Fairness of Exposure in Rankings

Ashudeep Singh
Cornell University
Ithaca, NY
ashudeep@cs.cornell.edu

Thorsten Joachims
Cornell University
Ithaca, NY
tj@cs.cornell.edu

ABSTRACT

Rankings are ubiquitous in the online world today. As we have transitioned from finding books in libraries to ranking products, jobs, job applicants, opinions and potential romantic partners, there is a substantial precedent that ranking systems have a responsibility not only to their users but also to the items being ranked. To address these often conflicting responsibilities, we propose a conceptual and computational framework that allows the formulation of fairness constraints on rankings. As part of this framework, we develop efficient algorithms for finding rankings that maximize the utility for the user while satisfying fairness constraints for the items. Since fairness goals can be application specific, we show how a broad range of fairness constraints can be implemented in our framework, including forms of demographic parity, disparate treatment, and disparate impact constraints. We illustrate the effect of these constraints by providing empirical results on two ranking problems.

CCS CONCEPTS

• **Information systems** → *Probabilistic retrieval models*; *Retrieval effectiveness*; *Presentation of retrieval results*;

KEYWORDS

fairness in rankings, fairness, algorithmic bias, position bias, equal opportunity

1 INTRODUCTION

Rankings have become one of the dominant forms with which online systems present results to the user. Far surpassing their conception in library science as a tool for finding books in a library, the prevalence of rankings now ranges from search engines and online stores, to recommender systems and news feeds. Consequently, it is no longer just books that are being ranked, but there is hardly anything that is *not* being ranked today – products, jobs, job seekers, opinions, potential romantic partners. Nevertheless, one of the guiding technical principles behind the optimization of ranking systems still dates back to four decades ago – namely the Probability Ranking Principle (PRP) [26]. It states that the ideal ranking should order items in the decreasing order of their probability of relevance, since this is the ranking that maximizes utility of the retrieval system to the user for a broad range of common utility measures in Information Retrieval. But is this uncompromising focus on utility to the users still appropriate when we are not ranking books in a library, but people, products and opinions?

There are now substantial arguments and precedent that many of the ranking systems in use today have responsibility not only to their users, but also to the items that are being ranked. In particular, the scarce resource that ranking systems allocate is the exposure of items to users, and exposure is largely determined by position in the ranking – and so is a job applicant's chances to be interviewed by an employer, an AirBnB host's ability to rent out their property, or a writer to be read. Disagreements about a fair allocation of exposure have already led to high-profile legal challenges such as the European Union antitrust violation fine on Google [28], and it has sparked a policy debate about search neutrality [12]. It is unlikely that there will be a universal definition of fairness that is appropriate across all applications, but we give three concrete examples where a ranking system may be perceived as unfair or biased in its treatment of the items that are being ranked, and where the ranking system may want to impose *fairness constraints* that guarantee some notion of fairness.

The main contribution of this paper is a conceptual and computational framework for formulating fairness constraints on rankings, and the associated algorithms for computing utility-maximizing rankings subject to such fairness constraints. This framework provides a flexible way for balancing fairness to the items being ranked with the utility the rankings provide to the users. In this way, we are not limited to a single definition of fairness, since different application scenarios probably require different trade-offs between the rights of the items and what can be considered an acceptable loss in utility to the user. We do show that a broad range of fairness constraints can be implemented in our framework, including forms of demographic parity, disparate treatment, and disparate impact constraints.

To motivate the need and range of situations where one may want to trade-off utility for some notion of fairness, we start with presenting the following three application scenarios. They make use of the concept of protected groups[1], where fairness is related to the differences in how groups are treated. The three examples illustrate how fairness can be related to a biased allocation of opportunity, misrepresentation of real-world distributions, and fairness as a freedom of speech principle.

Example 1: Fairly Allocating Economic Opportunity. Consider a web-service that connects employers (users) to potential employees (items). The following example demonstrates how small differences in item relevance can cause a large difference in exposure

Preprint, 2018,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

[1] Groups that are protected from discrimination by law, based on sex, race, age, disability, color, creed, national origin, or religion. We use a broader meaning of protected groups here that suits our domain.

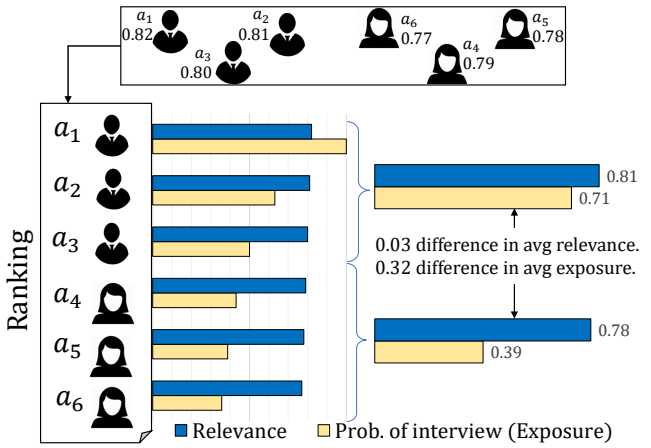


Figure 1: *Job seeker example* to illustrate how small a difference in relevance can lead to a large difference in exposure (an opportunity) for the group of females.

and therefore economic opportunity across groups. In this case, the web-service uses a ranking-based system to present a set of 6 applicants for a software engineering position to relevant employers (Figure 1). The set contains 3 males and 3 females. The male applicants have relevance of 0.80, 0.79, 0.78 respectively for the employers, while the female applicants have relevance of 0.77, 0.76, 0.75 respectively. Here we follow the standard probabilistic definition of relevance, where 0.77 means that 77% of all employers issuing the query find that applicant relevant. The Probability Ranking Principle suggests ranking these applicants in the decreasing order of relevance i.e. the 3 males at the top positions, followed by the females. What does this mean for exposure between the two groups? If we consider a standard exposure drop-off (i.e., position bias) of $1/\log(1+j)$, where j is the position in the ranking, as commonly used in the Discounted Cumulative Gain (DCG) measure, the female applicants will get 30% less exposure than the male applicants – even though the average difference in relevance between male and female applicants is just 0.03 (see Figure 1). Is this winner-take-all allocation of exposure fair in this context, even if the winner just has a tiny advantage in relevance? [2] It seems reasonable to distribute exposure more evenly, even if this may mean a small drop in utility to the employers.

Example 2: Fairly Representing a Distribution of Results. Sometimes the results of a query are used as a statistical sample – either explicitly or implicitly. For example, a user may expect that an image search for the query "CEO" returns roughly the right number of male and female executives, reflecting the true distribution of male vs. female CEOs in the world. If a search engine returns a highly disproportionate number of males as compared to females like in Figure 2, then the search engine may be perceived as biased. In fact, a study detected the presence of gender bias in image search results for a variety of occupations [17]. A biased information environment may affect users' perceptions and behaviors, and it was shown that such biases indeed affect people's belief about

[2]Note that this tiny advantage may come from 3% of the employers being gender biased, but this is not a problem we are addressing here.

Figure 2: Image search result for the query "CEO" showing a disproportionate number of male CEOs [4].

various occupations [17]. Note that the Probability Ranking Principle does not necessarily produce results that represent the relevance distribution in an unbiased way. This means that even if users' relevance distribution agrees with the true distribution of female CEOs, the optimal ranking according to the Probability Ranking Principle may still look like that in Figure 2. Instead of solely relying on the PRP, it seems reasonable to distribute exposure proportional to relevance, even if this may mean a drop in utility to the users.

Example 3: Giving Speakers Fair Access to Willing Listeners. Ranking systems play an increasing role as a medium for speech, creating a connection between bias and fairness in rankings and principles behind freedom of speech [12]. While the ability to produce speech and make this speech available on the internet has certainly created new opportunities to exercise freedom of speech for a speaker, there remains the question whether or not free speech makes its way to the interested listeners. Hence the study of the medium becomes necessary. Search engines are the most popular mediums of this kind and therefore have an immense capability of influencing user attention through their editorial policies, which has sparked a policy debate around search neutrality [11, 12, 14]. While no unified definition of search neutrality exists, many argue that search engines should have no editorial policies other than that their results are comprehensive, impartial, and solely ranked by relevance [24]. But does ranking solely on relevance necessarily imply the Probability Ranking Principle, or are there other relevance-based ranking principles that lead to a medium with a more equitable distribution of exposure and access to willing listeners?

2 RELATED WORK

Before introducing the algorithmic framework for formulating a broad range of fairness constraints on rankings, we first survey three related strands of prior work. First, this paper draws on concepts for algorithmic fairness of supervised learning in the presence of sensitive attributes. Second, we relate to prior work on algorithmic fairness for rankings. Finally, we contrast fairness with the well-studied area of diversified ranking in information retrieval.

2.1 Algorithmic Fairness

As algorithmic techniques, especially machine learning, find widespread applications, there is much interest in understanding its societal impacts. While algorithmic decisions can counteract existing biases, algorithmic and data-driven decision making affords new mechanisms for introducing unintended bias [2]. There have been numerous attempts to define notions of fairness in the supervised learning setting. The individual fairness perspective states that two individuals similar with respect to a task should be classified similarly [10]. Individual fairness is hard to define precisely because of the lack of agreement on task-specific similarity metrics for individuals. There is also a group fairness perspective for supervised learning that implies constraints like demographic parity and equalized odds. Demographic parity posits that decisions should be balanced around a sensitive attribute [5, 35] like gender or race. However, it has been shown that demographic parity causes a loss in the utility and infringes individual fairness [10], since even a perfect predictor typically does not achieve demographic parity. Equalized odds represents the equal opportunity principle for supervised learning and defines the constraint that the false positive and true positive rates should be equal for different protected groups [13]. Several recent works have focused on learning algorithms compatible with these definitions of fair classification [29, 31, 33], including causal approaches to fairness [18, 19, 21]. In this paper, we draw on many of the concepts introduced in the context of fair supervised learning but do not consider the problem of learning. Instead, we ask how to fairly allocate exposure in rankings based on relevance, independent of how these relevances may be estimated.

2.2 Fairness in Rankings

Several recent works have raised the question of group fairness in rankings. Yang and Stoyanovich [30] propose statistical parity based measures that compute the difference in the distribution of different groups for different prefixes of the ranking (top-10, top-20 and so on). The differences are then averaged for these prefixes using a discounted weighting (like in DCG). This measure is then used as a regularization term. Zehlike et al. [32] formulate the problem of finding a 'Fair Top-k ranking' that optimizes utility while satisfying two sets of constraints: first, in-group monotonicity for utility (i.e. more relevant items above less relevant within the group), second, a fairness constraint that the proportion of protected group items in every prefix of the *top-k* ranking is above a minimum threshold. Celis et al. [7] propose a constrained maximum weight matching algorithm for ranking a set of items efficiently under a fairness constraint indicating the maximum number of items with each sensitive attribute allowed in the top positions. Some recent approaches, like Asudeh et al. [1], have also looked at the task of designing fair scoring functions that satisfy desirable fairness constraints.

Most of the fairness constraints defined in the previous work reflect parity constraints restricting the fraction of items with each attribute in the ranking. The framework we propose goes beyond such parity constraints, as we propose a general algorithmic framework for efficiently computing optimal probabilistic rankings for a large class of possible fairness constraints.

2.3 Information Diversity in Retrieval

At first glance, fairness and diversity in rankings may appear related, since they both lead to more diverse rankings. However, their motivation and mechanisms are fundamentally different. Like the PRP, diversified ranking is entirely beholden to maximizing utility to the user, while our approach to fairness balances the needs of users and items. While both PRP and diversified ranking maximize utility for the user alone, their difference lies the utility measure that is maximized. Under extrinsic diversity [22], the utility measure accounts for uncertainty and diminishing returns from multiple relevant results [6, 23]. Under intrinsic diversity [22], the utility measure considers rankings as portfolios and reflects redundancy. And under exploration diversity [22], the aim is to maximize utility to the user in the long term through more effective learning. The work on fairness in this paper is fundamentally different in its motivation and mechanism, as it does not modify the utility measure for the user but instead introduces rights of the items that are being ranked.

3 A FRAMEWORK FOR RANKING UNDER FAIRNESS CONSTRAINTS

Acknowledging the ubiquity of rankings across applications, we conjecture that there is no single definition of what constitutes a fair ranking, but that fairness depends on context and application. In particular, we will see below that different notions of fairness imply different trade-offs in utility, which may be acceptable in one situation but not in the other. To address this range of possible fairness constraints, this section develops a framework for formulating fairness constraints on rankings, and then computing the utility-maximizing ranking subject to these fairness constraints.

For simplicity, consider a single query q and assume that we want to present a ranking r of a set of documents $\mathcal{D} = \{d_1, d_2, d_3 \ldots, d_N\}$. Denoting the utility of a ranking r for query q with $\mathrm{U}(r|q)$, the problem of optimal ranking under fairness constraints can be formulated as the following optimization problem:

$$\begin{aligned} r \quad = \quad & \operatorname{argmax}_r \mathrm{U}(r|q) \\ & \text{s.t. } r \text{ is fair} \end{aligned}$$

In this way, we generalize the goal of the Probabilistic Ranking Principle, which emerges as the special case of no fairness constraints. To fully instantiate and solve this optimization problem, we will specify the following four components. First, we define a general class of utility measures $\mathrm{U}(r|q)$ that contains many commonly used ranking metrics. Second, we address the problem of how to optimize over rankings, which are discrete combinatorial objects, by extending the class of rankings to probabilistic rankings. Third, we reformulate the optimization problem as an efficiently solvable linear program, which implies a convenient yet expressive language for formulating fairness constraints. And, finally, we show how a probabilistic ranking can be efficiently recovered from the solution of the linear program.

3.1 Utility of a Ranking

Virtually all utility measures used for ranking evaluation derive the utility of the ranking from the relevance of the individual items being ranked. For each user u and query q, $\mathrm{rel}(d|u, q)$ denotes the

binary relevance of the document d, i.e. whether the document is relevant to user u or not. Note that different users can have different $\text{rel}(d|u,q)$ even if they share the same q. To account for personalization, we assume that the query q also contains any personalization features and that $\mathcal{U}$ is the set of all users that lead to identical q. Beyond binary relevance, rel could also represent other relevance rating systems such as a Likert scale in movie ratings, or a real-valued score.

A generic way to express many utility measures commonly used in information retrieval is

$$\text{U}(r|q) = \sum_{u \in \mathcal{U}} P(u|q) \sum_{d \in \mathcal{D}} v(\text{rank}(d|r)) \lambda(\text{rel}(d|u,q)),$$

where v and λ are two application-dependent functions. The function $v(\text{rank}(d|r))$ models how much attention document d gets at rank $\text{rank}(d|r)$, and λ is a function that maps the relevance of the document for a user to its utility. In particular, the choice of v could be based on the position bias i.e. the fraction of users who examine the document shown at a particular position out of the total number of users who issue the query q. The choice of λ mapping relevance to utility is somewhat arbitrary. For example, a widely used evaluation measure, Discounted Cumulative Gain (DCG) [15] can be represented in our framework where $v(\text{rank}(d|r)) = \frac{1}{\log(1+\text{rank}(d|r))}$, and $\lambda(\text{rel}(d|u,q)) = 2^{\text{rel}(d|u,q)} - 1$ (or sometimes simply $\text{rel}(d|u,q)$):

$$DCG(r|q) = \sum_{u \in \mathcal{U}} P(u|q) \sum_{d \in \mathcal{D}} \frac{2^{\text{rel}(d|u,q)} - 1}{log(1 + \text{rank}(d|r))}$$

For a measure like $DCG@k(r|q)$, we can choose $v(\text{rank}(d|r)) = \frac{1}{\log(1+\text{rank}(d|r))}$ for $\text{rank}(d|r) \leq k$ and $v(\text{rank}(d|r)) = 0$ for $\text{rank}(d|r) > k$.

Since utility is linear in both v and λ, we can combine the individual utilities into an expectation

$$\begin{aligned}\text{U}(r|q) &= \sum_{d \in \mathcal{D}} v(\text{rank}(d|r)) \Bigg(\sum_{u \in \mathcal{U}} \lambda(rel(d|u,q))\, P(u|q) \Bigg) \\ &= \sum_{d \in \mathcal{D}} v(\text{rank}(d|r)) u(d|q),\end{aligned}$$

where

$$u(d|q) = \sum_{u \in \mathcal{U}} \lambda(rel(d|u,q))\, P(u|q)$$

is the expected utility of a document d for query q. In the case of binary relevances and λ as the identity function, $u(d|q)$ is equivalent to the probability of relevance. It is easy to see that sorting the documents by $u(d|q)$ leads to the ranking that maximizes the utility

$$\text{argmax}_r\, \text{U}(r|q) \equiv \text{argsort}_{d \in \mathcal{D}}\, u(d|q)$$

for any function v that decreases with rank. This is the insight behind the Probability Ranking Principle (PRP) [26].

3.2 Probabilistic Rankings

Rankings are combinatorial objects, such that naively searching the space of all rankings for a utility-maximizing ranking under fairness constraints would take time exponential in $|\mathcal{D}|$. To avoid such combinatorial optimization, we consider probabilistic rankings R instead of a single deterministic ranking r. A probabilistic ranking R is a distribution over rankings, and we can naturally extend the definition of utility to probabilistic rankings.

$$\begin{aligned}\text{U}(R|q) &= \sum_r R(r) \sum_{u \in \mathcal{U}} P(u|q) \sum_{d \in \mathcal{D}} v(\text{rank}(d|r)) \lambda(\text{rel}(d|u,q)) \\ &= \sum_r R(r) \sum_{d \in \mathcal{D}} v(\text{rank}(d|r))\, u(d|q)\end{aligned}$$

While distributions R over rankings are still exponential in size, we can make use of the additional insight that utility can already be computed from the marginal rank distributions of the documents. Let $\mathbf{P}_{i,j}$ be the probability that R places document d_i at rank j, then $\mathbf{P}$ forms a doubly stochastic matrix of size $N \times N$, which means that the sum of each row and each column of the matrix is equal to 1. In other words, the sum of probabilities for each position is 1 and the sum of probabilities for each document is 1, i.e. $\sum_i \mathbf{P}_{i,j} = 1$ and $\sum_j \mathbf{P}_{i,j} = 1$. With knowledge of the doubly stochastic matrix $\mathbf{P}$, expected utility for a probabilistic ranking can be computed as

$$\text{U}(\mathbf{P}|q) = \sum_{d_i \in \mathcal{D}} \sum_{j=1}^{N} \mathbf{P}_{i,j}\, u(d_i|q)\, v(j). \tag{1}$$

To make notation more concise, we can rewrite the utility of the ranking as a matrix product. For this, we introduce two vectors: $\mathbf{u}$ is a column vector of size N with $\mathbf{u}_i = u(d_i|q)$, and $\mathbf{v}$ is another column vector of size N with $\mathbf{v}_j = v(j)$. So, the expected utility (e.g. DCG) can be written as:

$$\text{U}(\mathbf{P}|q) = \mathbf{u}^T \mathbf{P} \mathbf{v} \tag{2}$$

3.3 Optimizing Fair Rankings via Linear Programming

We will see in Section § 3.4 that not only does R imply a doubly stochastic matrix $\mathbf{P}$, but that we can also efficiently compute a probabilistic ranking R for every doubly stochastic matrix $\mathbf{P}$. We can, therefore, formulate the problem of finding the utility-maximizing ranking under fairness constraints in terms of doubly stochastic matrices instead of distributions over rankings.

$$\begin{aligned}\mathbf{P} = \text{argmax}_{\mathbf{P}}\ & \mathbf{u}^T \mathbf{P} \mathbf{v} && \text{(expected utility)} \\ \text{s.t. }\ & \mathbb{1}^T \mathbf{P} = \mathbb{1}^T && \text{(sum of probabilities for each position)} \\ & \mathbf{P}\mathbb{1} = \mathbb{1} && \text{(sum of probabilities for each document)} \\ & 0 \leq \mathbf{P}_{i,j} \leq 1 && \text{(valid probability)} \\ & \mathbf{P} \text{ is fair} && \text{(fairness constraints)}\end{aligned}$$

Note that the optimization objective is linear in N^2 variables $\mathbf{P}_{i,j}$, $1 \leq i,j \leq N$. Furthermore, the constraints ensuring that $\mathbf{P}$ is doubly stochastic are linear as well, where $\mathbb{1}$ is the column vector of size N containing all ones. Without the fairness constraint and for any $\mathbf{v}_j$ that decreases with j, the solution is the permutation matrix that ranks the set of documents in decreasing order of utility (conforming to the PRP).

Now that we have expressed the problem of finding the utility-maximizing probabilistic ranking, besides the fairness constraint, as a linear program, a convenient language to express fairness constraints would be linear constraints of the form

$$\mathbf{f}^T \mathbf{P} \mathbf{g} = h.$$

One or more of such constraints can be added, and the resulting linear program can still be solved efficiently and optimally with standard algorithms like interior point methods. As we will show in Section § 4, the vectors $\mathbf{f}$, $\mathbf{g}$ and the scalar h can be chosen to implement a range of different fairness constraints. To give some intuition, the vector $\mathbf{f}$ can be used to encode group identity and/or relevance of each document, while $\mathbf{g}$ will typically reflect the importance of each position (e.g. position bias).

3.4 Sampling Rankings

The solution $\mathbf{P}$ of the linear program is a matrix containing probabilities of each document at each position. To implement this solution in a ranking system, we need to compute a probabilistic ranking R that corresponds to $\mathbf{P}$. From this probabilistic ranking, we can then sample rankings $r \sim R$ to present to the user[3].

Computing R from $\mathbf{P}$ can be achieved via the Birkhoff-von Neumann (BvN) decomposition [3], which provides a transformation to decompose a doubly stochastic matrix into a convex sum of permutation matrices. In particular, if $\mathbf{A}$ is a doubly stochastic matrix, there exists a decomposition of the form

$$\mathbf{A} = \theta_1 \mathbf{A_1} + \theta_2 \mathbf{A_2} + \cdots + \theta_n \mathbf{A_n}$$

where $1 \leq \theta_i \leq 1$, $\sum_i \theta_i = 1$, and where the $\mathbf{A_i}$ are permutation matrices [3]. In our case, the permutation matrices correspond to deterministic rankings of the document set and the coefficients correspond to the probability of sampling each ranking. According to the Marcus-Ree theorem, there exists a decomposition with no more than $(N-1)^2+1$ permutation matrices [20]. Such a decomposition can be computed efficiently in polynomial time using several algorithms [8, 9]. For the experiments in this paper, we use the implementation provided at https://github.com/jfinkels/birkhoff.

3.5 Summary of Algorithm

The following summarizes the algorithm for optimal ranking under fairness constraints. Note that we have assumed knowledge of the true relevances $u(d|q)$ throughout this paper, whereas in practice one would work with estimates $\hat{u}(d|q)$ from some predictive model.

(1) Set up the utility vector $\mathbf{u}$, the position discount vector $\mathbf{v}$, as well as the vectors $\mathbf{f}$ and $\mathbf{g}$, and the scalar h for the fairness constraints (see Section § 4).
(2) Solve the linear program from Section § 3.3 for $\mathbf{P}$.
(3) Compute the Birkhoff-von Neumann decomposition $\mathbf{P} = \theta_1 \mathbf{P}_1 + \theta_2 \mathbf{P}_2 + \cdots + \theta_n \mathbf{P}_n$.
(4) Sample permutation matrix $\mathbf{P}_i$ with probability proportional to θ_i and display the corresponding ranking r_i.

4 CONSTRUCTING GROUP FAIRNESS CONSTRAINTS

Now that we have established a framework for formulating fairness constraints and optimally solving the associated ranking problem, we still need to understand the expressiveness of constraints of the form $\mathbf{f}^T \mathbf{P} \mathbf{g} = h$. In this section, we explore how three concepts from algorithmic fairness – demographic parity, disparate treatment, and disparate impact – can be implemented in our framework. They all aim to fairly allocate exposure, which we now define formally. Let $\mathbf{v}_j$ represent the importance of position j, or more concretely the position bias at j, which is the fraction of users that examine the item at this position. Then we define exposure for a document d_i under a probabilistic ranking $\mathbf{P}$ as

$$\text{Exposure}(d_i|\mathbf{P}) = \sum_{j=1}^{N} \mathbf{P}_{i,j} \mathbf{v}_j \tag{3}$$

The goal is to allocate exposure fairly between groups G_k. Documents and items may belong to different groups because of some sensitive attributes – for example, news stories belong to different sources, products belong to different manufacturers, applicants belong to different genders. The fairness constraints we will formulate in the following implement different goals for allocating exposure between groups.

To illustrate the effect of the fairness constraints, we will provide empirical results on two ranking problems. For both, we use the average relevance of each document (normalized between 0 and 1) as the utility $\mathbf{u}_i = u(d_i|q)$ and set the position bias to $\mathbf{v}_j = \frac{1}{\log(1+j)}$ just like in the standard definition of DCG. More generally, one can also plug in the actual position-bias value, which can be estimated through an intervention experiment [16].

Job-seeker example. We come back to the job-seeker example from the introduction, and as illustrated in Figure 1. The ranking problem consists of 6 applicants with probabilities of relevance to an employer of $\mathbf{u} = (0.81, 0.80, 0.79, 0.78, 0.77, 0.76)^T$. Groups G_0 and G_1 reflect gender, with the first three applicants belonging to the male group and the last three to the female group.

News recommendation dataset. We use a subset the *Yow* news recommendation dataset [34] to analyze our method on a larger and real-world relevance distribution. The dataset contains explicit and implicit feedback from a set of users for news articles from different RSS feeds. We randomly sample a subset of news articles in the "people" topic coming from the top two sources. The sources are identified using RSS Feed identifier and used as groups G_0 and G_1. The 'relevant' field is used as the measure of relevance for our task. Since the relevance is given as a rating from 1 to 5, we divide it by 5 and add a small amount of Gaussian noise ($\epsilon = 0.05$) to break ties. The resulting $\mathbf{u}_i$ are clipped to lie between 0 and 1.

In the following, we formulate fairness constraints using three ideas for allocation of exposure to different groups. In particular, we will define constraints of the form $\mathbf{f}^T \mathbf{P} \mathbf{g} = h$ for the optimization problem in § 3.3. For simplicity, we will only present the case of a binary valued sensitive attribute i.e. two groups G_0 and G_1. However, these constraints may be defined for each pair of groups and for each sensitive attribute, and be included in the linear program.

4.1 Demographic Parity Constraints

Arguably the simplest way of defining fairness of exposure between groups is to enforce that the average exposure of the documents in both the groups is equal. Denoting average exposure in

[3]For usability reasons, it is preferable to make this sampling pseudo-random based on a hash of the user's identity, so that the same user receives the same ranking r if the same query is repeated.

a group with

$$\text{Exposure}(G_k|\mathbf{P}) = \frac{1}{|G_k|}\sum_{d_i \in G_k} \text{Exposure}(d_i|\mathbf{P}),$$

this can be expressed as the following constraint in our framework:

$$\text{Exposure}(G_0|\mathbf{P}) = \text{Exposure}(G_1|\mathbf{P}) \tag{4}$$

$$\Leftrightarrow \frac{1}{|G_0|}\sum_{d_i \in G_0}\sum_{j=1}^{N} \mathbf{P}_{i,j}\mathbf{v}_j = \frac{1}{|G_1|}\sum_{d_i \in G_1}\sum_{j=1}^{N} \mathbf{P}_{i,j}\mathbf{v}_j \tag{5}$$

$$\Leftrightarrow \sum_{d_i \in \mathcal{D}}\sum_{j=1}^{N}\left(\frac{\mathbb{1}_{d_i \in G_0}}{|G_0|} - \frac{\mathbb{1}_{d_i \in G_1}}{|G_1|}\right)\mathbf{P}_{i,j}\mathbf{v}_j = 0 \tag{6}$$

$$\Leftrightarrow \mathbf{f}^T P\mathbf{v} = 0 \qquad (\text{with } \mathbf{f}_i = \tfrac{\mathbb{1}_{d_i \in G_0}}{|G_0|} - \tfrac{\mathbb{1}_{d_i \in G_1}}{|G_1|})$$

In the last step, we obtain a constraint in the form $\mathbf{f}^T\mathbf{Pg} = h$ which one can plug it into the linear program from Section § 3.3. We call this a *Demographic Parity Constraint* similar to an analogous constraint in fair supervised learning [5, 35]. Similar to that setting, in our case also, such a constraint may lead to a big loss in utility in cases when the two groups are very different in terms of relevance distribution.

Experiments. We solved the linear program in § 3.3 twice – once without and once with the demographic parity constraint from above. For the job seeker example, Figure 3 shows the optimal rankings in terms of $\mathbf{P}$ without and with fairness constraint in panels (a) and (b) respectively. Color indicates the probability value.

Note that the fair ranking according to demographic parity includes a substantial amount of stochasticity. However, panels (c) and (d) show that the fair ranking can be decomposed into a mixture of two deterministic permutation matrices with the associated weights.

Compared to the DCG of the unfair ranking with 3.8193, the optimal fair ranking has slightly lower utility with a DCG of 3.8031. However, the drop in utility due to the demographic parity constraint could be substantially larger. For example, if we lowered the relevances for the female group to $\mathbf{u} = (0.82, 0.81, 0.80, 0.03, 0.02, 0.01)^T$, we would still get the same fair ranking as the current solution, since this fairness constraint is ignorant of relevance. In this ranking, roughly every second document has low relevance, leading to a large drop in DCG. It is interesting to point out that the effect of demographic parity in ranking is therefore analogous to its effect in supervised learning, where it can also lead to a large drop in classification accuracy [10].

We also conducted the same experiment on the news recommendation dataset. Figure 4 shows the optimal ranking matrix and the fair probabilistic ranking along with DCG for each. Note that even though the optimal unfair ranking places documents from G_1 starting at position 5, the constraint pushes the ranking of the news items from G_1 further up the ranking starting either at rank 1 or rank 2. In this case, the optimal fair ranking happens to be (almost) deterministic except at the beginning.

4.2 Disparate Treatment Constraints

Unlike demographic parity, the constraints we explore in this and the following section depend on the relevance of the items being ranked. In this way, these constraints have the potential to address the concerns for the job-seeker example from the introduction, where a small difference in relevance was magnified into a large difference in exposure. Furthermore, we saw that in the image-search example from the introduction that it may be desirable to have exposure be proportional to relevance to achieve some form of unbiased statistical representation. Denoting the average utility of a group with

$$\text{U}(G_k|q) = \frac{1}{|G_k|}\sum_{d_i \in G_k} \mathbf{u}_i,$$

this motivates the following type of constraint, which enforces that exposure of the two groups to be proportional to their average utility.

$$\frac{\text{Exposure}(G_0|\mathbf{P})}{\text{U}(G_0|q)} = \frac{\text{Exposure}(G_1|\mathbf{P})}{\text{U}(G_1|q)}$$

$$\Leftrightarrow \frac{\frac{1}{|G_0|}\sum_{d_i \in G_0}\sum_{j=1}^{N} \mathbf{P}_{i,j}\mathbf{v}_j}{\text{U}(G_0|q)} = \frac{\frac{1}{|G_0|}\sum_{d_i \in G_1}\sum_{j=1}^{N} \mathbf{P}_{i,j}\mathbf{v}_j}{\text{U}(G_1|q)} \tag{7}$$

$$\Leftrightarrow \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{\mathbb{1}_{d_i \in G_0}}{|G_0|\text{U}(G_0|q)} - \frac{\mathbb{1}_{d_i \in G_1}}{|G_1|\text{U}(G_1|q)}\right)\mathbf{P}_{i,j}\mathbf{v}_j = 0 \tag{8}$$

$$\Leftrightarrow \mathbf{f}^T P\mathbf{v} = 0 \qquad \left(\text{with } \mathbf{f}_i = \left(\tfrac{\mathbb{1}_{d_i \in G_0}}{|G_0|\text{U}(G_0|q)} - \tfrac{\mathbb{1}_{d_i \in G_1}}{|G_1|\text{U}(G_1|q)}\right)\right)$$

We name this constraint a *Disparate Treatment Constraint* because allocating exposure to a group is analogous to treating the two groups of documents. This is motivated in principle by the concept of *Recommendations as Treatments* [27], where recommending or exposing a document is considered as treatment and the user's click or purchase is considered the effect of the treatment.

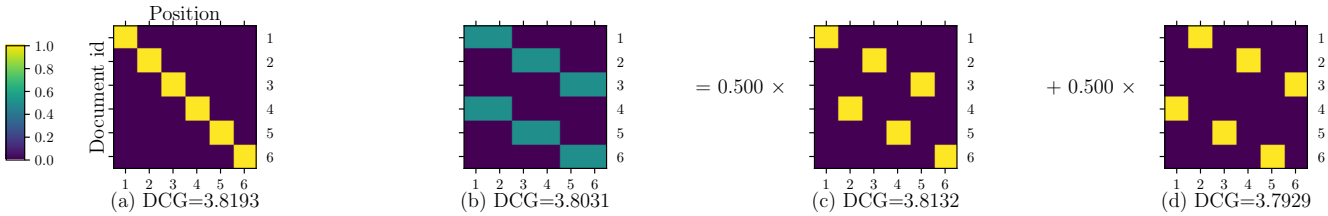
To quantify treatment disparity, we also define a measure called Disparate Treatment Ratio (DTR) to evaluate how unfair a ranking is in this respect i.e. how differently the two groups are treated.

$$\text{DTR}(G_0, G_1|\mathbf{P}, q) = \frac{\text{Exposure}(G_0|\mathbf{P})/\text{U}(G_0|q)}{\text{Exposure}(G_1|\mathbf{P})/\text{U}(G_1|q)}$$

Note that this ratio equals one if the disparate treatment constraint in Equation 7 is fulfilled. Whether the value is less than 1 or greater than 1 tells which group out of G_0 or G_1 is disadvantaged in terms of disparate treatment.

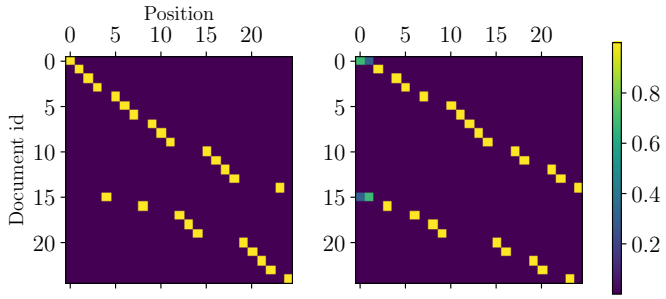
Experiments. We again compute the optimal ranking without fairness constraint, and with the disparate treatment constraint. The results for the job-seeker example are shown in Figure 5. The figure also shows the BvN decomposition of the resultant probabilistic ranking into three permutation matrices. As expected, the fair ranking has an optimal DTR while the unfair ranking has a DTR of 1.7483. Also expected is that the fair ranking has a lower DCG than the optimal deterministic ranking, but that it has higher DCG than the optimal fair ranking under demographic parity.

We conducted the same experiment for the news recommendation dataset. Figure 6 shows the optimal ranking matrix and the fair probabilistic ranking along with DCG for each. Here, the ranking computed without the fairness constraint happened to be almost



(a) DCG=3.8193 (b) DCG=3.8031 (c) DCG=3.8132 (d) DCG=3.7929

Figure 3: Job seeker example with demographic parity constraint. (a) Optimal unfair ranking that maximizes DCG. (b) Optimal fair ranking under demographic parity. (c) and (d) are the BvN decomposition of the fair ranking.



(a) DCG=5.2027 (b) DCG=5.1360

Figure 4: News recommendation dataset with demographic parity constraint. G_0: Document id. 0-14, G_1: 15-24 (a) Optimal unfair ranking that maximizes DCG. (b) Optimal fair ranking under demographic parity.

fair according to disparate treatment already, and the fairness constraint has very little impact on DCG.

4.3 Disparate Impact Constraints

In the previous section, we constrained the exposures (treatments) for the two groups to be proportional to their average utility. However, we may want to go a step further and define a constraint on the impact, i.e. the expected clickthrough or purchase rate, as this more directly reflects the economic impact of the ranking. In particular, we may want to assure that the clickthrough rates for the groups as determined by the exposure and relevance are proportional to their average utility. To formally define this, let us first model the probability of a document getting clicked according to the following simple click model [25]:

$$\begin{aligned} P(\text{click on document } i) &= P(\text{examining } i) \times P(i \text{ is relevant}) \\ &= \text{Exposure}(d_i|\mathbf{P}) \times P(i \text{ is relevant}) \\ &= \left(\sum_{j=1}^{N} \mathbf{P}_{i,j}\mathbf{v}_j\right) \times \mathbf{u}_i \end{aligned}$$

We can now compute the average clickthrough rate of documents in a group G_k as

$$\text{CTR}(G_k|\mathbf{P}) = \frac{1}{|G_k|} \sum_{i \in G_k} \sum_{j=1}^{N} \mathbf{P}_{i,j}\mathbf{u}_i\mathbf{v}_j.$$

The following *Disparate Impact Constraint* enforces that the expected clickthrough rate of each group is proportional to its average utility:

$$\frac{\text{CTR}(G_0|\mathbf{P})}{\text{U}(G_0|q)} = \frac{\text{CTR}(G_1|\mathbf{P})}{\text{U}(G_1|q)} \tag{9}$$

$$\Leftrightarrow \frac{\frac{1}{|G_0|}\sum_{i \in G_0}\sum_{j=1}^{N}\mathbf{P}_{i,j}\mathbf{u}_i\mathbf{v}_j}{\text{U}(G_0|q)} = \frac{\frac{1}{|G_1|}\sum_{i \in G_1}\sum_{j=1}^{N}\mathbf{P}_{i,j}\mathbf{u}_i\mathbf{v}_j}{\text{U}(G_1|q)} \tag{10}$$

$$\Leftrightarrow \sum_{i=1}^{N}\sum_{j=1}^{N}\left(\frac{\mathbb{1}_{d_i \in G_0}}{|G_0|\text{U}(G_0|q)} - \frac{\mathbb{1}_{d_i \in G_1}}{|G_1|\text{U}(G_1|q)}\right)\mathbf{u}_i\mathbf{P}_{i,j}\mathbf{v}_j = 0 \tag{11}$$

$$\Leftrightarrow \mathbf{f}^T P\mathbf{v} = 0 \qquad \left(\text{with } \mathbf{f}_i = \left(\frac{\mathbb{1}_{d_i \in G_0}}{|G_0|\text{U}(G_0|q)} - \frac{\mathbb{1}_{d_i \in G_1}}{|G_1|\text{U}(G_1|q)}\right)\mathbf{u}_i\right)$$

Similar to DTR, we can define the following Disparate Impact Ratio (DIR) to measure the extent to which the disparate impact constraint is violated:

$$\text{DIR}(G_0, G_1|\mathbf{P}, q) = \frac{\text{CTR}(G_0|\mathbf{P})/\text{U}(G_0|q)}{\text{CTR}(G_1|\mathbf{P})/\text{U}(G_1|q)}$$

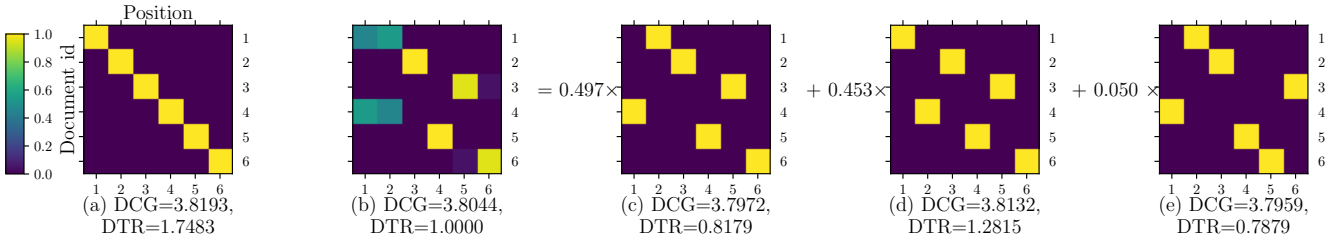
Note that this ratio equals one if the disparate impact constraint in Equation 11 is fulfilled. Similar to DTR, whether DIR is less than 1 or greater than 1 tells which group is disadvantaged in terms of disparate impact.

Experiments. We again compare the optimal rankings with and without the fairness constraint. The results for the job-seeker example are shown in Figure 7. Again, the optimal fair ranking has a BvN decomposition into three deterministic rankings, and it has a slightly reduced DCG. However, there is a large improvement in DIR from the fairness constraint, since the PRP ranking has a substantial disparate impact on the two groups.

The results for the news recommendation dataset are given in Figure 8, where we also see a large improvement in DIR. The DCG is lower than the unconstrained DCG and the DCG with disparate treatment constraint, but higher than the DCG with demographic parity constraint.

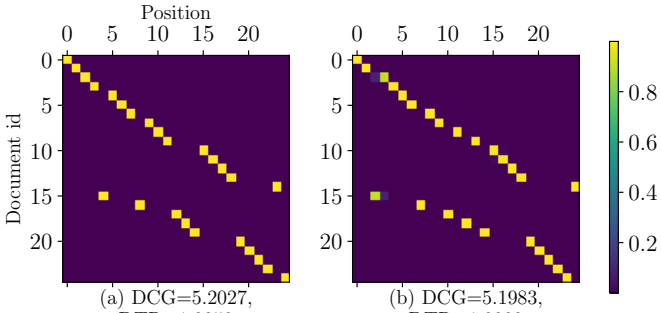
5 DISCUSSION

In the last section, we implemented three fairness constraints in our framework, motivated by the concepts of demographic parity, disparate treatment, and disparate impact. The main purpose was to explore the expressiveness of the framework, and we do not argue that these constraints are the only conceivable ones or the correct ones for a given application. In particular, it appears that fairness in rankings is inherently a trade-off between the utility of the users and the rights of the items that are being ranked, and that different applications require making this trade-off in different



(a) DCG=3.8193, DTR=1.7483 (b) DCG=3.8044, DTR=1.0000 (c) DCG=3.7972, DTR=0.8179 (d) DCG=3.8132, DTR=1.2815 (e) DCG=3.7959, DTR=0.7879

Figure 5: Job seeker example with disparate treatment constraint. (a) Optimal unfair ranking. (b) Fair ranking under disparate treatment constraint. (c), (d), (e) are the BvN decomposition of the fair ranking.



(a) DCG=5.2027, DTR=1.0859 (b) DCG=5.1983, DTR=1.0000

Figure 6: News recommendation dataset with disparate treatment constraint. (a) Optimal unfair ranking. (b) Fair ranking under disparate treatment constraint.

ways. For example, we may not want to convey strong rights to the books in a library when a user is trying to locate a book, but the situation is different when candidates are being ranked for a job opening. We, therefore, focused on creating a flexible framework that covers a substantial range of fairness constraints.

Group fairness vs. individual fairness. In our experiments, we observe that even though the constraints ensure that the rankings have no disparate treatment or disparate impact across groups, individual items within a group might still be considered to suffer from disparate treatment or impact. For example, in the job-seeker experiment for disparate treatment (Figure 5), the allocation of exposure to the candidates within group G_0 still follows the same exposure drop-off going down the ranking that we considered unfair according to the disparate treatment constraint. As a remedy, one could includ additional fairness constraints for other sensitive attributes, like race, disability, and national origin to further refine the desired notion of fairness. However, unless we rank items uniformly at random, it is not clear whether individual fairness can be achieved.

Using estimated utilities. In our definitions and experiments, we assumed that we have access to the true expected utilities (i.e. relevances) $u(d|q)$. In practice, these utilities are typically estimated via machine learning. This learning step is subject to other biases that may, in turn, lead to biased estimates $\hat{u}(d|q)$. Most importantly, biased estimates may be the result of selection biases in click data, but recent counterfactual learning techniques [16] have been shown to permit unbiased learning-to-rank despite biased click data.

Cost of fairness. Including the fairness constraints in the optimization problem comes at the cost of effectiveness as measured by DCG and other conventional measures. This loss in utility can be computed as $CoF = \mathbf{u}^T(\mathbf{P}^* - \mathbf{P})\mathbf{v}$, where $\mathbf{P}^*$ is the deterministic optimal ranking, and $\mathbf{P}$ represents the fair ranking. We have already discussed for the demographic parity constraint that this cost can be substantial. In particular, for demographic parity it is easy to see that the utility of the fair ranking approaches zero if all relevant documents are in one group, and the size of the other group approaches infinity.

Feasibility of fair solutions. The linear program from Section § 3.3 may not have a solution in extreme conditions, corresponding to cases where no fair solution exists. Consider the disparate treatment constraint

$$\frac{\text{Exposure}(G_0|\mathbf{P})}{\text{Exposure}(G_1|\mathbf{P})} = \frac{\text{U}(G_0|q)}{\text{U}(G_1|q)}.$$

We can adversarially construct an infeasible constraint by choosing the relevance so that the ratio on the RHS lies outside the range that LHS can achieve by varying $\mathbf{P}$. The maximum of the RHS occurs when all the documents of G_0 are placed above all the documents of G_1, and vice versa for the minimum.

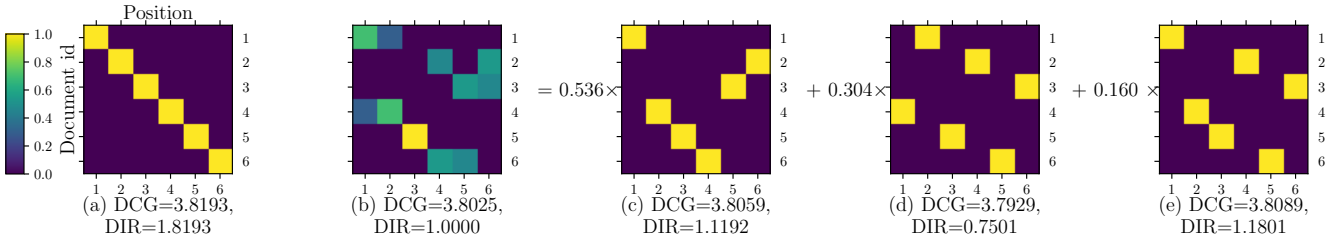
$$\max\left\{\frac{\text{Exposure}(G_0|\mathbf{P})}{\text{Exposure}(G_1|\mathbf{P})}\right\} = \frac{\sum_{j=1}^{|G_0|}\mathbf{v}_j}{\sum_{j=|G_0|+1}^{|G_0|+|G_1|}\mathbf{v}_j},$$
(all G_0 documents in top $|G_0|$ positions)

$$\min\left\{\frac{\text{Exposure}(G_0|\mathbf{P})}{\text{Exposure}(G_1|\mathbf{P})}\right\} = \frac{\sum_{j=|G_1|+1}^{|G_1|+|G_0|}\mathbf{v}_j}{\sum_{j=1}^{|G_1|}\mathbf{v}_j}$$
(all G_0 documents in bottom $|G_0|$ positions)

Hence, a fair ranking according to disparate treatment only exists if the ratio of average utilities lies within the range of possible values for the exposure:

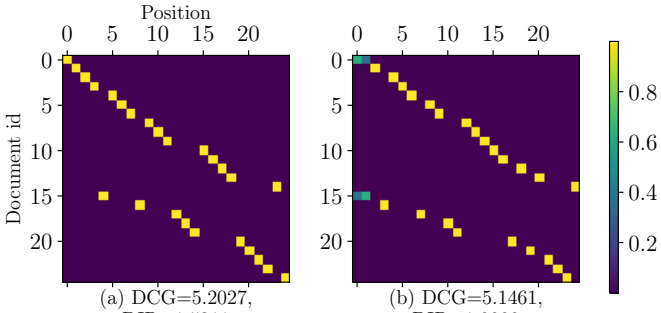
$$\frac{\sum_{j=|G_1|+1}^{|G_1|+|G_0|}\mathbf{v}_j}{\sum_{j=1}^{|G_1|}\mathbf{v}_j} \leq \frac{\text{U}(G_0|q)}{\text{U}(G_1|q)} \leq \frac{\sum_{j=1}^{|G_0|}\mathbf{v}_j}{\sum_{j=|G_0|+1}^{|G_0|+|G_1|}\mathbf{v}_j}$$

However, in such a scenario, the constraint can still be satisfied if we introduce more documents belonging to neither group (or the group with more relevant documents). This increases the range of the LHS, and the ranking doesn't have to give undue exposure to one of the groups.



(a) DCG=3.8193, DIR=1.8193 (b) DCG=3.8025, DIR=1.0000 (c) DCG=3.8059, DIR=1.1192 (d) DCG=3.7929, DIR=0.7501 (e) DCG=3.8089, DIR=1.1801

Figure 7: Job seeker example with disparate impact constraint. (a) Optimal unfair ranking. (b) Fair ranking under disparate impact constraint. (c), (d), (e) are the BvN decomposition of the fair ranking.



(a) DCG=5.2027, DIR=1.5211 (b) DCG=5.1461, DIR=1.0000

Figure 8: News recommendation dataset with disparate impact constraint. (a) Optimal unfair ranking. (b) Fair ranking under disparate impact constraint.

6 CONCLUSIONS

In this paper, we considered fairness of rankings through the lens of exposure allocation between groups. Instead of defining a single notion of fairness, we developed a general framework that employs probabilistic rankings and linear programming to compute the utility-maximizing ranking under a whole class of fairness constraints. To verify the expressiveness of this class, we showed how to express fairness constraints motivated by the concepts of demographic parity, disparate treatment, and disparate impact. We conjecture that the appropriate definition of fair exposure depends on the application, which makes this expressiveness desirable.

REFERENCES

[1] Abolfazl Asudehy, HV Jagadishy, Julia Stoyanovichz, and Gautam Das. 2017. Designing Fair Ranking Schemes. *arXiv preprint arXiv:1712.09752* (2017).

[2] Solon Barocas and Andrew D Selbst. 2016. Big data's disparate impact. *Cal. L. Rev.* 104 (2016), 671.

[3] Garrett Birkhoff. 1940. *Lattice theory*. Vol. 25. American Mathematical Soc.

[4] Amelia Butterly. 2015. Google Image search for CEO has Barbie as first female result. (2015). http://www.bbc.co.uk/newsbeat/article/32332603/google-image-search-for-ceo-has-barbie-as-first-female-result

[5] Toon Calders, Faisal Kamiran, and Mykola Pechenizkiy. 2009. Building classifiers with independency constraints. In *Data mining workshops, ICDMW*. 13–18.

[6] Jaime Carbonell and Jade Goldstein. 1998. The Use of MMR, Diversity-based Reranking for Reordering Documents and Producing Summaries. In *SIGIR*. 335–336. https://doi.org/10.1145/290941.291025

[7] L Elisa Celis, Damian Straszak, and Nisheeth K Vishnoi. 2017. Ranking with Fairness Constraints. *arXiv preprint arXiv:1704.06840* (2017).

[8] Cheng-Shang Chang, Wen-Jyh Chen, and Hsiang-Yi Huang. 1999. On service guarantees for input-buffered crossbar switches: a capacity decomposition approach by Birkhoff and von Neumann. In *IWQoS*. IEEE, 79–86.

[9] Fanny Dufossé and Bora Uçar. 2016. Notes on Birkhoff–von Neumann decomposition of doubly stochastic matrices. *Linear Algebra Appl.* 497 (2016), 108–115.

[10] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *ITCS*. 214–226.

[11] Laura A Granka. 2010. The politics of search: A decade retrospective. *The Information Society* 26, 5 (2010), 364–374.

[12] James Grimmelmann. 2011. Some skepticism about search neutrality. *The Next Digital Decade: Essays on the future of the Internet* (2011), 435. https://ssrn.com/abstract=1742444

[13] Moritz Hardt, Eric Price, and Nati Srebro. 2016. Equality of opportunity in supervised learning. In *NIPS*. 3315–3323.

[14] Lucas D Introna and Helen Nissenbaum. 2000. Shaping the Web: Why the politics of search engines matters. *The information society* 16, 3 (2000), 169–185.

[15] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)* 20, 4 (2002), 422–446.

[16] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2017. Unbiased learning-to-rank with biased feedback. In *WSDM*. 781–789.

[17] Matthew Kay, Cynthia Matuszek, and Sean A. Munson. 2015. Unequal Representation and Gender Stereotypes in Image Search Results for Occupations. In *CHI*. 3819–3828. https://doi.org/10.1145/2702123.2702520

[18] Niki Kilbertus, Mateo Rojas Carulla, Giambattista Parascandolo, Moritz Hardt, Dominik Janzing, and Bernhard Schölkopf. 2017. Avoiding discrimination through causal reasoning. In *NIPS*. 656–666.

[19] Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. 2017. Counterfactual fairness. In *NIPS*. 4069–4079.

[20] Marvin Marcus and Rimhak Ree. 1959. Diagonals of doubly stochastic matrices. *The Quarterly Journal of Mathematics* 10, 1 (1959), 296–302.

[21] Razieh Nabi and Ilya Shpitser. 2017. Fair inference on outcomes. *arXiv preprint arXiv:1705.10378* (2017).

[22] Filip Radlinski, Paul N Bennett, Ben Carterette, and Thorsten Joachims. 2009. Redundancy, diversity and interdependent document relevance. In *ACM SIGIR Forum*, Vol. 43. 46–52.

[23] Filip Radlinski, Robert Kleinberg, and Thorsten Joachims. 2008. Learning diverse rankings with multi-armed bandits. In *ICML*. ACM, 784–791.

[24] Addam Raff. 2009. Search, but You May Not Find. *New York Times* (2009). http://www.nytimes.com/2009/12/28/opinion/28raff.html

[25] Matthew Richardson, Ewa Dominowska, and Robert Ragno. 2007. Predicting Clicks: Estimating the Click-through Rate for New Ads. In *WWW*. 521–530. https://doi.org/10.1145/1242572.1242643

[26] Stephen E Robertson. 1977. The probability ranking principle in IR. *Journal of documentation* 33, 4 (1977), 294–304.

[27] Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. 2016. Recommendations as Treatments: Debiasing Learning and Evaluation. In *ICML*. 1670–1679.

[28] Mark Scott. 2017. Google Fined Record $2.7 Billion in E.U. Antitrust Ruling. *New York Times* (2017). https://www.nytimes.com/2017/06/27/technology/eu-google-fine.html

[29] Blake Woodworth, Suriya Gunasekar, Mesrob I Ohannessian, and Nathan Srebro. 2017. Learning non-discriminatory predictors. *arXiv preprint arXiv:1702.06081* (2017).

[30] Ke Yang and Julia Stoyanovich. 2017. Measuring fairness in ranked outputs. *SSDBM* (2017).

[31] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P Gummadi. 2017. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *WWW*. 1171–1180.

[32] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. FA* IR: A Fair Top-k Ranking Algorithm. *CIKM* (2017).

[33] Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi, and Cynthia Dwork. 2013. Learning fair representations. In *ICML*. 325–333.

[34] yi Zhang. 2005. *Bayesian Graphical Model for Adaptive Information Filtering*. PhD Dissertation. Carnegie Mellon University.

[35] Indre Zliobaite. 2015. On the relation between accuracy and fairness in binary classification. *FATML Workshop at ICML* (2015).