通过反事实去偏差改进图神经网络的公平性

Zengyi Wo
wozengyi1999@tju.edu.cn
Tianjin University
Tianjin, China

Chang Liu
Tianjin University
Tianjin, China

Yumeng Wang
Tianjin University
Tianjin, China

Minglai Shao*
shaoml@tju.edu.cn
Tianjin University
Tianjin, China

Wenjun Wang†
wjwang@tju.edu.cn
Tianjin University
Tianjin, China

摘要

图神经网络（GNNs）在建模图形结构数据方面取得了成功。然而，类似于其他机器学习模型，GNNs 可以基于种族和性别等属性表现出预测偏差。此外，GNNs 中的偏见可能会因图形结构和消息传递机制而加剧。最近的前沿方法提出通过过滤输入或表示中的敏感信息来减轻偏见，例如边缘丢弃或特征掩码。然而，我们主张这样的策略可能无意中消除非敏感特性，导致预测准确性与公平性之间的平衡受损。为了解决这一挑战，我们提出了一种新的利用反事实数据增强的方法来缓解偏见。该方法涉及在消息传递之前使用反事实创建多样化的邻居，促进从增强图形中学习无偏节点表示。随后，采用对抗判别器以减少传统 GNN 分类器中的预测偏见。我们提出的 Fair-ICD 技术确保了 GNNs 在适度条件下的公平性。通过三个 GNN 后端在标准数据集上的实验表明，Fair-ICD 显著提高了公平度量，同时保持了高预测性能。

CCS CONCEPTS

• **Computing methodologies** → **Neural networks**.

KEYWORDS

图表示学习；公平性；反事实增强

ACM Reference Format:

Zengyi Wo, Chang Liu, Yumeng Wang, Minglai Shao, and Wenjun Wang. 2024. 通过反事实去偏差改进图神经网络的公平性. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (KDD WorkShop)*. ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

*Corresponding author
†Corresponding author

KDD WorkShop, August 25-26, 2024, Barcelona, Spain

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

1 介绍

图结构数据，例如引用网络，在现实世界中变得越来越普遍。近年来，图神经网络（GNNs）因其在建模此类数据方面的卓越成就而获得了广泛关注。通过 GNN 学习的表示对于一系列下游任务（如节点分类 [1, 2] 和链接预测 [3, 4]）和实际应用（如药物发现 [5] 和异常检测 [6, 7]）至关重要，产生了巨大的社会影响。

尽管具有强大的表示能力，最近的研究 [8, 9] 发现图神经网络的预测缺乏对公平性的考量，可能导致歧视性和有偏见的决策。将此类不公平的预测应用于高风险决策任务（例如犯罪预测 [10] 和信用评估 [11]）可能会产生重要的伦理和道德问题。图神经网络预测中的偏差可以归因于它们倾向于继承并加剧历史数据中嵌入的偏见。*来自特征数据的偏差*。节点特征与敏感信息之间的统计相关性最终导致节点嵌入隐式编码了敏感属性信息。*来自图结构的偏差*。根据同质效应，具有相同敏感属性的节点相比不同敏感属性的节点更有可能连接在一起。*GNNs 的聚合机制*。图神经网络的聚合机制平滑了相邻节点的表示，进一步导致具有相同敏感属性的节点表示的收敛，最终加剧了从表示中得出的预测与敏感属性之间的相关性。

在过去几年中，有许多研究探索了增强 GNNs 公平性的问题。一般来说，这些方法的核心思想在于完全丢弃与敏感属性相关的信息（例如，边删除 [12]、特征遮罩 [9]），期望 GNNs 的预测结果变得独立于敏感属性以满足公平性要求。基于完全丢弃与敏感属性相关信息的概念，这种方法往往难以完全区分那些与下游任务无关但与敏感属性相关的信息和有效信息（真正对下游任务有帮助的信息）。因此，不可避免地会导致与下游任务有关的信息丢失，从而在表达的预测能力与公平性之间形成次优平衡。

为解决上述问题，我们建议采用一种新方法通过利用偏置补偿技术来增加节点邻域的多样性，并借助对抗性判别器来处理 GNN 分类中的偏置。我们的独特策略集中在反事实数据增强作为一种对抗偏置的方法上。该方法涉及在消息传递之前生成包含反事实的多样邻域，以从增强图中获取无偏的节点表示。此外，我们利用判别器的对抗训练来最小化标准 GNN 分类器中的有偏预测。我们的方法称为 Fair-ICD。

我们的贡献如下：(1) 我们提出了一种基于*偏置补偿*的新范式，用于学习公平的图神经网络（GNN），该方法利用数据增强来提高邻域的异质性以减轻敏感属性的影响。(2) 具体而言，我们通过反事实推理结合对抗判别器，在三种 GNN 主干中实现了公平的表示学习。(3) 实验结果表明，与最近最先进的方法相比，所提出的 Fair-ICD 可以在预测性能和公平性之间取得更好的平衡。

2 相关工作

2.1 图中的公平性

大多数机器学习模型缺乏对公平性的考虑，可能会导致有偏见和不公平的决策。图挖掘算法，例如 GNNs，也存在同样的问题。例如，工作推荐模型可能优先将机会分配给特定性别群体的申请人，即使来自不同性别群体的候选人具有与工作表现相关的相似资格。公平可以分为几种常见类型：*群体公平性* [13]，即算法既不应歧视也不应偏向任何敏感子组；*个体公平* [14]，即相似的个体获得相似的待遇；以及*反事实公平性* [15]，即公正的决策独立于敏感属性值。

EDITS[16] 是一种预处理方法，提出了一种模型不可知的框架，为任何 GNN 提供具有减少偏差的属性网络作为输入。NIFTY[12] 通过轻微扰动节点属性和边，并翻转敏感属性值来创建原始节点的反事实情况，生成两个增强图。Graphair[17] 提出了一种自动化增强模型，利用对抗学习和对比学习实现公平性和信息性，以生成增强图。FairVGNN[9] 考虑了特征传播后敏感信息泄漏的影响。它通过识别并屏蔽与敏感相关的信息通道以及自适应地钳制权重来自动学习公平视角。FairGNN[8] 确保生成的表示不通过对抗去偏处理泄露敏感信息，同时增加了协方差约束。如观察到的，*边缘扰动*、*特征遮罩*和*对抗去偏差*是增强公平性的常见且有效的方法。然而，大多数现有方法的核心思想是完全丢弃与敏感属性相关的信息。由于信息之间的复杂纠缠，在此过程中确保不会丢失下游任务的相关信息是一个挑战。

2.2 图中的反事实增强

在探索公平性的任务中，反事实 [12, 18–20] 节点是指具有不同敏感属性的原始节点的一个版本。具体来说，它是一个与原节点有相同标签（类似特征）但不同的

敏感属性值的节点。获得反事实节点的方法有三种：生成、建模和搜索。

生成。基于生成的方法通常通过翻转敏感属性来创建反事实。NIFTY[12] 通过扰动节点的敏感属性生成反事实增强图。虽然生成方法很简单，但它们依赖于一个不现实的假设，即敏感属性对其他属性和图形结构没有因果影响，导致生成的反事实并不真正存在。简单地翻转敏感属性可能会破坏原有的语义信息，从而影响最终性能。**建模**。基于模型的方法在翻转敏感属性后重构图结构与敏感属性之间的依赖关系。它们修改图形的结构或特征矩阵以获得反事实增强图。GEAR[19] 在扰动节点自身的敏感属性值及其邻居之后，通过 GraphVAE 生成两种类型的反事实增强图。尽管考虑了图结构和敏感属性之间的依赖关系，但这种方法得到的反事实仍然是非现实的。**搜索**。基于搜索的方法强调在训练数据中寻找合适的反事实。通过这种方法获得的反事实确实存在，并且它们包含的信息更加真实可靠。CAF[20] 利用标签和敏感属性作为指导，在表示空间中搜索潜在的反事实。

3 初步的

3.1 背景

3.1.1 **图形**. 设 $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathrm{A}, \mathrm{X})$ 表示一个带属性图，它由一组 N 节点 $\mathcal{V}$ 和一组边 $\mathcal{E}$ 组成。这里，$\mathrm{X} \in \mathbb{R}^{N \times D}$ 表示节点属性矩阵，$\mathrm{A} \in \mathbb{R}^{N \times N}$ 是邻接矩阵，在节点 v_i 和 v_j 之间存在边 $e_{ij} \in \mathcal{E}$ 时，$\mathrm{A}_{ij} = 1$，否则为 $\mathrm{A}_{ij} = 0$。每个节点 v_i 都有一个敏感属性 $s_i \in \{0, 1\}$。

3.1.2 **图神经网络（*GNNs*）**. 节点 v_i 的邻居集表示为 $\mathcal{N}(i) = \{v_j \mid (v_i, v_j) \in \mathcal{E}\}$。图神经网络（GNNs）旨在通过迭代传递来自相邻节点的信息来创建节点表示。在一个通用的消息传递方案中，节点 v_i 的表示更新如下：

$$z_i^{(k)} = \mathrm{UPDATE}(z_i^{(k-1)}, \mathrm{AGGREGATE}(z_j^{(k-1)} \mid j \in \mathcal{N}(i))), \tag{1}$$

其中函数 AGGREGATE(·) 和 UPDATE(·) 是可训练的，并分别负责邻居聚合和表示更新。

在图卷积网络 (**图卷积网络**)[21] 中，聚合过程采用平均聚合器，而更新函数涉及线性变换后跟非线性激活。**吉尼系数** [22] 采用求和聚合器来捕获完整的邻域信息，并使用多层感知机（MLP）作为更新函数。另一方面，**图森龄** [23] 利用多种聚合技术，如平均、LSTM 或池化聚合器，并在应用更新函数之前将节点自身的特征与聚合的邻域特征相结合。

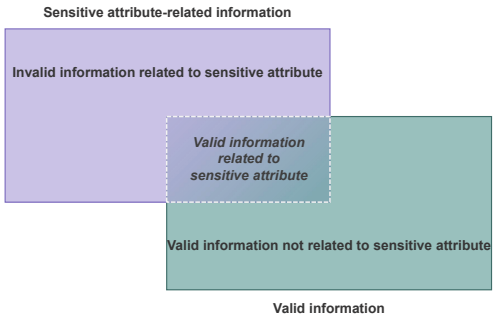

图 1: 敏感属性相关信息与有效信息之间的复杂纠缠。

3.1.3 **节点分类中的公平性评估**. **人口统计学上的公平性** [14] 。直观上，***DP*** 是一个常用指标，用于衡量两个敏感子群体接受率差异：

$$\mathrm{DP} = \left| P\left(\hat{Y} = 1 | S = 0\right) - P\left(\hat{Y} = 1 | S = 1\right) \right|. \tag{2}$$

机会平等 [8, 9, 16] 。该指标认识到在许多场景中，敏感特征可能与目标任务相关，需要对具有阳性真实标签的个体进行正面预测时独立于其敏感特征：

$$\mathrm{EO} = \left| P\left(\hat{Y} = 1 | Y = 1, S = 0\right) - P\left(\hat{Y} = 1 | Y = 1, S = 1\right) \right|. \tag{3}$$

3.2 GNNs 中的公平性

3.2.1 **动机**. 我们调查了三种数据增强策略在信息性和公平性方面的表现：边删除、特征掩码和偏置偏移，如表 1 所示。

- **边缘脱落** 是一种常见的数据增强和处理策略，涉及从图中删除一些边；
- **特征遮罩** 是一种数据增强技术，涉及隐藏某些节点特征，迫使模型依赖其可用的剩余信息。

表 1: 三种数据增强策略在 Pokec-n 数据集上的表现，涉及信息量和公平性。箭头（↗,↘）表示比基础版本表现更差或更好的性能。箭头（↑,↓）指示更好性能的方向。

Strategies	ACC (↑)	DP (↓)
vanilla	68.55±0.51	3.75±0.94
Edge-dropping	67.24±0.49 (↘)	1.22±0.94 (↗)
Feature-masking	66.10±1.45 (↘)	1.69±0.79 (↗)
偏移补偿	**69.06±0.6** (↗)	**0.67±0.39** (↗)

表 2: 原始图和反事实增强图在 Pokec-n 数据集上的偏差评估。设 $\mathcal{G}$ 表示原始图，$\mathcal{G}'$ 表示反事实增强图。箭头 (↑,↓) 指示更好的性能方向。

	$\mathcal{G}$	$\mathcal{G}'$
Avg. degree	16.53	16.53
Avg. heterogeneous degree (↑)	0.73	8.13
Nodes w/o heterogeneous neighbors (↓)	46134	8183

- **偏移补偿**，这鼓励中心节点更专注于聚合具有不同敏感属性的邻居节点，以试图抵消与敏感属性相关的偏见和无效信息，并减轻偏见。

结果表 1 显示了不同数据增强后的 GNN 性能：

- 表示为 **ACC**：**Bias-offsetting** > **Vanilla** > **Edge-dropping**, **Feature-masking** ；
- 表示成**动态规划**：**Bias-offsetting** < **Vanilla** < **Edge-dropping**, **Feature-masking** 。

表 2 显示了提高平均异质度，减少没有异质邻居的节点。我们计算了平均异质度来衡量图中的偏见程度。请注意，在此上下文中，异质性指的是具有不同敏感属性值的节点。

$$\text{Avg. heterogeneous degree} = \frac{\sum_{v_i \in \mathcal{V}} |\{v_j \in \mathcal{N}(i) | s_i \neq s_j\}|}{|\mathcal{V}|}, \tag{4}$$

平均异质度越低，相同敏感属性值的节点之间的连接程度越高，导致表示与敏感属性之间的相关性增加，从而增加了偏见的风险。**备注**根据我们的观察，我们确定边缘删除和特征掩码技术与标准方法相比，并不能一致地提高信息性和公平性。这一挑战源于敏感属性数据与相关信息之间的复杂相互作用，如图 1 所示。虽然边缘删除和特征掩码试图完全消除敏感属性数据，但这个过程也会丢弃对后续任务至关重要的相关重要信息，从而在信息性和公平性之间造成不充分的平衡。相反，我们的偏差补偿方法能够有效同时提升这两个维度。

3.2.2 **任务**. 信息量。在这项研究中，我们调查了半监督节点分类任务。具体而言，我们的目标是开发一个能够利用图 $\mathcal{G}$、节点特征 X 和标记的节点集 $\mathcal{V}^L \subseteq \mathcal{V}$ 来预测未标记节点集 $\mathcal{V}^U = \mathcal{V} \setminus \mathcal{V}^L$ 中每个节点标签 $\hat{y} \in \mathcal{Y}$ 的模型。一种常见的模型架构是将 GNN 编码器与分类器结合。分类性能评估涉及比较实际标签 Y 与预测标签 $\hat{Y}$。公平性。公平性的主要目标是尽量减少预测标签 $\hat{Y}$ 与敏感属性 S 之间的相关性，同时尽可能保持分类准确性，相当于减少节点表示和敏感属性之间的相关性。

4 方法论

概述 本节详细介绍了我们提出的方法，名为 Fair-ICD，旨在通过反事实数据增强来提高 GNN 学习表示的公平性。它由两个模块组成：无偏表示学习模块和对抗去偏模块。在无偏表示学习模块中，我们使用反事实方法增强原始图邻居的异质性，使从增强图中学习节点的无偏表示模式成为可能。在对抗去偏模块中，采用对抗训练来减少 GNN 分类器做出预测时的偏差。现在我们将详细介绍每个设计。

4.1 无偏表示学习

为了实现节点特征的公平（无偏）表示，我们提出了一种反事实增强方法。通过调查聚合过程中节点的公正性，我们的目标是增强节点邻居异质性的影响，从而使通过反事实干预来操控增强图生成成为可能。

4.1.1 **反事实数据增强**. 我们的目标是通过确保反事实具有对比的敏感属性，来减轻源于敏感属性信息的偏差。虽然一种直接的方法是翻转给定样本的敏感属性以创建反事实，但这种方法通常是无效的。为了克服这个局限性，如 3.2 节所述，我们引入了一种新颖的数据增强技术，该技术强调原始图中节点邻域的分布。

当处理具有不同敏感属性的节点时，我们保持边；相反，当节点共享相同的敏感属性时，我们识别邻域节点的反事实节点，并建立这些识别节点的反事实关系：

$$(v_a, v_b) = \underset{v_a, v_b \in \mathcal{V}}{\arg\min} \left[\|\boldsymbol{x}_a - \boldsymbol{x}_b\|_2^2 | s_a \neq s_b \right], \tag{5}$$

其中 v_b 代表 v_a 的反事实节点，也可以表示为 $v_a^c = v_b$。

反事实增强图 $\mathcal{G}'$ 定义如下：

$$\mathcal{G}' = \begin{cases} A_{ij}, & \text{if } s_i \neq s_j \\ A_{ik}, & \text{if } s_i = s_j \text{ and } v_j^c = v_k \end{cases}, \tag{6}$$

在实际实现中，我们最初通过计算节点特征之间的欧几里得距离来评估相似性。随后，我们识别出与每个节点最相似的 Top-k 节点，并在此集合中确定具有不同敏感信息的节点。通过应用上述反事实处理，我们增强了图结构，使中心节点周围的邻域表现出异质性。

4.1.2 **公平邻近捕获**. 由于存在无法找到高质量反事实节点的情况，我们使用一个简单的 MLP_θ 来学习增强的高质量异构邻域聚合模式：

$$\mathcal{L}_{\text{unbias}} = \arg\min \mathbb{E}_{i,j \sim \mathcal{V}} ||MLP_\theta(x_i) - \text{AGG}(x_i, A'_{ij})||, \tag{7}$$

在此背景下，$\text{AGG}(\cdot,\cdot)$ 表示均值聚合，而 A'_{ij} 表示增强图的邻接矩阵。

我们的目标是获得原始图中节点的无偏且公平的表示。由于原始特征包含大量有用的语义信息，我们将原始特征与无偏特征拼接，并将它们输入传统的 GNN 编码器 f_G 进行消息传递聚合：

$$\tilde{X}^k = X^k + \text{MLP}_\theta^k(X^k), X^{k+1} = f_G(\tilde{X}^k). \tag{8}$$

4.2 对抗去偏差

GNN 分类器 f_G 可能会做出有偏见的预测，因为学习到的表示 f_G 由于节点特征、图结构和 GNN 的聚合机制而表现出偏差。确保 f_G 公平性的一种方法是在最终层表示中消除偏差。为了实现这一目标，我们引入了一个对抗判别器来帮助 GNN 分类器去偏见。我们使用二元交叉熵损失为判别器提供约束：

$$\min \mathcal{L}_d = -\frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} [s_v \log \hat{s}_v + (1 - s_v) \log (1 - \hat{s}_v)]. \tag{9}$$

表 3: 数据集统计。

Dataset	Pokec-n
Nodes	66569
Edges	729129
Features	266
Lable	Working filed
sensitive attribute	Region

5 实验

5.1 实验设置

5.1.1 **数据集**. 我们遵循 [16] 中提出的方法，对我们的工作和基线方法在真实世界的基准数据集 Pokec-n[24] 上进行评估。这些数据集已在之前的图公平性学习研究中被广泛使用，并涵盖了一个多样化范围。我们在表 3 中提供了数据集统计信息。

5.1.2 **基线**. 我们将比较我们的方法与最新尖端技术以提高公平性。

- **原味的**：我们方法中的编码器基于三种广泛使用的图神经网络（GNN）：GCN[21]，GIN[22] 和 GraphSAGE[23]。
- **公平 GNN**[8]：该方法侧重通过对抗训练来减轻偏差。
- **编辑** [16]：通过修改图结构和节点属性来减少各种敏感群体之间歧视的方法。
- **NIFTY**[12]：一种结合特征扰动和边删除的技术，通过增强增强图与反事实图之间的相似性来确保公平性。
- **公平 VGNN**[9]：一个旨在通过隐藏相关渠道和自适应调整权重来防止敏感属性泄露的框架。

5.1.3 **评估协议**. 我们使用 F1 分数和准确性指标来评估下游分类任务的性能。在评估群体公平性时，根据

表 4: 节点分类 实验结果（均值与标准差）在数据集 Pokec-n 上的比较。

方法	模型	Pokec-n			
		F1	Acc	DP	EO
图卷积网络	Vanilla	**67.74±0.41**	68.55±0.51	3.75±0.94	2.93±1.15
	FairGNN	65.62±2.03	67.36±2.06	3.29±2.95	2.46±2.64
	EDITS	OOM	OOM	OOM	OOM
	NIFTY	64.02±1.26	67.24±0.49	1.22±0.94	2.79±1.24
	FairVGNN	64.85±1.17	66.10±1.45	1.69±0.79	1.78±0.70
	公平-ICD	67.55±0.41	**69.06±0.60**	**0.67±0.39**	**0.82±0.66**
GIN	Vanilla	67.87±0.70	69.25±1.75	3.71±1.20	2.55±1.52
	FairGNN	64.73±1.86	67.10±3.25	3.82±2.44	3.62±2.78
	EDITS	OOM	OOM	OOM	OOM
	NIFTY	61.82±3.25	66.37±1.51	3.84±1.05	3.24±1.60
	FairVGNN	68.01±1.08	68.37±0.97	1.88±0.99	1.24±1.06
	公平-ICD	**68.06±0.97**	**69.67±0.61**	**1.36±0.39**	**0.99±0.49**
GraphSAGE	Vanilla	67.15±0.88	69.03±0.77	3.09±1.29	2.21±1.60
	FairGNN	65.75±1.89	67.03±2.61	2.97±1.28	2.06±3.02
	EDITS	OOM	OOM	OOM	OOM
	NIFTY	61.70±1.47	68.48±1.11	3.84±1.05	3.90±2.18
	FairVGNN	67.40±1.20	68.50±0.71	**1.12±0.98**	1.13±1.02
	公平-ICD	**68.35±0.89**	**69.33±0.53**	1.22±0.46	**1.08±1.12**

先前的研究，我们考虑了 DP 和 EO。必须理解的是，较低的 DP 和 EO 值表明模型中的公平性水平更高。

5.1.4 **模型超参数和实现细节**. 我们在方法中利用多层感知器（MLP）来预测相邻实体的特征。我们的重点在于反事实数据增强，特别针对{3, 5, 10, 25}的 Top-k 值。对于 MLP 优化，我们使用了 Adam 优化器，并调整学习率在{0.1, 0.01, 0.001}范围内。我们的方法中的超参数系数在 [0, 10] 范围内进行了微调。GNN 编码器根据 [9] 中呈现的设置进行配置。结果以五次运行（使用不同的随机种子）的平均值和标准差形式展示。所有实验都在配备 24GB 内存的 GeForce GTX 3090 单个 GPU 单元上进行。

5.2 实验结果

我们展示了 Fair-ICD 的发现，以说明我们的方法——围绕反事实偏差缓解中心——可以达到比最先进(SOTA)方法更优的平衡。表 4 显示了 Fair-ICD 在各种 GNN 编码器上的整体分类性能和公平性方面表现出色。关于公平性，与表现最佳的基线相比，Fair-ICD 减少了 DP 和 EO。此外，在许多情况下，Fair-ICD 可以在准确性指标上超越标准编码器，这与其原理和模型架构相吻合。

6 结论

本文提出了一种基于反事实数据增强的创新策略，以实现偏差抵消，在消息传递之前使用反事实构建异质邻居，使得可以从增强图中学习节点的无偏表示。随后，采用对抗训练来减少传统 GNN 分类器预测中的偏差。我们称这种方法为 Fair-ICD，它在温和条件下确保了 GNN 的公平性。实验结果表明，在具有三种不同 GNN 骨干网络的基准数据集上，Fair-ICD 显著提高了公平度量指标，同时保持高预测精度。

ACKNOWLEDGMENTS

此项工作得到国家自然科学基金（编号 62272338）的支持。

REFERENCES

[1] Shunxin Xiao, Shiping Wang, Yuanfei Dai, and Wenzhong Guo. 2022. Graph neural networks in node classification: survey and evaluation. *Machine Vision and Applications* 33, 1 (2022), 4.

[2] Zengyi Wo, Minglai Shao, Wenjun Wang, Xuan Guo, and Lu Lin. 2024. Graph Contrastive Learning via Interventional View Generation. In *Proceedings of the ACM on Web Conference 2024*. 1024–1034.

[3] Muhan Zhang and Yixin Chen. 2018. Link prediction based on graph neural networks. *Advances in neural information processing systems* 31 (2018).

[4] Andrea Rossi, Denilson Barbosa, Donatella Firmani, Antonio Matinata, and Paolo Merialdo. 2021. Knowledge graph embedding for link prediction: A comparative analysis. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 15, 2 (2021), 1–49.

[5] Thomas Gaudelet, Ben Day, Arian R Jamasb, Jyothish Soman, Cristian Regep, Gertrude Liu, Jeremy BR Hayter, Richard Vickers, Charles Roberts, Jian Tang, et al. 2021. Utilizing graph machine learning within drug discovery and development. *Briefings in bioinformatics* 22, 6 (2021), bbab159.

[6] Xiaoxiao Ma, Jia Wu, Shan Xue, Jian Yang, Chuan Zhou, Quan Z Sheng, Hui Xiong, and Leman Akoglu. 2021. A comprehensive survey on graph anomaly detection with deep learning. *IEEE Transactions on Knowledge and Data Engineering* 35, 12 (2021), 12012–12038.

[7] Leman Akoglu, Hanghang Tong, and Danai Koutra. 2015. Graph based anomaly detection and description: a survey. *Data mining and knowledge discovery* 29 (2015), 626–688.

[8] 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*. 680–688.

[9] Yu Wang, Yuying Zhao, Yushun Dong, Huiyuan Chen, Jundong Li, and Tyler Derr. 2022. Improving Fairness in Graph Neural Networks via Mitigating Sensitive Attribute Leakage. In *SIGKDD*.

[10] Guangyin Jin, Qi Wang, Cunchao Zhu, Yanghe Feng, Jincai Huang, and Jiangping Zhou. 2020. Addressing crime situation forecasting task with temporal graph convolutional neural network approach. In *2020 12th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*. IEEE, 474–478.

[11] 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 systems with applications* 36, 2 (2009), 2473–2480.

[12] Chirag Agarwal, Himabindu Lakkaraju, and Marinka Zitnik. 2021. Towards a unified framework for fair and stable graph representation learning. In *Uncertainty in Artificial Intelligence*. PMLR, 2114–2124.

[13] Moritz Hardt, Eric Price, and Nati Srebro. 2016. Equality of opportunity in supervised learning. *Advances in neural information processing systems* 29 (2016).

[14] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*. 214–226.

[15] Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. 2017. Counterfactual fairness. *Advances in neural information processing systems* 30 (2017).

[16] Yushun Dong, Ninghao Liu, Brian Jalaian, and Jundong Li. 2022. Edits: Modeling and mitigating data bias for graph neural networks. In *Proceedings of the ACM web conference 2022*. 1259–1269.

[17] Hongyi Ling, Zhimeng Jiang, Youzhi Luo, Shuiwang Ji, and Na Zou. 2023. Learning fair graph representations via automated data augmentations. In *International Conference on Learning Representations (ICLR)*.

[18] Zhimeng Guo, Teng Xiao, Zongyu Wu, Charu Aggarwal, Hui Liu, and Suhang Wang. 2023. Counterfactual learning on graphs: A survey. *arXiv preprint arXiv:2304.01391* (2023).

[19] Jing Ma, Ruocheng Guo, Mengting Wan, Longqi Yang, Aidong Zhang, and Jundong Li. 2022. Learning fair node representations with graph counterfactual fairness. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 695–703.

[20] Zhimeng Guo, Jialiang Li, Teng Xiao, Yao Ma, and Suhang Wang. 2023. Towards fair graph neural networks via graph counterfactual. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 669–678.

[21] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[22] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2018. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826* (2018).

[23] Will Hamilton, Zhitao Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems* 30 (2017).

[24] Lubos Takac and Michal Zabovsky. 2012. Data analysis in public social networks. In *International scientific conference and international workshop present day trends of innovations*, Vol. 1.