87	1.29	1.49	1.31	-	-	**1.27**
Credit	ACC(↑)	63.72	68.12	66.19	67.31	70.17	68.81	57.62	60.93	62.30	**72.77**
	Balance(↑)	0.56	0.59	0.56	0.52	0.59	0.60	0.54	**0.80**	0.62	0.63
	IF(↓)	595.74	749.61	2051.75	647.78	407.85	1199.49	601.48	-	-	**36.98**
	GDIF(↓)	5.60	4.96	11.03	2.06	1.90	1.77	1.82	-	-	**1.74**
Income	ACC(↑)	11.07	12.26	12.04	11.47	12.28	12.35	**15.15**	11.81	11.10	12.40
	Balance(↑)	0.32	0.19	0.39	0.45	0.49	0.49	0.41	0.51	0.47	**0.54**
	IF(↓)	9033.38	370.01	3559.54	9454.47	2426.06	2422.90	3309.06	-	-	**162.81**
	GDIF(↓)	8.67	2.82	1.51	3.79	2.61	2.61	8.62	-	-	**1.30**
Bail	ACC(↑)	57.11	61.99	61.85	64.28	78.38	75.81	76.32	56.70	59.35	**80.23**
	Balance(↑)	0.85	0.80	0.94	0.92	0.91	0.86	0.89	0.94	0.92	**0.95**
	IF(↓)	8183.49	8746.30	5168.98	302.16	136.92	285.06	283.38	-	-	**39.76**
	GDIF(↓)	1.12	1.03±0.01	1.02	1.07	1.03	1.03	1.11	-	-	**1.01**

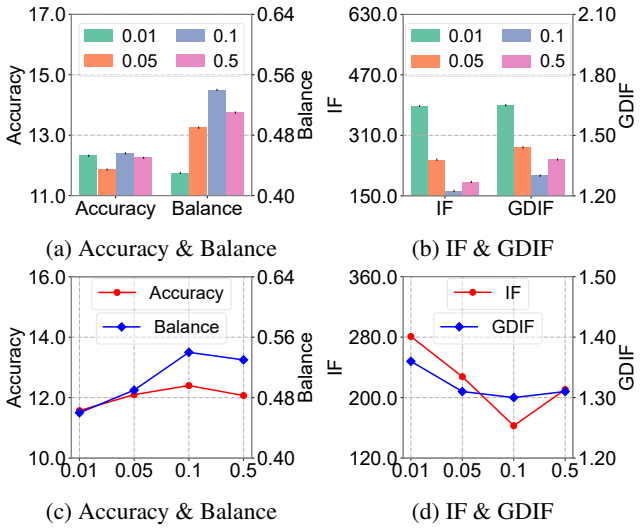
Table 1: The reported performance is the average of ten runs on four benchmark datasets, assessed by the mean of four metrics. Bold and underlined values indicate the best and second-best results.



(a) ACC on NBA (b) Balance on NBA

(c) ACC on Income (d) Balance on Income

Figure 3: Performance comparison w.r.t. different values of α and t on NBA and Income.



(a) Accuracy & Balance (b) IF & GDIF

(c) Accuracy & Balance (d) IF & GDIF

Figure 4: Performance comparison w.r.t. different values of β (top) and γ (bottom) on Income dataset.

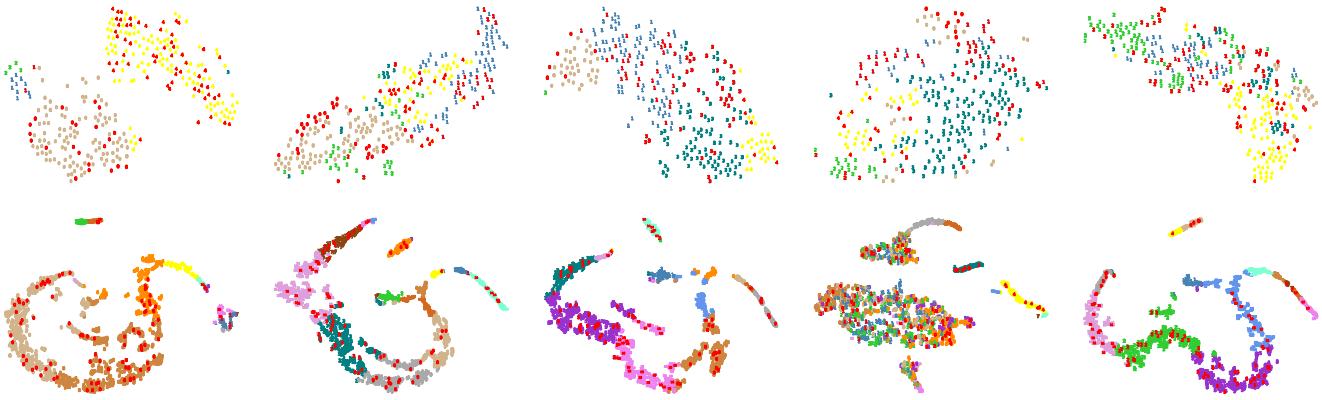
of the NBA dataset, where a larger random walk step might introduce noise for the node pairs.

Effect of the Component Weight. We further explore the weights of covariance constraint and adversarial debiasing, β and γ, within the range $\{0.01, 0.05, 0.1, 0.5\}$. As shown in Figure. 4, we have the following observations: ❶ Our FairGC is not sensitive to both β and γ in terms of clustering performance. ❷ Covariance constraint and adversarial learning contribute to improving fairness. Nevertheless, large values of β and γ may introduce perturbation during the training process, leading to decreased performance.

Ablation Study

We conduct ablation experiments from two perspectives. For individual fairness-aware contrastive learning, we examine the following variants: ❶ **variant 1:** removal of the multi-step transition matrix ($\tilde{\boldsymbol{R}}^t = \boldsymbol{I}$); ❷ **variant 2:** removal of the cross-view decoder. For group fairness-aware adversarial learning, we analyze these variants: ❶ **w/o ADV:** exclusion of the adversary module; ❷ **w/o COV:** elimination of the covariance constraint in the objective.

As shown in Table 2, we can find that the clustering performance has been slightly affected with the removal of individual and group fairness-aware learning. Specifically, the



(a) DCRN (b) HSAN (c) CCGC (d) MAGI (e) Ours

Figure 5: 2D t-SNE visualization of eight methods on the NBA (first row) and Income (second row) datasets. We randomly highlight 100 nodes in red to show their distributions across different clusters.

Dataset	Method	ACC(↑)	Balance(↑)	IF(↓)	GDIF(↓)
NBA	Variant 1	29.90	0.68	0.87	1.36
	Variant 2	29.16	0.71	0.35	1.34
	w/o ADV	30.60	0.58	0.18	1.29
	w/o COV	30.65	0.66	0.13	1.29
	Ours	**31.74**	**0.74**	**0.09**	**1.27**
Credit	Variant 1	70.37	0.60	94.08	1.94
	Variant 2	71.59	0.61	63.97	1.90
	w/o ADV	72.41	0.50	44.36	1.77
	w/o COV	72.04	0.58	42.12	1.74
	Ours	**72.77**	**0.63**	**36.98**	**1.73**
Income	Variant 1	10.74	0.52	834.63	1.64
	Variant 2	11.40	0.50	310.40	1.43
	w/o ADV	11.82	0.41	219.53	1.33
	w/o COV	12.04	0.46	224.37	1.34
	Ours	**12.40**	**0.54**	**162.81**	**1.30**
Bail	Variant 1	77.48	0.91	70.55	1.03
	Variant 2	76.67	0.90	61.45	1.04
	w/o ADV	79.76	0.83	41.54	1.01
	w/o COV	79.82	0.90	41.29	1.01
	Ours	**80.23**	**0.95**	**39.76**	**1.01**

Table 2: Ablation studies on both individual and group fairness-aware graph learning components.

global affinity graph and the cross-view decoder are critical to individual fairness, as they leverage individual similarity to enhance the capability to capture supervision information. Adversarial learning and covariance constraint, which contribute to effectively decoupling the relationship between sensitive attributes and features, also show a substantial impact in terms of group fairness. In total, the collaborative leverage of both individual and group fairness-aware learning enhances the overall performance of our `FairGC` in terms of clustering and fairness.

Visualization

To demonstrate the superiority of `FairGC`, we conduct 2D t-distributed stochastic neighbor embedding (t-SNE) to visualize the learned node representations on the NBA and Income. As shown in Figure. 5, we randomly select 100 nodes and highlight them in red to indicate that their nationality is overseas for the NBA dataset and their race is Black for the Income dataset. Compared to other baselines, our `FairGC` demonstrates more fairness-aware clustering results. For instance, the highlighted nodes in Income are concentrated in Clusters 5, 8, and 10 in CCGC, while our method shows a more evenly distributed performance. Meanwhile, it shows better separability between clusters, with tighter within-cluster aggregation and clearer distinctions between clusters.

Conclusion

In this paper, we propose a novel model termed `FairGC` to foster both individual and group fairness for deep graph clustering. And we believe this is the first study to address fairness-aware deep graph clustering. Specifically, we utilize Siamese encoders to construct two views with distinct semantics. Then, we compute view-specific affinity graphs, on which multi-step random walks are applied to capture high-order similarities of global node pairs. Building upon this, we reformulate contrastive learning with individual similarity to enhance individual fairness. Furthermore, we employ adversarial learning to ensure group fairness by mitigating the influence of sensitive information through the discriminator. Experimental results on four benchmark datasets demonstrate the efficacy of our `FairGC`. In future work, we plan to generalize our approach to more complex fairness settings, including scenarios with multiple sensitive attributes, as well as dynamic graphs.

Acknowledgments

This paper is partially supported by "the Fundamental Research Funds for the Central Universities" (N2217003), "the Fundamental Research Funds for the Central Universities" in the University of International Business and Economics (Grant No. 23QN02), the Liaoning Province Natural Science Foundation (2023010411-JH3/101) and the Key Laboratory of Optical Information and Simulation Technology of Liaoning Province.

References

Agarwal, C.; Lakkaraju, H.; and Zitnik, M. 2021. Towards a unified framework for fair and stable graph representation learning. In *Uncertainty in Artificial Intelligence*, 2114–2124.

Arjovsky, M.; and Bottou, L. 2017. Towards principled methods for training generative adversarial networks. *arXiv preprint arXiv:1701.04862*.

Bo, D.; Wang, X.; Shi, C.; Zhu, M.; Lu, E.; and Cui, P. 2020. Structural deep clustering network. In *Proceedings of the Web Conference*, 1400–1410.

Cai, L.; and Ji, S. 2020. A multi-scale approach for graph link prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 3308–3315.

Cao, M.; Song, J.; Yuan, J.; Zhang, B.; and Wang, C. 2023. FairHELP: Fairness-aware heterogeneous information network embedding for link prediction. In *International Conference on Database Systems for Advanced Applications*, 320–330. Springer.

Caton, S.; and Haas, C. 2024. Fairness in machine learning: A survey. *ACM Computing Surveys*, 56(7): 1–38.

Cui, G.; Zhou, J.; Yang, C.; and Liu, Z. 2020. Adaptive graph encoder for attributed graph embedding. In *Proceedings of the International ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 976–985.

Dai, E.; and Wang, S. 2021. Say no to the discrimination: Learning fair graph neural networks with limited sensitive attribute information. In *Proceedings of the International ACM Conference on Web Search & Data Mining*, 680–688.

Dong, Y.; Wang, S.; Ma, J.; Liu, N.; and Li, J. 2023. Interpreting unfairness in graph neural networks via training node attribution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 7441–7449.

Ghodsi, S.; Seyedi, S. A.; and Ntoutsi, E. 2024. Towards Cohesion-Fairness Harmony: Contrastive Regularization in Individual Fair Graph Clustering. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 284–296.

Gong, L.; Zhou, S.; Tu, W.; and Liu, X. 2022. Attributed Graph Clustering with Dual Redundancy Reduction. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 3015–3021.

Guo, D.; Chu, Z.; and Li, S. 2023. Fair Attribute Completion on Graph with Missing Attributes. In *Proceedings of the International Conference on Learning Representations*.

Guo, Z.; Li, J.; Xiao, T.; Ma, Y.; and Wang, S. 2023. Towards fair graph neural networks via graph counterfactual. In *Proceedings of the International Conference on Information and Knowledge Management*, 669–678.

Hartigan, J. A.; and Wong, M. A. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1): 100–108.

Jordan, K. L.; and Freiburger, T. L. 2015. The effect of race/ethnicity on sentencing: Examining sentence type, jail length, and prison length. *Journal of Ethnicity in Criminal Justice*, 13(3): 179–196.

Kang, J.; He, J.; Maciejewski, R.; and Tong, H. 2020. Inform: Individual fairness on graph mining. In *Proceedings of the International ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 379–389.

Kingma, D. P.; and Ba, J. 2015. Adam: A method for stochastic optimization. In *Proceedings of the International Conference on Learning Representations*.

Li, J.; Wang, Y.; and Merchant, A. 2023. Spectral Normalized-Cut Graph Partitioning with Fairness Constraints. In *Proceedings of the European Conference on Artificial Intelligence*, 1389–1397.

Li, Y.; Wang, X.; Xing, Y.; Fan, S.; Wang, R.; Liu, Y.; and Shi, C. 2024. Graph Fairness Learning under Distribution Shifts. In *Proceedings of the Web Conference*, 676–684.

Liu, J.; Cheng, J.; Han, R.; Tu, W.; Wang, J.; and Peng, X. 2025a. Federated Graph-Level Clustering Network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 18870–18878.

Liu, J.; Han, R.; Tu, W.; Wang, H.; Wu, J.; and Cheng, J. 2025b. Federated Node-Level Clustering Network with Cross-Subgraph Link Mending. In *Proceedings of the International Conference on Machine Learning*, 38540–38556.

Liu, Y.; Ding, K.; Liu, H.; and Pan, S. 2023a. Good-d: On unsupervised graph out-of-distribution detection. In *Proceedings of the International ACM Conference on Web Search & Data Mining*, 339–347.

Liu, Y.; Ding, K.; Lu, Q.; Li, F.; Zhang, L. Y.; and Pan, S. 2023b. Towards self-interpretable graph-level anomaly detection. In *Proceedings of the Conference on Neural Information Processing Systems*, 8975–8987.

Liu, Y.; Li, J.; Chen, Y.; Wu, R.; Wang, B.; Zhou, J.; Tian, S.; Shen, S.; Fu, X.; Meng, C.; Wang, W.; and Chen, L. 2024a. Revisiting Modularity Maximization for Graph Clustering: A Contrastive Learning Perspective. In *Proceedings of the International ACM SIGKDD Conference on Knowledge Discovery & Data Mining*.

Liu, Y.; Liang, K.; Xia, J.; Zhou, S.; Yang, X.; Liu, X.; and Li, Z. S. 2023c. Dink-Net: Neural Clustering on Large Graphs. In *Proceedings of the International Conference on Machine Learning*.

Liu, Y.; Tu, W.; Zhou, S.; Liu, X.; Song, L.; Yang, X.; and Zhu, E. 2022a. Deep graph clustering via dual correlation reduction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 7603–7611.

Liu, Y.; Xia, J.; Zhou, S.; Yang, X.; Liang, K.; Fan, C.; Zhuang, Y.; Li, S. Z.; Liu, X.; and He, K. 2022b. A Survey of Deep Graph Clustering: Taxonomy, Challenge, Application, and Open Resource. *arXiv preprint arXiv:2211.12875*.

Liu, Y.; Yang, X.; Zhou, S.; Liu, X.; Wang, Z.; Liang, K.; Tu, W.; Li, L.; Duan, J.; and Chen, C. 2023d. Hard sample aware network for contrastive deep graph clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 8914–8922.

Liu, Y.; Zhu, S.; Xia, J.; Ma, Y.; Ma, J.; Zhong, W.; Zhang, G.; Zhang, K.; and Liu, X. 2024b. End-to-end Learnable Clustering for Intent Learning in Recommendation. *arXiv preprint arXiv:2401.05975*.

Liu, Y.; Zhu, S.; Yang, T.; Ma, J.; and Zhong, W. 2024c. Identify Then Recommend: Towards Unsupervised Group Recommendation. In *Proceedings of the Conference on Neural Information Processing Systems*.

Lu, Y.; Lin, Y.; Yang, M.; Peng, D.; Hu, P.; and Peng, X. 2023. Decoupled Contrastive Multi-view Clustering with High-order Random Walks. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Luo, R.; Huang, H.; Yu, S.; Zhang, X.; and Xia, F. 2024. FairGT: A Fairness-aware Graph Transformer. In *Proceedings of the International Joint Conference on Artificial Intelligence*.

Newman, M. E. 2006. Finding community structure in networks using the eigenvectors of matrices. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 74(3): 036104.

Pan, S.; Hu, R.; Fung, S.-f.; Long, G.; Jiang, J.; and Zhang, C. 2019. Learning graph embedding with adversarial training methods. *IEEE transactions on cybernetics*, 50(6): 2475–2487.

Ren, T.; Zhang, H.; Wang, Y.; Ju, W.; Liu, C.; Meng, F.; Yi, S.; and Luo, X. 2025. MHGC: Multi-scale hard sample mining for contrastive deep graph clustering. *Information Processing & Management*, 62(4): 104084.

Song, W.; Dong, Y.; Liu, N.; and Li, J. 2022. Guide: Group equality informed individual fairness in graph neural networks. In *Proceedings of the International ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 1625–1634.

Tu, C.; Zeng, X.; Wang, H.; Zhang, Z.; Liu, Z.; Sun, M.; Zhang, B.; and Lin, L. 2018. A unified framework for community detection and network representation learning. *IEEE Transactions on Knowledge and Data Engineering*, 31(6): 1051–1065.

Tu, W.; Guan, R.; Zhou, S.; Ma, C.; Peng, X.; Cai, Z.; Liu, Z.; Cheng, J.; and Liu, X. 2024a. Attribute-Missing Graph Clustering Network. 15392–15401.

Tu, W.; Liao, Q.; Zhou, S.; Peng, X.; Ma, C.; Liu, Z.; Liu, X.; Cai, Z.; and He, K. 2024b. RARE: Robust Masked Graph Autoencoder. *IEEE Transactions on Knowledge and Data Engineering*, 36(10): 5340–5353.

Tu, W.; Zhou, S.; Liu, X.; Cai, Z.; Zhao, Y.; Liu, Y.; and He, K. 2025. WAGE: Weight-Sharing Attribute-Missing Graph Autoencoder. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7): 5760–5777.

Tu, W.; Zhou, S.; Liu, X.; Guo, X.; Cai, Z.; Zhu, E.; and Cheng, J. 2021. Deep fusion clustering network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 9978–9987.

Wang, C.; Pan, S.; Hu, R.; Long, G.; Jiang, J.; and Zhang, C. 2019. Attributed Graph Clustering: A Deep Attentional Embedding Approach. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 3670–3676.

Wang, J.; Lu, D.; Davidson, I.; and Bai, Z. 2023. Scalable spectral clustering with group fairness constraints. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 6613–6629.

Xia, W.; Wang, Q.; Gao, Q.; Yang, M.; and Gao, X. 2022. Self-consistent contrastive attributed graph clustering with pseudo-label prompt. *IEEE Transactions on Multimedia*.

Yang, X.; Liu, Y.; Zhou, S.; Wang, S.; Tu, W.; Zheng, Q.; Liu, X.; Fang, L.; and Zhu, E. 2023. Cluster-guided contrastive graph clustering network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 10834–10842.

Yeh, I.-C.; and Lien, C.-h. 2009. The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert systems with applications*, 36(2): 2473–2480.

Yi, S.; Ju, W.; Qin, Y.; Luo, X.; Liu, L.; Zhou, Y.; and Zhang, M. 2023. Redundancy-free self-supervised relational learning for graph clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15.

Zhan, D.; Guo, D.; Ji, P.; and Li, S. 2024. Bridging the Fairness Divide: Achieving Group and Individual Fairness in Graph Neural Networks. *arXiv*.

Zhang, H.; Ren, T.; Wang, Y.; Meng, F.; Ju, W.; and Tian, Y. 2025. Cluster-aware few-shot molecular property prediction with factor disentanglement. *IEEE Transactions on Neural Networks and Learning Systems*.

Zhang, M.; and Chen, Y. 2018. Link prediction based on graph neural networks. In *Proceedings of the Conference on Neural Information Processing Systems*, 5171–5181.