

GUIDE: Group Equality Informed Individual Fairness in Graph Neural Networks

ABSTRACT

Graph Neural Networks (GNNs) are playing increasingly important roles in critical decision-making scenarios due to their exceptional performance and end-to-end design. However, concerns have been raised that GNNs could make biased decisions against underprivileged groups or individuals. To remedy this issue, researchers have proposed various fairness notions including individual fairness that gives similar predictions to similar individuals. However, existing methods in individual fairness rely on Lipschitz condition: they only optimize overall individual fairness and disregard equality of individual fairness between groups characterized by sensitive attributes such as age and race. This leads to drastically different levels of individual fairness among groups, and hence certain demographic groups are discriminated. We tackle this problem by proposing a novel GNN framework GUIDE to achieve group equality informed individual fairness in GNNs. We aim to not only achieve individual fairness but also equalize the levels of individual fairness among groups. Specifically, the proposed framework operates on the similarity matrix to learn individually customized attention to achieve individual fairness without group level disparity. Furthermore, GUIDE is trained end-to-end and compatible with arbitrary GNN architectures. Comprehensive experiments on real-world datasets demonstrate GUIDE obtains good balance of group equality informed individual fairness and model utility.

KEYWORDS

Individual Fairness, Group Fairness, Graph Neural Networks

1 INTRODUCTION

Graph data is ubiquitous across a myriad academic and industrial domains, e.g., relationship graphs in social networks [22], user-content graphs in online shopping sites [28], biological interaction networks [20] and knowledge graphs [13]. In fact, graphs provide a useful abstraction to describe data and their inherent relations. To effectively gain deeper understanding from graph data, various algorithms have been proposed to tackle different graph mining tasks, including prediction [12], community detection [2], recommendation [8], and many more. Among them, Graph Neural Networks (GNNs) have attracted research attention in recent years due to their superior learning performances [10, 16, 23]. They are increasingly adopted in various tasks such as anomaly detection [25], social network recommendation [27], and graph classification [26]. Although GNNs have excelled in a diversity of tasks, concerns have been raised that directly adopting GNNs could empirically result in ethical and fairness issues [1, 4], such as racial or gender discrimination, which renders the adoption of GNNs in high-stake scenarios questionable. Generally, to analyze algorithmic fairness, researchers have developed multiple fairness notions [19], such as group fairness [11], which ensures equal outcome rates for members of different demographic subgroups; and

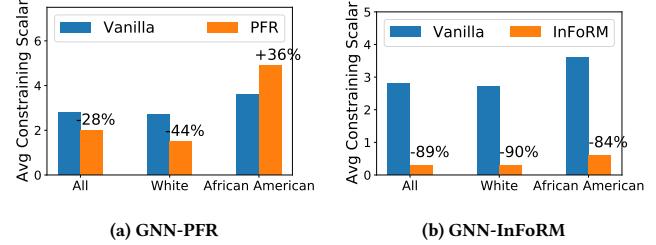


Figure 1: Average constraining scalar optimization based on PFR and INFoRM. Percentages denote the optimization results compared with vanilla values. Group disparity is exacerbated because the privileged group (White) receives significantly better optimization results compared with the disadvantaged group (African American).

individual fairness [7], which promotes treating similar individuals similarly. While group fairness has been widely studied in graph representation learning [17, 21] for GNNs, individual fairness still remains under-explored. Nevertheless, considering that individual fairness is able to enforce fairness at a finer granularity on the pairwise individual level compared with group fairness, it is a desirable notion of fairness to enforce in GNNs.

To model individual fairness in graphs, Lipschitz condition is the most commonly adopted mathematical foundation among existing works [7, 14, 18]. Specifically, for any pair of individuals (i, j) , there is a constraining scalar $\epsilon_{i,j}$ such that their output distance is bounded by their input distance multiplied by this scalar. Here, the input distance between individuals could be given by domain experts or oracle similarity matrix [5, 14]. And the largest constraining scalar across all pairs is named as the Lipschitz constant. Intuitively, if the Lipschitz constant is small, then the outcome distance is also constrained to be relatively small for similar individual pairs (i.e. pairs with small input distance) in the dataset. This implies that *similar people are treated similarly*, and vice versa. To enforce individual fairness, a commonly adopted approach¹ in existing works [14, 18] is to minimize the sum of output distance divided by input distance (or multiplied by similarity) for all pairs of individuals. Furthermore, this sum is bounded by the average constraining scalar ϵ for all pairs of individuals in the dataset, times the total number of pairs m . Therefore, as this loss is minimized, the average constraining scalar ϵ is also minimized and in this paper we use the terms *minimizing individual fairness loss* and *minimizing average constraining scalar* interchangably. Such technique has been empirically proved to be effective in enforcing the Lipschitz constant to be as small as possible. However, for different individual pairs in the population, the optimization of the constraining

¹Existing works use this loss form to optimize individual fairness: $\mathcal{L} = \sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}} \|Z[i, :] - Z[j, :]||_2^2 S[i, j]$ and $\mathcal{L} \leq m\epsilon$, where \mathcal{V} is the set of individuals, Z is the model output matrix, S is the similarity matrix, m is the number of pairwise comparisons, ϵ is the average constraining scalar.

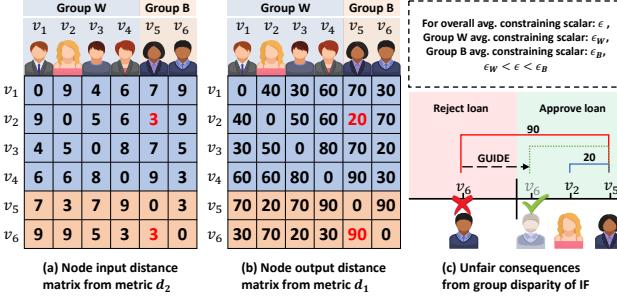


Figure 2: A toy example on the disparity of individual fairness between different groups in a loan approval system. Since Group W has lower average constraining scalar than Group B, its member v_2 could receive a lower outcome distance to v_5 compared to an individual v_6 from Group B (20 vs. 90) given equal input distances of 3. Such difference could place v_6 across the decision boundary, giving him different outcomes when he is equally similar to v_5 compared to v_2 .

scalar could be potentially influenced by the sensitive feature of the involved individuals such as gender or race. For example, the optimization results of some privileged demographic subgroups could be significantly better than that of the disadvantaged groups. We empirically show the extensive presence of such phenomenon in existing works through preliminary analysis below.

PFR [18] and INFoRM [14] are two representative works that optimize individual fairness in graphs. The effectiveness of these two approaches for the optimization of the average constraining scalar between the outcome and input distance for individual pairs has been empirically proved. Nevertheless, race could severely affect the optimization results in both approaches. For example, empirical explorations are shown in Fig. 1, where PFR and INFoRM are adapted to GNNs for node classification task on Income dataset [6]. Generally, the overall average constraining scalar is optimized towards a lower value, which indicates a lower level of the Lipschitz constant. However, such optimization effectiveness is largely attributed to the optimization of the constraining scalars for individual pairs involving white individuals. For individual pairs involving African Americans, they do not enjoy as much optimization as pairs involving white individuals in INFoRM, and their situation is even worse after the optimization of PFR.

It is worth mentioning that the group disparity of individual fairness optimization discussed above could lead to discrimination in real-world decision-making scenarios. Here we utilize an illustrative toy example in Fig. 2 to scrutinize how such group disparity leads to discrimination. Assume that two races (group W for white and group B for black) are involved in a loan approval system. Given the input distance matrix for individual pairs (i.e., the matrix in Fig. 2(a)), assume that the outcome distance (i.e., distance in the matrix of Fig. 2(b)) is already optimized through an existing individual fairness enforcing approach. In this example, the average constraining scalar for individual pairs involving members of group W (blue entries) is at a significantly lower level compared with that of group B (orange entries). Assume that there is a black (v_6) and a white (v_2) individual who are both with small input distances of 3 compared

to a third black individual v_5 who is already approved for the loan. Since the input distances of (v_2, v_5) and (v_6, v_5) are small, both pairs should be considered as similar pairs, and thus they should receive similar outcomes for loan approval. Nevertheless, as a consequence of being with larger constraining scalar, even though the black individual v_6 has the same input distance as v_2 to v_5 , v_6 still receives larger outcome distance to v_5 compared with v_2 (90 vs 20). A larger outcome distance could potentially put the black individual v_6 on the other side of the decision boundary, which makes v_6 more likely to receive a different outcome for his loan application.

To properly handle the problems we mentioned above, in this paper, we study a novel problem of enforcing group equality informed individual fairness. Specifically, we first design a metric to capture the disparity of individual fairness in groups. Then a novel GNN framework named GUIDE (GUIDE) is proposed to ensure not only overall individual fairness but also similar levels of individual fairness between groups after optimization, and moreover maintain model utility performance. GUIDE includes two main modules: a backbone GNN network learning from adjacency matrix to extract high-level node embeddings for downstream tasks, and an attention-based GNN layer learning from similarity matrix to achieve both group equality informed individual fairness and high utility performance. Extensive experiments demonstrate the effectiveness of the proposed framework. Our main contributions can be summarized as:

- **Problem Formulation.** We propose a metric to quantitatively measure group disparity of individual fairness with a solid theoretical foundation. Based on the metric, we formulate a novel problem of promoting group equality informed individual fairness in GNNs.
- **Algorithm Design.** We propose a novel framework GUIDE to relieve the disparity of individual fairness in different groups, optimize overall individual fairness and preserve prediction performance in downstream tasks.
- **Experimental Evaluation.** We conduct comprehensive experiments with baseline models and real-world datasets, which validates the superior performance of our framework.

2 PROBLEM FORMULATION

In this section, we introduce notations, preliminaries on existing individual fairness notions on graphs and present a new metric that can measure group disparity of individual fairness on graphs. Finally, we formulate our research problem.

2.1 Notations

In this paper, we use bold uppercase characters (e.g., \mathbf{A}) for matrices, bold lowercase characters (e.g., \mathbf{a}) for vectors, lowercase characters (e.g., a) for scalars, uppercase caligraphic characters (e.g., \mathcal{V}) for sets. Also, we represent the i -th row, j -th column, (i, j) -th entry of a matrix \mathbf{A} as $\mathbf{A}[i, :]$, $\mathbf{A}[:, j]$ and $\mathbf{A}[i, j]$, respectively. Additionally, we use lowercase bold vectors with index to represent the row vector of a matrix (e.g. $\mathbf{z}_i = \mathbf{Z}[i, :]$). The trace of matrix \mathbf{A} is $\text{Tr}(\mathbf{A})$. The ℓ_2 -norm of a vector $\mathbf{a} \in \mathbb{R}^d$ is $\|\mathbf{a}\|_2 = \sqrt{\mathbf{a}^\top \mathbf{a}}$. A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ consists of (1) \mathcal{V} : set of nodes ($|\mathcal{V}| = n$), (2) $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$: set of edges, and (3) $\mathbf{X} \in \mathbb{R}^{n \times d}$: node attributes where $\mathbf{x}_i \in \mathbb{R}^d$ is the attribute vector for i -th node. We assume there is a sensitive

Table 1: Symbols used in this paper.

Symbols	Definitions
$A \in \{0, 1\}^{n \times n}$	Adjacency matrix
$X \in \mathbb{R}^{n \times d}$	Features matrix
$H \in \mathbb{R}^{n \times h}$	Hidden state matrix
$Z \in \mathbb{R}^{n \times c}$	graph mining output matrix
$S \in \mathbb{R}^{n \times n}$	Similarity matrix
$L \in \mathbb{Z}^{n \times n}$	Laplacian of similarity matrix
$L_p \in \mathbb{Z}^{n \times n}$	Modified Laplacian for p^{th} group \mathcal{V}_p
\mathcal{V}	Set of all nodes in a graph
\mathcal{V}_p	p^{th} group
n	Total number of nodes
d	Total number of features
h	Dimension of hidden states
c	Dimension of output
m	Total number of nonzero similarities in S
m_p	Nonzero similarities for members in \mathcal{V}_p
ϵ	overall average constraining scalar
ϵ_p	average constraining scalar for \mathcal{V}_p

attribute set \mathcal{T} containing sensitive attributes for each individual and \mathcal{T} yields G disjoint groups in \mathcal{V} . We use \mathcal{V}_p to denote the set of individuals in group p . We use A , X for graph adjacency matrix and node feature matrix respectively. We also assume there is a similarity matrix S which contains pairwise node similarities according to domain knowledge or human judgement. It is worth noting that S may not be equal to A . The Laplacian L of the similarity matrix S is derived by subtracting S from the diagonal degree matrix of S . Most symbols are summarized in Table 1.

2.2 Preliminaries

In this section, we introduce the individual fairness loss commonly used in existing works [14, 18]. The intuition of this formulation is to minimize outcome distances for similar individuals (pairs with high similarity or low input distance).

DEFINITION 1. *The individual fairness loss $\mathcal{L}_{\text{ifair}}$ in a graph mining model can be measured by*

$$\mathcal{L}_{\text{ifair}} = \frac{\sum_{i \in \mathcal{V}} \sum_{j \in \mathcal{V}} \|Z[i, :] - Z[j, :] \|_2^2 S[i, j]}{2} = \text{Tr}(Z^T L Z), \quad (1)$$

and this loss satisfies

$$\mathcal{L}_{\text{ifair}} \leq m\epsilon, \quad (2)$$

where ϵ is the average constraining scalar for outcome distance for all individual pairs, m is the number of nonzero elements in S , Z is the graph mining output matrix, S is node similarity matrix, and L is the Laplacian of S . Here, the average constraining scalar is minimized as $\mathcal{L}_{\text{ifair}}$ is minimized.

2.3 Individual Fairness of a Group

As we presented in the analysis above, existing methods for individual fairness only minimizes the average constraining scalar ϵ for all pairwise comparisons. In a similar fashion, we can motivate a group specific average constraining scalar ϵ_p for pairwise comparisons between members of \mathcal{V}_p and all individuals in \mathcal{V} . It can be understood as examining a group in society, i.e. $\mathcal{V}_p \subseteq \mathcal{V}$, whether

its members are treated fairly compared to everyone in the society. Following Eq. (1), we define a metric U_p to represent the level of individual unfairness for group \mathcal{V}_p and it is closely related to the average constraining scalar ϵ_p for \mathcal{V}_p :

$$U_p = \frac{\sum_{i \in \mathcal{V}_p} \sum_{j \in \mathcal{V}} \|Z[i, :] - Z[j, :] \|_2^2 S[i, j]}{m_p} = \frac{\text{Tr}(Z^T L_p Z)}{m_p} \leq \epsilon_p, \quad (3)$$

where m_p is the number of nonzero pairwise similarities for members of \mathcal{V}_p against all individuals in \mathcal{V} , L_p is the modified Laplacian for group \mathcal{V}_p , and ϵ_p is the average constraining scalar for group \mathcal{V}_p . As for how to obtain L_p , we can define a modified similarity matrix S_p for group \mathcal{V}_p , where

$$S_p[i, j] = \begin{cases} 2S[i, j], & i, j \in \mathcal{V}_p \\ S[i, j], & i \in \mathcal{V}_p \text{ or } j \in \mathcal{V}_p \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Taking the Laplacian from this modified similarity matrix yields a L_p that satisfies Eq. (3).

2.4 Group Disparity of Individual Fairness

In this subsection, we first formally define the notion of group equality of individual fairness by leveraging the previous definitions. Then, we design a metric to measure group disparity of individual fairness (GDIF). Finally, we formulate our research problem.

DEFINITION 2. *Group equality of individual fairness is satisfied if the levels of individual unfairness for different groups are equal, i.e. $\forall \mathcal{V}_p, \mathcal{V}_q \subseteq \mathcal{V}, U_p = U_q$.*

As we discussed earlier that U_p represents the *level of individual unfairness* of group \mathcal{V}_p . If it is different across two groups, then the group with smaller U will on average have smaller output differences compared to group with larger U for individual pairs of equal input distances. Thus, differences in U can induce bias in model outcomes where sensitive attributes such as race, gender, or age affect model outcomes. We define a metric below, involving U_p and U_q , to measure the GDIF between two groups \mathcal{V}_p and \mathcal{V}_q :

$$GDIF_{p,q} = \max\left(\frac{U_p}{U_q}, \frac{U_q}{U_p}\right). \quad (5)$$

Here $GDIF_{p,q} \geq 1$. For two groups, $GDIF_{p,q} = 1$ means equal individual fairness between them. Then, we can measure the group disparity of individual fairness for all groups in the dataset as

$$GDIF = \sum_{p,q}^{1 \leq p < q \leq G} GDIF_{p,q}, \quad (6)$$

where G is total the number of groups.

Note that here we utilize ratios instead of differences to measure GDIF. It is intentional since ratios preserve the same disparity after equal amount of percentage reduction in unfairness. For example, if GDIF equals 2 and $U_p - U_q = 1$, a 80% reduction in both U_p and U_q will result in the same GDIF while the difference of U_p and U_q will reduce by 80% to be 0.2. This reduction in difference is arguably hiding inequality because 1) intuitively equal optimization

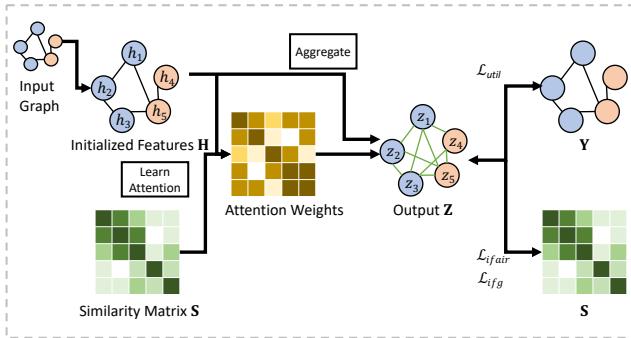


Figure 3: The overall framework of GUIDE.

(−80%) for two groups should not resolve group disparity and 2) the difference metric may be difficult to interpret since the it is hard to understand the meaning of 1 vs 0.2 while a ratio of 2 can be interpreted as group \mathcal{V}_p is twice fairer than group \mathcal{V}_q . Next, we formally define the research problem as follows.

PROBLEM 1. Promoting Group Equality Informed Individual Fairness in GNNs. Given a graph $G = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, ground truth labels \mathbf{Y} , a symmetric similarity matrix \mathbf{S} for nodes in \mathcal{V} , G disjoint groups classified by sensitive attributes ($\bigcup_{i=1}^G \mathcal{V}_i = \mathcal{V}$), our goal is to learn graph mining output \mathbf{Z} for a downstream task (such as node classification), where overall individual fairness loss (measured by Eq. (1)) is minimized and Group Disparity of Individual Fairness (measured by Eq. (6)) is minimized.

3 PROPOSED FRAMEWORK

In this section, we propose a novel GNN framework—GUIDE to solve Problem 1. Specifically, we first give an overview of the modules in GUIDE and subsequently give detailed descriptions of each module and lastly present the optimization objectives from Problem 1 as a total loss function to optimize the framework.

3.1 Framework Overview

The overview of our proposed framework is presented in Figure 3. The framework can be summarized as two modules. Each module utilizes a different input matrix, the adjacency matrix \mathbf{A} for node embedding initialization and the similarity matrix \mathbf{S} for fairness and utility optimization respectively. Both matrices are required in the framework as they may be different and contain different information that affect model outputs.

- **Backbone GNN for node initialization.** Arbitrary GNN is used to incorporate information from the adjacency matrix \mathbf{A} and feature matrix \mathbf{X} to initialize node embeddings.
- **Similarity informed attention for message passing.** This module uses fine-grained pairwise similarity information of individuals and learns personalized attention weights to aggregate node embeddings such that utility, individual fairness and GDIF objectives are optimized.

3.2 Backbone GNN for Node Initialization

We utilize a backbone GNN to initialize node embeddings \mathbf{H} as a preliminary step for GUIDE. This is performed because 1) the

adjacency matrix \mathbf{A} can often be different from similarity matrix \mathbf{S} , so it is important to incorporate its information into the framework and 2) our framework can integrate seamlessly to arbitrary GNN frameworks so it is adaptable to most existing GNN frameworks to promote group equality informed individual fairness. Specifically, an arbitrary GNN backbone network takes adjacency matrix \mathbf{A} and feature matrix \mathbf{X} as inputs to perform a given utility task e.g. node classifications by minimizing a cross-entropy loss of the output and ground truth labels. We extract the hidden layer representation $\mathbf{H} \in \mathbb{R}^{n \times h}$ before the fully connected layer from this backbone network as node embedding inputs for the next module.

3.3 Similarity Informed Attention

Our fairness objectives are two fold: overall individual fairness and group equality of individual fairness. Both objectives are evaluated using pairwise similarity metric for all nodes so the similarity matrix \mathbf{S} is crucial. To achieve our objectives, we use the similarity matrix as the input matrix for this GNN module since the message passing mechanism in GNNs aggregates neighbor nodes to derive output node representation and therefore making similar individuals have similar outputs. Furthermore, message passing on the similarity matrix sets stage for fine-grained optimization for every pair of similar individuals with respect to optimization objectives. This increased model capacity helps balance the optimization objective tradeoffs that we will elaborate in the following.

As we mentioned in 2.2, individual unfairness can be quantified and optimized from minimizing Eq. (1). As for GDIF objective, we aim to equalize levels of individual unfairness for all groups: $\forall \mathcal{V}_p, \mathcal{V}_q \subseteq \mathcal{V}, U_p = U_q$. To do so, we define a differentiable loss involving U_p and U_q for minimizing GDIF:

$$\mathcal{L}_{ifg} = \sum_{p,q}^{1 \leq p < q \leq G} \left(\frac{U_p}{U_q} - 1 \right)^2 + \left(\frac{U_q}{U_p} - 1 \right)^2. \quad (7)$$

where U_p and U_q are individual unfairness for group \mathcal{V}_p and group \mathcal{V}_q respectively. Note that this loss function is symmetrical to any given two groups such that it is the same regardless of the order. With this loss, we can effectively minimize GDIF end-to-end.

When the optimization objectives are combined (e.g. $\mathcal{L}_{ifair} + \mathcal{L}_{ifg}$), there is an empirical tradeoff of them which may be challenging to a generic GNN framework with fixed aggregation weights such as GCN [16]. In GCN, the only learnable parameter is the global shared weight matrix that transforms features while the aggregation weights for neighbors are fixed². This constraint proves too coarse for our framework because we are dealing with fine-grained pairwise comparisons of all individuals in the dataset and different pairs of individuals may contribute differently to optimization objectives based on their features, group membership, connectivity structures and the similarity values. Thus, we aim to obtain not only learnable feature transformation weights but also learnable aggregation weights for each node's neighbors. Our approach is to adopt the attention mechanism to yield fine-grained and individually personalized attention weights for node aggregation such that

²In GCN, a given node's representation is updated with the following operation $\mathbf{h}'_i = \mathbf{W} \sum_{j \in \mathcal{N}(v_i)} \frac{e_{j,i}}{\sqrt{d_j d_i}} \mathbf{h}_j$ where \mathbf{W} is the feature transform weights, $\mathcal{N}(v_i)$ is the neighbor set of node v_i , $e_{j,i}$ is fixed as either edge weight or 1, d_i is the degree of v_i and \mathbf{h}_j is the neighbor node embedding.

these attention weights reflect how important the neighbors are with respect to optimization objectives for each node. The personalized attention weights leverage the fine-grained pairwise similarity information given by the similarity matrix. Specifically, we treat the similarity values as base values for node aggregation where they are scaled by a learnable scalar. The products of such scalar and similarity value for all neighbors are then softmaxed to become the personalized aggregation weights. Since we are learning pairwise attention weights for all nonzero pairs on the similarity matrix, it is important to consider the time complexity. To make our framework scalable, we follow GAT[24] and compute similarity informed attention weights with two learnable weights \mathbf{W} and \mathbf{a} as follows:

$$\lambda_{i,j} = \frac{\exp(\phi(\mathbf{a}^T [\mathbf{W}\mathbf{h}_i || \mathbf{W}\mathbf{h}_j])\mathbf{S}[i,j])}{\sum_{j \in \mathcal{N}_i} \exp(\phi(\mathbf{a}^T [\mathbf{W}\mathbf{h}_i || \mathbf{W}\mathbf{h}_j])\mathbf{S}[i,j])}, \quad (8)$$

where $\lambda_{i,j}$ is the attention from node i to node j , $\mathbf{W} \in \mathbb{R}^{c \times h}$ is the weight matrix, $\mathbf{a} \in \mathbb{R}^{2c}$ is the attention weight vector, $\mathbf{h}_i \in \mathbb{R}^h$ is the input embedding of node i , $[\cdot || \cdot]$ is concatenation of two vectors, \mathcal{N}_i is the neighborhood set of node i , and ϕ is an activation function. And the operation for node aggregation is

$$\mathbf{z}_i = \sigma(\sum_{j \in \mathcal{N}_i} \lambda_{i,j} \mathbf{W}\mathbf{h}_j), \quad (9)$$

where $\mathbf{z}_i \in \mathbb{R}^c$ is the output embedding of node i and σ is an activation function.

To summarize the GUIDE framework, an arbitrary backbone GNN utilizes the adjacency matrix and feature matrix to initialize node embeddings as the input node embeddings for next module. Then, a GNN layer utilizes the similarity matrix as the input matrix and initialized node embeddings to learn individually personalized attention weights for each node against its neighbor in the similarity matrix to aggregate node embeddings such that the this module optimizes individual fairness without group disparity and achieves good performance for the utility objective.

3.4 Objective Function

We have introduced the \mathcal{L}_{ifair} for individual fairness optimization (Eq. 1) and \mathcal{L}_{ifg} for group equality of individual fairness optimization (Eq. 7). Last but not least, we define a utility loss to optimize the utility performance of our framework.

$$\mathcal{L}_{util} = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^K \mathbf{Y}_{ij} \log \hat{\mathbf{Y}}_{ij}, \quad (10)$$

where \mathbf{Y}_{ij} indicates the true label (in K class) of i^{th} node and $\hat{\mathbf{Y}}_{ij}$ is the prediction of i^{th} node.

In summary, the training process of GUIDE has three learning objectives: utility objective from Eq. (10), overall individual fairness from Eq. (1) and group equality of individual fairness from Eq. (7). Thus, the total loss function is a weighted sum of them, weighted by hyperparameters α and β :

$$\mathcal{L}_{total} = \mathcal{L}_{util} + \alpha \mathcal{L}_{ifair} + \beta \mathcal{L}_{ifg}. \quad (11)$$

Table 2: Dataset Statistics.

Dataset	Credit	Income	Pokec-n
# of nodes	30,000	14,821	66,569
# of node attributes	13	14	266
# of edges in A	304,754	100,483	1,100,663
# of edges in S	1,687,444	1,997,641	32,837,463
Group ratio	11.2	3.16	21.0
Group avg degree ratio	12.6	2.8	58.8
Sensitive Attribute	age	race	age

4 EXPERIMENTS

In this section, we conduct extensive experiments on real-world datasets to validate the effectiveness of GUIDE. Specifically, we aim to answer the following research question:

- **RQ1:** How well does GUIDE balance utility, individual fairness and group equality of individual fairness objectives compared to baselines?
- **RQ2:** Are attention weights learned from GUIDE differentiated with respect to different pairs of nodes' influences on optimization objectives such as GDIF?

4.1 Datasets

We utilize real-world datasets that are used in critical decision systems and social networks: Credit, Income, and Pokec-n.

Credit: Credit graph dataset is constructed on 30,000 individuals. They are connected based on features such as spending and payment habits [31]. The task is to predict if an individual will default on credit card payment and the sensitive attribute is age.

Income: Income graph dataset is constructed on 14,821 individuals who are sampled from the *Adult Data Set* [6]. We keep only two races *White* and *Black* in the sampling process and individuals are connected based on their features. The supervised task is to predict if a person makes over \$50K a year.

Pokec-n: Pokec-n is a sampled dataset from Slovakia's most popular social network [22]. The dataset contains 66,569 individuals and they are connected by friend relationships. We treat age (≥ 40) as the binary sensitive attribute and the supervised task is to predict the working field of the users.

Note here that although these datasets have binary sensitive attributes, it is straightforward to extend training to multi-group cases because GDIF is defined for multi-group datasets.

4.2 Experimental Settings

To validate the effectiveness of the proposed framework, we conduct experiments against the following baseline models:

- **FairGNN** [4]: uses adversarial learning such that GNNs make fair node classifications that satisfy group fairness metrics. We directly apply it to various GNN backbones and analyze if it optimizes GDIF in the dataset.
- **NIFTY** [1]: optimizes counterfactual fairness and stability by perturbing nonsensitive and sensitive attributes and performs contrastive learning in training. We directly adopt it and explore if GDIF is optimized through this baseline.
- **PFR** [18]: aims to learn fair node embeddings that comply with individual fairness in downstream tasks by utilizing two relationship matrices: W_X for feature similarities derived

Table 3: Experiment results on Credit, Income and Pokec-n datasets. Model indicates the debiasing algorithm and Vanilla represents no debiasing is performed. \uparrow denotes the larger, the better; \downarrow means the opposite. Best performances are in bold. Individual (un)fairness numbers are reported in thousands. All entries are averages and standard deviations.

Model	Credit								
	GCN			GIN			Jumping Knowledge		
	AUC(\uparrow)	IF(\downarrow)	GDIF(\downarrow)	AUC(\uparrow)	IF(\downarrow)	GDIF(\downarrow)	AUC(\uparrow)	IF(\downarrow)	GDIF(\downarrow)
Vanilla	0.68 \pm 0.04	39.02 \pm 3.78	1.32 \pm 0.07	0.71 \pm 0.00	120.02 \pm 15.42	1.75 \pm 0.21	0.64 \pm 0.11	31.06 \pm 13.90	1.32 \pm 0.06
FairGNN	0.68 \pm 0.01	23.33 \pm 12.59	1.33 \pm 0.10	0.68 \pm 0.02	77.32 \pm 48.47	2.18 \pm 0.19	0.66 \pm 0.02	2.99 \pm 1.68	1.53 \pm 0.40
NIFTY	0.69\pm0.00	30.80 \pm 1.39	1.24 \pm 0.02	0.70 \pm 0.01	56.43 \pm 37.85	1.63 \pm 0.27	0.69\pm0.00	26.44 \pm 2.39	1.24 \pm 0.03
PFR	0.64 \pm 0.13	36.58 \pm 6.91	1.41 \pm 0.08	0.71 \pm 0.01	162.58 \pm 103.87	2.40 \pm 1.23	0.67 \pm 0.05	36.30 \pm 18.22	1.35 \pm 0.03
InFoRM	0.61 \pm 0.11	30.51 \pm 17.47	1.30 \pm 0.12	0.72\pm0.00	29.32 \pm 1.48	1.46 \pm 0.05	0.63 \pm 0.16	21.39 \pm 1.56	1.46 \pm 0.04
GUIDE	0.66 \pm 0.02	0.28\pm0.11	1.00\pm0.00	0.67 \pm 0.01	0.39\pm0.04	1.00\pm0.00	0.66 \pm 0.02	0.27\pm0.15	1.00\pm0.00
Income									
	GCN			GIN			Jumping Knowledge		
Vanilla	0.77 \pm 0.00	369.11 \pm 0.03	1.29 \pm 0.00	0.81\pm0.01	2815.59 \pm 1047.33	1.87 \pm 0.48	0.80\pm0.00	488.73 \pm 166.83	1.18 \pm 0.16
FairGNN	0.76 \pm 0.00	249.73 \pm 87.53	1.17 \pm 0.04	0.79 \pm 0.00	1367.93 \pm 875.64	3.30 \pm 1.18	0.77 \pm 0.00	219.30 \pm 42.92	1.30 \pm 0.12
NIFTY	0.73 \pm 0.00	42.14 \pm 5.83	1.38 \pm 0.04	0.79 \pm 0.01	608.98 \pm 314.83	1.17 \pm 0.26	0.73 \pm 0.02	48.25 \pm 10.48	1.39 \pm 0.09
PFR	0.75 \pm 0.00	245.97 \pm 0.58	1.32 \pm 0.00	0.79 \pm 0.00	2202.64 \pm 445.24	2.36 \pm 1.17	0.73 \pm 0.13	327.57 \pm 155.49	1.12 \pm 0.23
InFoRM	0.78\pm0.00	195.61 \pm 0.01	1.36 \pm 0.00	0.80 \pm 0.01	308.45 \pm 13.92	1.62 \pm 0.30	0.79 \pm 0.00	192.58 \pm 12.87	1.35 \pm 0.11
GUIDE	0.73 \pm 0.01	33.19\pm10.17	1.00\pm0.00	0.74 \pm 0.02	83.88\pm20.29	1.00\pm0.00	0.74 \pm 0.01	42.49\pm21.93	1.00\pm0.00
Pokec-n									
	GCN			GIN			Jumping Knowledge		
Vanilla	0.77\pm0.00	951.72 \pm 37.28	6.90 \pm 0.12	0.76\pm0.01	4496.47 \pm 1535.62	8.35 \pm 1.24	0.79\pm0.00	1631.27 \pm 93.94	8.47 \pm 0.45
FairGNN	0.69 \pm 0.03	363.73 \pm 78.38	6.21 \pm 1.28	0.68 \pm 0.03	391.07 \pm 562.93	5.25 \pm 3.52	0.70 \pm 0.00	807.97 \pm 281.26	11.68 \pm 2.89
NIFTY	0.74 \pm 0.00	85.25 \pm 10.55	5.06 \pm 0.29	0.76\pm0.01	2777.36 \pm 346.29	9.28 \pm 0.28	0.73 \pm 0.01	477.31 \pm 165.68	8.20 \pm 1.33
PFR	0.53 \pm 0.00	98.25 \pm 9.44	15.84 \pm 0.03	0.60 \pm 0.01	628.27 \pm 85.89	6.20 \pm 0.79	0.68 \pm 0.00	729.77 \pm 74.62	15.66 \pm 5.47
InFoRM	0.77\pm0.00	733.10 \pm 23.54	6.86 \pm 0.18	0.76\pm0.01	1499.55 \pm 149.10	7.59 \pm 0.35	0.79\pm0.00	1099.12 \pm 66.12	7.32 \pm 0.41
GUIDE	0.73 \pm 0.02	55.05\pm30.87	1.11\pm0.03	0.74 \pm 0.01	120.65\pm17.33	1.12\pm0.03	0.75 \pm 0.02	83.09\pm18.70	1.13\pm0.02

from k-nearest-neighbor and W_F from human judgement for pairwise similarities. In order to compare it with other baselines, we use A and S for them respectively. We use the output embeddings from PFR as inputs for GNN backbones.

- **InFoRM** [14]: formulates individual fairness loss in a graph based on Lipschitz condition and performs optimization in preprocessing, processing and postprocessing stages of the graph mining model. We add the proposed individual fairness loss to GNN backbone training.

We use cosine similarity to instantiate the computation of similarity matrix S. In this way, the entries of S will be in $[0, 1]$. The performance of GUIDE on model utility is evaluated with AUCROC (AUC). We also evaluate individual (un)fairness (IF) and group disparity of individual fairness (GDIF), respectively. All baselines and GUIDE can use arbitrary GNN backbones. In order to present extensive comparison, we conduct experiments on 3 different GNN backbones: GCN [16], GIN [29], and JumpingKnowledge [30]. All hidden layer dimension is set as 16, number of training epochs is 3000, Adam optimizer [15] is used with learning rate at 1e-3, and weight decay is set as 1e-5. All experiments are run 5 times individually. The experiment results are shown in Table 3.

In terms of hyperparameter details for reproducibility, for PFR, γ is 0.5 for Credit, 0.5 for Income and 0.25 for Pokec-n. For InFoRM, α is 5e-7 for Credit, 1e-7 for Income, and 1e-8 for Pokec-n. For FairGNN, α is 4 and β is 1000 for Credit, 4 and 10 for Income and 4 and 100 for Pokec-n. For NIFTY, λ is 0.5 across all datasets. For

GUIDE, α is 2.5e-5 and β is 10 for Credit, 1e-7 and 0.25 for Income, and 2.5e-7 and 0.05 for Pokec-n.

4.3 Effectiveness of GUIDE

Here we aim to answer **RQ1** and experiment results are presented in Table 3. Our goal is to minimize individual unfairness and group disparity of individual fairness without sacrificing utility performance. We answer **RQ1** by drawing the following observations.

- In any vanilla version of GNNs, we observe high individual unfairness and high group disparity of individual fairness. We also see the GDIF metric is the worst in Pokec-n which has high sensitive group imbalance.
- Baseline models do reduce individual unfairness but none of them minimized group disparity of individual fairness and often times leaving the disparity unchanged or exacerbated.
- The highly expressive GNN backbone, GIN, achieves the best utility performance but also has the highest individual unfairness and group disparity of individual fairness. Fortunately, we notice that GUIDE has the most debiasing power for GIN, reducing individual unfairness by over 90% across all datasets and lower GDIF to 1 or almost 1.
- GUIDE obtains the lowest GDIF and IF in all experiments while achieving comparable node classification performances as vanilla and baseline models, thus accomplishing three optimization goals proposed in Problem 1.

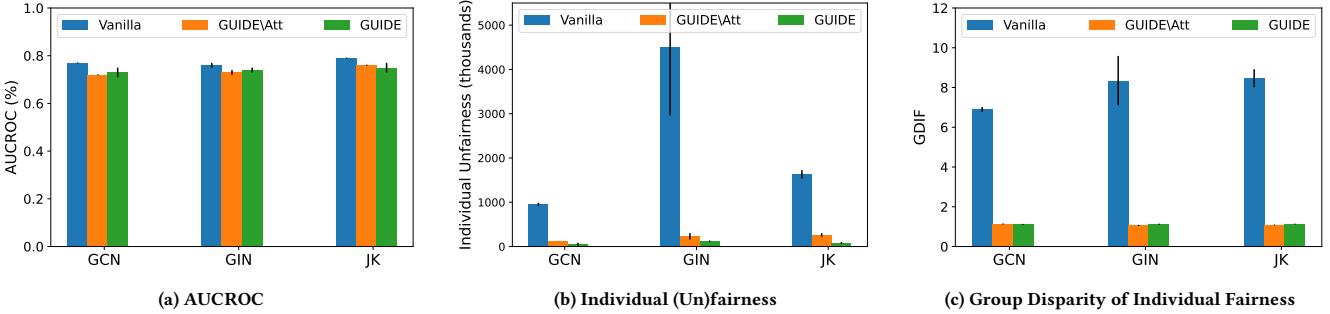


Figure 4: Ablation results of vanilla GNN, GUIDE and its variant using all three GNN backbones. (a) Node classification performance comparison across vanilla GNN, GUIDE and its variant; (b) Individual fairness promotion comparison across vanilla GNN, GUIDE and its variant; (c) GDIF optimization comparison across vanilla GNN, GUIDE and its variant.

4.4 Ablation Study

We then present ablation studies for our framework. Due to space limit, we only present results on Pokec-n with GCN, GIN, and JumpingKnowledge GNN backbones in Figure 4. The reason for this dataset is because it is the largest of the three datasets in terms of node and edge counts so it shows that our framework has scalable performances in balancing utility and fairness objectives. We compare GUIDE to GNN backbones that use fixed aggregation weights instead of personalized aggregation weights to optimize utility and fairness objectives. We name it GUIDE\Att. Both GUIDE\Att and GUIDE use the same loss function in Eq. (3.4) where fairness objectives in individual unfairness and group disparity of individual fairness are weighted by hyperparameters α and β and added to utility loss. In subsection 3.3, we discussed that fixed aggregation weights are typically adopted in GNN frameworks such as GCN, GIN and JumpingKnowledge. However, in the case of reducing GDIF, we prefer personalized aggregation weights to 1) utilize rich pairwise similarity information and 2) derive fine-grained and customized attention weights for each node so model capacity is increased and yield better tradeoff of optimization objectives. We explore the utility performance (measured by AUCROC), and fairness performances in individual unfairness and group disparity of individual fairness and make the following observations

- From (4a), it shows that node classification AUCROC of GUIDE\Att and GUIDE are comparable to vanilla with GUIDE scoring slightly higher than GUIDE\Att.
- From (4b) and (4c), both GUIDE\Att and GUIDE minimized group disparity of individual fairness to around 1. They also both reduced individual unfairness substantially. However, GUIDE is more superior in its ability to reduce individual unfairness while maintaining the same GDIF.

4.5 Attention Weights Analysis

For RQ2, we explore whether learned attention weights from GUIDE are differentiated with respect to different pairs of nodes' influences on GDIF. Our approach can be summarized as conducting correlation test between node pairs' GDIF influence and their corresponding attention weights changes $\Delta\lambda$ from optimizing the GDIF objective. To simplify analysis, we only compare magnitude of these values by using absolute value of differences because the

Table 4: Two-tailed Correlation Test.

Dataset	Credit	Income
Correlation of $C_{i,j}$ and $\Delta\lambda_{i,j}$	0.036	0.044
Two-tailed p-value	0.00	0.00
# of samples	1,687,444	1,997,641

goal is to explore if learned attention weights can differentiate node pairs that cause big changes to GDIF. We first approximate the influence on GDIF from different pairs of nodes by leave-one-out calculations. We obtain model outputs from GUIDE trained with only utility and individual fairness objectives ($\beta = 0$) so a vanilla benchmark GDIF can be calculated for this set of outputs denoted as $GDIF_{benchmark}$. Next, for a specific node pair (i, j) , we leave them out of the GDIF calculation, i.e. $\|z_i - z_j\|_2^2 S[i, j]$ and $\|z_j - z_i\|_2^2 S[j, i]$ are removed from the calculation of individual unfairness of corresponding groups ($U_1 \dots U_G$). We denote this GDIF as $GDIF_{-(i,j)}$. Finally, we define the influence of node pair (i, j) as:

$$C_{i,j} = C_{j,i} = |GDIF_{-(i,j)} - GDIF_{benchmark}|. \quad (12)$$

We then calculate the absolute value of change in attention weights between GUIDE trained with and without GDIF objective ($\beta = 0$ and $\beta \neq 0$) to represent how attention weights changed when GDIF objective is optimized.

$$\Delta\lambda_{i,j} = |\lambda_{i,j}^{\beta \neq 0} - \lambda_{i,j}^{\beta=0}|. \quad (13)$$

The two-tailed correlation test results for Credit and Income datasets are listed in Table 4. We see positive correlations between the magnitude of influence on GDIF and magnitude of change in attention weight for node pairs. The significant p-values are indicative that GUIDE can yield personalized attention weights for node pairs with respect to their influences on GDIF.

5 RELATED WORK

Algorithmic fairness. Researchers have formulated a variety of fairness notions and they can be broadly categorized as group fairness, counterfactual fairness and individual fairness. *Group fairness* is defined as enforcing equal outcome statistics such as true positives across different groups. Zafar et al. [32] propose *demographic parity* which requires equal likelihood of positive outcome regardless of group membership. Hardt et al. [11] present *equal*

opportunity which argues people from different groups should have equal true positive rates. *Counterfactual fairness* promotes fixed model outcomes for individuals regardless of what their sensitive attributes are in reality or counterfactual scenarios [1]. For *Individual Fairness*, Dwork et al. [7] propose individual fairness which requires *treating similar individuals similarly*. They formulate it as an optimization problem involving pairwise individual similarity and Lipschitz condition. Lahoti et al. [18] treat individual fairness as a low-rank representation learning problem by minimizing output distances multiplied by individual similarity. García-Soriano et al. [9] minimize the amount of individual unfairness from enforcing group fairness by optimizing a max-min ranking problem. Majority of these models rely on Lipschitz condition and we have found this formulation could result in different levels of individual fairness for different groups which leads to discrimination against certain demographic subgroups. To our best knowledge, we are the first to investigate this issue in individual fairness and provide a solution.

Fairness in graph mining. As graph mining models are increasingly adopted for many learning tasks, numerous solutions have been proposed to correct potential unfairness in graph mining algorithms. For group fairness, Rahman et al. [21] propose the notion of equality of representation which extends statistical parity to the node2vec model. Bose et al. [3] propose a compositional adversarial method to remove the influence of sensitive attributes in learned embeddings. Dai et al. [4] develop a similar adversarial framework but debiasing is performed in end-to-end GNN predictions. For individual fairness, Kang et al. [14] optimize individual fairness in by deriving a individual fairness loss on graph datasets and reduce it before, during and after training of the mining model. Dong et al. [5] treat optimization of individual fairness in GNNs as a ranking problem which bypasses the limitation of Lipschitz condition. Our approach differs from these cited works in that we not only optimize individual fairness but also explicitly equalize the levels of fairness across groups such that individuals in different groups are treated equally fair in pairwise comparisons.

6 CONCLUSION

Graph Neural Networks have shown superior performances in a variety of tasks and are increasingly adopted in high-stakes decision systems. However, there has been heightened concerns that GNNs could generate unfair decisions for underprivileged groups or individuals without fairness constraints. Out of various proposed algorithmic fairness notions, individual fairness has fine granularity on the individual level and promotes *treating similar individuals similarly*. However, in our analysis of several works on individual fairness, we have found that their formulation from Lipschitz condition could lead to different levels of individual fairness for different groups, thus creating discrimination on the group level. We tackle this problem by developing a novel and flexible GNN framework GUIDE which incorporates an arbitrary GNN for utility maximization and a fairness network that learns individually custom attention weights for debiasing. We conduct extensive experiments on real-world datasets to demonstrate the effectiveness of our proposed framework and the results show GUIDE substantially remove group disparity of individual fairness, achieve overall individual fairness and maintain prediction performance.

REFERENCES

- [1] Chirag Agarwal, Himabindu Lakkaraju, and Marinka Zitnik. 2021. Towards a Unified Framework for Fair and Stable Graph Representation Learning. *CoRR* abs/2102.13186 (2021), arXiv:2102.13186 <https://arxiv.org/abs/2102.13186>
- [2] Punam Bedi and Chhavi Sharma. 2016. Community detection in social networks. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 6, 3 (2016), 115–135.
- [3] Avishhek Joey Bose and William L. Hamilton. 2019. Compositional Fairness Constraints for Graph Embeddings. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9–15 June 2019, Long Beach, California, USA (Proceedings of Machine Learning Research, Vol. 97)*. PMLR, 715–724.
- [4] Enyan Dai and Suhang Wang. 2021. 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 (Virtual Event, Israel) (WSDM '21)*. Association for Computing Machinery, New York, NY, USA, 680–688. <https://doi.org/10.1145/3437963.3441752>
- [5] Yushun Dong, Jian Kang, Hanghang Tong, and Jundong Li. 2021. Individual Fairness for Graph Neural Networks: A Ranking based Approach. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 300–310.
- [6] Dheeru Dua and Casey Graff. 2017. UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml>
- [7] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard S. Zemel. 2012. Fairness through awareness. In *Innovations in Theoretical Computer Science 2012, Cambridge, MA, USA, January 8–10, 2012*. ACM, 214–226.
- [8] Wengi Fan, Yao Ma, Qing Li, Yuan He, Yihong Eric Zhao, Jiliang Tang, and Dawei Yin. 2019. Graph Neural Networks for Social Recommendation. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13–17, 2019*. ACM, 417–426.
- [9] David García-Soriano and Francesco Bonchi. 2021. Maxmin-Fair Ranking: Individual Fairness under Group-Fairness Constraints. *CoRR* abs/2106.08652 (2021). <https://arxiv.org/abs/2106.08652>
- [10] William L. Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive Representation Learning on Large Graphs. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4–9, 2017, Long Beach, CA, USA*. 1024–1034.
- [11] Moritz Hardt, Eric Price, and Nati Srebro. 2016. Equality of Opportunity in Supervised Learning. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5–10, 2016, Barcelona, Spain*. 3315–3323.
- [12] Mohammad Al Hasan and Mohammed J. Zaki. 2011. A Survey of Link Prediction in Social Networks. In *Social Network Data Analytics*, Charu C. Aggarwal (Ed.). Springer, 243–275.
- [13] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. 2020. A Survey on Knowledge Graphs: Representation, Acquisition and Applications. *CoRR* abs/2002.00388 (2020). <https://arxiv.org/abs/2002.00388>
- [14] Jian Kang, Jingrui He, Ross Maciejewski, and Hanghang Tong. 2020. InFoRM: Individual Fairness on Graph Mining. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23–27, 2020*. ACM, 379–389.
- [15] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). <http://arxiv.org/abs/1412.6980>
- [16] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings*. OpenReview.net.
- [17] Matthias Kleindessner, Samira Samadi, Pranjal Awasthi, and Jamie Morgenstern. 2019. Guarantees for Spectral Clustering with Fairness Constraints. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9–15 June 2019, Long Beach, California, USA (Proceedings of Machine Learning Research, Vol. 97)*. PMLR, 3458–3467.
- [18] Preethi Lahoti, Krishna P. Gummadi, and Gerhard Weikum. 2019. Operationalizing Individual Fairness with Pairwise Fair Representations. *Proc. VLDB Endow.* 13, 4 (2019), 506–518.
- [19] Nihareeb Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.* 54, 6, Article 115 (July 2021), 35 pages. <https://doi.org/10.1145/3457607>
- [20] Georgios A. Pavlopoulos, Maria Secrier, Charalampos N. Moschopoulos, Theodoros G. Soldatos, Sophia Kossida, Jan Aerts, Reinhard Schneider, and Pantelis G. Bagos. 2011. Using graph theory to analyze biological networks. *BioData Min.* 4 (2011), 10.
- [21] Tahleen A. Rahman, Bartłomiej Surma, Michael Backes, and Yang Zhang. 2019. Fairwalk: Towards Fair Graph Embedding. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10–16, 2019*. ijcai.org, 3289–3295.

- [22] L. Takac and Michal Zábovský. 2012. Data analysis in public social networks. *International Scientific Conference and International Workshop Present Day Trends of Innovations* (01 2012), 1–6.
- [23] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.
- [24] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. 2018. Graph Attention Networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net. <https://openreview.net/forum?id=rXMpikCZ>
- [25] Daixin Wang, Yuan Qi, Jianbin Lin, Peng Cui, Quanhui Jia, Zhen Wang, Yanming Fang, Quan Yu, Jun Zhou, and Shuang Yang. 2019. A Semi-Supervised Graph Attentive Network for Financial Fraud Detection. In *2019 IEEE International Conference on Data Mining, ICDM 2019, Beijing, China, November 8-11, 2019*. IEEE, 598–607.
- [26] Yiwei Wang, Wei Wang, Yuxuan Liang, Yujun Cai, and Bryan Hooi. 2021. Cur-Graph: Curriculum Learning for Graph Classification. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*. ACM / IW3C2, 1238–1248. <https://doi.org/10.1145/3442381.3450025>
- [27] Qitian Wu, Hengrui Zhang, Xiaofeng Gao, Peng He, Paul Weng, Han Gao, and Guihai Chen. 2019. Dual Graph Attention Networks for Deep Latent Representation of Multifaceted Social Effects in Recommender Systems. In *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*. ACM, 2091–2102. <https://doi.org/10.1145/3308558.3313442>
- [28] Shiwen Wu, Wentao Zhang, Fei Sun, and Bin Cui. 2020. Graph Neural Networks in Recommender Systems: A Survey. *CoRR* abs/2011.02260 (2020). <https://arxiv.org/abs/2011.02260>
- [29] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2019. How Powerful are Graph Neural Networks?. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.
- [30] Keyulu Xu, Chengtao Li, Yonglong Tian, Tomohiro Sonobe, Ken-ichi Kawarabayashi, and Stefanie Jegelka. 2018. Representation Learning on Graphs with Jumping Knowledge Networks. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholm, Sweden, July 10-15, 2018 (Proceedings of Machine Learning Research, Vol. 80)*. PMLR, 5449–5458.
- [31] I-Cheng Yeh and Che-hui Lien. 2009. The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients. *Expert Syst. Appl.* 36, 2 (2009), 2473–2480.
- [32] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez-Rodriguez, and Krishna P. Gummadi. 2017. Fairness Constraints: Mechanisms for Fair Classification. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20-22 April 2017, Fort Lauderdale, FL, USA (Proceedings of Machine Learning Research, Vol. 54)*. PMLR, 962–970.

7 SUPPLEMENTAL MATERIALS