Applying Fairness Constraints on Graph Node Ranks Under Personalization Bias

Emmanouil Krasanakis(✉), Symeon Papadopoulos, and Ioannis Kompatsiaris

CERTH-ITI, 57001 Thessaloniki, Greece
{maniospas,papadop,ikom}@iti.gr

Abstract. In this work we address algorithmic fairness concerns that arise when graph nodes are ranked based on their structural relatedness to a personalized set of query nodes. In particular, we aim to mitigate disparate impact, i.e. the difference in average rank between nodes of a sensitive attribute compared to the rest, while also preserving node rank quality. To do this, we introduce a personalization editing mechanism that helps ranking algorithms achieve different trade-offs between fairness constraints and rank changes. In experiments across four real-world social graphs and two base ranking algorithms, our approach outperforms baseline and existing methods in uniformly mitigating disparate impact, even when personalization suffers from extreme bias. In particular, it achieves better trade-offs between fairness and node rank quality under disparate impact constraints.

Keywords: Node ranking · Personalized ranking · Algorithmic fairness · Disparate impact mitigation

1 Introduction

Machine learning has been widely adopted in systems that affect important aspects of people's lives, from recommending social media friends to assisting jurisdictional or employment decisions. Since these systems often learn to replicate human-generated and systemic biases, fairness concerns arise when automated decisions end up correlated to sensitive attributes, such as gender or ethnicity [1,2]. Fairness is commonly defined as similar assessment between sensitive and non-sensitive groups of data samples under a statistical measure [1,3–5]. In this work, we focus on disparate impact elimination [6–9], which requires (approximate) statistical parity between sensitive and non-sensitive positive predictions.

Node ranking refers to a class of methods that organize relational data into graphs and score the structural relatedness of their nodes to a set of query ones. This process can be *personalized*, in the sense that query nodes share an attribute, such as their political views, in which case scores (ranks) can be used as estimators for that attribute [10–12]. If no personalization takes place and all nodes become queries, ranks reflect their structural importance [13,14].

R. M. Benito et al. (Eds.): COMPLEX NETWORKS 2020, SCI 944, pp. 610–622, 2021.
https://doi.org/10.1007/978-3-030-65351-4_49

Although graph node ranking is an important machine learning discipline, remarkably little work has been done to make it fair. In fact, the first -to our knowledge- principled understanding of node rank fairness was only recently proposed by Tsioutsiouliklis et al. [15], who explore disparate impact mitigation for the node ranks of Google's PageRank algorithm [16]. We now initiate a discussion on the fairness of *personalized* node ranking algorithms. Contrary to the non-personalized case, where node rank quality is tied to ad hoc definitions of structural importance, there exist objective notions of personalized node rank quality that fairness-aware approaches should ideally respect. For example, ranking social graph nodes to recommend friends should aim for relevance yet its outcome should not be influenced by sensitive attributes, such as race.

In this work we leverage a model that weights training samples to make classifiers fair [3] and adapt it to estimate an unbiased personalization that yields fairer node ranks. Our new adaptation can be trained towards a variety of fairness objectives, such as fully or partially eliminating disparate impact while minimizing rank edits. We demonstrate its effectiveness by comparing it to baseline and existing practices across two ranking algorithms and four real-world graphs with both unbiased and extremely biased personalization.

Our contribution lies in initiating a discussion on fairness-aware personalized ranking algorithms, where we address the trade-off between biased personalization and the preservation of prediction-related node rank quality. Furthermore, we investigate whether approaches uniformly introduce fairness in the sense that they do so for both the whole graph and an evaluation subset of nodes.

2 Background

2.1 Personalized Node Ranking Algorithms

Personalized node ranking starts from a set of query nodes sharing an attribute of interest and scores nodes v per some notion of structural proximity to query ones. We organize node scores, which are called *ranks* and not to be confused with ordinalities, into vectors r with elements $r[v] \geq 0$. We similarly consider a user-provided personalization vector p with elements $p[v] \in [0, 1]$ reflecting the importance of nodes v being used as queries (0 corresponds to non-query nodes).

Ranking algorithms are often expressed as graph filters [17,18]. These use a normalization W of the graph's adjacency matrix, whose elements $W[u, v]$ define transitions from nodes u to v. Then, given that propagating the personalization n hops away can be written as $W^n p$, they weight different propagation distances:

$$r = H(W)p \qquad H(W) = \sum_{n=0}^{\infty} h_n W^n \tag{1}$$

where $H(W)$ is called a graph filter. Different filters can be obtained for different weights h_n and methods of calculating W. For example, the graph's adjacency matrix M can be normalized column-wise $W = MD^{-1}$ or symmetrically $D^{-\frac{1}{2}}MD^{-\frac{1}{2}}$, where $D = diag\big(\big[\sum_u M[v, u]\big]_v\big)$ is the diagonal table of node degrees. Two well-known graph filters are Personalized PageRank [19,20] and Heat Kernels [21], which respectively arise from hop weights $h_n = (1-a)a^n$ and $h_n = e^{-t}t^n/n!$ for parameters $a \in [0,1]$ and $t \in \{1,2,3,\dots\}$.

2.2 Sweeping Node Ranks

The sweep procedure [22,23] utilizes node ranking algorithms to identify congregations of nodes that are tightly-knit together and well-separated from the rest of the graph, a concept known as low subgraph conductance [24]. This procedure assumes that a base ranking algorithm R with strong locality [25], such as personalized PageRank and Heat Kernels, yields ranks $R(p)$ for a personalization p that comprises structurally close query nodes. It then compares ranks with their non-personalized counterparts $R(\mathbf{1})$, where $\mathbf{1}$ is a vector of ones:

$$r_{sweep} = \frac{R(p)[v]}{R(\mathbf{1})[v]} \tag{2}$$

From now on, we will refer to this post-processing as the *sweep ratio*.

The sweep procedure orders all nodes based on their sweep ratio and cuts the graph into two partitions so that conductance is minimized. This practice statistically yields well-separated partitions for a variety of node ranking algorithms [22–24]. From a high-level perspective, this indicates that the sweep ratio tends to improve node rank quality.

2.3 Algorithmic Fairness and Graph Mining

Algorithmic fairness is broadly understood as parity between sensitive and non-sensitive group samples over a chosen statistical property. Three popular fairness-aware objectives [1,3–5] are disparate treatment elimination, disparate impact elimination and disparate mistreatment elimination. These correspond to not using the sensitive attribute in predictions, preserving statistical parity between the fraction of sensitive and non-sensitive positive labels and achieving identical predictive performance on the two groups under a measure of choice.

In this work, we focus on mitigating disparate impact unfairness [1,6–9]. An established measure that quantifies this fairness objective is the *pRule* [6]; denoting as $R[v]$ the binary outputs of a system R for samples v, S the sensitive group, S' the non-sensitive group (which is the complement of S) and $P(a|b)$ the probability of a conditioned on b, this is defined as:

$$pRule = \frac{\min(p_S, p_{S'})}{\max(p_S, p_{S'})} \in [0,1] \qquad \begin{array}{l} p_S = P(R[v]=1|p \in S) \\ p_{S'} = P(R[v]=1|p \notin S) \end{array} \tag{3}$$

The higher the pRule, the fairer a system is. There is precedence [6] for considering 80% pRule or higher as fair. Calders-Verwer disparity $|p_S - p_{S'}|$ [7] is a correlated measure optimized at the same point, but is less descriptive in that it biases fairness assessment against high fractions of positive predictions.

In domains related to ranking, fairness has been defined for the order of recommended items [26–29] as equity in the ranking positions between sensitive and non-sensitive items. However, these notions of fairness are not applicable to the more granular understanding provided by node ranks.

In graphs, the notion of achieving fair node embeddings has been proposed [30,31]. These are the first approaches that introduce fair random walks, a stochastic process modeled by personalized PageRank. However, the fairness of these walks is only implicitly asserted through embedding fairness. A more advanced understanding has been achieved recently in the more general domain of graph neural networks [32], which can be trained to produce fair recommendations, even under partial knowledge of the sensitive attribute.

Last, a recent work by Tsioutsiouliklis et al. [15] has initiated a discourse on node rank fairness. Although focused on non-personalized ranking, it first recognizes the need of optimizing a trade-off between fairness and preserving rank quality. Furthermore, it provides a first definition of node rank fairness, called ϕ-fairness. Under a stochastic interpretation of node ranks, where they are proportional to the probability of nodes assuming positive labels, ϕ-fairness becomes equivalent to disparate impact elimination when $\phi = \frac{|S|}{|S|+|S'|}$.

In this work we consider the similar objectives of a) trading-off deviation from the original ranks and high pRule and b) preserving rank quality under fairness constraints. The pRule is calculated according to the above-mentioned stochastic interpretation of ranks through:

$$
\begin{aligned}
p_S &= P(R[v]=1|p\in S) = \tfrac{1}{|S|}\sum_{v\in S} L_\infty(r)[v] \\
p_{S'} &= P(R[v]=1|p\notin S) = \tfrac{1}{|S'|}\sum_{v\notin S} L_\infty(r)[v]
\end{aligned}
\tag{4}
$$

where $L_\infty(r)$ is a normalization that divides ranks with their maximum value and R is a stochastic process with probability $P(R[v]=1) = \frac{r[v]}{\max_u r[u]} = L_\infty(r)[v]$.

2.4 The CULEP Model

In previous work [3], we tackled the problem of making black box calibrated binary classifiers fair by pre-processing training data. To this end, we proposed a Convex Underlying Error Permutation (CULEP) model that weighs the importance of training samples to treat unfairness similarly to how an ideal but unobserved distribution of fair training labels would. To do this, we theorized that unfairness correlates to misclassification error (i.e. the difference between binary classification labels and calibration probabilities) and whether samples are sensitive. Furthermore, we recognized that strongly misclassified samples could exhibit different degrees of bias from correctly classified ones.

Under these considerations the CULEP model tries to promote fairness by introducing a type of parameterized balancing between these sources of unfairness. In particular, after a stochastic analysis similar to the one we will later conduct in this work, training samples i with misclassification error err_i are assigned weights proportional to $\alpha_i E_{\beta_i}(err_i) + (1-\alpha_i)E_{\beta_i}(-err_i)$ where the values of $\alpha_i \in [0,1], \beta_i \geq 0$ depend only on whether samples are sensitive or not and $E_\beta(\cdot)$ is a function asymmetric around 0, such as the exponential $E_\beta(x) = e^{\beta x}$. These parameters can be tuned to satisfy various fairness objectives, including trade-offs between preserving accuracy (when $\beta_i = 0$) and improving the pRule (when β_i are large enough and α_i balance towards mitigating positive label disparity).

3 Our Approach: Fair Personalizer

We theorize that there exist two types of potential node rank bias: stationary and rank-related. The first arises when ranks end up multiplied with a fixed bias-related quantity for each node. Whereas the second depends on the personalization, which transfers either its own or graph edge bias to the ranks. Of the two, stationary bias is easier to treat, as it does not depend on the personalization and only attacks the ranking algorithm's outcome. In fact, the sweep ratio eliminates it, as it ends up dividing node ranks with their bias term.

On the other hand, rank-related bias is harder to tackle. To see why, let us consider an invertible graph filter, such as the closed form of personalized PageRank $H(W) = (1-a)(I-aW)^{-1}$, and a personalization vector p that yields ranks $r = H(W)p$. We assume that there exist ideal ranks r_{fair} satisfying a fairness-aware objective, such as minimizing the following trade-off between preserving ranks and improving the pRule with weight w_{pRule} up to sup_{pRule}:

$$\text{minimize } \frac{1}{|V|}\|L_\infty(r_{fair}) - L_\infty(r)\|_1 - w_{pRule} \cdot \min\{pRule(r_{fair}), sup_{pRule}\} \quad (5)$$

where $pRule(r_{fair})$ calculates the pRule of those ranks across all graph nodes V and $\|\cdot\|_1$ is the L_1 norm that sums the absolute values of vector elements. Then, the graph's structure (e.g. edge sparseness) may cause $H(W)$ to be near-singular and hence propagate back small fair rank changes as large differences between the original personalization and its fair counterpart $p_{fair} = H^{-1}(W)r_{fair}$.

Setting aside the potential intractability of optimally editing the personalization, we argue that this practice should be preferred to postprocessing ranks, as it respects underlying structural characteristics exposed when the graph filter diffuses the personalization through edges. To keep this upside, we propose that searching for fairness-inducing personalization edits can be made easier if these are expressed through parametric models of only few parameters to learn.

The CULEP model could in theory fit this role, since it depends on only four parameters (α_i and β_i can each assume only two values, depending on whether i are sensitive). However, it can not be ported to graph ranking as-is, since weighting zero elements of the personalization vector through multiplication does not affect ranks at all and there is no rank validation set on which to tune its parameters. To address these issues, we adapt this model to perform non-linear edits on the penalization vector and use the original personalization as a rough one-class validation set of known positive examples.

We start from a stochastic interpretation of ranks that snaps them to 1 with probability $P(\cdot)$ proportional to their value and to 0 otherwise. We also consider an edited vector personalization p_{est} that estimates ranks $r_{est} = H(W)p_{est}$ of similar fairness to some unobserved ideal ones r_{fair}. For ease of notation, in the rest of this section we consider all vector operations (including multiplication) to be applied element-wise.

We first analyse whether estimated ranks match the ideal fair ones:

$$P(r_{fair} = r_{est}) = P(r_{fair} = r_{est}|p = r_{est})P(p = r_{est}) \\ + P(r_{fair} = r_{est}|p \neq r_{est})P(p \neq r_{est})$$

Borrowing CULEP's theorization, the probabilities of estimated node ranks being fair given that they approximate well the original personalization are correlated with the probability of personalization being fair $P(p_{fair} = p_{est})$ and the same holds true given that estimated node ranks do *not* approximate well the original personalization. Furthermore, if one of these two types of probabilities becomes larger for a node the other should become smaller and conversely. Finally, we consider an exponential-based approximation (whose ability to achieve fairness has been experimentally demonstrated [3]) of how these types of probabilities differ from their correlated fair personalization. This approximation depends on rank and personalization differences and whether nodes are sensitive:

$$P(r_{fair} = r_{est}|p = r_{est}) \approx K\, P(p_{fair} = p_{est})\, e^{-b(L_\infty(r)-p)} \\ P(r_{fair} = r_{est}|p \neq r_{est}) \approx K\, P(p_{fair} = p_{est})\, e^{b(L_\infty(r)-p)}$$

where b is a vector of real values such that $b[v] = \{b_S \text{ if } v \in S, b_{S'} \text{ otherwise}\}$ and $K > 0$ is a normalization constant that makes probabilities sum to 1.

We further assume that selecting sensitive and non-sensitive nodes as part of the personalization is done with fixed probabilities a_S and $a_{S'}$ pertaining to the personalization bias and organize those into a vector $a = P(p = r_{est})$ with elements $a[v] = \{a_S \text{ if } v \in S, a_{S'} \text{ otherwise}\} \in [0, 1]$.

Given the above analysis, we select a fair personalization estimation p_{fair} based on the self-consistency criterion that, when it approaches fairness-inducing personalization, estimated fair ranks should also approach the ideal fair ones:

$$p_{est} = P(r_{fair} = r_{est}|p_{fair} = p_{est}) \approx \frac{P(r_{fair} = r_{est})}{P(p_{fair} = p_{est})} \\ \propto a e^{-b(L_\infty(r)-p)} + (1-a)e^{b(L_\infty(r)-p)} \tag{6}$$

4 Experiment Setup

4.1 Graphs

To assess whether our approach can achieve fairness while preserving node rank quality, we experiment on four graphs: two Facebook friendship graphs [33], one Twitter graph of political retweets [34] and one Amazon graph of frequent product co-purchases [35]. The first three comprise real-world fairness sensitive attributes, but not adequately many nodes and edges to calculate ranks of high quality. On the other hand, the Amazon graph is not annotated with fairness-related attributes but is large enough for ranking algorithms to boast high quality, which our approach aims to maintain.

The Facebook graphs each start from a given user and record social relations between them and their friends, including relations between friends. Ten such graphs are available in the source material, out of which we randomly select two to experiment on. These are denoted as FacebookX, where X is their starting user. We select the anonymized binary 'gender' attribute as sensitive and the first anonymized binary 'education' attribute as the prediction label. The Twitter graph comprises only one anonymized sensitive attribute of binary political opinions (left or right). The Amazon graph does not contain sensitive information and we consider the product category 'Book' to be sensitive. Due to lack of predictive attributes for the Twitter and Amazon graphs, we define predictions for the sensitive attribute's binary complement, which makes those graphs exhibit what we later dub as extreme unfairness.

These graphs are overviewed in Table 1. Columns correspond to graph names, number of nodes, number of edges, fraction of nodes with positive labels, number of nodes designated as sensitive and *pRule* value of their positive labels.

Table 1. Experiment graph characteristics

Graph	Nodes	Edges	Positive%	Sensitive%	pRule
Facebook0	347	5,038	68%	36%	.91
Facebook686	170	3,312	55%	46%	.91
Twitter	18,470	48,365	61%	39%	0
Amazon	334,863	925,872	>99%	<1%	0

4.2 Compared Methods

In our experiments we investigate methods that bring fairness to personalized PageRank and Heat Kernels. These are run with parameters $a = .99$ and $t = 5$ for the Amazon graph to diffuse the personalization many hops away [36] and the frequently used $a = .85$ and $t = 3$ for the graphs with fewer nodes. In all cases we deploy symmetric normalization, which a preliminary investigation revealed to yield higher AUC (see Subsection 4.3). Ranks were computed to a numerical precision of 10^{-9} using the *pygrank*[1] graph ranking library. We compare the following fairness-aware schemes on the two base node ranking algorithms:

None. The base ranking algorithm.

[1] https://pypi.org/project/pygrank/.

Mult [baseline]. A simple post-processing baseline that multiplies ranks across the sensitive and non-sensitive groups with a different constant each, so that disparate impact is fully mitigated. If r are the base ranking algorithm's node ranks, this method yields ranks:

$$r_{Mult}[v] = \left(\frac{\phi s[v]}{\sum_{u\in S} s[u]r[u]} + \frac{(1-\phi)(1-s[v])}{\sum_{u\notin S} s[u]r[u]}\right) r[v]$$

where $\phi = \frac{|S|}{|S|+|S'|}$ is the fraction of graph nodes that are sensitive and $s[u] = \{1 \text{ if } u \in S, 0 \text{ otherwise}\}$. It holds that $\sum_{v\in S} r_{Mult}[v] = \sum_{v\notin S} r_{Mult}[v]$.
FairWalk [31]. The graph filter equivalent of random walks previously used for fair node embeddings, in which the adjacency matrix is adjusted to yield the same total weight for hopping to sensitive and non-sensitive neighbors.
LFPRO [15]. Near-optimal redistribution of ranks causing disparate impact.
Sweep [22,23]. Postprocessing node ranks with the sweep ratio of Eq. 2.
SLFPRO [baseline]. Applying LFPRO on the outcome of Sweep.
FP [this work]. The model of Eq. 6 whose parameters $a_S, a_{S'} \in [0,1]$ and exponentials $b_S, b_{S'} \in [-10, 10]$ are trained with the *pygrank* library's coordinate descent optimization towards Eq. 5 for $w_{pRule} = 1$ and $sup_{pRule} = 1$.
CFP [this work]. Constraining the FP model to prioritize improving the pRule but only up to 80% by optimizing Eq. 5 for $w_{pRule} = 10$ and $sup_{pRule} = .8$.
SweepFP [this work]. Applying FP on the outcome of Sweep.
SweepCFP [this work]. Applying CFP on the outcome of Sweep.

4.3 Evaluation

To compare the different fairness-aware methods, we randomly split graph nodes into training and evaluation sets, where the former comprise a fraction among $[10\%, 20\%, 30\%]$ of graph nodes, uniformly sampled without repetition. This mimics real-world usage of node ranking algorithms, where not many labels are known. For each split fraction we sample one training set for the Amazon and 5 for the other graphs (we pass the same sets to each ranking algorithm) and average the following measures across the respective evaluation sets:
AUC. The area under curve of the receiver operating characteristics [37], which is often used to measure the quality of rank-based recommendations given known binary labels. 50% AUC indicates random node ranks, whereas 100% AUC perfect rank quality. We stress that fairness-aware methods are tasked with preserving but not improving potentially low node rank quality.
WR. We propose this novel measure of fairness that captures the worst pRule between the ranks of all graph nodes and the ranks of evaluation nodes. Our motivation is that some fairness-aware algorithms are designed to yield perfect disparate impact elimination (i.e. 100% pRule) when considering all graph nodes, but subsets of nodes should also exhibit increased fairness. For example, if a method achieves 100% and 1% pRule on all graph and evaluation nodes respectively, it should not be considered fair. For WR to accurately assess whether

disparate impact treatment are uniformly spread across the graph, we avoid directly optimizing towards the pRule of evaluation nodes.

We also consider cases of biased personalization, in which sensitive nodes are underrepresented. This is probable to occur when sensitive nodes are disproportionately few or when query node selection is biased against them, for example due to people with sensitive attributes being reluctant to share information [38]. To simulate this behavior, we repeat experiments with extreme personalization bias, in which sensitive nodes are removed from the personalization.

5 Experiments

We first conduct experiments under unbiased personalization, where training nodes are uniformly sampled. In Table 2 we detail the outcome of applying fairness-aware schemes on the personalized PageRank and Heat Kernel algorithms. We omit results for the Twitter and Amazon graphs, which by definition follow the extreme personalization covered in subsequent experiments.

In this first series of experiments, the high pRule of Facebook graph labels is transferred through uniform sampling to the personalization and lets base ranking algorithms comfortably exceed 80% pRule. Nevertheless, no method yields perfect fairness for both all nodes and their evaluation subset. As can be seen from WR assessments, FP and SweepFP dominate other approaches in satisfying both fairness terms. SweepFP also maintains equal or better AUC compared to base ranking algorithms. CFP and SweepCFP do not improve fairness as much, since their 80% pRule constraint is already satisfied, yet yield similar or higher rank quality and fairness compared to base algorithms and non-FP approaches.

Table 2. Experiments for unbiased personalization

	Personalized PageRank				Heat Kernels			
	Facebook0		Facebook686		Facebook0		Facebook686	
	AUC	WR	AUC	WR	AUC	WR	AUC	WR
None	.54	.90	.55	.92	.53	.85	.56	.83
Mult	.53	.95	.55	.94	.53	.89	.55	.85
FairWalk	.52	.79	.55	.95	.52	.77	.55	.85
LFPRO	.53	.94	.55	.92	.52	.81	.55	.74
Sweep	.55	.92	.58	.94	.54	.86	.58	.80
SweepLFPRO	.54	.94	.57	.93	.53	.81	.57	.77
FP	.50	.94	.53	.96	.49	.95	.51	.96
CFP	.53	.92	.53	.91	.52	.89	.51	.92
SweepFP	.56	.94	.55	.95	.56	.95	.54	.92
SweepCFP	.56	.95	.55	.91	.57	.88	.54	.88

We now experiment with extreme personalization bias, where there is no sensitive query node. Extreme bias ends up being too unfair for base ranking algorithms to reach 80% WR. Furthermore, only methods that involve personalization editing assist them in consistently doing so. Although these sometimes significantly reduce node rank quality, they avoid settling for minimal rank changes that fail to reach meaningful levels of fairness, as other approaches often do.

Significant node rank quality reductions for personalization editing occur only on the Twitter and Amazon graphs. In these, sensitive and predictive labels are complementary and, from a classification viewpoint, improving fairness comes at the direct cost of erroneously identifying sensitive nodes as positive. However, in Amazon graph the granularity provided by ranks sometimes suffers less from this problem for SweepFP and SweepCFP; these affect ranks just enough to improve fairness but not so much that AUC is impacted.

An interesting finding is that the sweep ratio detrimentally affects FP and CFP on the Twitter graph. This indicates that its success in aiding fairness-aware approaches in other cases can be attributed to higher quality node ranks proving more leeway for personalization editing to improve fairness trade-offs.

Table 3. Experiments for extreme personalization bias

	Personalized PageRank								Heat Kernels							
	Facebook0		Face/k686		Twitter		Amazon		Facebook0		Face/k686		Twitter		Amazon	
	AUC	WR	AUC	WR	AUC	WR	AUC	WR	AUC	WR	AUC	WR	AUC	WR	AUC	WR
None	.53	.69	.54	.74	.58	0	.96	.39	.54	.37	.55	.39	.58	0	.92	.08
Mult	.51	.75	.52	.73	.49	.25	.50	.98	.50	.40	.52	.38	.56	.11	.67	.63
FairWalk	.52	.78	.54	.78	.57	.03	1	.17	.52	.49	.55	.42	.58	.02	.93	.06
LFPRO	.51	.75	.52	.70	.54	.53	.42	.98	.42	.48	.50	.48	.57	.53	.16	.65
Sweep	.55	.67	.56	.74	.58	0	1	.34	.54	.35	.56	.38	.58	0	.93	.07
SLFPRO	.51	.72	.52	.68	.55	.52	.48	.98	.44	.44	.49	.43	.58	.53	.13	.64
FP	.53	.95	.52	.93	.49	.93	.50	.99	.47	.81	.52	.90	.44	.96	.49	.98
CFP	.52	.90	.52	.82	.53	.80	.67	.80	.49	.82	.52	.81	.45	.80	.69	.80
SweepFP	.54	.91	.54	.92	.27	.96	.52	1	.52	.78	.52	.80	.38	.94	.76	.77
SweepCFP	.54	.88	.54	.84	.43	.80	.99	.80	.53	.80	.54	.83	.48	.81	.85	.80

Across all experiments, FairWalk yields fairer ranks compared to both Mult, which perfectly balances ranks of all nodes, and LFPRO, despite the latter being known to achieve balance between rank retention and fairness results in non-personalized settings [15]. This corroborates our assumption that non-personalized fairness when considering all graph nodes does not necessarily carry over to personalization and uniform notions of fairness.

Overall, SweepFP achieves similar or better levels of uniform disparate impact mitigation and node rank quality trade-offs compared to other methods and is only sometimes outperformed by other FP-based methods and only under extreme unfairness. We hence suggest using this method when disparate impact mitigation is the most important objective of node ranking algorithms.

On the other hand, SweepCFP and CFP always achieve their objective of reaching 80% WR. From this we surmise that they successfully prevent overfitting towards non-uniform notions of fairness and should be preferred when ranking needs to satisfy only a predetermined fairness level. SweepCFP further maintains node rank quality close to the base algorithm under most settings.

The broader success of personalization editing approaches can be attributed to their pre-processing nature, which addresses the catastrophic effects of bias before being entangled with produced ranks through complex network dynamics.

6 Conclusions and Future Work

In this work we tackled the problem of mitigating disparate impact while preserving the quality of graph node ranks and explored personalization editing as a means to do so. Our approach derives a personalization editing mechanism whose parameters can be adjusted to trade-off rank preservation and fairness objectives. Experimenting on two ranking algorithms and four real-world social graphs, we found that, when combined with the rank post-processing of the sweep procedure, our approach significantly outperforms existing and baseline methods in uniformly mitigating bias across ranks while in large part preserving their quality, even under cases of extreme unfairness.

For future work, we are interested in exploring the efficacy of our approach on more graphs and node ranking algorithms. We are especially interested in graphs of real-world sensitive attributes where ranking exhibits high predictive capabilities and algorithms that are not graph filters. Furthermore, the FP model or an adjustment could be used to mitigate other types of unfairness, such as disparate mistreatment, or do so under partial knowledge of sensitive attributes.

Acknowledgements. This work was partially funded by the European Commission under contract numbers H2020-860630 NoBIAS and H2020-825585 HELIOS.

References

1. Chouldechova, A.: Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. Big data **5**(2), 153–163 (2017)
2. Kleinberg, J., Ludwig, J., Mullainathan, S., Rambachan, A.: Algorithmic fairness. AEA Papers Proc. **108**, 22–27 (2018)
3. Krasanakis, E., Spyromitros-Xioufis, E., Papadopoulos, S., Kompatsiaris, Y.: Adaptive sensitive reweighting to mitigate bias in fairness-aware classification. In: Proceedings of the 2018 World Wide Web Conference, pp. 853–862 (2018)
4. Zafar, M.B., Valera, I., Gomez-Rodriguez, M., Gummadi, K.P.: Fairness constraints: a flexible approach for fair classification. J. Mach. Learn. Res. **20**(75), 1–42 (2019)
5. Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., et al.: Bias in data-driven artificial intelligence systems-an introductory survey. Wiley Interdiscip. Rev.: Data Min. Knowl. Disc. **10**(3), e1356 (2020)

6. Biddle, D.: Adverse Impact and Test Validation: A Practitioner's Guide To Valid and Defensible Employment Testing. Gower Publishing, Ltd. (2006)
7. Calders, T., Verwer, S.: Three Naive Bayes approaches for discrimination-free classification. Data Min. Knowl. Disc. **21**(2), 277–292 (2010)
8. Kamiran, F., Calders, T.: Data preprocessing techniques for classification without discrimination. Knowl. Inform. Syst. **33**(1), 1–33 (2012)
9. Feldman, M., Friedler, S.A., Moeller, J., Scheidegger, C., Venkatasubramanian, S.: Certifying and removing disparate impact. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 259–268 (2015)
10. Haveliwala, T.H.: Topic-sensitive pagerank: a context-sensitive ranking algorithm for web search. IEEE Trans. Knowl. Data Eng. **15**(4), 784–796 (2003)
11. Xiang, L., Yuan, Q. Zhao, S., Chen, L., Zhang, X., Yang, Q., Sun, J.: Temporal recommendation on graphs via long-and short-term preference fusion. In: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 723–732 (2010)
12. Yang, J., Leskovec, J.: Defining and evaluating network communities based on ground-truth. Knowl. Inform. Syst. **42**(1), 181–213 (2015)
13. Ding, Y., Yan, E., Frazho, A., Caverlee, J.: Pagerank for ranking authors in co-citation networks. J. Am. Soc. Inform. Sci. Technol. **60**(11), 2229–2243 (2009)
14. Heidemann, J., Klier, M., Probst, F.: Identifying key users in online social networks: a pagerank based approach (2010)
15. Tsioutsiouliklis, S., Pitoura, E., Tsaparas, P., Kleftakis, I., Mamoulis, N.: Fairness-aware link analysis (2020). arXiv preprint: arXiv:2005.14431
16. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Technical report. Stanford InfoLab (1999)
17. Segarra, S., Marques, A.G., Ribeiro, A.: Optimal graph-filter design and applications to distributed linear network operators. IEEE Trans. Signal Process. **65**(15), 4117–4131 (2017)
18. Ortega, A., Frossard, P., Kovačević, J., Moura, J.M., Vandergheynst, P.: Graph signal processing: overview, challenges, and applications. Proc. IEEE **106**(5), 808–828 (2018)
19. Andersen, R., Chung, F., Lang, K.: Local partitioning for directed graphs using pagerank. In: Bonato, A., Chung, F.R.K. (eds.) Algorithms and Models for the Web-Graph: WAW 2007. Lecture Notes in Computer Science, vol. 4863, pp. 166–178. Springer, Berlin, Heidelberg (2007). https://doi.org/10.1007/978-3-540-77004-6_13
20. Bahmani, B., Chowdhury, A., Goel, A.: Fast incremental and personalized pagerank. In: Proceedings of the VLDB Endowment, 4(3) (2010)
21. Kloster, K., Gleich, D.F.: Heat kernel based community detection. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1386–1395 (2014)
22. Andersen, R., Chung, F., Lang, K.: Local graph partitioning using pagerank vectors. In: 2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), pp. 475–486. IEEE (2006)
23. Andersen, R., Chung, F., Lang, K.: Using pagerank to locally partition a graph. Internet Math. **4**(1), 35–64 (2007)
24. Chalupa, D.: A memetic algorithm for the minimum conductance graph partitioning problem (2017). arXiv preprint: arXiv:1704.02854

25. Wu, X.-M., Li, Z., So, A.M., Wright, J., Chang, S.-F.: Learning with partially absorbing random walks. In: Advances in Neural Information Processing Systems, pp. 3077–3085 (2012)
26. Beutel, A., Chen, J., Doshi, T., Qian, H., Wei, L., Wu, Y., Heldt, L., Zhao, Z., Hong, L., Chi, E.H., et al.: Fairness in recommendation ranking through pairwise comparisons. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2212–2220 (2019)
27. Biega, A.J., Gummadi, K.P., Weikum, G.: Equity of attention: amortizing individual fairness in rankings. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 405–414 (2018)
28. Yang, K., Stoyanovich, J.: Measuring fairness in ranked outputs. In: Proceedings of the 29th International Conference on Scientific and Statistical Database Management, pp. 1–6 (2017)
29. Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., Baeza-Yates, R.: Fa* ir: a fair top-k ranking algorithm. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 1569–1578 (2017)
30. Bose, A., Hamilton, W.: Compositional fairness constraints for graph embeddings. In: International Conference on Machine Learning, pp. 715–724 (2019)
31. Rahman, T.A., Surma, B., Backes, M., Zhang, Y.: Fairwalk: towards fair graph embedding. In: IJCAI, pp. 3289–3295 (2019)
32. Dai, E., Wang, S.: FairGNN: Eliminating the discrimination in graph neural networks with limited sensitive attribute information. arXiv preprint: arXiv:2009.01454 (2020)
33. J. Leskovec and J. J. Mcauley, "Learning to discover social circles in ego networks," in *Advances in neural information processing systems*, pp. 539–547, 2012
34. Rossi, R.A., Ahmed, N.K.: The network data repository with interactive graph analytics and visualization. In: AAAI (2015)
35. Leskovec, J., Adamic, L.A., Huberman, B.A.: The dynamics of viral marketing. ACM Trans. Web (TWEB) **1**(1), 5 (2007)
36. Krasanakis, E., Schinas, E., Papadopoulos, S., Kompatsiaris, Y., Symeonidis, A.: Boosted seed oversampling for local community ranking. Inform. Process. Manag. **57**(2), 102053 (2020)
37. Bradley, A.P.: The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn. **30**(7), 1145–1159 (1997)
38. Cassel, C., Bindman, A.: Risk, benefit, and fairness in a big data world. JAMA **322**(2), 105–106 (2019)