Recent Advances in Fairness Analysis of User Profiling Approaches in E-Commerce with Graph Neural Networks

Erasmo Purificato[1,2], Ludovico Boratto[3] and Ernesto W. De Luca[1,2]

[1]Otto von Guericke University Magdeburg, Faculty of Computer Science, 39106 Magdeburg, Germany
[2]Leibniz Institute for Educational Media | Georg Eckert Institute, 38118 Brunswick, Germany
[3]University of Cagliari, Department of Mathematics and Computer Science, 09124 Cagliari, Italy

Abstract
User profiling is a critical procedure for e-commerce applications that captures online users' attributes, understands user models, supports the provision of tailor-made goods and services, and improves user satisfaction. With the advent of novel technologies like Graph Neural Networks (GNNs), the performance of user profiling approaches has improved by leaps and bounds, in step with the growing concern about data and algorithmic fairness. This paper provides an overview of recent advances in the fairness analysis of GNN-based models for user profiling in the e-commerce domain. We present the results of our recent works addressing the need for an accurate analysis of state-of-the-art models and the lack of a unified tool for enabling any user to perform a fairness analysis on a specific dataset by leveraging the most performing models in this context. Our goal is to foster discussions on the potential implications of our work within the community, not only from a technical view but also from domain experts' perspective.

Keywords
User Profiling, Algorithmic Fairness, Graph Neural Networks, E-Commerce

1. Background and Motivation

In the past few years, there has been a significant rise in the amount of personal information people share on a daily basis while interacting with artificial intelligence (AI) systems, notably with information retrieval (IR) systems and recommenders (RSs). In particular, one of the domains in which AI research has always found fertile ground by exploiting this enormous amount of user data is *e-commerce*. RSs have a well-established tradition of application in e-commerce platforms (e.g. Amazon, Alibaba, JD) with the aim of helping consumers find the most appropriate products to purchase [1, 2, 3]. These systems use product knowledge (either "hand-coded" knowledge provided by experts or "mined" knowledge learned from the behaviour of consumers) to guide consumers through the often-overwhelming task of locating products

22nd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2023) - Discussion Papers, November 06–09, 2023, Rome, Italy
erasmo.purificato@ovgu.de (E. Purificato); ludovico.boratto@acm.org (L. Boratto); ernesto.deluca@ovgu.de (E. W. De Luca)
https://erasmopurif.com/ (E. Purificato); https://www.ludovicoboratto.com/ (L. Boratto); https://ernestodeluca.eu/ (E. W. De Luca)
0000-0002-5506-3020 (E. Purificato); 0000-0002-6053-3015 (L. Boratto); 0000-0003-3621-4118 (E. W. De Luca)

they will like [1]. These systems have proved to be beneficial to both service providers and users [2].

In this context, one of the most important phases for the design and implementation of RSs is **user profiling**, which aims to generate profile vectors that represent users' interests by retrieving their personal information and historical interactions [4]. These profiles can be used for candidate generation [5], click-through rate prediction [6], conversion rate prediction [7], and long-term user engagement optimisation [8, 9]. Initial methods for profiling users concentrated solely on the examination of fixed attributes and were known as *explicit user profiling* [10]. Data for this approach was often obtained through online surveys or forms. However, these methods have been shown to be ineffective as users are reluctant to provide their personal information directly. As a result, modern systems have shifted their focus towards *implicit user profiling*, which profiles users based on their actions and interactions. This approach is also commonly referred to as *behavioural user profiling* [11].

An effective form to represent these behaviours in a natural manner is through the use of graphs, where nodes represent users and edges depict their interactions. The technologies that have lately demonstrated to be the most performing in dealing with such data structures are **Graph Neural Networks** (GNNs) [12, 13, 14, 15, 16]. As in many other domains, GNNs have become one of the state-of-the-art methods for e-commerce applications due to their powerful capability on topological feature extraction and relational reasoning [17]. Several promising GNN-based models for user profiling have been recently proposed and evaluated on e-commerce datasets. Chen et al. [18] presented a Heterogeneous Graph Attention Network (HGAT) for learning user representations taking into account the graph structure and the attention mechanism to determine the importance of each node's neighbour. Chen et al. [19] introduced a GCN-based model which reveals the benefits of improving the node representation before executing user profiling tasks. Yan et al. [20] proposed a Heterogeneous Graph Network (HGN) to enhance prediction performances by considering multiple types of relations and entities for user profiling, in contrast to previous works only based on single types. Typically, existing approaches evaluate user profiling techniques based on their ability to correctly classify a user's personal characteristics, such as gender or age [18]. However, the ease of accessing, manipulating, and mining such user-generated data raises concerns about data and algorithmic biases and fairness [21]. As with every machine learning (ML) system trained on historical data, GNNs risk replicating biases learned in such data and even amplifying them in their output. This is mainly due to the topology of graph structures and the typical message-passing process of GNNs, as nodes of the same sensitive attribute are more likely to be linked to each other than those different [22]. Unfair practices hidden in this class of models can be dangerous, especially in the e-commerce domain. Indeed, if modelling is more effective for specific demographic groups, they would inevitably systematically obtain less effective services.

Given the described situation of constantly emerging GNN-based models for user profiling in the e-commerce domain and the related concerns about fair outcomes, there are two specific challenges that require adequate consideration among the numerous issues that can arise from the depicted background:

- a rigorous analysis of potential discrimination provided by the state-of-the-art models in the field to understand how beneficial and safe their usage is in this domain;

- the development of a unified tool which allows users of any kind, such as researchers or practitioners, to evaluate the accuracy and fairness performances of any GNN-based model without the need for in-depth technical knowledge of the same.

Our contributions. In this paper, we provide an overview of the recent advances in the fairness analysis of GNN-based models for user profiling in the e-commerce domain to address the challenges previously illustrated. In particular, we describe our two recent works already published in prestigious venues (i.e. CIKM'22 and SIGIR'23) with the aim of bringing them to the attention of this community and fostering the discussion on potential implications with domain experts, not only from a technical point of view.

The paper is organised as follows. In Section 2, we present the fairness analysis conducted on the state-of-the-art GNN-based models for user profiling in terms of *disparate impact* and *disparate mistreatment* [23]. In Section 3, we describe the implementation of *FairUP* [24], a unified framework developed for evaluating these particular types of models. Finally, we will discuss yet open challenges and future work in the context (Section 4), which starts from the foundations laid by the research illustrated in this paper.

2. Fairness Analysis of State-of-the-Art GNN-based Models for User Profiling

The topic of **algorithmic fairness** has gained significant attention in recent times, particularly in light of the growing usage of automated decision-making systems. A substantial amount of literature has been produced on the overall approaches for identifying and mitigating bias in ML models [25, 26], specific user-related scenarios [27] and RSs [28, 29]. Only a few works have been published to evaluate fairness on GNNs (e.g. [30, 31]), but none of them estimated potential bias in state-of-the-art GNN-based models for user profiling tasks.

In the work [23], we evaluated the two most performing GNNs in the field of user profiling: *CatGCN* and *RHGN*.

CatGCN [19] is a Graph Convolutional Network (GCN) model tailored for graph learning on categorical node features. This model improves the initial node representation by integrating two types of explicit interaction modelling into its learning process: a local multiplication-based interaction on each pair of node features and a global addition-based interaction on an artificial feature graph. The proposed method demonstrates the effectiveness of performing feature interaction modelling before graph convolution.

RHGN [20] is a Relation-aware Heterogeneous Graph Network designed to model multiple relations on a heterogeneous graph between different kinds of entities. The core parts of this model are a transformer-like multi-relation attention, used to learn the node importance and uncover the meta-relation significance on the graph, and a heterogeneous graph propagation network employed to gather information from multiple sources. This approach outperforms several GNN-based models on user profiling tasks.

We performed two user profiling tasks by performing a binary classification on two real-world datasets from popular e-commerce platforms, namely *Alibaba* and *JD*.

Alibaba dataset[1] contains click-through rates data about ads displayed on Alibaba's Taobao platform and has been adopted in both [19] and [20] for evaluation.

JD dataset[2] consists of users and items from the retailer company of the same name having *click* and *purchase* relationships, already used in [20].

The fairness analysis constitutes the core of the contributions. We defined the fairness metrics adopted in our work considering $y \in \{0, 1\}$ as the binary target label and $\hat{y} \in \{0, 1\}$ as the prediction of the user profiling model $f : x \to y$. The sensitive attribute was denoted with $s \in \{0, 1\}$. Our focus in this paper was the assessment of the fairness of the previously-described GNNs in terms of **disparate impact**, which is a condition of indirect and often unintentional discrimination that emerges when procedures or systems seem to apparently treat people equally [32]. It affects situations where the model disproportionately discriminates against particular groups, even if the model does not explicitly utilise the sensitive attribute to make predictions but rather on some proxy attributes [33]. This is precisely what happens in the analysed GNNs, where the user models are generated by aggregating information from customers, and the sensitive attribute is not explicitly considered during classification. The concept of disparate impact is helpful when there is not a clear linkage in training data between the predicted label and the sensitive attribute [34].

Instead, when it is difficult to determine the correctness of a prediction related to sensitive attribute values, a complete fairness inspection should always include the viewpoint of **disparate mistreatment**. This notion considers the *misclassification rates* for user groups having different values of the sensitive attribute rather than considering the corrected predictions [34]. Furthermore, disparate mistreatment is valuable where misclassification costs rely on the group affected by the error.

We selected three metrics to evaluate disparate impact (i.e. *statistical parity*, *equal opportunity* and *overall accuracy equality*) and one metric for disparate mistreatment (i.e. *treatment equality*).

Statistical parity (or *demographic parity*) [35, 36] describes fairness as an equal probability for each group of being assigned to the positive class, i.e. predictions independent with sensitive attributes.

$$P(\hat{y} = 1 | s = 0) = P(\hat{y} = 1 | s = 1) \tag{1}$$

Equal opportunity [37] requires the probability of a user in a positive class to be classified with the positive outcome should be equal for each group.

$$P(\hat{y} = 1 | y = 1, s = 0) = P(\hat{y} = 1 | y = 1, s = 1) \tag{2}$$

Overall accuracy equality [38] defines fairness as the equal probability of a subject from either a positive or negative class to being assigned to its respective class.

$$\begin{aligned} P(\hat{y} = 0 | y = 0, s = 0) + P(\hat{y} = 1 | y = 1, s = 0) = \\ = P(\hat{y} = 0 | y = 0, s = 1) + P(\hat{y} = 1 | y = 1, s = 1) \end{aligned} \tag{3}$$

Treatment equality [38] demands the error ratio made by the classifier to be equal across different groups.

$$\frac{P(\hat{y} = 1 | y = 0, s = 0)}{P(\hat{y} = 0 | y = 1, s = 0)} = \frac{P(\hat{y} = 1 | y = 0, s = 1)}{P(\hat{y} = 0 | y = 1, s = 1)} \tag{4}$$

[1] https://tianchi.aliyun.com/dataset/dataDetail?dataId=56
[2] https://github.com/guyulongcs/IJCAI2019_HGAT

According to [39] and [31], to quantitatively estimate the fairness scores of the analysed models, we operationalise the metrics defined by Eqs. (1)-(4) as follows:

$$\Delta_{SP} = |P(\hat{y} = 1|s = 0) - P(\hat{y} = 1|s = 1)|, \tag{5}$$

$$\Delta_{EO} = |P(\hat{y} = 1|y = 1, s = 0) - P(\hat{y} = 1|y = 1, s = 1)|, \tag{6}$$

$$\begin{aligned} \Delta_{OAE} = |&P(\hat{y} = 0|y = 0, s = 0) + P(\hat{y} = 1|y = 1, s = 0)- \\ &- P(\hat{y} = 0|y = 0, s = 1) + P(\hat{y} = 1|y = 1, s = 1)|, \end{aligned} \tag{7}$$

$$\Delta_{TE} = \left| \frac{P(\hat{y} = 1|y = 0, s = 0)}{P(\hat{y} = 0|y = 1, s = 0)} - \frac{P(\hat{y} = 1|y = 0, s = 1)}{P(\hat{y} = 0|y = 1, s = 1)} \right| \tag{8}$$

Through an extensive set of experiments[3], we derived the following observations about the analysed models, correlating their different user profiling paradigms with the fairness metrics scores to create a baseline for future assessment considering GNN-based models for user profiling in e-commerce:

1. The ability of *RHGN* to represent users through multiple interaction modelling gains better values in terms of fairness than a model only relying on binary associations between users and items, as *CatGCN*, which also amplifies discrimination by modelling users' local interactions.
2. Even though *RHGN* demonstrates to be a fairer model than *CatGCN*, a debiasing process is equally needed in order to exploit the user models produced by both GNNs while deeming them as fair.
3. In scenarios where the correctness of a decision on the target label w.r.t. the sensitive attributes are not well defined or where there is a high cost for misclassified instances, a complete fairness assessment should always take into account disparate mistreatment evaluation since disparate impact results could be misleading for these specific contexts.

3. FairUP Framework

To tackle the second challenge discussed in Section 1, in our most recent work [24], we developed **FairUP**, a framework for the fairness analysis of GNN-based user profiling models (we made publicly available the source code[4], the live web application[5] and a demonstration video[6]).

This framework, whose simplified architecture is shown in Figure 1, is founded on our previous fairness analysis, illustrated in Section 2. FairUP entrusts researchers and practitioners to examine the accuracy performance and fairness scores of the included models simultaneously. Specifically, it is composed of several components that allow users to:

- compute the fairness of the input dataset;
- mitigate the potential biases in the dataset by applying different well-known debiasing approaches, namely *sampling* [40], *reweighting* [40] and *disparate impact remover* [36];

[3]Source code of [23] available at https://link.erasmopurif.com/CIKM22-code.
[4]Source code of [24] available at https://link.erasmopurif.com/FairUP-source-code.
[5]Web application of [24] available at https://link.erasmopurif.com/FairUP.
[6]Demo video of [24] available at https://link.erasmopurif.com/FairUP-demo-video.

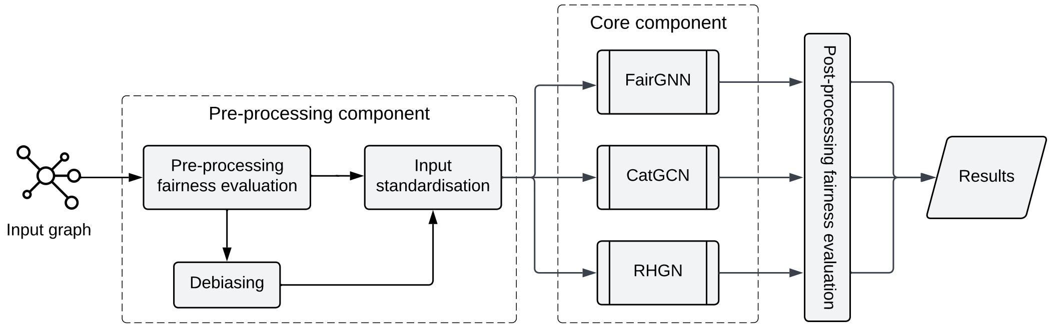

Figure 1: Logic architecture of the FairUP framework.

- train one or more GNNs, i.e. *CatGCN* [19], *RHGN* [20] and *FairGNN* [31], by specifying the parameters for each of them;
- evaluate post-hoc fairness by leveraging the four fairness metrics already adopted in our previous work [23] and described in Section 2, i.e. *statistical parity*, *equal opportunity*, *overall accuracy equality* and *treatment equality*.

In the **pre-processing component**, the main objective is to properly prepare the input data for the different models available in the framework.

The first prototype version of the framework includes the possibility for the users to select one out of four datasets from various domains. For e-commerce, there are two datasets, and they are the same adopted in [23], i.e. *Alibaba* and *JD*.

The pre-processing component itself is divided into two main modules: the optional *pre-processing fairness evaluation* (which also contains the *debiasing* methods) followed by the *input standardisation*. In particular, after evaluating the dataset fairness, if biases are found, the user can decide to apply a pre-processing debiasing approach. Three methods are supported: *sampling*, *reweighting* and *disparate impact remover*.

Sampling [40] tries to re-sample the dataset so that the discrimination is mitigated or removed. After partitioning the dataset into four groups, it computes for each class label and sensitive attribute the expected sizes if the given dataset had been non-discriminatory. Finally, a *uniform* or *preferential sampling* algorithm is applied.

Reweighting [40] attempts to mitigate bias in the dataset by setting different weights to the dataset entries. Specifically by giving the unfavourable sensitive attributes higher weights than favourable ones. To perform this, it estimates the *expected* and *observed* probability for a given sensitive attribute label and class label. If the expected probability is higher than the observed probability, there is a bias towards the opposite class label. To overcome that, lower weights are assigned to entries that are favoured.

Disparate impact remover [36] has been particularly designed to remove disparate impact bias from a dataset. This is done by revising the sensitive attribute features so that the correlation between those features and the prediction class is reduced and balanced for all prediction classes of the dataset.

Figure 2: Example of the initial page of the FairUP UI.

Since each GNN model requires the input to be structured in a specific form, the **input standardisation** module converts the original dataset as expected by the selected models. Being the framework extensible, when a new GNN is added, the related standardisation procedure can be developed.

The state-of-the-art GNN-based models included in the first release of FairUP are three: *CatGCN*, *RHGN* and *FairGNN*, and the first two have already been described in Section 2.

FairGNN [31] is a GNN framework proposed to mitigate bias in model predictions using an in-processing debiasing approach. The framework is divided into three models: a classifier (which can be either a Graph Convolutional Network (GCN) [16] or a Graph Attention Network (GAT) [13]), a sensitive attribute estimator and an adversary. This model uses a min-max game approach for mitigating unfairness between the classifier and the adversary, where the adversary tries to estimate sensitive attributes in the dataset from the predicted node representation by the classifier, and the classifier aims to learn node representations that fool the adversary into making wrong predictions.

The final **post-processing fairness evaluation** component aims to evaluate the fairness scores of the trained models by exploiting the metrics illustrated in Section 2: *statistical parity*, *equal opportunity*, *overall accuracy equality*, and *treatment equality*.

To make the framework easily accessible for any type of user, from researchers to simple practitioners or domain experts without a technical background, FairUP is equipped with an intuitive user interface (UI) through which everyone can analyse the performance of the given GNN-based user profiling modes on e-commerce datasets. An example of how the initial page of the FairUP UI is displayed in Figure 2.

4. Conclusion and Future Work

To summarize, this paper has presented an overview of the latest developments in the fairness analysis of GNN-based models for user profiling, with a specific focus on e-commerce. Our objective was to address the challenges highlighted in the introductory section. Specifically, we have discussed two of our recent works (i.e. [23, 24]) with the intention of bringing them to the attention of this community. We hope to encourage discussions on potential implications, not only from a technical perspective but also from the viewpoint of domain experts.

In future work, we will address another open challenge in the fairness analysis of most application domains, which affects e-commerce among others, which is the situation where the fairness metrics are applied in classification scenarios where both the target class and the sensitive attribute are binary, leading to an incorrect evaluation of models' biases, potentially distorting the original data conditions.

References

[1] J. B. Schafer, J. A. Konstan, J. Riedl, E-commerce recommendation applications, Data mining and knowledge discovery 5 (2001) 115–153.

[2] R. E. Bawack, S. F. Wamba, K. D. A. Carillo, S. Akter, Artificial intelligence in e-commerce: a bibliometric study and literature review, Electronic markets 32 (2022) 297–338.

[3] P. M. Alamdari, N. J. Navimipour, M. Hosseinzadeh, A. A. Safaei, A. Darwesh, A systematic study on the recommender systems in the e-commerce, Ieee Access 8 (2020) 115694–115716.

[4] Y. Gu, Z. Ding, S. Wang, D. Yin, Hierarchical user profiling for e-commerce recommender systems, in: Proceedings of the 13th International Conference on Web Search and Data Mining, 2020, pp. 223–231.

[5] Z. Wang, Z. Jiang, Z. Ren, J. Tang, D. Yin, A path-constrained framework for discriminating substitutable and complementary products in e-commerce, in: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018, pp. 619–627.

[6] G. Zhou, N. Mou, Y. Fan, Q. Pi, W. Bian, C. Zhou, X. Zhu, K. Gai, Deep interest evolution network for click-through rate prediction, in: Proceedings of the AAAI conference on artificial intelligence, volume 33, 2019, pp. 5941–5948.

[7] T. Bai, L. Zou, W. X. Zhao, P. Du, W. Liu, J.-Y. Nie, J.-R. Wen, Ctrec: A long-short demands evolution model for continuous-time recommendation, in: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2019, pp. 675–684.

[8] X. Zhao, L. Xia, L. Zhang, Z. Ding, D. Yin, J. Tang, Deep reinforcement learning for page-wise recommendations, in: Proceedings of the 12th ACM Conference on Recommender Systems, 2018, pp. 95–103.

[9] L. Zou, L. Xia, Z. Ding, J. Song, W. Liu, D. Yin, Reinforcement learning to optimize long-term user engagement in recommender systems, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 2810–2818.

[10] D. Poo, B. Chng, J.-M. Goh, A hybrid approach for user profiling, in: 36th Annual Hawaii

International Conference on System Sciences, 2003. Proceedings of the, IEEE, 2003, pp. 9–13.
[11] S. Kanoje, S. Girase, D. Mukhopadhyay, User profiling trends, techniques and applications, arXiv preprint arXiv:1503.07474 (2015).
[12] W. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, Advances in neural information processing systems 30 (2017).
[13] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, Y. Bengio, Graph attention networks, arXiv preprint arXiv:1710.10903 (2017).
[14] C. Zhang, D. Song, C. Huang, A. Swami, N. V. Chawla, Heterogeneous graph neural network, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019, pp. 793–803.
[15] T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: 5th International Conference on Learning Representations, ICLR 2017, Conference Track Proceedings, 2017.
[16] L. Yao, C. Mao, Y. Luo, Graph convolutional networks for text classification, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, 2019, pp. 7370–7377.
[17] C. Feng, Y. He, S. Wen, G. Liu, L. Wang, J. Xu, B. Zheng, Dc-gnn: Decoupled graph neural networks for improving and accelerating large-scale e-commerce retrieval, in: Companion Proceedings of the Web Conference 2022, 2022, pp. 32–40.
[18] W. Chen, Y. Gu, Z. Ren, X. He, H. Xie, T. Guo, D. Yin, Y. Zhang, Semi-supervised user profiling with heterogeneous graph attention networks, in: Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019, pp. 2116–2122.
[19] W. Chen, F. Feng, Q. Wang, X. He, C. Song, G. Ling, Y. Zhang, Catgcn: Graph convolutional networks with categorical node features, IEEE Transactions on Knowledge and Data Engineering (2021).
[20] Q. Yan, Y. Zhang, Q. Liu, S. Wu, L. Wang, Relation-aware heterogeneous graph for user profiling, in: Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Association for Computing Machinery, New York, NY, USA, 2021, pp. 3573–3577.
[21] F. Anzum, A. Z. Asha, M. L. Gavrilova, Biases, fairness, and implications of using ai in social media data mining, in: 2022 International Conference on Cyberworlds (CW), IEEE, 2022, pp. 251–254.
[22] T. Rahman, B. Surma, M. Backes, Y. Zhang, Fairwalk: towards fair graph embedding, in: Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019, pp. 3289–3295.
[23] E. Purificato, L. Boratto, E. W. De Luca, Do graph neural networks build fair user models? assessing disparate impact and mistreatment in behavioural user profiling, in: Proceedings of the 31st ACM International Conference on Information & Knowledge Management, 2022, pp. 4399–4403.
[24] M. Abdelrazek, E. Purificato, L. Boratto, E. W. De Luca, Fairup: A framework for fairness analysis of graph neural network-based user profiling models, in: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 3165–3169. doi:10.1145/3539618.3591814.

[25] S. Barocas, M. Hardt, A. Narayanan, Fairness and Machine Learning, fairmlbook.org, 2019. http://www.fairmlbook.org.
[26] S. Caton, C. Haas, Fairness in machine learning: A survey, arXiv preprint arXiv:2010.04053 (2020).
[27] E. Purificato, F. Lorenzo, F. Fallucchi, E. W. De Luca, The use of responsible artificial intelligence techniques in the context of loan approval processes, International Journal of Human–Computer Interaction (2022) 1–20.
[28] B. Edizel, F. Bonchi, S. Hajian, A. Panisson, T. Tassa, Fairecsys: mitigating algorithmic bias in recommender systems, International Journal of Data Science and Analytics 9 (2020) 197–213.
[29] Y. Li, Y. Ge, Y. Zhang, Tutorial on fairness of machine learning in recommender systems, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 2654–2657.
[30] Y. Dong, J. Kang, H. Tong, J. Li, Individual fairness for graph neural networks: A ranking based approach, in: Proceed. of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021, pp. 300–310.
[31] E. Dai, S. Wang, Say no to the discrimination: Learning fair graph neural networks with limited sensitive attribute information, in: Proceed. of the 14th ACM International Conference on Web Search and Data Mining, 2021, pp. 680–688.
[32] S. Hajian, F. Bonchi, C. Castillo, Algorithmic bias: From discrimination discovery to fairness-aware data mining, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 2125–2126.
[33] M. Wan, D. Zha, N. Liu, N. Zou, Modeling techniques for machine learning fairness: A survey, arXiv preprint arXiv:2111.03015 (2021).
[34] M. B. Zafar, I. Valera, M. Gomez Rodriguez, K. P. Gummadi, Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment, in: Proceedings of the 26th International Conference on World Wide Web, 2017, pp. 1171–1180.
[35] C. Dwork, M. Hardt, T. Pitassi, O. Reingold, R. Zemel, Fairness through awareness, in: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference, 2012, pp. 214–226.
[36] M. Feldman, S. A. Friedler, J. Moeller, C. Scheidegger, S. Venkatasubramanian, Certifying and removing disparate impact, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 259–268.
[37] M. Hardt, E. Price, N. Srebro, Equality of opportunity in supervised learning, Advances in neural information processing systems 29 (2016).
[38] R. Berk, H. Heidari, S. Jabbari, M. Kearns, A. Roth, Fairness in criminal justice risk assessments: The state of the art, Sociological Methods & Research 50 (2021) 3–44.
[39] A. Beutel, J. Chen, Z. Zhao, E. H. Chi, Data decisions and theoretical implications when adversarially learning fair representations, arXiv preprint arXiv:1707.00075 (2017).
[40] F. Kamiran, T. Calders, Data preprocessing techniques for classification without discrimination, Knowledge and information systems 33 (2012) 1–33.