IFGNN: An Individual Fairness Awareness Model for Missing Sensitive Information Graphs

Kejia Xu[1], Zeming Fei[2], Jianke Yu[2], Yu Kong[1], Xiaoyang Wang[1(✉)], and Wenjie Zhang[1]

[1] University of New South Wales, Sydney NSW 2052, Australia
{kejia.xu,xiaoyang.wang1,wenjie.zhang}@unsw.edu.au
{yu.kong}@student.unsw.edu.au
[2] University of Technology Sydney, Ultimo NSW 2007, Australia
{zeming.fei,jianke.yu}@student.uts.edu.au

Abstract. Graph neural networks (GNNs) provide an approach for analyzing complicated graph data for node, edge, and graph-level prediction tasks. However, due to societal discrimination in real-world applications, the labels in datasets may have certain biases. This bias is magnified as GNNs iteratively obtain information from neighbourhoods through message passing and aggregation, generating unfair embeddings that implicitly affect the prediction results. In real-world datasets, missing sensitive attributes is common due to incomplete data collection and privacy concerns. However, research on the fairness of GNNs in incomplete graph data is limited and mainly focuses on group fairness. Addressing individual unfairness in GNNs when the sensitive attributes are missing remains unexplored. To solve this novel problem, we introduce a model named IFGNN, which leverages a GNN-based encoder and a decoder to generate node embeddings. Additionally, IFGNN adopts the Lipschitz condition to ensure individual fairness. Through comprehensive experiments on four real-world datasets compared with baseline models in node classification tasks, the results demonstrate that IFGNN can achieve individual fairness while maintaining high prediction accuracy.

Keywords: Individual fairness · Sensitive attribute · GNN

1 Introduction

Graph-structured data has been employed to represent real-world complex systems, such as social networks [32, 33, 36], financial networks [5, 6], knowledge graphs [7], etc. GNNs have appeared as a solution for various downstream tasks, including node classification, link prediction, community detection, and graph search. GNNs provide an analytical approach to comprehend graph data [2, 35]. For example, GNN-based recommender systems can personalize recommendations based on user preferences. The message-passing mechanism within GNNs plays an important role in enabling its powerful learning ability, capturing and aggregating the graph structure and node attribute information to generate node embeddings [13].

However, the message-passing mechanism is prone to be influenced by sensitive attributes, resulting in biased representations for downstream tasks [1]. During the message-passing process, nodes incorporate neighbourhood information through diverse mechanisms, such as graph convolutional and attention mechanism [27, 30]. Consequently, sensitive attributes (e.g., race, gender and age) from neighbour nodes are inherently included and amplified during the process, leading to bias within GNNs [3, 28, 37]. This bias becomes particularly concerning in high-risk decision-making and classification applications, as it can result in unfair implicit preferences towards privileged individuals, potentially leading to social discrimination [11, 21, 24]. For instance, if a particular age group has more positive labels in the dataset than other age groups, GNNs are inclined to produce positive predictions for individuals within that age group while ignoring the influence of other attributes.

To address the discrimination problem caused by sensitive attributes, researchers introduced the concepts of group fairness and individual fairness [24]. Group fairness ensures equal treatment across different demographic subgroups. For example, when considering different groups determined by race, the model should produce the same predictions for nodes within the same group [17]. On the other hand, individual fairness aims to mitigate the model's bias towards each individual and yield the same treatment to similar individuals [10, 12]. Given that individual fairness imposes finer-grained constraints on the model, it is essential to consider how to minimize individual unfairness [34]. Particularly, when considering the missing sensitive attributes in graph data, individual fairness may be the only fairness criterion [37].

In many real-world applications, graphs usually suffer from missing information due to incomplete data collection and privacy concerns [4]. For instance, users may choose the 'prefer not to tell' option in social network applications, leading to missing sensitive attributes such as income and age. Existing research primarily falls into assuming that the graph data is complete without missing sensitive attributes or employing GNNs to predict the missing information [4, 14, 22]. FairGNN [9] is a notable model that utilizes an independent Graph Convolutional Networks (GCN) estimator to predict the limited missing sensitive attributes and subsequently evaluates the fairness of the node classification task. However, this approach of completing information may introduce noises and errors that significantly affect the model's accuracy and fairness [16, 22]. While FairAC [14] mainly focuses on mitigating feature and topology-level unfairness by excluding sensitive features when the entire node attributes are missing. Although these two studies discuss the mitigation of GNNs fairness in an incomplete graph, they do not aim to solve the individual unfairness issue. Thus, they cannot be treated as baselines. To our best knowledge, no research has focused on individual fairness without constraining the number of missing sensitive attributes.

In this paper, our focus lies in ensuring the model's individual fairness while preserving the prediction accuracy without predicting the missing sensitive attributes, especially under various rates of missing sensitive attributes. To address

this problem, we propose a novel GNN-based model named Individual Fairness Graph Neural Network (IFGNN). The IFGNN model incorporates the Lipschitz condition, a well-established and widely used mathematical approach for achieving individual fairness [18]. Unlike existing works, we introduce a GNN-based encoder and a decoder that obtain the latent neighbourhood information to generate node embeddings, thereby solving the issue of missing sensitive information.

Contributions. The main contributions of this paper are shown as follows.

- We present a novel problem of addressing individual unfairness and ensuring prediction performance for graph data with unlimited missing sensitive attributes.
- We propose a GNN-based encoder and a decoder with Lipschitz constraint model IFGNN, to solve the problem without predicting the missing information.
- Our proposed model outperforms baseline models regarding individual fairness and prediction accuracy on four real-world datasets.

Roadmap. The rest of the paper is organized as follows. In Section 2, we give a comprehensive overview of existing related work. In Section 3, we introduce the preliminaries and the formal problem definition. In Section 4, we present the details of our proposed model IFGNN. In Section 5, we conduct comparative experiments and the ablation study and conclude the paper in Section 6.

2 Related Work

In this section, we introduce the related works of fairness within graph data. Specifically, we introduce approaches dealing with incomplete graph data. Approaches to promote fairness can be categorized into three strategies: pre-processing, in-processing and post-processing. In the pre-processing phase, biased attributes within the original dataset are explicitly modified or fair node embeddings are directly introduced [8]. For instance, Rahman et al.'s Fairwalk framework [26] employs node2vec to generate fair node embeddings through a fairness-aware embedding method. In the in-processing phase, fairness is maintained by integrating fairness constraint objective functions into the model [39]. Zhang et al. proposed an adversarial learning framework [41], which balances the model's fairness by optimizing the prediction function while weakening the adversary function. The post-processing phase is processed after the model's execution, alleviating the discrimination found in the output embedding [17]. In the work of Masrour et al. [23], a dyadic-level fairness criterion was introduced, generating supplementary links to mitigate the problem of Internet users receiving excessively sensitive-attributes-dominated information.

These three strategies encompass both conventional and GNN-based approaches. Among conventional approaches, Zemel et al.'s framework [40] achieves bias-reduced data encoding by obfuscating non-sensitive attributes of individuals. Another conventional technique is the ifair framework [20], which maps

personal attributes to low-rank representation while ensuring data integrity. In recent years, GNNs have experienced significant development [31, 42, 43]. A representative GNN-based technique is the GFairHint model [37], which generates fairness awareness graph data by determining if two nodes are similar and employs GNNs to generate fair node embeddings.

Inspired by the fairness research within non-graph data, introducing the Lipschitz condition as a fairness constraint is a common strategy in existing research [12]. A representative framework that employs Lipschitz conditions to solve fairness is InFoRM [18]. Furthermore, the GUIDE model [28], a GNN-based model that considers both individual and group fairness, achieves individual fairness in subgroups by minimizing the Lipschitz constant.

Fairness research within incomplete graphs remains limited. The most relevant to our work is the FairGNN mentioned above [9]. FairGNN uses a sensitive attribute estimator to predict the missing sensitive attributes and introduces it into adversarial learning to generate fair node embeddings [9]. However, FairGNN is only applicable when the number of missing sensitive attributes is limited, and it is designed to ensure group fairness instead of individual fairness [9]. It is worth considering that predicting sensitive attributes may not only affect the model performance but also violate privacy policies [16, 22]. In contrast, our model effectively addresses the individual unfairness problem in incomplete graphs without having to know users' hidden sensitive attributes, and it remains unaffected by the number of missing sensitive attributes. Therefore, our proposed model has broader applicability for solving real-world problems.

3 Preliminaries

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{X})$ denote an undirected attributed graph, where $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$ is the set of N nodes. For each node $v \in \mathcal{V}$, it is associated with a set of M attributes $\{x_v^1, x_v^2, \ldots, x_v^M\}$. Following the general setting, among the attributes, one attribute $x_s \in \{0, 1\}$ is considered as the sensitive attribute. We use $\mathcal{V}_S \subseteq \mathcal{V}$ to denote the set of nodes with the missing sensitive attribute. r_s is the missing proportion, which is the ratio of the number of nodes in $\mathcal{V}_S$ to that in $\mathcal{V}$. $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges and $\mathcal{X} \in \mathbb{R}^{N \times M}$ is the node attribute matrix where M is the dimension of attributes. The value of the matrix $S \in \mathbb{R}^{N \times N}$ represents the cosine similarity between input feature vectors of pairs of nodes. $\mathcal{Y} = \{y_1, y_2, \ldots, y_N\}$ denotes the node labels where $y_N \in \{0, 1\}$.

Problem statement. Given a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{X})$ with the sensitive attribute x_s, which includes missing values, in this paper, we aim to learn a fair model f for node label prediction. It can be formulated as follows:

$$f(\mathcal{G}, x_s) \rightarrow \hat{\mathcal{Y}}. \tag{1}$$

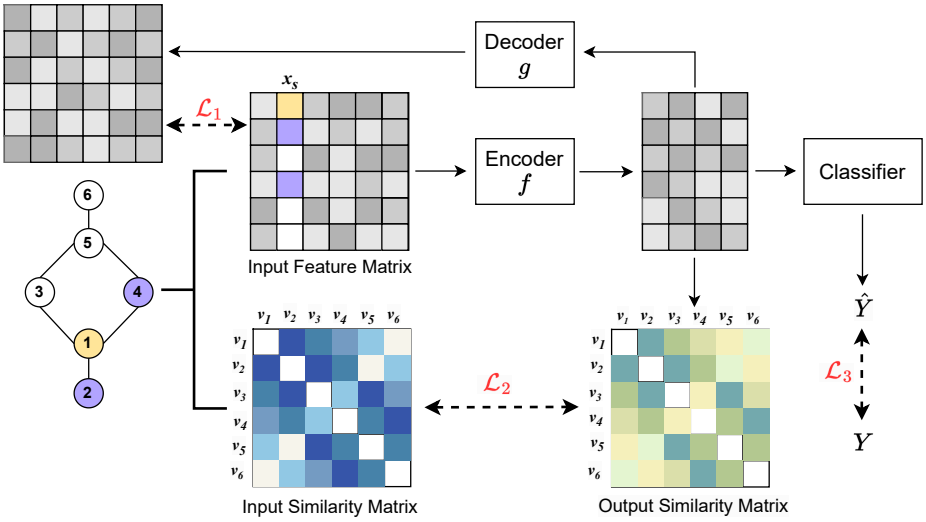

Fig. 1. The overview of the proposed IFGNN.

4 Proposed Model

We propose to solve the problem with a node representation learning model IFGNN. Figure 1 shows the general framework of our proposed model, which consists of a GNN-based encoder, a multi-layer perceptron (MLP) decoder, an individual fairness mechanism and a label classifier. The encoder generates node embeddings, while the decoder assists the encoder in learning more expressive node embeddings. The fairness mechanism addresses unfairness issues by minimizing the distance between the input pairwise similarity matrix and the output pairwise similarity matrix.

4.1 Encoder and Decoder

Figure 1 shows an example of an incomplete graph with missing sensitive attribute x_s. For demonstration, we intentionally set the rate of missing sensitive attributes r_s to 0.5. In the example graph, nodes v_3, v_5 and v_6 have missing sensitive attributes, while the sensitive attributes ($x_s \in \{0, 1\}$) of other nodes are represented by distinct colours. Intuitively, when there are missing node attributes, an estimator can be deployed to predict the missing information.

However, utilizing a GNN-based estimator for completing the node information might result in the amplification of error and noise in the subsequent operations (e.g., fairness-aware processing). To prevent this issue, it is imperative to leverage the characteristic of the message-passing mechanism in GNNs, which enables the information aggregation from neighbour nodes to generate node embeddings [38]. The missing sensitive information including the latent structural information will be implicitly compensated [16]. The GNN-based encoder f takes

the original informative node feature matrix denoted by $\mathcal{X} \in \mathbb{R}^{N \times M}$. For each node v, the embedding at the l-th layer will be:

$$h_v^l = Combine(h_v^{l-1}, Aggregate(h_u^l, \forall u \in \mathcal{N}(v))), \tag{2}$$

where h_v^l is the output embedding of node v at l-th layer and $\mathcal{N}(v)$ denotes the set of neighbour nodes of node v. The $Aggreagte(\cdot)$ operation iteratively aggregates and obtains the embeddings of neighbour nodes and the $Combine(.,.)$ operation combines the updated embedding of neighbour node u with the own embedding of node v from the previous layer.

The GNN-based encoder will generate the output embedding matrix $\mathrm{H} = f(\mathcal{X})$, where $\mathrm{H} \in \mathbb{R}^{N \times d}$. It is important to note that any model with GNN as a backbone, such as GCN and Graph Attention Networks (GAT), can be utilized as an encoder.

We employ the MLP with a nonlinear activation layer as the decoder g. The output of the decoder is defined as $\mathcal{X}' = g(\mathrm{H})$. $\mathcal{X}'$ is the reconstructed attribute matrix, where $\mathcal{X}' \in \mathbb{R}^{N \times M}$. The encoder and decoder are trained to minimize the following:

$$\mathcal{L}_1 = \frac{1}{N}(\mathcal{X}' - \mathcal{X})^2, \tag{3}$$

By minimizing the difference between $\mathcal{X}'$ and $\mathcal{X}$, the model is enabled to re-utilize the input feature matrix during the training process. The Eq. (3) not only helps the encoder obtain more representative node feature embeddings but also improves the overall training stability, thereby strengthening the robustness of the entire model.

4.2 Fairness Mechanism

To achieve individual fair node embeddings within the encoder, we adopted the Lipschitz condition. It is initially proposed by [12], indicating that any two similar individuals should receive the same outputs.

Definition 1. *Given any pair of nodes v_i and v_j, Lipschitz constraint ϵ_{ij} restricts the output distance between v_i and v_j by:*

$$D_{out}(f(v_i), f(v_j)) \leq \epsilon_{ij} \cdot D_{in}(v_i, v_j), \forall v_i, v_j \in \mathcal{V}, \tag{4}$$

where $D_{out}(.,.)$ denotes the output similarity distance of a pair of nodes, $D_{in}(.,.)$ denotes the input similarity distance.

The intention is that a model constrained by the Lipschitz condition will assign a smaller output similarity distance to any pair of nodes with a higher input similarity, aiming to treat similar node pairs equally, even in the existence of sensitive attributes. The Eq. (4) in Definition 1 can be written as follows:

$$D_{out}(f(v_i), f(v_j)) \cdot \frac{1}{D_{in}(v_i, v_j)} \leq \epsilon_{ij}, \forall v_i, v_j \in \mathcal{V}. \tag{5}$$

Given the input similarity distance of all node pairs, the sum of the total output similarity distance is minimized by minimizing the sum of ϵ_{ij} of all node pairs. Therefore, we propose to enhance the model's fairness awareness that minimizes:

$$\mathcal{L}_2 = \sum_{v_i \in \mathcal{V}} \sum_{v_j \in \mathcal{V}} \cos(\mathrm{H}[i,:], \mathrm{H}[j,:]) S[i:j], \tag{6}$$

where $\mathrm{H}[i,:]$ and $\mathrm{H}[j,:]$ is the encoder output node embeddings for node v_i and v_j, $S[i,j]$ is the corresponding similarity value in the input feature similarity matrix S, and $\cos(.,.)$ denotes the cosine similarity distance of the node pair. Note that any differentiable similarity metric can be used for similarity calculation. We employ cosine similarity for its efficiency in evaluation [25]. Training the encoder with the objective function $\mathcal{L}_2$, we can achieve the goal of tackling unfairness in the model, as it minimizes the sum of the Lipschitz constraint ϵ_{ij}.

4.3 Node Classifier

The node classifier consists of an MLP layer and a nonlinear activation layer. It takes the node attribute embedding matrix H as the input, which is generated by the encoder and mitigated unfairness by $\mathcal{L}_2$, to predict the node labels $\hat{\mathcal{Y}}$. The output of the MLP layer will be:

$$\hat{y}_v = \sigma(W h_v + b). \tag{7}$$

Here, we use binary cross-entropy with logits loss, which implicitly applies the nonlinear activation function, as the objective function for the classifier. It should be noted that the threshold τ should be set to obtain the label of each node. The objective function of the classifier is to minimize:

$$\mathcal{L}_3 = -[y_v \cdot log\hat{y}_v + (1 - y_v) \cdot log(1 - \hat{y}_v)], \tag{8}$$

where y_v is the label of node v and σ is the sigmoid function.

4.4 IFGNN Learning Objective

To train IFGNN, the final learning objective is to minimize:

$$\mathcal{L} = \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_2 + \lambda_3 \mathcal{L}_3, \tag{9}$$

where λ_1, λ_2 and λ_3 are hyperparameters and the sum of them is 1. Given that the $\mathcal{L}_3$ function is the primary loss function, while $\mathcal{L}_1$ and $\mathcal{L}_2$ are employed as auxiliary assurance for accuracy and individual fairness in the node classification task, we judiciously decrease the weights of $\mathcal{L}_1$ and $\mathcal{L}_2$ to mitigate the risk of overfitting without dampening the influence of $\mathcal{L}_3$.

By minimizing the overall objective function, IFGNN is optimized to generate node embeddings that capture the latent information with the missing sensitive attribute and deal with individual unfairness.

Table 1. Experiment results when the missing rate of the sensitive attribute is 0.3.

Dataset	Metrics	IFGNN	GCN	GAT	GraphSAGE
Credit	ACC(↑)	**0.7992**	0.7939	0.7772	0.7868
	AUC(↑)	0.7297	**0.7331**	0.6967	0.7077
	IUF (↓)	**0.0043**	0.0600	0.0739	0.0443
German	ACC(↑)	**0.7360**	0.6760	0.7200	0.7000
	AUC(↑)	0.6674	0.6257	0.6569	**0.6753**
	IUF (↓)	**0.0120**	0.0560	0.6398	0.0560
Income	ACC(↑)	**0.8325**	0.8122	0.7928	0.8214
	AUC(↑)	**0.8520**	0.7891	0.7778	0.8288
	IUF (↓)	**0.0084**	0.0267	0.0302	0.0340
Recidivism	ACC(↑)	**0.7722**	0.7449	0.7315	0.7353
	AUC(↑)	**0.8692**	0.8328	0.8470	0.8231
	IUF (↓)	**0.0182**	0.0481	0.0619	0.0612

5 Experiments

In this section, we conduct experiments on four real-world graph datasets to evaluate the performance of our proposed IFGNN. We aim to answer the following research questions:

- **RQ1** Compared to several baseline models, can the proposed IFGNN maintains the best individual fairness and prediction accuracy under different sensitive attribute missing rates?
- **RQ2** How well the GNN-based encoder and the encoder affect the prediction accuracy of the proposed IFGNN?
- **RQ3** Can the fairness mechanism help tackle individual unfairness in the IFGNN?

5.1 Experimental Setup

In our experimental setup, we implemented the random data split, taking 50% for the training set, 25% for the validation set, and the remaining 25% for the test set. We set the threshold $\tau = 0.5$. The proposed IFGNN is compared with three baseline models on datasets with different rates of the missing sensitive attribute. For each dataset, we set three missing rate values r_s as 0.3, 0.5 and 0.7, to simulate real-world scenarios where sensitive attributes are missing. We randomly select $|r_s N|$ nodes from the dataset and set the corresponding sensitive attribute value to -1. To reduce the training time for large datasets and demonstrate the utility of the fairness mechanism in addressing individual unfairness, we sampled only 20% of the nodes from the training set to participate in the process outlined in Eq. (6), such as the generation of node pair-wise similarity matrices.

Table 2. Experiment results when the missing rate of the sensitive attribute is 0.5.

Dataset	Metrics	IFGNN	GCN	GAT	GraphSAGE
Credit	ACC(↑)	**0.7992**	0.7904	0.7765	0.7856
	AUC(↑)	**0.7381**	0.7162	0.7057	0.7095
	IUF (↓)	**0.0031**	0.0372	0.0407	0.0425
German	ACC(↑)	**0.7280**	0.6680	0.7240	0.6720
	AUC(↑)	**0.6520**	0.6249	0.6398	0.6317
	IUF (↓)	**0.0180**	0.0600	0.0480	0.0520
Income	ACC(↑)	**0.8281**	0.8160	0.7976	0.8170
	AUC(↑)	**0.8511**	0.8265	0.7776	0.8274
	IUF (↓)	**0.0108**	0.0208	0.0240	0.0379
Recidivism	ACC(↑)	**0.7703**	0.7391	0.7052	0.7347
	AUC(↑)	0.8719	0.8564	**0.8812**	0.8275
	IUF (↓)	**0.0125**	0.0320	0.0521	0.0562

Datasets. The datasets including Credit Defaulter (Credit), German Credit (German), Income and Recidivism. The summaries of dataset details are presented as follows:

- Credit. The Credit dataset is comprised of 3,000 nodes, each representing individuals who are using credit card payments. There are 13 attributes for each node, including education level and other bills and payments-related features. The classification task is to predict whether an individual will choose to set a credit card as the default payment method for the next month, considering age as the sensitive attribute.
- German. The German dataset is constructed on 1,000 nodes, where each node represents individuals who are customers of a German bank. Individuals are described with the 30 attributes, such as employment statement, loan amounts and the number of loans. The gender attribute is considered a sensitive attribute. The classification task is to determine whether an individual is a creditworthy customer based on all attributes.
- Income. The Income dataset consists of 14,821 nodes, where each node represents a person with income. There are 14 attributes for each node, and we choose race as the sensitive attribute. Other attributes include working hours per week and work classes. The task is to predict if the income of an individual is over $\$50,000$ annually.
- Recidivism. The Recidivism dataset has 18,876 nodes, each representing a defendant who got released on bail during 1990-2009 [21]. There are 18 attributes for each node, such as age, education level and marital status. The task is to determine if a defendant would receive bail based on all attributes of the defendant. The sensitive attribute used here is race.

Table 3. Experiment results when the missing rate of the sensitive attribute is 0.7.

Dataset	Metrics	IFGNN	GCN	GAT	GraphSAGE
Credit	ACC(↑)	**0.7992**	0.7904	0.7765	0.7648
	AUC(↑)	**0.7992**	0.7236	0.6775	0.7200
	IUF (↓)	**0.0023**	0.0312	0.0360	0.0275
German	ACC(↑)	0.7230	0.6960	**0.7360**	0.7040
	AUC(↑)	0.6498	0.6316	0.6515	**0.6746**
	IUF (↓)	**0.0080**	0.0480	0.0320	0.0440
Income	ACC(↑)	**0.8230**	0.8133	0.7955	0.8194
	AUC(↑)	**0.8468**	0.7902	0.7616	0.8177
	IUF (↓)	**0.0067**	0.0197	0.0189	0.0351
Recidivism	ACC(↑)	**0.7654**	0.7179	0.7004	0.7417
	AUC(↑)	**0.8716**	0.8424	0.8622	0.7926
	IUF (↓)	**0.0119**	0.0301	0.0403	0.0430

Evaluation metrics The model will be evaluated on utility (node classification performance) and individual fairness performance. The evaluation metrics are presented as follows.

- Utility. For evaluating the model's utility in node classification tasks, we employed widely used metrics: Accuracy (ACC) and AUCROC (AUC) scores.
- Individual fairness. An individual fairness awareness model should not be affected by the sensitive attribute and generate the same predictions for similar nodes, as highlighted by Dwork et al. [12]. Building upon this concept, we introduce a novel metric termed Individual Fairness (IUF), designed to measure individual fairness. In cases where the sensitive attribute is not considered, node predictions depend only on non-sensitive attributes. Consequently, the model should produce the same predictions for nodes with similar non-sensitive attributes. Therefore, when the sensitive attribute is considered, a model with individual fairness should yield predictions the same as those generated by the model that do not consider the sensitive attribute. IUF aims to quantify the inconsistency between predictions in these two cases. A lower IUF value implies enhanced individual fairness in the model.

Baseline. To our best knowledge, there is no research focusing on individual fairness in graph data with missing sensitive attributes, the proposed IFGNN is compared against three baseline models: GCN [19], GAT [29] and GraphSAGE [15].

5.2 Performances and Discussions

To solve **RQ1**, we conduct experiments under different r_s: 0.3, 0.5, and 0.7, to compare the performance of the proposed IFGNN with three baseline models.

The results of these experiments are presented in Table 1, 2, and 3. The results reveal that variations in r_s do not affect the effectiveness of the proposed IFGNN in maintaining individual fairness, as it consistently outperforms the baseline models significantly.

Moreover, the proposed IFGNN shows superior performance in terms of both ACC and AUC compared to the three baseline models in most cases. This observation indicates that the encoder and decoder enable the model to learn more representative node embeddings. It is worth noting that when the dataset contains a small number of nodes, such as the German dataset, the ACC and AUC of all the compared models are generally lower, including the proposed IFGNN. This can be attributed to the insufficient information in the dataset, which makes it a challenge for models to learn sufficient feature representations. However, the proposed decoder assists the encoder in enhancing feature learning, thereby leading to higher ACC and AUC for the proposed IFGNN compared to the baseline models. Although in rare cases, the three baseline models slightly outperform the proposed IFGNN in terms of ACC and AUC, they show severe individual discrimination at the same time. It indicates that the proposed IFGNN effectively addresses individual unfairness while maintaining satisfactory utility performance.

As for the baseline models, we observed that they do not show remarkable differences in terms of ACC and AUC. It implies that the aggregation processes among GCN, GAT and GraphSAGE do not significantly impact the utility performance in the node classification task. Therefore, the encoder in the proposed IFGNN can be replaced with any model using GNN as a backbone.

Table 4. Experiment results in the ablation study of IFGNN.

	Metrics	Credit	German	Income	Recidivism
IFGNN w/o $\mathcal{L}_1$	ACC (↑)	0.7756	<u>0.7320</u>	0.8200	<u>0.7300</u>
	AUC (↑)	0.7129	0.6352	0.8362	0.8326
	IUF (↓)	<u>0.0131</u>	<u>0.0120</u>	<u>0.0121</u>	<u>0.0303</u>
IFGNN w/o $\mathcal{L}_2$	ACC (↑)	<u>0.7785</u>	**0.7360**	<u>0.8214</u>	0.7196
	AUC (↑)	<u>0.7131</u>	<u>0.6472</u>	<u>0.8392</u>	<u>0.8382</u>
	IUF (↓)	0.0200	0.0160	0.0167	0.0476
IFGNN	ACC (↑)	**0.7992**	**0.7360**	**0.8325**	**0.7722**
	AUC (↑)	**0.7297**	**0.6674**	**0.8520**	**0.8692**
	IUF (↓)	**0.0043**	**0.0120**	**0.0084**	**0.0182**

5.3 Ablation Study

In this section, we aim to answer **RQ2** and **RQ3** on how the encoder, decoder and fairness mechanism affect the performance of IFGNN. In ablation experiments, we set the missing sensitive attribute rate to 0.3 for the four datasets. We

explore the performance under two cases by excluding the objective functions: $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively. These cases are denoted as IFGNN w/o $\mathcal{L}_1$ and IFGNN w/o $\mathcal{L}_2$. The results of the experiments are presented in Table 4.

When the IFGNN is without the learning objective function $\mathcal{L}_1$, the model does not employ the encoder and decoder to learn node feature embeddings for the node classification task. The ACC and AUC, which are used to evaluate the performance of the binary classification model, become unsatisfactory compared to the results of IFGNN w/o $\mathcal{L}_2$ and the IFGNN model. The encoder and decoder are employed to gather information from neighbourhoods, thereby alleviating the effect of missing sensitive attributes. The unsatisfactory prediction performance indicates that the IFGNN w/o $\mathcal{L}_1$ model loses the ability to obtain sufficient structure and node attribute information on the graph to generate representative node embeddings. In contrast, IFGNN achieves the highest utility performance with the help of $\mathcal{L}_1$.

When the IFGNN is without the learning objective function $\mathcal{L}_2$, the model does not under the constraint of the Lipschitz condition. As a result, the model can only produce biased results using the encoder, decoder, and node classifier, leading to higher IUF values compared to the results of IFGNN w/o $\mathcal{L}_1$ and the IFGNN model. While IFGNN consistently achieves the lowest IUF values across all four datasets. It indicates the superiority of our overall learning objective of IFGNN, which effectively enhances the model's individual fairness awareness.

6 Conclusion

In this paper, we study a novel problem of maintaining individual fairness while ensuring prediction performance in graph data with unlimited missing sensitive attributes. To solve this problem, we propose a novel model IFGNN which incorporates a GNN-based encoder and a decoder for learning node feature embeddings when the sensitive attribute is missing. To address the individual unfairness within IFGNN, we employ the Lipschitz condition that helps mitigate the discrimination caused by the sensitive attribute. We conduct extensive experiments on four real-world datasets to demonstrate that IFGNN outperforms the baseline models in terms of enhancing individual fairness awareness and prediction accuracy.

References

1. Agarwal, C., Lakkaraju, H., Zitnik, M.: Towards a unified framework for fair and stable graph representation learning. In: Uncertainty in Artificial Intelligence. pp. 2114–2124. PMLR (2021)
2. Awasthi, A., Garov, A.K., Sharma, M., Sinha, M.: Gnn model based on node classification forecasting in social network. In: AISC. pp. 1039–1043 (2023)
3. Beutel, A., Chen, J., Zhao, Z., Chi, E.H.: Data decisions and theoretical implications when adversarially learning fair representations. arXiv preprint arXiv:1707.00075 (2017)

4. Chen, X., Chen, S., Yao, J., Zheng, H., Zhang, Y., Tsang, I.W.: Learning on attribute-missing graphs. IEEE transactions on pattern analysis and machine intelligence **44**(2), 740–757 (2020)
5. Cheng, D., Chen, C., Wang, X., Xiang, S.: Efficient top-k vulnerable nodes detection in uncertain graphs. IEEE Transactions on Knowledge and Data Engineering (2021)
6. Cheng, D., Wang, X., Zhang, Y., Zhang, L.: Risk guarantee prediction in networked-loans. In: IJCAI (2020)
7. Cheng, D., Yang, F., Wang, X., Zhang, Y., Zhang, L.: Knowledge graph-based event embedding framework for financial quantitative investments. In: SIGIR. pp. 2221–2230 (2020)
8. Choudhary, M., Laclau, C., Largeron, C.: A survey on fairness for machine learning on graphs. arXiv preprint arXiv:2205.05396 (2022)
9. Dai, E., Wang, S.: Say no to the discrimination: Learning fair graph neural networks with limited sensitive attribute information. In: WSDM. pp. 680–688 (2021)
10. Dong, Y., Kang, J., Tong, H., Li, J.: Individual fairness for graph neural networks: A ranking based approach. In: KDD. pp. 300–310 (2021)
11. Du, M., Yang, F., Zou, N., Hu, X.: Fairness in deep learning: A computational perspective. IEEE Intelligent Systems **36**(4), 25–34 (2020)
12. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., Zemel, R.: Fairness through awareness. In: Proceedings of the 3rd innovations in theoretical computer science conference. pp. 214–226 (2012)
13. Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E.: Neural message passing for quantum chemistry. In: International conference on machine learning. pp. 1263–1272. PMLR (2017)
14. Guo, D., Chu, Z., Li, S.: Fair attribute completion on graph with missing attributes. arXiv preprint arXiv:2302.12977 (2023)
15. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. Advances in neural information processing systems **30** (2017)
16. Hao, Y., Cao, X., Sheng, Y., Fang, Y., Wang, W.: Ks-gnn: Keywords search over incomplete graphs via graphs neural network. Advances in Neural Information Processing Systems **34**, 1700–1712 (2021)
17. Hardt, M., Price, E., Srebro, N.: Equality of opportunity in supervised learning. Advances in neural information processing systems **29** (2016)
18. Kang, J., He, J., Maciejewski, R., Tong, H.: Inform: Individual fairness on graph mining. In: KDD. pp. 379–389 (2020)
19. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
20. Lahoti, P., Gummadi, K.P., Weikum, G.: ifair: Learning individually fair data representations for algorithmic decision making. In: ICDE. pp. 1334–1345 (2019)
21. Loveland, D., Pan, J., Bhathena, A.F., Lu, Y.: Fairedit: Preserving fairness in graph neural networks through greedy graph editing. arXiv preprint arXiv:2201.03681 (2022)
22. Mansoor, H., Ali, S., Alam, S., Khan, M.A., Hassan, U.U., Khan, I.: Impact of missing data imputation on the fairness and accuracy of graph node classifiers. In: IEEE International Conference on Big Data (Big Data). pp. 5988–5997 (2022)
23. Masrour, F., Wilson, T., Yan, H., Tan, P.N., Esfahanian, A.: Bursting the filter bubble: Fairness-aware network link prediction. In: AAAI. vol. 34, pp. 841–848 (2020)

24. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A.: A survey on bias and fairness in machine learning. ACM computing surveys (CSUR) **54**(6), 1–35 (2021)
25. Pieterse, J., Mocanu, D.C.: Evolving and understanding sparse deep neural networks using cosine similarity. arXiv preprint arXiv:1903.07138 (2019)
26. Rahman, T., Surma, B., Backes, M., Zhang, Y.: Fairwalk: Towards fair graph embedding (2019)
27. Sarkar, D., Roy, S., Malakar, S., Sarkar, R.: A modified gnn architecture with enhanced aggregator and message passing functions. Engineering Applications of Artificial Intelligence **122**, 106077 (2023)
28. Song, W., Dong, Y., Liu, N., Li, J.: Guide: Group equality informed individual fairness in graph neural networks. In: KDD. pp. 1625–1634 (2022)
29. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903 (2017)
30. Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y., et al.: Graph attention networks. stat **1050**(20), 10–48550 (2017)
31. Wang, H., Yu, J., Wang, X., Chen, C., Zhang, W., Lin, X.: Neural similarity search on supergraph containment. IEEE Transactions on Knowledge and Data Engineering (2023)
32. Wang, X., Zhang, Y., Zhang, W., Lin, X.: Efficient distance-aware influence maximization in geo-social networks. IEEE Transactions on Knowledge and Data Engineering **29**(3), 599–612 (2016)
33. Wang, X., Zhang, Y., Zhang, W., Lin, X., Chen, C.: Bring order into the samples: A novel scalable method for influence maximization. IEEE Transactions on Knowledge and Data Engineering **29**(2), 243–256 (2016)
34. Wang, X., Gu, T., Bao, X., Chang, L., Li, L.: Individual fairness for local private graph neural network. Knowledge-Based Systems **268**, 110490 (2023)
35. Wu, W., Li, B., Luo, C., Nejdl, W.: Hashing-accelerated graph neural networks for link prediction. In: Proceedings of the Web Conference 2021. pp. 2910–2920 (2021)
36. Wu, Y., Zhao, J., Sun, R., Chen, C., Wang, X.: Efficient personalized influential community search in large networks. Data Science and Engineering **6**(3), 310–322 (2021)
37. Xu, P., Zhou, Y., An, B., Ai, W., Huang, F.: Gfairhint: Improving individual fairness for graph neural networks via fairness hint. arXiv preprint arXiv:2305.15622 (2023)
38. Yu, J., Wang, H., Wang, X., Li, Z., Qin, L., Zhang, W., Liao, J., Zhang, Y.: Group-based fraud detection network on e-commerce platforms. In: KDD. pp. 5463–5475 (2023)
39. Zafar, M.B., Valera, I., Rogriguez, M.G., Gummadi, K.P.: Fairness constraints: Mechanisms for fair classification. In: Artificial intelligence and statistics. pp. 962–970. PMLR (2017)
40. Zemel, R., Wu, Y., Swersky, K., Pitassi, P.T., Dwork, C.: Learning fair representations. In: Proceedings of the 30th International Conference on Machine Learning. vol. 28, pp. 325–333 (2013)
41. Zhang, B.H., Lemoine, B., Mitchell, M.: Mitigating unwanted biases with adversarial learning. In: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society. pp. 335–340 (2018)
42. Zhang, X., Wang, H., Yu, J., Chen, C., Wang, X., Zhang, W.: Polarity-based graph neural network for sign prediction in signed bipartite graphs. World Wide Web **25**(2), 471–487 (2022)

43. Zhang, X., Wang, H., Yu, J., Chen, C., Wang, X., Zhang, W.: Bipartite graph capsule network. World Wide Web (WWW) **26**(1), 421–440 (2023)