Judiciously Reducing Sub-group Comparisons for Learning Intersectional Fair Representations

Clara Rus[1(✉)], Andrew Yates[2], and Maarten de Rijke[1]

[1] University of Amsterdam, Amsterdam, The Netherlands
c.a.rus@uva.nl, m.derijke@uva.nl
[2] Johns Hopkins University, Baltimore, MD, USA
andrew.yates@jhu.edu

Abstract. Ensuring fairness in ranking systems is critical to avoid discriminatory outcomes towards minority groups in high stakes domains such as recruitment. Most fairness interventions only address fairness for one or more binary groups without accounting for intersectional fairness. We study the problem of achieving intersectional fairness in ranking systems, where individuals may face compounded disadvantages. We adapt and extend pre-processing fairness intervention methods to optimize for intersectional group fairness. As the number of intersectional sub-groups grows exponentially with the number of attributes, optimization becomes computationally expensive. We propose to reduce the number of sub-group comparisons when optimizing for intersectional fairness, based on the highest disparities between sub-groups. We show that limiting sub-group comparisons achieves comparable or better intersectional fairness. We validate this on three real-world datasets and a simulated setup designed to test robustness to intersectional fairness challenges.

Keywords: Intersectional · Fairness · Ranking

1 Introduction

Ranking systems influence what a user is exposed to and the likelihood of an item being selected by the user. When the items being ranked are people (e.g., on a recruitment platform), one should be sensitive to the potential biases present in a ranking system that could lead to negative outcomes for minority groups. In high-risk domains this can have real-life negative implications, e.g., unequal medical treatment [22], denial of parole [7] or of access to high paying jobs [28].

Work on fairness often focuses on achieving fairness towards one binary attribute [31,38,39,42]. Some interventions are designed to account for multiple attributes (e.g., gender and race) [5,20,40]. But even with fairness towards multiple attributes, a system might not satisfy intersectional fairness [36]. Intersectional discrimination arises when an individual faces discrimination due to their membership in multiple groups at the same time. E.g., Black women in the legal system may face challenges that neither Black men nor white women experience, as their discrimination is shaped by both their race and gender [9].

R. Campos et al. (Eds.): ECIR 2026, LNCS 16485, pp. 319–334, 2026.
https://doi.org/10.1007/978-3-032-21324-2_26

Intersectional Fairness. From a computer science perspective, intersectional fairness can be described as follows: the system should be fair towards all sub-groups defined by the intersection of the sensitive attributes. But intersectionality should not be reduced merely to achieving fairness for sub-groups [18]. It is equally important to consider the societal context that gives rise to the unique inequalities faced by individuals subject to intersectional discrimination. Pre-processing fairness interventions account for this context by addressing societal biases embedded in the data and correcting for them, therefore, we operationalize our approach using this class of methods. Addressing biases in data aligns with EU and US non-discrimination law requirements [19,30], especially the EU AI Act,[1] which includes obligations about investigating and correcting biases in data [1]. Moreover, pre-processing methods are model-agnostic, making them practical for integration in real-world systems.

Challenges. Optimizing for intersectional fairness poses several [15,34]: (i) *scalability:* the number of intersectional sub-groups grows exponentially with the number of sensitive attributes, making the optimization computationally expensive and possibly making it hard to find an optimal solution to account for so many groups; (ii) *redundancy:* redundant or unnecessary comparisons between certain sub-groups can introduce inefficiencies without improving fairness outcomes; and (iii) *data scarcity:* the number of items belonging to one intersectional group can get very small, possibly affecting the reliability of the fairness improvements as well as the fairness assessments for these groups.

Our Proposal. We address the above identified challenges with intersectional fairness. First, we propose to reduce the number of sub-group comparisons, using the relationships between intersectional sub-groups by prioritizing comparisons, based on the highest disparities between sub-groups. We show that this reduces the computational complexity that comes with intersectional fairness, without sacrificing the effectiveness of the fairness interventions, achieving comparable or better improvements, compared to a pairwise comparison, in terms of intersectional fairness on the downstream ranking task. Second, we show that reducing sub-groups comparisons can be a practical solution for scenarios where some intersectional groups are too small to support reliable optimization. Third, we propose an evaluation set-up on simulated scenarios to test the robustness of fairness methods, under reduced sub-group optimization, given the challenges that come with intersectional fairness. To operationalize our approach, we propose two pre-processing methods, **xLFR** and **gFair**, to support intersectional fairness, and evaluate them across three real-world datasets for ranking people in high-risk domains. Our code is available on GitHub.[2]

Theory of Change. We address the problem of intersectional fairness in ranking systems for high-stakes domains, where people belonging to multiple marginalized groups might face intersectional discrimination. Our findings are particularly relevant for real-world applications with significant societal impli-

[1] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L_202401689.
[2] https://github.com/ClaraRus/Learning_Intersectional_Fairness_in_Rankings.

cations, where scalability, redundancy, and data scarcity are persistent challenges. This work challenges the current pairwise approach to intersectional fairness, providing evidence that intersectional fairness can be enforced efficiently rather than being dismissed as infeasible. The design decisions in this work were informed by interacting with recruitment companies and legal consultants, ensuring its practicability. Pre-processing methods are easy to be plugged into the existing recruitment systems, and are more inline with legal considerations [19]. They assume access to sensitive information at training time, which should be acquired in compliance with GDPR and the AI Act [1,2], and group labels should ensure inclusivity [34] (e.g., accounting for non-binary gender identities and avoiding residual "other" categories that obscure marginalized groups.) Moreover, candidates belonging to marginalized communities are less likely to share such information out of fear of discrimination [27]. Pre-processing methods cannot fully debias the data, as there can be biases for which the method did not account, thus, we advice against such claims that can lead to fairness washing. Moreover, the proposed approach should be tested in real-life systems.

2 Related Work

Most work on intersectional fairness [15], including our work, reduces the problem to achieving fairness towards all intersectional sub-groups. This view has been criticized [18,23] as it simplifies the complex dynamics of intersectionality and fails to address issues such as context, power imbalances, and relationality, but it does provide an initial framework for exploring intersectional fairness. Due to its complex nature, there is a lack of fairness interventions focused on intersectionality, especially in ranking. Most interventions focus on achieving fairness towards one attribute with some being designed to account for multiple attributes [5,12,20,40,41]. iFair [20] is a pre-processing method focused on individual fairness; Zehlike et al. [41] propose a post-processing method that re-ranks items given some fairness constraints. Devic et al. [12] show that uncertainty-aware ranking functions [32] can also achieve multi-attribute fairness. Though not designed for intersectional fairness, these methods can be adapted by defining a single attribute representing all intersectional sub-groups, turning the task into a multi-group problem. CIFRank [37] is a pre-processing method for intersectional fairness, but it requires access to sensitive information at inference time and its changes on the data might not be sufficient to promote fairness in the output ranking [29,30]. Pastor and Bonchi [24] propose a re-ranking method that accounts for intersectional fairness while dynamically identifying the statistically significant intersectional groups that are disadvantaged.

Others address intersectional fairness in a classification setting by debiasing the input data [8,14,16,17]. Most methods rely on a pairwise comparison between intersectional sub-groups. E.g., Dzakpasu et al. [14] focus on debiasing the input data such that the equality of opportunity between all intersectional sub-groups is minimized. While these comparisons are comprehensive, they suffer from scalability issues due to the exponential growth of

intersectional sub-groups [26]. Diana et al.[13] focus on minimizing the worst of fairness of a sub-group, still requiring to compare all sub-groups. In this work, we propose a set of optimization strategies for achieving intersectional fairness in rankings while reducing the number of required sub-group comparisons. Moreover, we operationalize this approach using pre-processing methods, which we argue account for the societal context by mitigating biases encoded in the data.

3 Fairness Methods

3.1 Preliminaries and Notation

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of n candidates and $S = \{S_1, S_2, \ldots, S_m\}$ a set of m sensitive attributes (e.g. gender). Define a ranking function $f : X \to R$ that assigns a relevance score to each item. The ranking task is to produce a permutation π over X such that: $\pi(x_i) < \pi(x_j)$ if, and only if, $f(x_i) > f(x_j)$, where $\pi(x)$ denotes the rank of item x. Given the original data X, we apply a pre-processing method F' to produce debiased data X', then the ranking function will be $f : X' \to R$. The optimization objective can be defined:

$$F'(X) = \arg\min_{\tilde{X}} \ \lambda_r L_{\text{recon}}(\tilde{X}, X) + \lambda_f L_{\text{fair}}(\tilde{X}), \tag{1}$$

where $L_{\text{recon}} = (X - X')^2$ is the reconstruction loss, preserving task-relevant information, L_{fair} is the fairness loss, and λ_r, λ_f hyperparameters balancing utility and fairness. To define L_{fair} we formalize the sensitive data as follows. For each attribute $S_i \in S$, G_{S_i} represents the groups, i.e., the set of values that S_i can take, and $G_{S_i,j}$ denotes the j-th value of G_{S_i}. The set $I = G_{S_1} \times G_{S_2} \times \cdots \times G_{S_m}$ represents all unique sub-groups formed by the intersection of the sensitive attributes in S, where $m = |S|$. For example, if $S = \{\text{gender, nationality}\}$, then $I = \{\text{male EU, male non-EU, female EU, female EU}\}$.

3.2 Intersectional Fairness Optimization

Next, we introduce the problem of intersectional fairness optimization and our contribution, which is a set of strategies that reduce the number of sub-group comparisons, lowering the computational complexity. Generally, the problem of optimizing for intersectional group fairness, can be addressed by **independently** optimizing for each attribute:

$$L_{\text{fair}} = \sum_{i=1}^{|S|} \sum_{\substack{x,y=1 \\ x \neq y}}^{|G_{S_i}|} g(G_{S_i,x}, G_{S_i,y}), \tag{2}$$

or by reducing it to a **pairwise** optimization across intersectional sub-groups:

$$L_{\text{fair}} = \sum_{\substack{x,y=1 \\ x \neq y}}^{|I|} g(I_x, I_y), \tag{3}$$

where g is a fairness function. One strategy [13] to simplify the optimization is to **dynamically** identify the biggest gap between all intersectional sub-groups $L_{\text{fair}} = \max_{x,y}^{|I|} g(I_x, I_y)$. This simplifies the optimization problem, but in order to compute the biggest gap we still need to compare all intersectional sub-groups. As a solution, we propose a set of optimization strategies for intersectional fairness that reduce the sub-group comparisons and computational complexity.

Control Comparison. According to this strategy, we reduce the number of comparisons by comparing all sub-groups with a control group during fairness optimization: $L_{\text{fair}} = \sum_{x=1\ x\neq c}^{|I|} g(I_x, I_c)$, where I_c is the control group. This reduces the computational complexity from $\mathcal{O}(|I|^2)$ to $\mathcal{O}(|I|)$. We consider two ways of selecting the control group. **(i) Control A**: to be the socially **advantaged** group. This highlights the inequalities with the privileged class, making the representations change such that those inequalities are reduced across all sub-groups. However, this could be interpreted that the norm is considered to be that of the privileged group, which is a view that has been critiqued [25]. And **(ii) Control D**: to be the socially **disadvantaged** group. The comparison with the socially disadvantaged group highlights the inequalities between all sub-groups and the least privileged group.

Extremes. According to this strategy, we compare only two opposite intersectional groups, such that they differ on all attributes (gender, race, etc.): $L_{\text{fair}} = g(I_a, I_b)$, where, e.g., I_a is the non-European female group and I_b is the European male group. This reduces the complexity to $\mathcal{O}(1)$ and focuses on the most contrasting sub-groups, where bias is often most pronounced. The choice of extremes can be guided by selecting the most privileged and least privileged groups. However, in practice, the least privileged group may have too few samples, making this optimization unstable. To address this, we propose as a practical solution, **extremes***, which uses proxy groups that are still opposites in terms of attributes but may not reflect the largest observed disparity; these proxies can allow for more stable optimization while still capturing sub-group level biases. Although extremes optimization considers only two sub-groups, potentially ignoring other sub-groups, we hypothesize that due to the relationship between the intersectional biases, it should positively impact the other groups.

3.3 Pre-processing Methods

We extend two pre-processing methods used for ranking tasks, to optimize for intersectional group fairness: LFR [42] and iFair [20]. They work on the same underlying principle: the transformation of the candidate's representation is formulated in terms of a probabilistic mapping to a set of prototypes (points in the input space). The model can be defined as a discriminative clustering model, as each candidate data point is assigned to the prototypes, which act as the clusters (V), with a certain probability (P). The transformation of the data point (x) representing the candidate in the feature space is defined as $x' = \sum_{k=1}^{K} P_k V_k$,

where P_k is the probability of x being mapped to prototype V_k. The probability is computed using a distance function between the data point (x) and the prototype (V_k). Thus, both LFR and iFair aim to find the best K prototypes (V) given some fairness optimization constraints.

xLFR is our extension of LFR that accounts for intersectional fairness with non-binary values. It achieves group fairness by creating representations such that the probability of a random data point from group u has the same probability as a random data point from group v to be mapped to a prototype (V_k): $g(u,v) = \sum_{k=1}^{K} |P_k^u - P_k^v|$, then the optimization can be performed using Eq. 3. Additionally, it adds a utility loss to Eq. 1: $L_{util} = -y\log(y') - (1-y)\log(1-y')$, where y is the label and y' is the transformed label. As the original LFR was designed for a binary classification task, our proposed loss is adapted to distinguish between relevant and not-relevant items in the ranking task.

gFair is our extension of iFair to achieve group fairness, instead of individual fairness. iFair achieves individual fairness by computing the pairwise distance between all data points and making sure that the new representations preserve this distance in the new space, independent of the sensitive attributes. Thus, similar individuals should be close to each other in the feature space, with the similarity measured using a distance function between task related features. This can be defined as: $\sum_{i,\ j \in N} \left(d(x'_i, x'_j) - d(x^*_i, x^*_j)\right)^2$, where x^* the original representation excluding the sensitive attributes. Expanding on this idea, gFair optimizes for group fairness by enforcing that individuals from group u should be close to each other in the feature space to similar individuals from group v independent of their sensitive attributes. This can be defined as: $g(u,v) = \left(d(x'_u, x'_v) - d(x^*_u, x^*_v)\right)^2$, where x^* is the original representation excluding the sensitive attributes. In order to create representations that are independent of the sensitive information, the model uses a learnable weight (α) in the distance function ($d(x_i, x_j) = [\sum_n^N (\alpha_n (x_i, n - x_j, n)^p]^{1/p}$) that indicates the importance of each feature in computing the similarity between individuals and the mapping to the prototypes. Ideally, features that strongly proxy sensitive attributes will have α values near zero, offering interpretable insight into which features act as proxies. Additionally, we add an in-group-fairness (IGF) loss to Eq. 1, to make sure that the distance between candidates belonging to the same group is still preserved in the new representation space: $L_{\text{IGF}} = \sum_{i \in G} \sum_{y,\ z \in G_i} \left(d(x'_y, x'_z) - d(x^*_y, x^*_z)\right)^2$.

4 Experimental Setup

Parameters and Settings. We apply pre-processing methods to reduce bias in the training data, to improve fairness in the resulting ranking. We use RankNet [4], a pairwise learning-to-rank model, using the Ranklib library [33]. Experiments are run on a Linux machine, Intel Xeon Gold 5118 CPU (2.30GHz) with 5 random splits, fixed random seed (42), 70/30 train-test using stratified sampling across intersectional groups for negative and positive samples. Negative

samples are those with scores below a dataset specific threshold. The hyperparameters λ_f, λ_r, λ_u, and λ_{igf} were tuned over the set $1, 0.1, 0.01, 0.001, 0$.

Evaluation Measures. We focus on utility and group fairness, meaning members of different groups should be treated the same. Most fairness measures support one binary attribute, with some supporting multi-groups [35], making them suitable for measuring intersectional fairness. The inverse *Jensen-Shannon* **(JS)** divergence measures the difference between the percentage distribution of intersectional sub-groups in the top-k and an ideal distribution, sub-group proportion of positive samples. As improvements in JS do not guarantee a fair representation of the most underreprested sub-group, we measure the *percentage of the most disadvantaged group* (**%D**) in the top-k. We use the *normalized discounted cumulative gain* **(NDCG)** to measure the loss in utility between the new ranking and the ground truth original one based on the original scores.

Pre-processing Baselines. We test our proposed pre-processing approaches, under reduced sub-group optimization, against **iFair** [20], which achieves **individual** fairness while obfuscating the sensitive information, supporting multi-group optimization, and **CIFRank** [36], which is designed to handle intersectional fairness, using **causal** estimation to create counterfactual representations such that all candidates look like they are part of the same group.

Real-World Datasets. Our experiments are conducted on 3 real-world datasets for ranking people with significant societal implications in high-risks domains. **COMPAS** [3], for judiciary domain, includes criminal history, the probability of recidivism, and sensitive attributes: gender and race. Following [37] we use a subset of the COMPAS dataset, with 4,163 samples, where candidates are ranked from low to high, prioritizing them for release or supportive interventions.

BIOS [11], a recruitment dataset, with features extracted from the text biography: term frequency of the occupation in the biography, length and number of words, cosine similarity between the occupation and biography, and sensitive features: gender and nationality (inferred from the name as recruiter might when reviewing a resume.) We selected 27 queries (occupations) with max 5K samples.

MEPS 2022[3] a medical dataset providing information on Americans' healthcare psychological distress, mental and physical health scores, reflecting perceived health status which may underestimate risk for individuals who overreport their health, and sensitive attributes: age, gender and race. The 10,766 candidates are ranked by their total medical usage, as a proxy for healthcare need.

Simulated Data. We propose a simulated set-up to test the robustness to the challenges that come with intersectional fairness. It simulates a recruitment scenario where each sample is a candidate to be ranked for a job opening, represented by two features that are sampled from a normal distribution, which can be interpreted as number of years of education (X1) and number of years of experience (X2) [38]. The candidate's fitness to the job opening, is computed

[3] https://meps.ahrq.gov/data_stats/download_data/pufs/h243/h243doc.shtml.

as the weighted sum between X1 and X2, where the weights are randomly sampled. Intersectional bias is simulated following the cumulative additive model [6], where implicit biases compound across multiple attributes. Unless specified otherwise, each intersectional sub-group is composed of 200 positive samples and 40 negative samples. We propose several settings to test the robustness towards: **(i) Degree of bias:** The bias varies by having low bias (0.2) and strong bias (0.4) equally towards both 2 attributes. Consistent with the category dominance model [21], where a single social category has a dominant influence, we simulate a scenario where the bias is imbalanced (0.2 for one attribute; 0.4 for the other). **(ii) Number of samples:** In this setting we simulate 2 attributes with low equal bias (0.2). The number of positive samples within each intersectional sub-group varies as follows: $|I_i| - |I_i| * b_i$, where b_i is the cumulative bias score for the ith intersectional group, resulting into reduced sub-group proportion (0.2) and strong reduced sub-group proportion (0.4). This simulates a scenario where the most disadvantaged group has less samples than the advantaged groups. This setup lets us examine how methods handle cases where intersectional groups become increasingly small (e.g., when we have many attributes), testing their robustness under limited sample sizes.

(iii) Number of attributes: In this setting we vary the number of attributes by having 2, 3, 4 attributes. Public datasets often have max 2 attributes available [10]. Each attribute has low bias between the groups.

5 Results

Table 1 shows the results of the intersectional pre-processing fairness methods on 3 real-world datasets: COMPAS, BIOS and MEPS, while Fig. 1 describes the trade-off between fairness (JS, %D) and NDCG over the simulated set-ups. Since MEPS includes three sensitive attributes, we conduct experiments using both two and three attributes for intersectional optimization. We evaluate fairness as the percentage of each group in the top-k, reflecting selection for outcomes such as release (COMPAS), recruitment screening (BIOS), and healthcare support (MEPS). Following [36] we use top-500 for COMPAS and top-100 for MEPS. For BIOS and the simulated set-ups, we evaluate at the top-10, as only a small pool of candidates can be considered for an interview. For all datasets, the most disadvantaged group is the least represented in the top-k.

5.1 Reducing Intersectional Comparisons

Achieving intersectional fairness poses several challenges, including scalability and redundancy. The number of intersectional sub-groups increases exponentially with the number of sensitive attributes making it both computationally expensive and hard to find an optimal solution to account for so many groups. Unlike traditional multi-group fairness settings, intersectional fairness benefits from inherent relationships between biases affecting different sub-groups. Building on this assumption, we explore the following research question: **(RQ1)**

Table 1. Intersectional optimization on real-world datasets. Time measured in seconds (**lowest**), Utility (U) as NDCG. (c)-causal, (I)-individual, (p)-pairwise, (d)-dynamic, (i)-independent, (D)-control D, (A)-control A, (e)-extremes, (*)-proxy.

		COMPAS				BIOS				MEPS				MEPS (3 attr)			
Method	Opt	JS↑	%D↑	U↑	Time↓	JS↑	%D↑	U↑	Time↓	JS↑	%D↑	U↑	Time↓	JS↑	%D↑	U↑	Time↓
Original	-	0.51	0.10	1.00	-	0.66	0.11	0.71	-	0.76	0.17	0.75	-	0.59	0.01	0.75	-
CIFRank	(c)	0.51	0.11	0.97	0.43	0.66	0.11	0.71	1.60	0.70	0.16	0.30	0.69	0.60	0.01	0.60	1.24
iFair	(I)	0.56	0.15	0.97	224	0.67	0.13	0.72	1208	0.81	0.22	0.76	927	0.67	0.03	0.75	946
xLFR	(p)	0.55	0.10	0.99	491	0.68	0.14	0.71	2502	0.78	0.20	0.74	2111	0.61	0.04	0.71	2521
	(d)	0.54	0.11	0.84	489	0.66	0.12	0.72	2502	0.80	0.23	0.77	2111	0.61	0.02	0.44	2521
	(i)	0.55	0.10	0.90	455	0.66	0.11	0.71	2432	0.76	0.18	0.66	2173	0.66	0.02	0.74	2475
	(D)	0.55	0.10	0.97	418	0.68	0.11	0.72	2149	0.81	0.24	0.77	2035	0.63	0.05	0.77	2146
	(A)	0.55	0.10	0.99	441	0.67	0.11	0.71	2214	0.79	0.22	0.76	2119	0.61	0.05	0.44	2130
	(e)	0.56	0.11	0.86	**412**	0.66	0.13	0.71	**2038**	0.80	0.23	0.76	**2029**	0.61	0.06	0.48	**2019**
	(e)*	0.49	0.09	0.94	**412**	0.66	0.12	0.71	**2038**	0.77	0.21	0.84	**2029**	0.64	0.09	0.72	**2019**
gFair	(p)	0.50	0.10	0.83	587	0.67	0.14	0.72	5896	0.80	0.20	0.72	3670	0.65	0.02	0.79	3609
	(d)	0.51	0.10	0.78	621	0.67	0.13	0.72	5998	0.80	0.20	0.73	3899	0.66	0.01	0.67	3520
	(i)	0.51	0.11	0.90	980	0.67	0.11	0.72	11794	0.80	0.20	0.81	7457	0.68	0.03	0.72	10571
	(D)	0.52	0.12	0.99	217	0.66	0.14	0.72	3448	0.80	0.20	0.75	2166	0.65	0.01	0.78	1534
	(A)	0.51	0.10	0.78	226	0.67	0.13	0.72	3557	0.80	0.20	0.73	2656	0.64	0.02	0.80	2288
	(e)	0.51	0.11	0.82	**166**	0.66	0.13	0.72	**2024**	0.80	0.21	0.73	**1037**	0.65	0.02	0.73	**98**
	(e)*	0.51	0.11	0.99	**166**	0.66	0.12	0.72	**2024**	0.80	0.22	0.77	**1037**	0.64	0.01	0.80	**98**

Are all pairwise sub-group comparisons necessary in optimizing pre-processing fairness interventions for intersectional fairness in rankings? To answer this, we conduct a comparative analysis of intersectional optimization strategies while reducing the sub-group comparisons.

Real-World Data. On the real-world datasets (Table 1) with fairness evaluated over two protected attributes, there is no clear pattern into an optimization strategy that consistently yields the best fairness outcomes across all settings. However, the results clearly indicate that reducing the number of sub-group comparisons is an effective strategy for improving intersectional fairness. The extremes and dynamic optimizations consistently improve %D while maintaining or improving the JS divergence. Hence, optimizing for the biggest discrepancy between groups indeed improves the outcomes for the most disadvantaged group, without hurting the outcomes of the other groups. Moreover, it has the potential of making overall improvements for the other intersectional groups, balancing the distribution among the groups in the top-k. The control optimizations consistently improved the JS. This is an expected outcome, as comparing all sub-groups to a single control group better captures the overall sub-group disparities. Of the control groups, control D, which emphasizes discrepancies relative to the most disadvantaged group, further prioritizes improvements for this group, as it shows to obtain stronger gains in %D. Prior work [36] showed that optimizing attributes independently fails to ensure fairness for intersectional groups in a

post-processing context. Our results show that the independent optimization is not effective in a pre-processing context. CIFRank consistently underperforms across all datasets, while iFair achieves consistent improvements in terms of both JS and %D, regardless of its optimization being designed for individual fairness. Overall, in terms of NDCG, all fairness methods, regardless of the optimization strategy, except xLFR independent, show no loss or a small loss in utility compared to the original ranking across all datasets. The fairness methods successfully promote disadvantaged candidates to the top of the ranking while selecting the strongest individuals from each sub-group.

Simulated Setup. On the simulated setup (Fig. 1), optimizations tend to cluster by pre-processing method, indicating that the choice of pre-processing method has a greater impact than the type of intersectional optimization. Unlike the real-world datasets, the simulated setup imposes a uniform bias on all candidates, making it more difficult to achieve fairness improvements by promoting disadvantaged individuals to the top of the ranking. This results in having bigger discrepancies between the extreme groups. In contrast, real-world datasets often contain structural biases that affect groups unevenly. Nonetheless, all intersectional optimization strategies improve JS on the downstream ranking task. Among the gFair optimizations, extremes seems to be the most effective across all simulated set-ups, suggesting that fairness improvements can be achieved without comparing all groups, especially in a setting where the extreme group discrepancy is significantly bigger than the discrepancies of the other sub-groups. Overall, gFair obtains a better trade-off between JS and NDCG, while xLFR achieves better improvements in terms of JS, but with a higher cost in terms of utility. xLFR promotes more candidates from the disadvantaged group, which could lower the NDCG. However, when gFair promotes disadvantaged candidates, their average relevance is higher than those promoted by xLFR. As gFair aims to preserve the order within a group, this can lead to a better utility-fairness trade-off. Similar to the real-world datasets, CIFRank is underperforming in most settings, achieving slight improvements. Unlike the real-world setting, iFair obtains unstable improvements.

Computation Time. Table 1 shows the average training *Time* in seconds, over the 5 runs, of training the pre-processing methods under our proposed intersectional optimization strategies. Reducing the number of comparisons (control A, control D, extremes) maintains consistently lower training times. In contrast, more complex methods, such as independent, pairwise, and dynamic, show higher computation times. The contrast is more visible among the gFair variants. This highlights the trade-off between computational efficiency and method complexity. Among all methods, CIFRank is the fastest, however, it does require access to sensitive information during inference time, making it less suitable in a real-environment. Overall, iFair is faster than most xLFR and gFair variants, but without guaranteeing group fairness as it optimizes for individual fairness.

Answer to RQ1. A pairwise optimization is not necessary for improving the fairness on the downstream ranking task. Reducing the sub-group comparison achieves comparable or better performance on both the real-world datasets and

in the simulated set-ups. There is no clear pattern indicating which type of reduced intersectional optimization performs best overall, however, in most settings the extremes optimization shows to obtain consistent improvements, especially in terms of %D, without hurting the JS. Reducing the number of comparisons reduces the computation time required to train the pre-processing fairness methods, making the extremes optimization obtain a good trade-off between fairness enhancement and computational costs.

5.2 Robustness of Intersectional Approaches

Intersectionality poses several challenges: varying degrees of bias, complex intersectional relationships, increased number of attributes, and the cardinality of sub-groups. We conduct experiments on simulated datasets (Fig. 1) to answer **(RQ2) How robust are pre-processing fairness interventions to intersectional fairness challenges under reduced sub-group comparisons?**

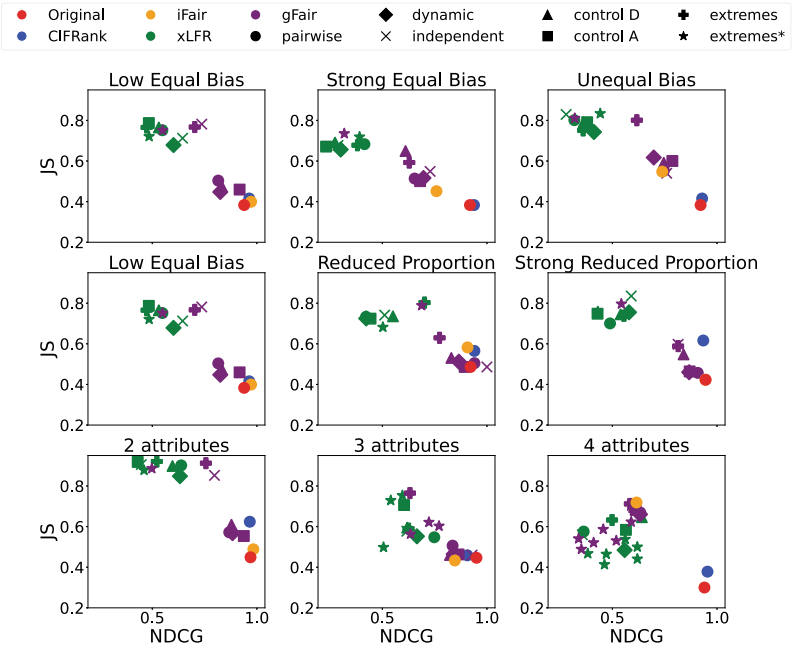
Degree of Bias (row 1). Overall, xLFR and gFair seem robust against variations in the strength of bias, and to unequal bias distribution. Notably, out of gFair variants, the extremes (purple +) is consistently among the top, being one of the few improving %D.

Number of Samples (row 2). Similarly, xLFR seems to be robust against the reduced number of samples in each intersectional sub-group, as well as gFair with the extremes optimizations being consistently among the best. However, when checking the %D gFair fails to make improvements under reduced proportions.

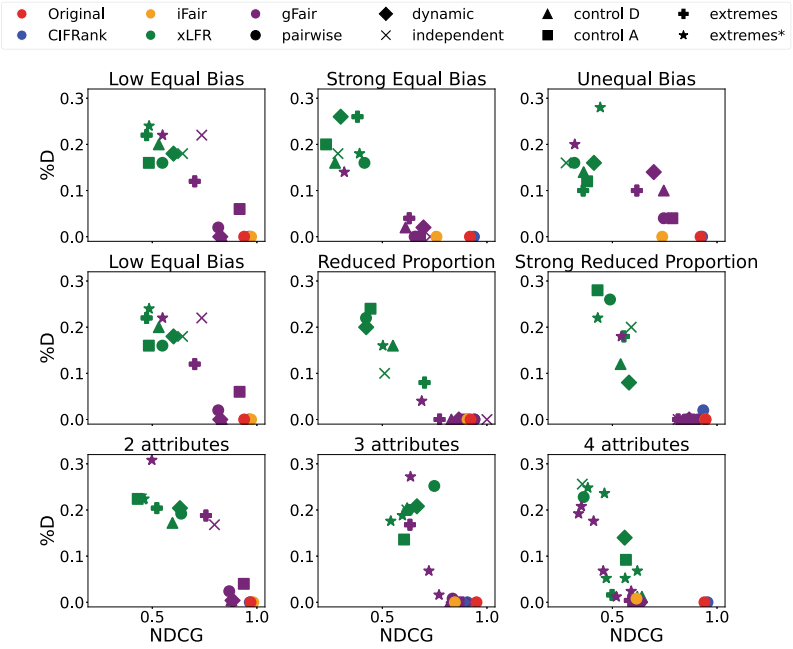
Number of Attributes (row 3). Having 4 attributes implies 16 sub-groups, so the evaluation is performed at top-50 here. Both xLFR and gFair improve the JS, with gFair extremes being constantly in the top. LFR consistently improves %D, while most gFair variants fail to do so, except extremes showing gains, but failing as the number of attributes increases to 4.

MEPS (Table 1). As MEPS has 3 attributes, we can test the methods against a real-world dataset as well. In this dataset, as the granularity of the sub-groups increases, the number of samples in the intersectional group decreases. All xLFR variants improve both JS and %D, while all gFair variants improve JS, with overall higher values than xLFR. xLFR achieves better improvements in %D. Out of the gFair variants extremes, control A, independent and pairwise show improvements. As in the simulated set-up, extremes optimization is stable as the number of attributes increases.

Answer to RQ2. The choice of pre-processing method matters more than the choice of optimization strategy, with xLFR being robust against simulated intersectional challenges, regardless of the choice of optimization. For gFair, the extremes optimization seems to be more robust than the other optimizations; however, it is highly affected by reducing the number of samples and increasing the number of attributes to 4, as it does not manage to improve %D.



(a) NDCG vs. JS in the simulated setting.



(b) NDCG vs. %D in the simulated setting.

Fig. 1. Resultes obtained in the simulated set-up.

5.3 Intersectional Fairness with Small Groups

A core challenge of intersectionality is data scarcity, the reduced number of samples of intersectional sub-groups, especially for the most disadvantaged groups. We address **(RQ3) Can proxy comparisons between larger intersectional sub-groups improve intersectional fairness?** via the *extreme* optimization, which compares opposite groups with larger sample sizes as proxies for the true extremes. Table 1 - *extremes** shows results on real-world datasets. For MEPS (3 attr), the most disadvantaged "young" group is represented through a larger proxy group. Our solution, achieves comparable improvements or no improvements in comparison with considering the real extremes. Notably, on the MEPS dataset (3 attr) xLFR obtained better improvements with the proxy variant while keeping a good trade-off with utility. However, on COMPAS xLFR extremes* shows a decrease in fairness. In the simulated setup (Fig. 1, "*" markers), we evaluate all proxy combinations for higher-attribute settings. Overall, our proxy solution shows comparable or better improvements in fairness across the simulated set-up. Notably, where gFair fails to improve %D in the reduced sample size setting, the proxy approach improves. Overall, our proxy solution helps, especially in settings with reduced sample size, but, the proxy groups should be chosen with consideration as in some settings of real-world datasets, such as on the COMPAS dataset, where the bias is not uniform and compound, it shows to harm the fairness.

6 Conclusion

We focus on the problem of unfairness in ranking systems for high-stakes domains, where people belonging to multiple marginalized groups might face intersectional discrimination. We address pre-existing biases in the data, including the societal context that gives rise to inequalities, using pre-processing fairness methods that debias the data to produce fairer rankings for intersectional groups. We challenge the current pairwise comparison paradigm and propose several optimization strategies that reduce the number of sub-group comparisons required during the optimization of pre-processing fairness methods. We propose a simulated set-up for evaluating the robustness of fairness enhancing methods, against the challenges of intersectionality. Our experiments on real-world datasets and simulated scenarios showed that reducing sub-group comparisons can improve computational efficiency while maintaining or improving fairness towards intersectional groups, with the extremes optimization being the most efficient with consistent fairness improvements. This provides concrete evidence that intersectional fairness can be enforced efficiently rather than being dismissed as infeasible. Future work could explore theoretical fairness guarantees across intersectional sub-groups, given the optimization strategies. As intersectional optimization assumes that sub-groups contain a significant number of candidates, which marginalized groups often lack, we propose the extremes* proxy solution, tackling the data scarcity problem. We show empirically that this solution leads to improvements, however, when the bias is not compound, this

strategy might fail. Future work could focus on testing the robustness of this approach and extending the simulated setup beyond compound bias to more complex scenarios. Intersectionality is a complex social concept, and while our work simplifies the problem, it serves as a starting point.

Acknowledgments. We thank the reviewers for their feedback. This research was supported by EU's Horizon Europe program (No. 101070212 and 101201510), and by NWO (VI.Vidi.223.166, 024.004.022, NWA.1389.20.183, and KICH3.LTP.20.006). All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. van Bekkum, M.: Using sensitive data to de-bias AI systems: Article 10 (5) of the EU AI Act. Comput. Law Sec. Rev. **56**, 106115 (2025)
2. van Bekkum, M., Borgesius Zuiderveen, F.: Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception? Comput. Law Sec. Rev. **48**, 105770 (2023)
3. Bias, M.: There's software used across the country to predict future criminals and it's biased against blacks. ProPublica (2016). https://www.propublica.org/article/machine-bias-risk-assessments-in-criminalsentencing
4. Burges, C., et al.: Learning to rank using gradient descent. In: Proceedings of the 22nd International Conference on Machine Learning, pp. 89–96 (2005)
5. Celis, L.E., Straszak, D., Vishnoi, N.K.: Ranking with fairness constraints. arXiv preprint arXiv:1704.06840 (2017)
6. Connor, P., Weeks, M., Glaser, J., Chen, S., Keltner, D.: Intersectional implicit bias: evidence for asymmetrically compounding bias and the predominance of target gender. J. Pers. Soc. Psychol. **124**(1), 22 (2023)
7. Corbett-Davies, S., Pierson, E., Feller, A., Goel, S.: A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear. Washington Post **17** (2016)
8. Creager, E., et al.: Flexibly fair representation learning by disentanglement. In: International Conference on Machine Learning, pp. 1436–1445, PMLR (2019)
9. Crenshaw, K.: Demarginalizing the intersection of race and sex: a black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. In: Feminist Legal Theories, pp. 23–51, Routledge (2013)
10. Criscuolo, C., Martinenghi, D., Piccirillo, G.: A tutorial on intersectionality in fair rankings. arXiv preprint arXiv:2502.05333 (2025)
11. De-Arteaga, M., et al.: Bias in bios: A case study of semantic representation bias in a high-stakes setting. In: Proceedings of the Conference on Fairness, Accountability, and Transparency, pp. 120–128 (2019)
12. Devic, S., Korolova, A., Kempe, D., Sharan, V.: Stability and multigroup fairness in ranking with uncertain predictions. arXiv preprint arXiv:2402.09326 (2024)

13. Diana, E., Gill, W., Kearns, M., Kenthapadi, K., Roth, A.: Minimax group fairness: Algorithms and experiments. In: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society, pp. 66–76 (2021)
14. Dzakpasu, D.Q., Liu, J., Li, J., Liu, L.: Integrating fair representation learning with fairness regularization for intersectional group fairness. In: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, pp. 560–569 (2024)
15. Gohar, U., Cheng, L.: A survey on intersectional fairness in machine learning: Notions, mitigation, and challenges. arXiv preprint arXiv:2305.06969 (2023)
16. Kang, J., Xie, T., Wu, X., Maciejewski, R., Tong, H.: Infofair: Information-theoretic intersectional fairness. In: 2022 IEEE International Conference on Big Data (Big Data), pp. 1455–1464, IEEE (2022)
17. Kobayashi, K., Nakao, Y.: One-vs.-one mitigation of intersectional bias: a general method to extend fairness-aware binary classification. arXiv preprint arXiv:2010.13494 (2020)
18. Kong, Y.: Are "intersectionally fair" AI algorithms really fair to women of color? A philosophical analysis. In: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 485–494 (2022)
19. Kumar, D., Grosz, T., Rekabsaz, N., Greif, E., Schedl, M.: Fairness of recommender systems in the recruitment domain: an analysis from technical and legal perspectives. Front. Big Data **6**, 1245198 (2023)
20. Lahoti, P., Gummadi, K.P., Weikum, G.: iFair: Learning individually fair data representations for algorithmic decision making. In: 2019 IEEE 35th International Conference on Data Engineering, pp. 1334–1345, IEEE (2019)
21. Macrae, C.N., Bodenhausen, G.V., Milne, A.B.: The dissection of selection in person perception: inhibitory processes in social stereotyping. J. Pers. Soc. Psychol. **69**(3), 397 (1995)
22. Nelson, A.: Unequal treatment: confronting racial and ethnic disparities in health care. J. Natl Med. Assoc. **94**(8), 666 (2002)
23. Ovalle, A., Subramonian, A., Gautam, V., Gee, G., Chang, K.W.: Factoring the matrix of domination: a critical review and reimagination of intersectionality in ai fairness. In: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society, pp. 496–511 (2023)
24. Pastor, E., Bonchi, F.: Intersectional fair ranking via subgroup divergence. Data Min. Knowl. Disc. **38**(4), 2186–2222 (2024)
25. Purdie-Vaughns, V., Eibach, R.P.: Intersectional invisibility: the distinctive advantages and disadvantages of multiple subordinate-group identities. Sex Roles **59**, 377–391 (2008)
26. Ramachandranpillai, R., Sampath, K., Mohammad, A., Alikhani, M.: Fairness at every intersection: Uncovering and mitigating intersectional biases in multimodal clinical predictions. arXiv preprint arXiv:2412.00606 (2024)
27. Rosales, C., Buslón, N., Curi, F., Jorge, R.: Beyond the algorithm: Expanding the understanding of fairness and non-discrimination in algorithmic hiring – The case for Latin American migrants seeking employment opportunities in Spain. Tech. rep., FINDHR Expert Reports (2023). https://findhr.eu/wp-content/uploads/2024/01/FINDHR-Expert-report_by-Cesar-Rosales-et-al.pdf
28. Rus, C., Luppes, J., Oosterhuis, H., Schoenmacker, G.H.: Closing the gender wage gap: Adversarial fairness in job recommendation. arXiv preprint arXiv:2209.09592 (2022)

29. Rus, C., de Rijke, M., Yates, A.: Counterfactual representations for intersectional fair ranking in recruitment. In: RecSys in HR'23: The 3rd Workshop on Recommender Systems for Human Resources, in conjunction with the 17th ACM Conference on Recommender Systems (2023)
30. Rus, C., Yates, A., de Rijke, M.: A study of pre-processing fairness intervention methods for ranking people. In: Goharian, N., et al. (eds.) European Conference on Information Retrieval, pp. 336–350, Springer, Cham (2024). https://doi.org/10.1007/978-3-031-56066-8_26
31. Singh, A., Joachims, T.: Policy learning for fairness in ranking. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems, pp. 5426–5436 (2019)
32. Singh, A., Kempe, D., Joachims, T.: Fairness in ranking under uncertainty. Adv. Neural. Inf. Process. Syst. **34**, 11896–11908 (2021)
33. The Lemur Project: http://lemurproject.org (2023)
34. Wang, A., Ramaswamy, V.V., Russakovsky, O.: Towards intersectionality in machine learning: Including more identities, handling underrepresentation, and performing evaluation. In: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 336–349 (2022)
35. Wang, Y., Ma, W., Zhang, M., Liu, Y., Ma, S.: A survey on the fairness of recommender systems. ACM Trans. Inf. Syst. **41**(3), 1–43 (2023)
36. Yang, K., Gkatzelis, V., Stoyanovich, J.: Balanced ranking with diversity constraints. arXiv preprint arXiv:1906.01747 (2019)
37. Yang, K., Loftus, J.R., Stoyanovich, J.: Causal intersectionality for fair ranking. arXiv preprint arXiv:2006.08688 (2020)
38. 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 Conference on Information and Knowledge Management, pp. 1569–1578, ACM (2017)
39. Zehlike, M., Castillo, C.: Reducing disparate exposure in ranking: a learning to rank approach. In: Proceedings of the Web Conference 2020, pp. 2849–2855 (2020)
40. Zehlike, M., Hacker, P., Wiedemann, E.: Matching code and law: achieving algorithmic fairness with optimal transport. Data Min. Knowl. Disc. **34**(1), 163–200 (2020)
41. Zehlike, M., Sühr, T., Baeza-Yates, R., Bonchi, F., Castillo, C., Hajian, S.: Fair top-k ranking with multiple protected groups. Inf. Process. Manage. **59**(1), 102707 (2022)
42. Zemel, R., Wu, Y., Swersky, K., Pitassi, T., Dwork, C.: Learning fair representations. In: International Conference on Machine Learning, pp. 325–333. PMLR (2013)