Fairness-Aware Graph Representation Learning with Limited Demographic Information

Zichong Wang[1], Zhipeng Yin[1], Liping Yang[2],
Jun Zhuang[3], Rui Yu[4], Qingzhao Kong[5], and Wenbin Zhang✉[1]

[1] Florida International University, Miami, USA
[2] University of New Mexico, Albuquerque, USA
[3] Boise State University, Boise, USA
[4] University of Louisville, Louisville, USA
[5] Jimei University, Xiamen, China
{ziwang,wenbin.zhang}@fiu.edu

Abstract. Ensuring fairness in node classification with Graph Neural Networks is fundamental to promoting trustworthy and socially responsible machine learning systems. In response, numerous fair graph learning methods have been proposed in recent years. However, most of them assume full access to demographic information, a requirement rarely met in practice due to privacy, legal, or regulatory restrictions. To this end, this paper introduces a novel fair graph learning framework that mitigates bias in graph learning under limited demographic information. Specifically, we first propose a mechanism guided by available demographic information to generate proxies for demographic information and then design a strategy to ensure consistent node embeddings across demographic groups. Additionally, we propose an adaptivity confidence strategy that dynamically adjusts each node's contribution to fairness and utility based on prediction confidence. Through extensive experiments on multiple datasets and fair graph learning frameworks, we demonstrate the framework's effectiveness in both mitigating bias and maintaining model utility.

Keywords: Fairness · Fair representation learning · Graph neural networks.

1 Introduction

Graph Neural Networks (GNNs) have become a prevalent approach for handling complex real-world applications, such as healthcare [1], social network analysis [30], and recommendation systems [16]. The success of GNNs relies on message-passing mechanisms, which aggregate information from neighboring nodes, effectively capturing both graph structural information and node attribute information [32,61]. However, despite their successes, GNNs tend to inherit and even exacerbate existing biases from graph data [34], propagating and amplifying unfair patterns embedded in network topology and features. This inadvertent amplification of societal biases and the potential for discriminatory outcomes have highlighted the urgent need to develop strategies that promote fairness within these systems. To this end, a number of approaches [40,64,37] have been proposed in recent years, with most relying on complete demographic information to guide fair graph learning.

However, this requirement often does not align with realistic situations, as collecting or explicitly utilizing demographic information (*e.g.*, gender, race) can be restricted or prohibited due to privacy concerns, legal constraints, ethical considerations, or social sensitivity [17,18]. For example, in many real-world graph datasets such as academic collaboration networks or online social platforms, demographic information are often missing, incomplete, or intentionally withheld to protect user privacy, with studies showing that less than 30% of users voluntarily disclose demographic information [22]. Consequently, existing fairness-aware graph methods, which typically assume full availability of demographic information, become impractical when only limited demographic information is available. This gap between theoretical fairness requirements and real-world constraints creates a critical gap that significantly limits the applicability and practical deployment of current graph fairness solutions.

To fill this gap, a few works [48,18,26] have begun to explore achieving fairness without full demographics. However, these pioneering approaches focus on i.i.d. data settings and overlook the unique characteristics of graph-structured data. Consequently, existing methods cannot be easily adapted to graph-structured data, which appears widely in many real-world scenarios. This limitation has left fair graph learning with limited demographic information as a highly open research area with several unique challenges: **i) Difficulty of identifying missing demographic information from limited demographic labels:** In many real applications, only a small subset of nodes reveal their demographics. These disclosed labels can over-represent favored groups or cluster in areas of the graph with different link patterns. The uneven coverage makes it difficult to train reliable predictors for the missing demographic information; using the limited labels without care can intensify bias instead of reducing it. **ii) Complexity of mitigating interconnected biases in graph data:** Graph data presents unique fairness challenges because biases exist in multiple interconnected forms between node attributes and graph structure. These biases interact through message-passing mechanisms, making them particularly difficult to identify and mitigate. Without full demographic labels, it is hard to distinguish whether patterns in the graph represent biases or not. This makes achieving fair node representations exceptionally challenging, since we cannot directly observe or measure the demographic disparities that need to be mitigated in the learned embeddings. **iii) Balancing model utility and fairness:** A major challenge in fairness work is maintaining model utility while improving fairness. Enhancing fairness typically requires the model to pay more attention to samples from deprived groups, which can reduce performance for favored groups. This challenge is further compounded by the absence of demographic information, which creates uncertainty about subgroup membership.

To address aforementioned challenges, this paper proposes *Demographic-agnostic Fair Graph Representation (DFGR)*, which is designed to reduce bias in graph learning algorithms when only limited demographic information is available. *To the best of our knowledge, this is the first work that designs to achieve fair graph learning without full demographic information while preserving maximum task-related information.* Specifically, DFGR uses the limited demographic labels in the training set to guide an encoder, built according to our causal analysis to generate proxies for demographic information. Armed with these identified demographic proxies, DFGR then enforces

three constraints aimed at ensuring learned node representations remain invariant to demographic information while retaining as much task-related information as possible. Additionally, by incorporating the proposed adaptivity confidence strategy, DFGR imposes fairness constraints only on samples with high confidence, reducing performance loss while improving model fairness and helping the model to better learn from samples with low confidence. The main contributions of our work can be summarized as:

- We address the largely unattended challenge of achieving fair graph learning with incomplete demographic information. We propose a novel method to generate proxies for demographic information and leverage these proxies as a foundation for our fairness framework, ensuring consistent node embeddings across different demographic groups through three designed constraints.
- We introduce a novel adaptivity confidence strategy that improves fairness while minimizing utility loss by dynamically adjusting the weight of each sample's contribution to the fairness loss based on the classification confidence level.
- We conduct extensive experiments on four real-world graph datasets that demonstrate DFGR's effectiveness in mitigating bias while maintaining comparable utility to state-of-the-art methods.

2 Related Work

2.1 Fair Graph Learning

In recent years, extensive research has been conducted to improve the fairness of GNNs by mitigating biases from training data [53,20,39] or training GNNs with fairness-aware frameworks [8,53,64]. The core idea behind most of these approaches is the removal of demographics-related information, thereby enforcing GNNs to make decisions independent of the demographic information [54]. In other words, it aims to achieve algorithmic decisions that do not discriminate against or favor certain groups defined by the demographic information. Despite their great success, most existing fair GNNs assume access to predefined demographic information during training, which is impractical in most real-world socially sensitive applications due to privacy, legal, or regulatory restrictions [2]. In addition, a few works make initial explorations of fair graph learning with missing demographics. Specifically, FairGNN [8] aims to learn fair GNNs with limited demographics. To achieve this goal, FairGNN employs the demographic estimator to predict the demographics while improving fairness via adversarial learning. In addition, FairAC [12] embeds nodes with observed attributes, then employs an attention mechanism to aggregate neighbor features for nodes with missing attributes. However, both methods overlook that different groups may differ in their willingness to share demographic information. Members of a favored group may be more willing to disclose their data, while individuals from a deprived group may withhold it due to fear of discrimination. As a result, these methods are less effective when demographic information is highly limited or unevenly available.

2.2 Fairness with Incomplete Demographic Information

The research community is paying growing attention to fair machine learning models that do not rely on fully known demographic labels, because many socially sensitive applications permit only limited or no direct access to demographic information [14]. Existing approaches can be divided into two main types: those employing proxy demographics [11,26,60] and those adhering to minimax fairness principles [3,21,28]. Specifically, the core idea behind proxy-based methods is that proxy demographic information can be obtained when certain features correlate with real demographic information or when partial demographic information is available. In scenarios where direct access to demographic information is impossible, proxy fairness notions rely on these correlated or predicted features to approximate the real demographic information, while minimax fairness is based on John Rawls's difference principle [27], which is designed to minimize the model loss for the least advantaged subgroup. However, existing fairness work with incomplete demographic information is focused on independent and identically distributed (i.i.d.) data, which is unable to mitigate the bias exhibited by the relational information (*i.e.*, graph structure information) and cannot be easily extended to graph data.

Different from the above works, DFGR addresses a new fair graph learning research problem where demographic information is incomplete, yet fairness and equity in graph-based decision-making systems remain essential. In addition, we introduce an adaptive confidence strategy that focuses on high-confidence prediction nodes, enabling DFGR to enhance fairness while minimizing the fairness constraint's effect on model utility.

3 Preliminaries

3.1 Notations

For clarity in writing, we describe our method and accompanying proofs under the setting of a node classification task with binary demographic information and binary labels. We represent a graph as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, where $|\mathcal{V}| = n$ is the number of nodes and $|\mathcal{E}| = r$ is the number of edges. The matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$ contains d-dimensional feature vectors, with the u^{th} row corresponding to node u. The adjacency matrix $\mathbf{A} \in \{0,1\}^{n \times n}$ has entries $\mathbf{A}_{u,k} = 1$ if there is an edge $e_{u,k} \in \mathcal{E}$ between nodes u and k, and $\mathbf{A}_{u,k} = 0$ otherwise. We let $S \in \{0,1\}^{n \times 1}$ denote the demographic information, and write s_u for the value of u. We define $S_d = \{\, u \mid s_u = 0\}$ as the deprived group (for example, female), and $S_f = \{\, u \mid s_u = 1\}$ as the favored group (for example, male). Each node u also has a one-hot ground-truth label y_u, and $\hat{y}_u$ is its predicted label. We let $y_u = 1$ indicate a granted label and $y_u = 0$ indicate a rejected label.

3.2 Fair Causal analysis

Existing works have selected correlated non-demographic attributes as proxies for missing demographic information based on prior knowledge [14]. However, when dealing with high-dimensional attributes, it becomes challenging to accurately determine the

proxies for demographic information. At the same time, selecting appropriate proxies is crucial for promoting fairness. To this end, we conduct a causal analysis of the underlying mechanisms in the observed graph to help identify missing demographic information. Without loss of generality, in this work, we focus on the fair node classification without full demographic information and construct a Structural Causal Model [24] (SCM) as shown in Figure 1. It presents the causal relationships among six variables: Demographic Information (S), Ground-Truth Label (Y), Graph Structure (A), Node Features (X), Ego-graph (G), and Node Representation (h). Each connection in the SCM represents a causal relation. Specifically, S is typically determined at birth; it does not have a parent variable in the causal graph and solely acts as a cause influencing other variables, including X and A. In addition, S, X and A will all affect the final node representation through the GNN message passing mechanism, while it also should contain important information for downstream node classification tasks and ego graph reconstruction.

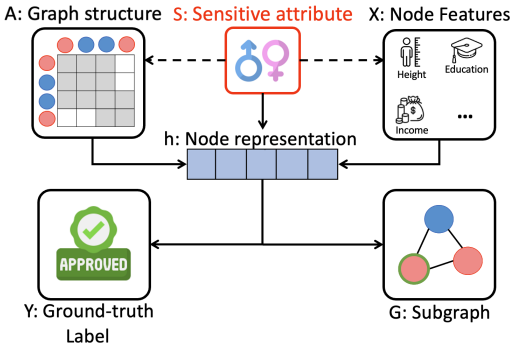

Fig. 1: Structural Causal Model for DFGR.

4 The Proposed Framework - DFGR

4.1 DFGR: In a Nutshell

In this section, we propose a novel framework, DFGR, which aims to learn fair node representations without full demographic information. As the illustration shown in Figure 2, DFGR mainly includes three key components: i) demographic information identification module; ii) fair node representation learning module; iii) adaptivity confidence strategy module. In the demographic information identification module, DFGR obtained the demographics proxy by incorporating node representations of both the graph structure and non-sensitive attributes. By using the identity proxy of demographic information in the fair node representation learning module, DFGR aims to learn fair node representations to minimize the identifiability of demographic information in node representation while preserving as much label-related information as feasible by establishing three constraints. Finally, DFGR adjusts the weight for each proxy to enhance

model fairness stability while facilitating gradual learning from simpler to more complex instances. Each of these modules will be elaborated upon in the subsequent discussion.

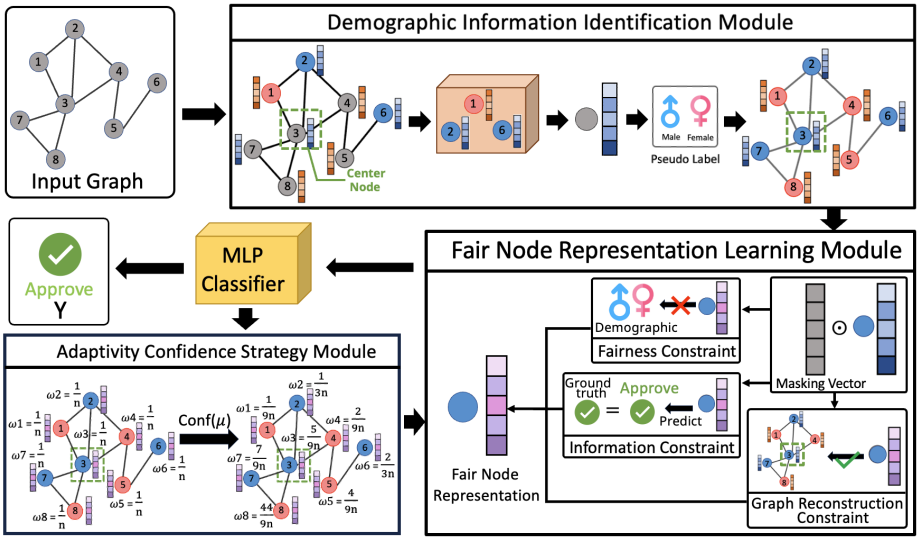

Fig. 2: An illustration of the proposed framework DFGR.

4.2 Demographic Information Identification Module

In this subsection, we introduce a demographic information identification module, designed to infer the missing demographic labels by using the observed graph data (*i.e.*, X, A, and Y) together with the demographic labels that are available. As demographic information is assumed to influence both graph connectivity and non-demographic attributes, and to model this effect, we construct a proxy by integrating representations of the graph structure with the non-demographic features. Specifically, a graph encoder is employed to generate this proxy, as it is capable of capturing complex patterns in high-dimensional data and combining signals from links and features, even when only a limited number of demographic labels are available. Formally, the encoder is defined as follows:

$$\mathbf{h}_u^{(l)} = \xi\, \mathbf{h}_u^{(l-1)} + \sum_{k \in \mathcal{N}(u)} \alpha_{u,k}^{(l)}\, \mathrm{ReLU}\Big(\mathbf{W}^{(l)}\, \mathbf{h}_k^{(l-1)}\Big), \quad \alpha_{u,k}^{(l)} = \frac{\exp\big(e_{u,k}^{(l)}\big)}{\sum_{k \in \mathcal{N}(u)} \exp\big(e_{u,k}^{(l)}\big)} \tag{1}$$

where $\mathbf{h}_u^{(l)}$ represents the embedding of node u at layer l, and $\mathbf{h}_u^{(l-1)}$ is the embedding from the previous layer. The parameter ξ is a learnable scalar that controls how much of

the previous representation is retained at layer l. The matrix $\mathbf{W}^{(l)}$ is a learnable weight matrix, and $\mathrm{ReLU}(\cdot)$ is the nonlinear activation function. The set $\mathcal{N}(u)$ denotes the neighborhood of the center node u, representing all nodes directly connected to u in the graph. The attention coefficient $\alpha_{u,k}^{(l)}$ reflects the relative importance of neighbor k to node u in the aggregation process.

Based on the encoder architecture, the module is trained via a supervised classification task aimed at predicting missing demographic information. The encoder parameters are learned by minimizing the cross-entropy loss between predicted and observed demographic labels. This supervision encourages the encoder to map nodes with similar demographics to nearby points in the representation space, even under partial label availability. Once training converges, the encoder parameters are fixed and subsequently used as a feature extractor that transforms high-dimensional node features into compact, informative embeddings. These low-dimensional representations function as proxies for demographic information in later components of the framework. This approach enables the model to indirectly capture and regulate the effects of missing demographic attributes, despite the absence of complete ground-truth labels.

4.3 Fair Node Representation Learning Module

Building on the demographic information identified in the previous module, we now introduce the fair node representation learning module, which aims to mitigate bias in node embeddings and prevent demographic-related information from influencing downstream tasks. As discussed earlier, nodes sharing the same demographic characteristics often tend to be more densely connected. The message-passing process can then smooth the representations of such nodes, further separating them from nodes of different demographic subgroups and causing predictions to become overly dependent on demographic information [62]. To address this issue, we propose mapping each node's embedding into a new representation space. This transformation conceals any cues that might reveal membership in a deprived subgroup, while preserving as much of the task-related information as possible. Specifically, we introduce three constraints: the fairness constraint, the information constraint, and the graph reconstruction constraint.

Fairness Constraint. The fairness constraint is designed to remove any demographic-related information (*i.e.*, whether a node is in the favored or deprived subgroup) in node representations that could introduce bias into downstream predictions. As illustrated in Figure 2, the original node embedding $\mathbf{h}_u$ may clearly reflect demographic information, such as whether u belongs to the male or female subgroup. After applying our fairness constraint, this information becomes obscured in the new embedding $\mathbf{h'}_u$. As shown in Figure 2, node representations $\mathbf{h}$ that distinguish between male and female are represented by red and blue colors, respectively. These distinct representations then converge toward a more uniform representation in the fair node embedding space $\mathbf{h}'$, depicted in purple, thereby obscuring the demographic information. Specifically, we transform each node representation $\mathbf{h}_u$ into a new space that makes it impossible to deduce whether node u belongs to a specific demographic subgroup.

In this new space, each node's information is represented using a set of prototypical probabilistic mappings. Let ρ be a multinomial random variable over prototypes $\{\mathbf{h'}_1, \mathbf{h'}_2, \ldots, \mathbf{h'}_m\}$, each of which has the same dimensionality as $\mathbf{h}_u$ and m is the

number of prototypical. A node representation does not retain demographic group information if it appears with the same probability in both the deprived subgroup (S_d) and the favored subgroup (S_f). Formally, this condition can be written as:

$$P(\rho \mid u \in S_d) = P(\rho \mid u \in S_f) \tag{2}$$

where $P(\rho \mid u \in S_d)$ represents the probability distribution of the prototypical mapping ρ for nodes in the deprived subgroup, and $P(\rho \mid u \in S_f)$ represents the same for nodes in the favored subgroup. In addition, the probability of assigning a node representation $\mathbf{h}_u$ to prototype j can be written as:

$$P(\rho = j|\mathbf{h}) = \frac{\exp(-d(\mathbf{h}, \mathbf{h'}_j))}{\sum_{j=1}^{m} \exp(-d(\mathbf{h}, \mathbf{h'}_j))} \tag{3}$$

where $d(\cdot)$ is a distance function (*e.g.*, Euclidean distance).

Building on this, we measure the difference in the probability of each prototype in different sensitive groups. Hence, the *group probability disparity* (GPD) is formally defined as $GPD = |GP_{S_d} - GP_{S_f}|$, where *group probability* (GP) is defined as follows:

$$\begin{cases} GP_{S_d} = \dfrac{1}{|\mathcal{V}_{S_d}|} \displaystyle\sum_{u \in S_d} \sum_{j=1}^{m} P(\rho = \mathbf{h'}_j|\mathbf{h}_u) \\ GP_{S_f} = \dfrac{1}{|\mathcal{V}_{S_f}|} \displaystyle\sum_{u \in S_f} \sum_{j=1}^{m} P(\rho = \mathbf{h'}_j|\mathbf{h}_u) \end{cases} \tag{4}$$

In addition, to prevent the fair embedding from discarding important information, *i.e.*, task-related information, identity information, along with demographic-related information, we introduce a reconstruction term that penalizes large deviations from the original node representations. Therefore, we add a term to measure the squared difference between $\mathbf{h}$ and $\mathbf{h}'$, ensuring that each node's transformed embedding remains sufficiently close to the original. Finally, the fairness constraint is defined as follows:

$$\mathcal{L}_F = GPD + \sum_{u=1}^{n} \|\mathbf{h}_u - \mathbf{h'}_u\|^2 \tag{5}$$

where $\mathbf{h'}_u$ is the fair node representations of $\mathbf{h}_u$. This constraint encourages the model to encode all information contained within the raw features except for any information that could lead to biased learning. By enforcing this constraint, we ensure that the node representations in the new space do not contain information that could be used to discriminate between demographic subgroups.

Information Constraint. For each node u, the transformed representation $\mathbf{h'}_u$ should preserve essential features and structural information, ensuring its usefulness for downstream tasks. In other words, the model should be able to make accurate label predictions using node representations (*i.e.*, $\mathbf{h'}_u \rightarrow y_u$). As illustrated in Figure 2, the information constraint ensures the retention of the task-related information to accurately predict the label in both the fair node embedding $\mathbf{h'}_u$ and the original node embedding $\mathbf{h}_u$. Hence, the objective of the information constraint is to minimize the loss of the prediction model, as shown in Equation 6:

$$\mathcal{L}_I = \frac{1}{|\mathcal{V}_L|} \sum_{v_u \in \mathcal{V}_L} -(y_u \log(\hat{y_u}) + (1 - y_u) \log(1 - \hat{y_u})) \tag{6}$$

where y_i is the one-hot encoding of the ground-truth label of node u.

Graph Reconstruction Constraint. For each node u, another objective is to ensure that its node representations accurately represent the node itself. This requirement is fulfilled by accurately reconstructing the ego-graph $\mathcal{G}_u$ from the new node embedding $\mathbf{h'}_u$. Formally, we define the graph reconstruction constraint as a graph structure reconstruction loss, $\mathcal{L}_G$:

$$\mathcal{L}_G = \frac{1}{|\mathcal{E}_{S_d}| + |\mathcal{E}_{S_f}|} \sum_{e_{u,k} \in \mathcal{E}} L\big(e_{u,k}, \hat{e}_{u,k}\big) \tag{7}$$

where $\mathcal{E}_{S_d}$ and $\mathcal{E}_{S_f}$ are sets of sampled edges connecting nodes from deprived and favored subgroups, respectively, and $L(\cdot)$ is the cross-entropy loss. The term $e_{u,k}$ denotes the actual connection status between nodes u and k, whereas $\hat{e}_{u,k} = \sigma(\mathbf{h'}_u \mathbf{h'}_k^\top)$ is the predicted probability of a link, with $\sigma(\cdot)$ representing the sigmoid function. Because positive edges are relatively sparse, we randomly sample one negative edge for each positive edge to balance the training data.

In essence, the introduction of the graph reconstruction constraint serves as a precaution against noise infiltration into node representations. This ensures that the reconstructed ego-graph $\mathcal{G}_u$ remains faithful to the original graph structure, thereby preserving important structural information while removing demographic bias.

4.4 Adaptivity Confidence Strategy Module

Following our Fair Node Representation Learning Module, this section introduces an Adaptivity Confidence Strategy Module that adjusts the weight for each node based on the model's confidence. The key insight is that if the model is highly confident about a node's prediction, it is more important to ensure that demographic information is fully masked in the node's representation. On the other hand, for low-confidence nodes, the model is essentially guessing the node label, so a strict fairness penalty is less important. Intuitively, if the classifier already has low confidence in determining a node's label prediction, the risk of embedding leaked demographic information that drives biased outcomes is smaller. Conversely, for nodes with high classification confidence, we need to ensure that this confidence does not derive from demographic information. For example, preventing the model from predicting that someone will get an interview at a software company simply because the person is male.

To address this, we propose the Adaptivity Confidence Strategy Module, which dynamically weights the fairness loss of each node based on classification confidence. Let $\mathrm{conf}(u)$ be the confidence score for node u's predicted label, and let τ be a threshold that separates high-confidence from low-confidence samples. If $\mathrm{conf}(u) \geq \tau$, we consider the node's classification to be reliable, and thus apply a larger penalty whenever its demographic information is not sufficiently "masked." Conversely, if $\mathrm{conf}(u) < \tau$, the node's classification is already uncertain (close to random guessing), so the urgency

of masking its demographic information is reduced. Formally, the group probability can be updated as:

$$\begin{cases} GP_{S_d} = \dfrac{\sum_{u \in S_d} w_u \sum_{j=1}^{m} P(\rho = \mathbf{h'}_j \mid \mathbf{h}_u)}{\sum_{u \in S_d} w_u}, \\ GP_{S_f} = \dfrac{\sum_{u \in S_f} w_u \sum_{j=1}^{m} P(\rho = \mathbf{h'}_j \mid \mathbf{h}_u)}{\sum_{u \in S_f} w_u}. \end{cases} \tag{8}$$

where $w_u = \frac{\text{conf}(u)}{\sum_{z=1}^{n} \text{conf}(z)}$ is the weight of node u.

In this way, nodes with high-confidence predictions receive more stringent fairness treatment, ensuring that their label prediction does not induce bias. Meanwhile, nodes with low-confidence predictions incur a smaller fairness penalty, acknowledging that the classifier's uncertainty already diminishes the likelihood of discrimination arising from their representations while also helping the model increase the confidence of its predictions.

To sum up, the adaptivity confidence strategy module helps the model concentrate on fully masking demographic information for nodes where the proxy of demographic information is more reliably recognized, thereby efficiently allocating fairness constraints where they are most needed and promoting both fairness and prediction accuracy.

4.5 Overall Learning Object

To jointly optimize utility and fairness, we define a unified loss function for training our demographic-agnostic fair graph representation (DFGR) framework. Specifically, we combine four terms: i) information loss for preserving task-related information, ii) graph reconstruction loss to reconstruct the graph topology, and iii) fairness loss for removing demographics-related information. The final objective function is:

$$\begin{aligned} \min \quad \mathcal{L}_{\text{total}} &= \mathcal{L}_I \; + \; a\,\mathcal{L}_G \; + \; b\,\mathcal{L}_F \\ &= \frac{1}{|\mathcal{V}_L|} \sum_{v_u \in \mathcal{V}_L} \Big[- \big(y_u \log(\hat{y}_u) \; + \; (1 - y_u) \log\big(1 - \hat{y}_u\big) \big) \Big] \\ &= \text{GPD} \; + \; \sum_{u=1}^{n} \big\| \mathbf{h}_u - \mathbf{h}'_u \big\|^2 \\ &= \frac{1}{|\mathcal{E}_{S_d}| + |\mathcal{E}_{S_f}|} \sum_{e_{u,k} \in \mathcal{E}} L\big(e_{u,k},\, \hat{e}_{u,k}\big) \end{aligned} \tag{9}$$

where a and b are tunable hyperparameters to balance the contributions of the various elements in the overall objective function. The terms $\mathcal{L}_I$, $\mathcal{L}_G$, and $\mathcal{L}_F$ correspond to the utility loss, the graph reconstruction loss, and the fairness loss, respectively.

5 Experiment

In this section, we describe the experimental design used to comprehensively evaluate our proposed framework, DFGR. We first introduce the datasets utilized in the experiments, followed by a description of the baselines selected for comparison. Next, we outline the evaluation metrics adopted to assess prediction and fairness performance. Finally, we present and analyze the results of the experiments.

5.1 Experimental Settings

Datasets. We evaluate DFGR on four widely used real-world datasets, *i.e.*, the **Credit** dataset [50], **Pokec-z** and **Pokec-n** datasets [31], and the **NBA** dataset [8]. The **Credit** dataset consists of credit card holders represented as nodes, connected by edges based on similarities in spending and payment behaviors. Each node includes transaction-related features. The **Pokec-z** and **Pokec-n** datasets originate from a popular social network in Slovakia, corresponding to two distinct provincial sub-networks. Nodes represent users characterized by attributes such as gender, age, and interests, while edges represent friendships. The prediction task involves classifying users' occupational fields. The **NBA** dataset models professional basketball players as nodes, connected based on similarity in performance metrics. The prediction task is to determine if a player's salary exceeds the league average. The detailed statistics of these datasets are shown in Table 1. In all datasets, isolated nodes are removed before experiments. The data is partitioned into training (50%), validation (20%), and testing (30%) sets. To evaluate the effectiveness of our method under scenarios with incomplete demographic information, we mask 40% demographic information in the training and validation sets while maintaining complete labeling of demographic information in the test set.

Table 1: Summary of the datasets in the experiments.

Dataset	**Credit**	**Pokec-z**	**Pokec-n**	**NBA**
Vertices	30,000	67,797	66,569	403
Edges	137,377	882,765	729,129	16,570
Feature Dimension	13	65	65	97
Demographics	Age	Region	Region	Country

Baselines. We compare the proposed framework with several state-of-the-art methods, categorized into two groups: **i) Vanilla Graph Model**: GCN [15], which utilizes spectral graph convolutions without explicit fairness constraints; **ii) Fairness-aware Methods**: FairKD [3] addresses fairness by first overfitting a teacher model to generate soft labels, which then guide a student model via knowledge distillation. KSMOTE [48] leverages clustering to assign proxy demographic labels, balancing subgroup representation through synthetic oversampling. FairRF [60] promotes fairness by explicitly exploring and mitigating feature-related biases, eliminating reliance on demographic

information. FairAC [12] extends fairness considerations beyond i.i.d. data to graph data; it generates node embeddings based on observed attributes and employs an attention mechanism to aggregate neighbor information for nodes with missing attributes. FairGKD [63] enhances fairness in graph scenarios through graph-based knowledge distillation, transferring fair representations learned by a teacher GNN to a student model, and knowledge sharing. Finally, FairGNN [8] aims to learn fair GNNs with limited demographic information by employing a demographic information estimator to predict the demographics while improving fairness via adversarial learning. For methods not originally designed for graph data (FairKD, KSMOTE, and FairRF), we adapt them to work with our method backbone using the authors' original implementations.

Evaluation Metrics. We evaluated the proposed framework with respect to two key aspects: prediction performance and fairness performance. To evaluate prediction performance, we chose two metrics for node classification, *i.e.*, accuracy and F1-Score [29], where higher scores indicate better prediction results. For fairness assessment, we utilize two commonly used metrics: Demographic parity (Δ_{DP}) [19] and Equal Opportunity (Δ_{EO}) [13]. These fairness metrics measure the disparity in predictions between different demographic groups, where values closer to zero indicate higher fairness.

5.2 Experimental Results

RQ1: How does DFGR perform in balancing utility and fairness across real-world graph datasets? To answer this question, Table 2 summarizes the comparisons between our proposed method, DFGR, and the baseline methods. Specifically, two key observations emerge: i) DFGR achieves superior fairness when demographic information is missing. Across all evaluated datasets, DFGR consistently demonstrates better fairness performance than baseline methods. This advantage stems from DFGR's ability to effectively leverage node features and graph structure to accurately generate proxy of demographic information, establishing a solid foundation for bias mitigation. Furthermore, DFGR mitigates multiple forms of bias in graph data, better preventing demographic information from leaking into downstream classification tasks. ii) DFGR demonstrates comparable predictive performance compared with existing fairness methods. Unlike existing approaches that impose uniform fairness constraints on all nodes, DFGR dynamically adjusts each node's contribution to the fairness loss through the adaptivity confidence strategy, enabling better learning for nodes with low confidence. Overall, these results highlight DFGR's advantage in effectively balancing predictive performance and fairness.

RQ2: How Do the Hyper-parameters a and b Impact the Trade-off Between Utility and Fairness in DFGR? We investigate the sensitivity of DFGR to two key hyper-parameters, *i.e.*, a and b. As shown in Figure 3, as a increases, the model achieves better prediction performance and fairness. However, if it passes a certain threshold, both prediction performance and fairness stabilize. For parameter b, as shown in Figure 3, we observe three distinct phases: when b is very small, the fairness constraints have minimal impact. As b increases, fairness steadily improves, though prediction accuracy gradually declines due to stronger regularization. Beyond a threshold (*e.g.*, e^1 for Credit/NBA, e^3 for Pokec-z/Pokec-n), fairness performance stabilizes or slightly deteriorates because excessive regularization restricts the model's representational capacity.

Table 2: Comparison of DFGR with baseline methods on real-world datasets. The best-performing result in each row is highlighted in bold, and the second-best result is underlined.

Dataset	Metrics \ Methods	GCN	KSMOTE	FairKD	FairRF	FairAC	FairGKD	FairGNN	DFGR
Credit	Accuracy (↑)	**0.781 ± 0.016**	0.736 ± 0.009	0.711 ± 0.012	0.735 ± 0.017	0.748 ± 0.026	0.743 ± 0.028	0.687 ± 0.012	0.743 ± 0.032
	F1-Score (↑)	**0.868 ± 0.023**	0.817 ± 0.012	0.796 ± 0.023	0.809 ± 0.022	0.831 ± 0.018	0.834 ± 0.013	0.783 ± 0.043	0.825 ± 0.018
	$\Delta_{DP}(\downarrow)$	0.117 ± 0.013	0.071 ± 0.003	0.094 ± 0.036	0.067 ± 0.017	0.047 ± 0.015	0.038 ± 0.011	0.123± 0.036	**0.034 ± 0.015**
	$\Delta_{EO}(\downarrow)$	0.096 ± 0.017	0.055 ± 0.013	0.075 ± 0.042	0.057 ± 0.018	0.041 ± 0.014	0.037 ± 0.021	0.115 ± 0.042	**0.030 ± 0.013**
Pokec-z	Accuracy (↑)	**0.699 ± 0.024**	0.697 ± 0.024	0.673 ± 0.021	0.690 ± 0.014	0.655 ± 0.031	0.660 ± 0.025	0.689 ± 0.071	0.671 ± 0.041
	F1-Score (↑)	**0.622 ± 0.024**	0.611 ± 0.018	0.592 ± 0.013	0.617 ± 0.019	0.603 ± 0.013	0.618 ± 0.009	0.603 ± 0.021	0.620 ± 0.032
	$\Delta_{DP}(\downarrow)$	0.075 ± 0.025	0.037 ± 0.017	0.045 ± 0.014	0.032 ± 0.012	0.032 ± 0.018	**0.029 ± 0.021**	0.038 ± 0.022	0.031 ± 0.013
	$\Delta_{EO}(\downarrow)$	0.062 ± 0.013	0.039 ± 0.010	0.048 ± 0.009	0.034 ± 0.012	0.029 ± 0.014	0.030 ± 0.018	0.033 ± 0.029	**0.027 ± 0.015**
Pokec-n	Accuracy (↑)	0.689 ± 0.015	0.669 ± 0.013	0.663 ± 0.016	0.673 ± 0.013	0.675 ± 0.028	0.681 ± 0.021	0.675 ± 0.028	**0.689 ± 0.024**
	F1-Score (↑)	**0.631 ± 0.022**	0.611 ± 0.018	0.603 ± 0.023	0.616 ± 0.032	0.621 ± 0.026	0.628 ± 0.029	0.619 ± 0.032	0.630 ± 0.029
	$\Delta_{DP}(\downarrow)$	0.084 ± 0.013	0.061 ± 0.010	0.067 ± 0.015	0.056 ± 0.009	0.026 ± 0.013	0.025 ± 0.015	0.036 ± 0.012	**0.021 ± 0.018**
	$\Delta_{EO}(\downarrow)$	0.078 ± 0.019	0.066 ± 0.013	0.064 ± 0.013	0.061 ± 0.016	0.025 ± 0.027	0.027 ± 0.030	0.044 ± 0.020	**0.023 ± 0.010**
NBA	Accuracy (↑)	0.668 ± 0.025	0.654 ± 0.023	0.671 ± 0.036	0.664 ± 0.033	**0.673 ± 0.028**	0.670 ± 0.024	0.658 ± 0.027	0.723 ± 0.024
	F1-Score (↑)	0.703 ± 0.022	0.685 ± 0.038	0.681 ± 0.023	0.687 ± 0.012	0.699 ± 0.038	0.706 ± 0.033	0.694 ± 0.032	**0.711 ± 0.029**
	$\Delta_{DP}(\downarrow)$	0.063 ± 0.043	0.057 ± 0.033	0.042 ± 0.025	0.044 ± 0.038	0.034 ± 0.004	0.040 ± 0.067	0.036 ± 0.021	**0.032 ± 0.036**
	$\Delta_{EO}(\downarrow)$	0.074 ± 0.043	0.065 ± 0.033	0.055 ± 0.014	0.042 ± 0.026	0.037 ± 0.017	0.032 ± 0.010	0.034 ± 0.025	**0.028 ± 0.005**

To sum up, these results highlight the trade-off between fairness and task performance. Thus, careful tuning of a and b is essential for optimal model performance.

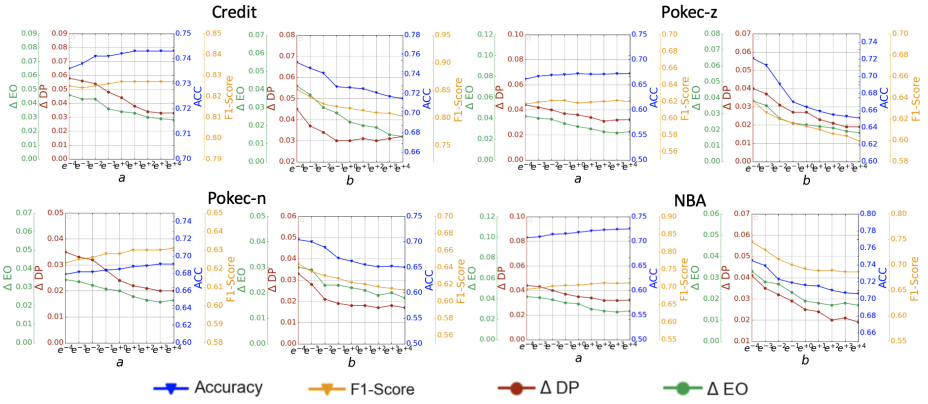

Fig. 3: Study on Hyper-parameters sensitivity analysis.

RQ3: What is the Impact of Each Component on the DFGR on its utility and fairness? We conducted ablation studies to assess the contributions of each module within the DFGR framework. DFGR consists of three key modules: the Demographic Information Identification Module, the Fair Node Representation Learning Module, and the Adaptivity Confidence Strategy Module. Notable, we did not create a variant without the Demographic Information Identification Module because this module is foundational to DFGR's operation. Without it, the framework cannot identify proxies for the missing demographic information, which are essential inputs for the subsequent modules. Therefore, removing this component would render the entire framework inoperable, making such an ablation study impractical.

For the Fair Node Representation Learning Module, we created two variants: DFGR-NF (without the Fairness Constraint) and DFGR-NG (without the Graph Recon-

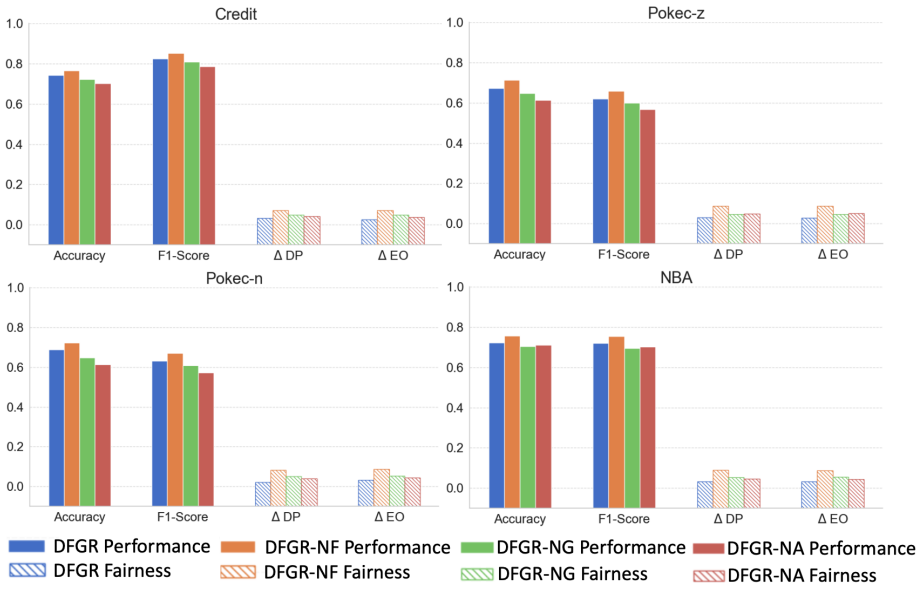

Fig. 4: Ablation study results for DFGR, DFGR-NF, DFGR-NG and DFGR-NA.

struction Constraint). As shown in Figure 4, the fairness metrics of DFGR-NF dropped significantly. This occurs because without the Fairness Constraint, demographic information in node representations directly passes to downstream classification tasks, leading to discriminatory decisions. The DFGR-NG variant shows better fairness metrics than DFGR-NF but still demonstrates a slight decrease compared to the full DFGR model, along with reduced performance metrics. This degradation occurs because without the Graph Reconstruction Constraint, node representations fail to capture important structural information, resulting in decreased graph representational power.

We also examined the impact of the Adaptivity Confidence Strategy Module by creating the DFGR-NA variant (without adaptive confidence). As shown in Figure 4, DFGR-NA shows reduced performance compared to the complete DFGR model. This is because applying fairness constraints with equal strength to all nodes makes it more difficult for the model to learn from samples with low confidence, thereby reducing the overall accuracy.

RQ4: What Effect of Different Number of Prototypes m Values on DFGR's Fairness and Utility? Similar to the previous RQs, we conducted experiments with a variety of values for m in {5, 10, 15, 20, 25}, keeping all other training factors the same. We compare DFGR's utility and fairness under different settings. As shown in Figure 5, as m increases, the model exhibits enhanced fairness with an increasing number of prototypes; however, this improvement plateaus beyond a certain threshold (*e.g.*, smaller than 10 for Credit, 10 for Pokec-z, 15 for Pokec-n, and 20 for NBA). Conversely, predictive performance diminishes as the number of prototypes rises. Overall, the optimal balance between performance and fairness seems to be achieved when the number of prototypes is within the 10 to 15 range.

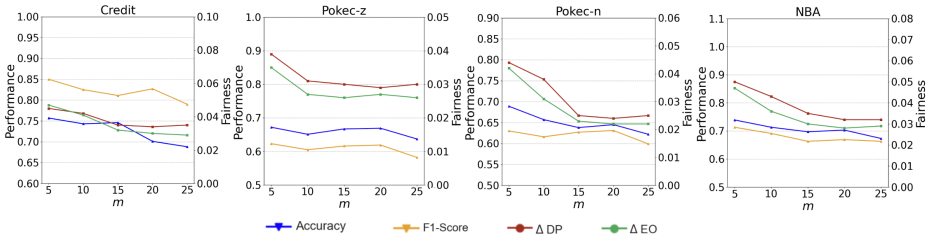

Fig. 5: Study the choice of the number of prototypes.

6 Conclusion

In this paper, we tackled the limitation of existing fairness-aware graph learning methods, which require complete demographic information. Recognizing that real-world scenarios frequently involve missing or restricted access to demographic information due to privacy, ethical, or regulatory concerns, we proposed a novel framework to achieve fairness in graph learning without relying on fully available demographics. Moreover, by integrating adaptive confidence strategies, our method effectively balances fairness and utility, mitigating bias while minimizing degradation in predictive performance. The proposed modules are readily extensible to existing fairness-aware graph learning frameworks. Experiments on four real-world datasets demonstrate that DFGR outperforms all baseline methods in terms of both fairness and utility metrics.

Acknowledgements

This work was supported in part by the National Science Foundation (NSF) under Grant No. 2404039.

References

1. An, Q., Rahman, S., Zhou, J., Kang, J.J.: A comprehensive review on machine learning in healthcare industry: classification, restrictions, opportunities and challenges. Sensors **23**(9), 4178 (2023)
2. Ashurst, C., Weller, A.: Fairness without demographic data: A survey of approaches. In: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization. pp. 1–12 (2023)
3. Chai, J., Jang, T., Wang, X.: Fairness without demographics through knowledge distillation. Advances in Neural Information Processing Systems **35**, 19152–19164 (2022)
4. Chinta, S.V., Fernandes, K., Cheng, N., Fernandez, J., Yazdani, S., Yin, Z., Wang, Z., Wang, X., Xu, W., Liu, J., et al.: Optimization and improvement of fake news detection using voting technique for societal benefit. In: 2023 IEEE International Conference on Data Mining Workshops (ICDMW). pp. 1565–1574. IEEE (2023)
5. Chinta, S.V., Wang, Z., Yin, Z., Hoang, N., Gonzalez, M., Quy, T.L., Zhang, W.: Fairaied: Navigating fairness, bias, and ethics in educational ai applications. arXiv preprint arXiv:2407.18745 (2024)

6. Chinta, S.V., Wang, Z., Zhang, X., Viet, T.D., Kashif, A., Smith, M.A., Zhang, W.: Ai-driven healthcare: A survey on ensuring fairness and mitigating bias. arXiv preprint arXiv:2407.19655 (2024)
7. Chu, Z., Wang, Z., Zhang, W.: Fairness in large language models: A taxonomic survey. ACM SIGKDD Explorations Newsletter, 2024 pp. 34–48 (2024)
8. Dai, E., Wang, S.: Say no to the discrimination: Learning fair graph neural networks with limited sensitive attribute information. In: Proceedings of the 14th ACM International Conference on Web Search and Data Mining. pp. 680–688 (2021)
9. Z., Wang, A., Palikhe, Z., Yin, Z., Zhang, W.: Fairness definitions in language models explained. arXiv preprint arXiv:2407.18454 (2024)
10. Dzuong, J., Wang, Z., Zhang, W.: Uncertain boundaries: Multidisciplinary approaches to copyright issues in generative ai. arXiv preprint arXiv:2404.08221 (2024)
11. Grari, V., Lamprier, S., Detyniecki, M.: Fairness without the sensitive attribute via causal variational autoencoder. arXiv preprint arXiv:2109.04999 (2021)
12. Guo, D., Chu, Z., Li, S.: Fair attribute completion on graph with missing attributes. arXiv preprint arXiv:2302.12977 (2023)
13. Hardt, M., Price, E., Srebro, N.: Equality of opportunity in supervised learning. Advances in neural information processing systems **29** (2016)
14. Kenfack, P.J., Kahou, S.E., Aïvodji, U.: A survey on fairness without demographics. Transactions on Machine Learning Research (2024)
15. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
16. Ko, H., Lee, S., Park, Y., Choi, A.: A survey of recommendation systems: recommendation models, techniques, and application fields. Electronics **11**(1), 141 (2022)
17. Krumpal, I.: Determinants of social desirability bias in sensitive surveys: a literature review. Quality & quantity **47**(4), 2025–2047 (2013)
18. Lahoti, P., Beutel, A., Chen, J., Lee, K., Prost, F., Thain, N., Wang, X., Chi, E.: Fairness without demographics through adversarially reweighted learning. Advances in neural information processing systems **33**, 728–740 (2020)
19. Le Quy, T., Roy, A., Iosifidis, V., Zhang, W., Ntoutsi, E.: A survey on datasets for fairness-aware machine learning. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **12**(3), e1452 (2022)
20. Ling, H., Jiang, Z., Luo, Y., Ji, S., Zou, N.: Learning fair graph representations via automated data augmentations. In: International Conference on Learning Representations (ICLR) (2023)
21. Martinez, N.L., Bertran, M.A., Papadaki, A., Rodrigues, M., Sapiro, G.: Blind pareto fairness and subgroup robustness. In: International Conference on Machine Learning. pp. 7492–7501. PMLR (2021)
22. Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., Beaton, M.: Teens, social media, and privacy. Pew Research Center **21**(1055), 2–86 (2013)
23. Ni, H., Han, L., Chen, T., Sadiq, S., Demartini, G.: Fairness without sensitive attributes via knowledge sharing. In: The 2024 ACM Conference on Fairness, Accountability, and Transparency. pp. 1897–1906 (2024)
24. Pearl, J., et al.: Models, reasoning and inference. Cambridge, UK: CambridgeUniversityPress **19**(2), 3 (2000)
25. Peers, S., Hervey, T., Kenner, J., Ward, A.: The EU Charter of fundamental rights: a commentary. Bloomsbury Publishing (2021)
26. Pelegrina, G.D., Couceiro, M., Duarte, L.T.: A statistical approach to detect sensitive features in a group fairness setting. arXiv preprint arXiv:2305.06994 (2023)
27. Rawls, A.: Theories of social justice (1971)

28. Sagawa, S., Koh, P.W., Hashimoto, T.B., Liang, P.: Distributionally robust neural networks for group shifts: On the importance of regularization for worst-case generalization. arXiv preprint arXiv:1911.08731 (2019)
29. Sokolova, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. Inf. Process. Manage. **45**(4), 427–437 (Jul 2009). https://doi.org/10.1016/j.ipm.2009.03.002, https://doi.org/10.1016/j.ipm.2009.03.002
30. Tabassum, S., Pereira, F.S., Fernandes, S., Gama, J.: Social network analysis: An overview. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **8**(5), e1256 (2018)
31. Takac, L., Zabovsky, M.: Data analysis in public social networks. In: International scientific conference and international workshop present day trends of innovations. vol. 1 (2012)
32. Wang, Z., Chu, Z., Blanco, R., Chen, Z., Chen, S.C., Zhang, W.: Advancing graph counterfactual fairness through fair representation learning. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 40–58. Springer Nature Switzerland (2024)
33. Wang, Z., Chu, Z., Doan, T.V., Wang, S., Wu, Y., Palade, V., Zhang, W.: Fair graph u-net: A fair graph learning framework integrating group and individual awareness. In: proceedings of the AAAI conference on artificial intelligence. vol. 39, pp. 28485–28493 (2025)
34. Wang, Z., Narasimhan, G., Yao, X., Zhang, W.: Mitigating multisource biases in graph neural networks via real counterfactual samples. In: 2023 IEEE International Conference on Data Mining (ICDM). pp. 638–647. IEEE (2023)
35. Wang, Z., Qiu, M., Chen, M., Salem, M.B., Yao, X., Zhang, W.: Toward fair graph neural networks via real counterfactual samples. Knowledge and Information Systems pp. 1–25 (2024)
36. Wang, Z., Saxena, N., Yu, T., Karki, S., Zetty, T., Haque, I., Zhou, S., Kc, D., Stockwell, I., Wang, X., et al.: Preventing discriminatory decision-making in evolving data streams. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency. pp. 149–159 (2023)
37. Wang, Z., Ulloa, D., Yu, T., Rangaswami, R., Yap, R., Zhang, W.: Individual fairness with group constraints in graph neural networks. In: 27th European Conference on Artificial Intelligence (2024)
38. Wang, Z., Wallace, C., Bifet, A., Yao, X., Zhang, W.: Fg^2an: Fairness-aware graph generative adversarial networks. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 259–275. Springer Nature Switzerland (2023)
39. Wang, Z., Yin, Z., Zhang, Y., Yang, L., Zhang, T., Pissinou, N., Cai, Y., Hu, S., Li, Y., Zhao, L., et al.: Fg-smote: Towards fair node classification with graph neural network. ACM SIGKDD Explorations Newsletter **26**(2), 99–108 (2025)
40. Wang, Z., Yin, Z., Zhang, Y., Yang, L., Zhang, T., Pissinou, N., Cai, Y., Hu, S., Li, Y., Zhao, L., et al.: Graph fairness via authentic counterfactuals: Tackling structural and causal challenges. ACM SIGKDD Explorations Newsletter **26**(2), 89–98 (2025)
41. Wang, Z., Hoang, N., Zhang, X., Bello, K., Zhang, X., Iyengar, S. S., Zhang, W.: Towards Fair Graph Learning without Demographic Information. In: The 28th International Conference on Artificial Intelligence and Statistics, vol. 258, pp. 2
42. Wang, Z., Zhang, W.: Group fairness with individual and censorship constraints. In: 27th European Conference on Artificial Intelligence (2024)
43. Wang, Z., Zhou, Y., Qiu, M., Haque, I., Brown, L., He, Y., Wang, J., Lo, D., Zhang, W.: Towards fair machine learning software: Understanding and addressing model bias through counterfactual thinking. arXiv preprint arXiv:2302.08018 (2023)
44. Wang, Z., Wu, A., Moniz, N., Hu, S., Knijnenburg, B., Zhu, Q., Zhang, W.: Towards Fairness with Limited Demographics via Disentangled Learning. In: Proceedings of the 34th International Joint Conference on Artificial Intelligence (IJCAI), (2025)

45. Wang, Z., Zhang, W.: FDGen: A Fairness-Aware Graph Generation Model. In: Proceedings of the 42nd International Conference on Machine Learning (ICML). PMLR, (2025)
46. Wang, Z., Liu, F., Pan, S., Liu, J., Saeed, F., Qiu, M., Zhang, W.: fairGNN-WOD: Fair Graph Learning Without Complete Demographics. In: Proceedings of the 34th International Joint Conference on Artificial Intelligence (IJCAI), (2025)
47. Wang, Z., Yin, Z., Yap, R., Zhang, X., Hu, S., Zhang, W.: Redefining Fairness: A Multi-dimensional Perspective and Integrated Evaluation Framework. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, (2025)
48. Yan, S., Kao, H.t., Ferrara, E.: Fair class balancing: Enhancing model fairness without observing sensitive attributes. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. pp. 1715–1724 (2020)
49. Yazdani, S., Saxena, N., Wang, Z., Wu, Y., Zhang, W.: A comprehensive survey of image and video generative ai: Recent advances, variants, and applications (2024)
50. Yeh, I.C., Lien, C.h.: 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 (2009)
51. Yin, Z., Agarwal, S., Kashif, A., Gonzalez, M., Wang, Z., Liu, S., Liu, Z., Wu, Y., Stockwell, I., Xu, W., et al.: Accessible health screening using body fat estimation by image segmentation. In: 2024 IEEE International Conference on Data Mining Workshops (ICDMW). pp. 405–414. IEEE (2024)
52. Yin, Z., Wang, Z., Xu, W., Zhuang, J., Mozumder, P., Smith, A., Zhang, W.: Digital forensics in the age of large language models. arXiv preprint arXiv:2504.02963 (2025)
53. Yin, Z., Wang, Z., Zhang, W.: Improving fairness in machine learning software via counterfactual fairness thinking. In: Proceedings of the 2024 IEEE/ACM 46th International Conference on Software Engineering: Companion Proceedings. pp. 420–421 (2024)
54. Zhang, W., Hernandez-Boussard, T., Weiss, J.: Censored fairness through awareness. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 14611–14619 (2023)
55. Zhang, W., Ntoutsi, E.: Faht: an adaptive fairness-aware decision tree classifier. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence (2019)
56. Zhang, W., Wang, Z., Kim, J., Cheng, C., Oommen, T., Ravikumar, P., Weiss, J.: Individual fairness under uncertainty. In: 26th European Conference on Artificial Intelligence. pp. 3042–3049 (2023)
57. Zhang, W., Weiss, J.C.: Longitudinal fairness with censorship. In: proceedings of the AAAI conference on artificial intelligence. vol. 36, pp. 12235–12243 (2022)
58. Zhang, W., Zhou, S., Walsh, T., Weiss, J.C.: Fairness Amidst Non-IID Graph Data: A Literature Review. AI Magazine, vol. 46, no. 1, article e12212 (2025)
59. Zhang, W.: AI Fairness in Practice: Paradigm, Challenges, and Prospects. AI Magazine, vol. 45, no. 3, pp. 386–395 (2024)
60. Zhao, T., Dai, E., Shu, K., Wang, S.: Towards fair classifiers without sensitive attributes: Exploring biases in related features. In: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. pp. 1433–1442 (2022)
61. Zheng, X., Wang, Y., Liu, Y., Li, M., Zhang, M., Jin, D., Yu, P.S., Pan, S.: Graph neural networks for graphs with heterophily: A survey. arXiv preprint arXiv:2202.07082 (2022)
62. Zhu, H., Fu, G., Guo, Z., Zhang, Z., Xiao, T., Wang, S.: Fairness-aware message passing for graph neural networks. arXiv preprint arXiv:2306.11132 (2023)
63. Zhu, Y., Li, J., Chen, L., Zheng, Z.: The devil is in the data: Learning fair graph neural networks via partial knowledge distillation. In: Proceedings of the 17th ACM International Conference on Web Search and Data Mining. pp. 1012–1021 (2024)
64. Zhu, Y., Li, J., Zheng, Z., Chen, L.: Fair graph representation learning via sensitive attribute disentanglement. In: Proceedings of the ACM Web Conference 2024. pp. 1182–1192 (2024)