User Fairness, Item Fairness, and Diversity for Rankings in Two-Sided Markets

Lequn Wang
lw633@cornell.edu
Department of Computer Science, Cornell University
Ithaca, New York, USA

Thorsten Joachims
tj@cs.cornell.edu
Department of Computer Science, Cornell University
Ithaca, New York, USA

ABSTRACT

Ranking items by their probability of relevance has long been the goal of conventional ranking systems. While this maximizes traditional criteria of ranking performance, there is a growing understanding that it is an oversimplification in online platforms that serve not only a diverse user population, but also the producers of the items. In particular, ranking algorithms are expected to be fair in how they serve all groups of users — not just the majority group — and they also need to be fair in how they divide exposure among the items. These fairness considerations can partially be met by adding diversity to the rankings, as done in several recent works. However, we show in this paper that user fairness, item fairness and diversity are fundamentally different concepts. In particular, we find that algorithms that consider only one of the three desiderata can fail to satisfy and even harm the other two. To overcome this shortcoming, we present the first ranking algorithm that explicitly enforces all three desiderata. The algorithm optimizes user and item fairness as a convex optimization problem which can be solved optimally. From its solution, a ranking policy can be derived via a novel Birkhoff-von Neumann decomposition algorithm that optimizes diversity. Beyond the theoretical analysis, we investigate empirically on a new benchmark dataset how effectively the proposed ranking algorithm can control user fairness, item fairness and diversity, as well as the trade-offs between them.

CCS CONCEPTS

• **Information systems → Information retrieval diversity**; **Probabilistic retrieval models**; **Retrieval effectiveness**.

KEYWORDS

Fairness, Diversity, Algorithmic Bias, Social Welfare, Ranking, Recommender System, Two-sided Market

ACM Reference Format:

Lequn Wang and Thorsten Joachims. 2021. User Fairness, Item Fairness, and Diversity for Rankings in Two-Sided Markets. In *Proceedings of the 2021 ACM SIGIR International Conference on the Theory of Information Retrieval (ICTIR '21), July 11, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3471158.3472260

ICTIR '21, July 11, 2021, Virtual Event, Canada

ACM ISBN 978-1-4503-8611-1/21/07…$15.00
https://doi.org/10.1145/3471158.3472260

1 INTRODUCTION

We consider ranking problems that involve two-sided markets of producers and consumers. Such two-sided markets are widespread in online platforms — movie producers and audiences in a streaming platform, job seekers and employers in a resume database, or news agencies and information seekers in a news-feed app. In these two-sided markets, the items compete with each other for exposure to the users, while the users gain utility from the recommender system by finding items they like. The platform mediates this market through the ranking algorithm, with great influence on which users get exposed to which items.

Conventional ranking algorithms maximize the average utility to the users by following the probability ranking principle (PRP) [52]. However, there is growing understanding that this is an oversimplification in online platforms that mediate a two-sided market. First, the objective of maximizing the average utility can unfairly marginalize minority user groups, decreasing how useful the ranking system is to them in order to better serve a majority user group [65]. Second, the items compete with each other for exposure to the users, and there is the need to divide the exposure between the items in a fair way. It was shown that maximizing the average utility to the users can be unfair to the items, and that it can lead to winner-takes-all dynamics that amplify existing inequities [56]. Violating user and/or item fairness is not only ethically fraught for many applications, it may also drive users and items from the platform [25], or violate anti-discrimination law [44], anti-trust law [55], or freedom of speech principles [28].

Diversification of search results has often been employed to address these concerns, as well as related issues like super-star economics [41], perpetuation of stereotypes [8, 34], ideological polarization [6] and spread of misinformation [62, 64]. However, while diversification appears related to fairness at first glance, it is not clear whether standard formalizations of diversity [39, 50] actually achieve fairness and vice versa.

In this paper we provide the first theoretical study of the interplay between user fairness, item fairness and diversity for rankings in two-sided markets. To enable this theoretical analysis, we quantify the three desiderata in an intent-aware setup [3, 21, 23] where users have different intents and items have varying relevance to different intents. In particular, we formalize user fairness as an economic social-welfare objective where user groups differ in their intent distributions, and relate this to submodular diversity objectives. For item fairness, we adapt the disparate treatment constraints proposed in [56] for the intent-aware setup to ensure the exposure is fairly allocated to the item groups based on their merit.

Through this theoretical analysis, we show that user fairness, item fairness and diversity are independent goals. Specifically, algorithms that optimize any one of the desiderata can fail to satisfy and even harm the other two.

To address this problem, we present the first ranking algorithm that explicitly enforces all three desiderata — called TSFD Rank for Two-Sided Fairness and Diversity. The algorithm optimizes user and item fairness as a convex optimization problem which can be solved optimally. From its solution, a ranking policy can be derived via a novel Birkhoff-von Neumann decomposition [11] algorithm that optimizes diversity.

In addition to the theoretical analysis, we constructed the first benchmark dataset with annotations for intents, user groups and item groups. On this dataset, we empirically evaluate the proposed TSFD Rank with ablation studies quantifying the dependencies between user fairness, item fairness and diversity.

2 RELATED WORK

As algorithmic techniques, especially machine learning, are increasingly used to make decisions that directly impact people's life, there is growing interest in understanding their societal impact. Many works proposed mathematical desiderata to test algorithmic fairness in binary classification [2, 22, 29, 37]. These desiderata often operationalize definitions of fairness from political philosophy and sociology. We study the societal impact of the less explored problem of ranking which, unlike binary classification, is a structured output prediction problem with an exponentially large output space. Since users have different preferences and items compete for exposure in the rankings, the fairness definitions from binary classification do not directly translate to ranking problems.

Unfairness in rankings typically comes from two sources. Some works focus on the endogenous design of the fair ranking systems [56, 67]. They answer what a fair ranking system is and how to achieve fairness assuming all the system information such as the relevance and the position bias is known. The second source of unfairness are the exogenous factors such as biases in the data [12, 31, 51, 68] and biases during relevance estimation [16, 69]. Some works take both into consideration [66, 72]. We focus on the endogenous design of fair and diverse rankings in two-sided markets, which is orthogonal to exogenous factors.

Most existing works on fairness in rankings consider item fairness. They can be classified into three types: (1) composition-based item fairness which ensures statistical parity of where the items are ranked [5, 19, 20, 27, 60, 67, 71]; (2) pairwise-comparison-based item fairness which aims for statistical parity of pairwise ranking errors between item groups [9, 33, 43]; and (3) merit-exposure-based item fairness which explicitly quantifies the amount of exposure an item gets in a ranking and allocates exposure to the items based on their merit [10, 41, 42, 54, 56, 57]. We adopt the third type of item fairness since (1) unlike composition-based item fairness, it can allocate exposure based on merit; (2) unlike pairwise-comparison-based item fairness, it takes position bias into consideration; and (3) it explicitly quantifies the amount of exposure an item gets in a ranking which enables the quantifiable study of the relationship of item fairness to user fairness and diversity.

Fewer works consider user fairness in rankings and they often consider user unfairness problems that result from achieving item fairness across different queries. Some works [7, 45] propose to fairly share the utility drop among the user groups when achieving item fairness across queries. Patro et al. [46] regard the drastic change of exposure to the items during the policy updates to be unfair, and propose an online update algorithm to smoothly update the policy so that the exposure to the items changes gradually while ensuring a minimum utility for the users during the policy updates. In contrast, we identify user unfairness problems originating from the user intent difference and uncertainty for an individual query, which exists even when we do not consider item fairness. Some works consider user fairness in group recommendation, where a recommendation needs to satisfy a group of users with different preferences [65]. They assume the relevance of each item to each user is known. We model the user preferences and the associated uncertainty in an intent-aware setup.

Diversity in rankings and recommendations also challenges the PRP. The key mechanism behind diversity is to model the utility as a function that is not modular (i.e. linearly additive) in the set of ranked items, but that exhibits a diminishing-returns property — most commonly in the form of a submodular set function [21, 50, 70]. In extrinsic diversity [18, 50, 73], this is used to hedge against the uncertainty about the user's information need; and in intrinsic diversity [23, 49] this is used to model complementarity and substitution in a sense of portfolio optimization. Since we are dealing with uncertain user intents, our goal is to achieve extrinsic diversity for the rankings.

Two-sided platforms are modeled as matchmakers that reduce the friction between the two sides of the market. The key to the success of two-sided platforms is to ensure a critical mass of participants on both sides, since they are in need of each other. Literature in economics [4, 17, 25, 53] focuses on the effect of business strategies, primarily about pricing, on the two-sided markets, but typically does not model the effect that the platform's ranking algorithm has on the interactions between users and items. Recently, some works [1, 15, 58] advocate viewing recommendation problem in the context of two-sided markets and discussed fairness issues on both sides. But neither mathematical definitions nor theoretical characterizations of fairness are provided.

The algorithmic study of two-sided matching markets dates back to the stable matching algorithm analyzed by Gale and Shapley [26]. Some works propose algorithms in this context to select a fair stable matching from a set of stable matchings [35, 40]. Recently, Sühr et al. [61] consider fairness concerns in ride-hailing platforms and propose an online matching algorithm to ensure salary fairness for the drivers amortized over time. In reciprocal recommendation problems, the success is measured by the satisfaction on both sides of the market such as in online dating platforms [47]. We consider problems where the items have no preferences over the users and where there are no supply constraints on the items.

One key aspect of fair rankings is the fair division of the exposure to the users among the items. Fair division [14, 48, 59] has been studied for decades where the goal is to allocate a set of resources to the agents. Two of the classic desiderata for fair division are (1) proportionality i.e. every agent receives its "fair share" of the utility, and (2) envy-freeness i.e. no agent wishes to swap her allocation with another agent. In the proposed TSFD Rank, the optimization of user and item fairness can be thought of as ensuring proportionality for the users and the items, and the optimization for diversity can be interpreted as reducing the envy of the users.

3 RANKING IN A TWO-SIDED MARKET

As a basis for our theoretical analysis, as well as the derivation of the TSFD Rank algorithm, we first formalize the problem of intent-aware ranking in two-sided markets. This includes definitions of utility, user fairness, item fairness and diversity.

3.1 Intent-Aware Setup and Utility

We consider the problem of ranking a set of items $\mathbb{D}^q = \{d_1, d_2, d_3, ...\}$ to present to a user with query q. A query can be a text query (e.g. "Schwarzenegger") or any other context for ranking (e.g. "featured movies today"). In the intent-aware setup [3, 21, 23], each user has an unobserved intent $i \in \mathbb{I}$ that further refines the query (e.g. preferred movie genre). We denote with $\mathcal{I}_{\mathcal{U}^q}$ the intent distribution of the user population $\mathcal{U}^q$ for query q. Each item has varying relevance to different intents (e.g. a movie has varying relevance to different genres), and we denote the relevance of an item d to an intent i as $r(d, i)$. A wide range of existing methods can be used to learn this relevance function, and we merely assume that relevance estimates $r(d, i)$ are available. A ranking σ is a permutation of the set of items and a ranking policy $\pi(\cdot|q)$ is a probability distribution over all possible permutations, where deterministic ranking policies are a special case. Focusing on additive ranking metrics, the utility of a ranking policy π is

$$U(\pi|q) = \mathbb{E}_{i\sim\mathcal{I}_{\mathcal{U}^q},\sigma\sim\pi(\cdot|q)}\left[\sum_{d\in\mathbb{D}^q} e(\sigma(d))r(d,i)\right], \tag{1}$$

where $\sigma(d)$ is the rank of the item d and e maps this rank to the exposure d will receive in the position-based model [24]. Since a user has limited attention for each ranking, we assume the total exposure is bounded i.e. $0 < \sum_m e(m) < \infty$. While (1) involves an expectation in the exponential space of rankings, $U(\pi|q)$ can equivalently be written in terms of a marginal rank probability matrix $\Sigma^{\pi,q}$ with

$$\Sigma^{\pi,q}_{m,n} = \mathbb{E}_{\sigma\sim\pi(\cdot|q)}\left[\mathbb{1}_{\{\sigma(d_m)=n\}}\right] \quad \forall m, n \tag{2}$$

where each entry represents the probability of ranking item d_m at rank n under policy π

$$U(\pi|q) = \left(\mathbf{r}^{\mathcal{U}^q}\right)^\top \Sigma^{\pi,q}\mathbf{e}. \tag{3}$$

$\mathbf{r}^{\mathcal{U}^q}$ is the vector containing the expected relevance of each item to the whole user population with $\mathbf{r}^{\mathcal{U}^q}_m = \mathbb{E}_{i\sim\mathcal{I}_{\mathcal{U}^q}}[r(d_m, i)]$, and $\mathbf{e}$ is the exposure vector with $\mathbf{e}_n = e(n)$. The marginal rank probability matrix is doubly stochastic [11] since the sum of each row and each column is 1, i.e. $\sum_m \Sigma^{\pi,q}_{m,n} = 1$ for all n and $\sum_n \Sigma^{\pi,q}_{m,n} = 1$ for all m.

3.2 User Fairness

Overall utility as defined in Eq. (1) reflects an average over all users. However, different user groups $UG \in \mathbb{UG}$ (e.g. male vs. female users) can have different intent distributions $\mathcal{I}_{\mathcal{UG}^q}$ for a query q, and suboptimal ranking performance for a minority group may get drowned out. Since group membership is typically not known for privacy reasons [30], a ranking policy needs to make sure that it does not unfairly provide disparate levels of utility to the user groups. We define the utility of a ranking policy π for a user group UG as the expected utility for the users in this group

$$\begin{aligned} U(\pi|UG,q) &= \mathbb{E}_{i\sim\mathcal{I}_{\mathcal{UG}^q},\sigma\sim\pi(\cdot|q)}\left[\sum_{d\in\mathbb{D}^q} e(\sigma(d))r(d,i)\right] \\ &= \left(\mathbf{r}^{\mathcal{UG}^q}\right)^\top \Sigma^{\pi,q}\mathbf{e}, \end{aligned} \tag{4}$$

where $\mathbf{r}^{\mathcal{UG}^q}$ is the vector containing the expected relevance of each item to the user group UG with $\mathbf{r}^{\mathcal{UG}^q}_m = \mathbb{E}_{i\sim\mathcal{I}_{\mathcal{UG}^q}}[r(d_m, i)]$. A fair ranking ensures that each user group gets an equitable amount of utility. In economics, the goal of providing an equitable amount of utility across groups is typically formalized through a social-welfare function [63], which is maximized to optimize fairness. We adopt

$$UF(\pi|q) = \sum_{UG\in\mathbb{UG}} \rho^q_{UG} f(U(\pi|UG,q)) \tag{5}$$

as our class of social-welfare functions, where f is an increasing concave function (e.g. log) that models the diminishing return property. This social-welfare objective provides larger return for increasing the utility of a user group with little utility compared to increasing the utility of a user group that already receives a large utility. Thus it encourages the ranking policy to provide more equal utility to each user group. f can be chosen based on application requirements. ρ^q_{UG} denotes the group proportion of UG, i.e. the probability that a user sampled from the whole user distribution $\mathcal{U}^q$ belongs to user group UG. Since $UF(\pi|q)$ is a convex combination of concave functions of $\Sigma^{\pi,q}$, user fairness is a concave function of $\Sigma^{\pi,q}$.

Since the intent distribution of the overall user population is a convex combination of the intent distribution of each user group $\mathcal{I}_{\mathcal{U}^q} = \sum_{UG\in\mathbb{UG}} \rho^q_{UG}\mathcal{I}_{\mathcal{UG}^q}$, user fairness $UF(\pi|q)$ is a lower bound of the overall utility in (1) after transformation through the inverse of the user fairness function f^{-1}

$$\begin{aligned} &f^{-1}(UF(\pi|q)) \\ =&f^{-1}\left(\sum_{UG\in\mathbb{UG}} \rho^q_{UG} f(U(\pi|UG,q))\right) \\ \le&f^{-1}\left(f\left(\sum_{UG\in\mathbb{UG}} \rho^q_{UG} U(\pi|UG,q)\right)\right) \\ =&U(\pi|q). \end{aligned}$$

The inequality holds because f is concave. This shows that maximizing user fairness is maximizing a lower bound of the overall utility from Eq. (1).

3.3 Item Fairness

Fairness to the items is akin to a fair-division problem. Specifically, items compete for exposure to the users, since exposure is a prerequisite for items to derive utility (e.g. revenue) from the ranking. We adopt the disparate treatment constraints proposed in [56] for our theoretical and empirical analysis. The disparate treatment constraints ensure that each item group DG gets an amount of exposure $E(\pi|DG, q)$ that is proportional to its merit $M(DG, q) > 0$

$$\frac{E(\pi|DG_m, q)}{M(DG_m, q)} = \frac{E(\pi|DG_n, q)}{M(DG_n, q)} \quad \forall m, n. \tag{6}$$

Further specifying $E(\pi|DG, q)$ and $M(DG, q)$, the average exposure of an item group is defined as

$$E(\pi|DG,q) = \mathbb{E}_{\sigma\sim\pi(\cdot|q)}\left[\frac{1}{|DG|}\sum_{d\in DG} e(\sigma(d))\right] = \frac{\left(\mathbf{l}^{DG}\right)^\top \Sigma^{\pi,q}\mathbf{e}}{|DG|}, \tag{7}$$

where $\mathbf{l}^{DG}$ is the label vector that denotes whether an item belongs to item group DG with $\mathbf{l}^{DG}_m = \mathbb{1}_{\{d_m \in DG\}}$. For the empirical evaluation, we adopt the average relevance of the items in the item group as the merit function

$$M(DG,q) = \mathbb{E}_{i\sim\mathcal{I}_{\mathcal{U}^q}}\left[\frac{1}{|DG|}\sum_{d\in DG} r(d,i)\right]. \tag{8}$$

In practice, the merit function can be chosen based on application-specific requirements.

Finally, to quantify that items also draw utility from the rankings, we define the utility of an item group DG as

$$U(\pi|DG,q) = \mathbb{E}_{i\sim\mathcal{I}_{\mathcal{U}^q},\sigma\sim\pi(\cdot|q)}\left[\sum_{d\in DG} e(\sigma(d))r(d,i)\right] = \left(\mathbf{l}^{DG}\circ\mathbf{r}^{\mathcal{U}^q}\right)^\top \Sigma^{\pi,q}\mathbf{e}, \tag{9}$$

where $\circ$ denotes the element-wise product. In the position-based click model [24], $U(\pi|DG, q)$ corresponds to the sum of the click-through rates of the items in item group DG under ranking policy π.

3.4 Diversity

The original and dominant motivation for diversity in ranking arises from the uncertainty about the user's intent [18]. To hedge against this uncertainty, a diversified ranking aims to provide utility no matter what the unknown intent of the user is (i.e. extrinsic diversity [49]). To formalize this goal, we first define the utility of a ranking σ for an intent i with an additive metric analogous to the overall utility in Eq. (1) as

$$U(\sigma|i,q) = \sum_{d\in\mathbb{D}^q} e(\sigma(d))r(d,i). \tag{10}$$

Similar to user fairness, the diversity $D(\sigma|q)$ of a ranking σ is typically quantified using an increasing concave function g that encourages each ranking in the ranking policy to cover multiple intents [3, 21, 50, 70]

$$D(\sigma|q) = \mathbb{E}_{i\sim\mathcal{I}_{\mathcal{U}^q}}\left[g(U(\sigma|i,q))\right]. \tag{11}$$

Consequently, for a ranking policy π, the expected diversity is

$$D(\pi|q) = \mathbb{E}_{\sigma\sim\pi(\cdot|q),i\sim\mathcal{I}_{\mathcal{U}^q}}\left[g(U(\sigma|i,q))\right]. \tag{12}$$

Diversity and user fairness differ in two fundamental ways. First, user fairness aggregates over user groups, while diversity aggregates over intents. Second, user fairness amortizes over intents and draws from π as input to the concave function, while diversity takes the expectation over intents after the concave transformation. This adds emphasis on optimizing each individual ranking in the diversity objective. It also implies that diversity $D(\pi|q)$ can not be written as a linear function of $\Sigma^{\pi,q}$. Furthermore, unlike utility and user fairness, two ranking policies π and π' that both produce the same marginal rank probability matrix $\Sigma^{\pi,q} = \Sigma^{\pi',q}$ can have different diversity $D(\pi|q) \neq D(\pi'|q)$.

Similar to user fairness, diversity is also a lower bound on the overall utility from Eq. (1) after transformation with the inverse function g^{-1}

$$\begin{aligned} g^{-1}(D(\pi|q)) &= g^{-1}\left(\mathbb{E}_{i\sim\mathcal{I}_{\mathcal{U}^q},\sigma\sim\pi(\cdot|q)}\left[g(U(\sigma|i,q))\right]\right) \\ &\leq g^{-1}\left(g\left(\mathbb{E}_{i\sim\mathcal{I}_{\mathcal{U}^q},\sigma\sim\pi(\cdot|q)}\left[U(\sigma|i,q)\right]\right)\right) \\ &= U(\pi|q), \end{aligned}$$

where the inequality holds because g is concave. This indicates that maximizing diversity also maximizes a lower bound on the overall utility. Diversity maximization can be expressed as a submodular maximization problem with two matroid constraints [38], and we will optimize it using the standard greedy approximation algorithm in our experiments. For completeness, the algorithm is detailed in the appendix.

4 THEORETICAL ANALYSIS

In this section, we analyze the interplay between utility, user fairness, item fairness and diversity. First, we provide worse-case analyses showing that individual user groups, item groups, or intents can needlessly receive zero utility if their interests are not explicitly represented in the ranking objective. This indicates that user fairness, item fairness and diversity are fundamentally different objectives and achieving one of them does not automatically satisfy another. Second, we develop a new form of utility-efficiency analysis to show that achieving one of user fairness, item fairness and diversity might fail to satisfy the utility efficiency of the others. This suggests that the utility efficiencies of the three desiderata are in conflict with each other and achieving one of them might harm the other two.

4.1 Zero-Utility Analysis

Our zero-utility analysis investigates whether a user group, item group or intent can receive a utility of zero, even if a non-zero solution exists. For clarity, we first define a class of non-degenerate ranking problems, to focus our theoretical analysis on non-degenerate cases where non-zero solutions exist.

Definition 1. *(Non-degenerate ranking problem) A ranking problem represented by a tuple* ($\mathbb{I}$, $\mathbb{UG}$, $\mathbb{DG}$, $\mathbb{D}^q$ $\mathcal{U}^q$, r, e) *is non-degenerate if (1) every user group* UG *has positive group proportion* $\rho^q_{UG} > 0$*; (2) every intent* i *has positive probability mass in the user intent distribution* $\mathcal{I}_{\mathcal{U}^q}(i) > 0$*; (3) for every intent* i*, there exists an item* d *that has positive relevance for the intent* $r(d,i) > 0$*; and (4) for every item group* DG*, there exists an item* $d \in DG$ *such that the expected relevance of the item is positive* $\mathbb{E}_{i\sim\mathcal{I}_{\mathcal{U}^q}}[r(d,i)] > 0$.

For user groups and item groups, we investigate whether every group achieves non-zero utility as defined in Eqs. (4) and (9) under different policies. For the intents, since diversity is a function of individual rankings and each single ranking might not be able to provide non-zero utility for every intent due to limited number of non-zero exposure positions, we define the amount of intent covered by a ranking σ as the amount of intent that has non-zero

Table 1: Summary of zero-utility analysis.

The policy optimizing	Non-zero utility for every user group?	Non-zero utility for every item group?	Rankings cover maximum intent?
Utility	×	×	×
User fairness	✓	×	×
Item fairness	×	✓	×
Diversity	×	×	✓
TSFD Rank	✓	✓	×

utility

$$\sum_{i \in \{i | i \in \mathbb{I}, U(\sigma|i,q) > 0\}} [\mathcal{I}_{\mathcal{U}^q}(i)],$$

and investigate whether each ranking sampled from a ranking policy covers the maximum amount of intent covered by any ranking

$$max_\sigma \sum_{i \in \{i | i \in \mathbb{I}, U(\sigma|i,q) > 0\}} [\mathcal{I}_{\mathcal{U}^q}(i)].$$

We present two example theorems for this zero-utility analysis. Theorem 1 shows that there exist ranking problems where maximizing overall utility needlessly provides zero utility for some user groups.

Theorem 1. *There exist non-degenerate ranking problems such that any ranking policy π maximizing overall utility $U(\pi|q)$ has utility $U(\pi|UG, q) = 0$ for some user group UG.*

Proofs of all theorems are provided in the appendix. The disparate treatment identified in Theorem 1 is not necessary, since the following Theorem 2 shows that directly maximizing user fairness can always ensure non-zero utility for every user group.

Theorem 2. *For any non-degenerate ranking problem, there exists a user fairness function f such that if a ranking policy π maximizes user fairness $UF(\pi|q)$, then every user group has non-zero utility under this ranking policy π.*

Table 1 summarizes the other formal results, which are detailed in the appendix. For the sake of brevity, we use "maximize item fairness" to refer to the more appropriately descriptive "maximize utility subject to the disparate treatment constraints". We provide an intuition of the analysis through the ranking problem in Figure 1.

First, maximizing overall utility can lead to zero utility for a user group and/or for an item group, and it can fail to cover the maximum amount of intent. Since items in d_{3*} have strictly larger expected relevance to the whole user population than all the other items, a ranking policy that maximizes utility will rank items in d_{3*} over all the other items. If only 3 positions have non-zero exposure, items in d_{3*} will occupy all the 3 non-zero exposure positions and thus leave zero exposure for the other items. This leads to a ranking policy with zero utility for UG_1, DG_2, i_1 and i_2. We can easily construct a ranking that covers all the intents by selecting three items that are relevant to the three intents respectively and putting them in the 3 non-zero exposure positions. Thus any ranking sampled from the policy maximizing utility fails to cover the maximum amount of intent, since the ranking will only cover intent i_3.

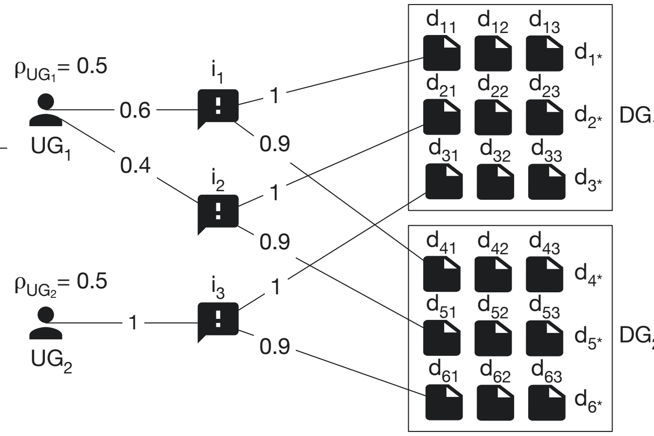

Figure 1: A ranking problem that illustrates the zero-utility analysis. The ranking problem consists of 2 user groups, 3 intents and 2 item groups. 6 sets of items are partitioned into the 2 item groups as indicated by the squares. Each set consists of 3 items with exactly the same relevance to every intent and we denote with d_{m*} the m^{th} set. The numbers on the edges between user groups and intents represent the intent distribution. The numbers on the edges between intents and items represent relevance. For clarity, we omit edges with 0 probability or 0 relevance.

Second, enforcing item fairness can lead to zero utility for a user group and fail to cover the maximum amount of intent. Similarly, maximizing item fairness would rank items in d_{3*} over items in d_{1*}, d_{2*} and rank items in d_{6*} over items in d_{4*}, d_{5*}, since items in d_{3*} and d_{6*} have the largest expected relevance to the whole user population within each item group. If only 3 positions have non-zero exposure, items in d_{3*} and d_{6*} will occupy the 3 non-zero exposure positions. This leads to a ranking policy with zero utility for UG_1, i_1 and i_2. As discussed in the last paragraph, the maximum amount of intent that can be covered by a ranking is 1. Thus maximizing item fairness fails to cover the maximum amount of intent since the rankings sampled from the policy maximizing item fairness only cover i_3.

Third, maximizing user fairness can lead to zero utility for an item group and fail to cover the maximum amount of intent. Again, any ranking policy that maximizes user fairness would rank items in d_{1*} over items in d_{2*}, d_{4*}, d_{5*} and rank items in d_{3*} over items in d_{6*}, since items in d_{1*} and d_{3*} have the largest expected relevance for the two user groups respectively, and all the items have positive relevance to only one user group. If only 3 positions have non-zero exposure, then items in d_{1*} and d_{3*} will occupy all the 3 non-zero exposure positions. This leads to a ranking policy with zero utility for DG_2 and i_2. Thus the rankings sampled from the policy maximizing user fairness fail to cover the maximum amount of intent.

Fourth, maximizing diversity can lead to zero utility for a user group and/or for an item group. If only 1 position has non-zero exposure, maximizing diversity will always put one item from d_{3*} in that non-exposure position since items in d_{3*} have the largest relevance to i_3, the intent with the largest density. This leads to zero utility for DG_2 and UG_1.

This worst-case analysis indicates that it is necessary to optimize each of user fairness, item fairness, and diversity, since any one criterion does not even provide the guarantee of non-zero utility for the others.

4.2 Utility-Efficiency Analysis

Our utility-efficiency analysis investigates if optimizing for one of user fairness, item fairness, or diversity can provide a utility-efficient solution for any of the other desiderata. To answer this kind of questions, we first introduce the precise meaning of utility efficiency for user groups, item groups and intents.

For the user groups, we focus on utility Pareto efficiency of ranking policies. We begin by defining a dominance relation between two policies with respect to a multi-objective optimization problem. The objectives are the utilities of a ranking policy for the user groups $U(\pi|UG, q)$ from Eq. (4).

Definition 2. *(Dominance of ranking policies for the user groups) For a non-degenerate ranking problem, a ranking policy π dominates another ranking policy π' for the user groups $\mathbb{UG}$ if $U(\pi|UG, q) \geq U(\pi'|UG, q)$ for all $UG \in \mathbb{UG}$ and there exists $UG \in \mathbb{UG}$ such that $U(\pi|UG, q) > U(\pi'|UG, q)$.*

The Pareto efficiency of a ranking policy for the user groups is then defined as follows.

Definition 3. *(Pareto efficiency of ranking policies for the user groups) For a non degenerate ranking problem, a ranking policy π is Pareto efficient for the user groups $\mathbb{UG}$ if π is not dominated by any ranking policy π' for $\mathbb{UG}$.*

For the intents, since diversity emphasises the performance of each ranking, we analyze the utility Pareto efficiency of rankings for the intents.

Definition 4. *(Dominance of rankings for the intents) For a non-degenerate ranking problem, a ranking σ dominates another ranking σ' for the intents $\mathbb{I}$ if $U(\sigma|i, q) \geq U(\sigma'|i, q)$ for all $i \in \mathbb{I}$ and there exists $i \in \mathbb{I}$ such that $U(\sigma|i, q) > U(\sigma'|i, q)$.*

Definition 5. *(Pareto efficiency of rankings for the intents) For a non-degenerate ranking problem, a ranking σ is Pareto efficient for the intents $\mathbb{I}$ if σ is not dominated by any ranking σ' for $\mathbb{I}$.*

For the item groups, utility efficiency is achieved when items are ranked by their expected relevance to the whole user population within each item group, since otherwise we can switch the two items that are not ranked by their expected relevance to get larger utility for the item group they belong to without changing the exposure allocation among the item groups.

Definition 6. *(Items ranked by expected relevance within each item group) For a non-degenerate ranking problem and a ranking policy π, the items are ranked by their expected relevance to the whole user population within each item group under π when for any σ with $\pi(\sigma|q) > 0$ and for all $DG \in \mathbb{DG}, d_m, d_n \in DG$, if $\mathbb{E}_{i \sim \mathcal{I}_{\mathcal{U}q}}[r(d_m, i)] > \mathbb{E}_{i \sim \mathcal{I}_{\mathcal{U}q}}[r(d_n, i)]$, then $e(\sigma(d_m)) \geq e(\sigma(d_n))$.*

Achieving utility efficiency can be interpreted as not picking a solution that could easily be improved upon. Thus, if a procedure fails the test of utility efficiency, it clearly provides a suboptimal solution to the user groups, item groups or the intents. We present two example theorems that characterize the utility efficiency of optimizing overall utility, user fairness, item fairness, and diversity on the utility of users, items, and intents. Theorem 3 shows that maximizing user fairness is Pareto efficient for the user groups.

Table 2: Summary of utility efficiency analysis.

The policy optimizing	Ranking policy Pareto efficient for the users?	Items ranked by expected relevance within each item group?	Each ranking Pareto efficient for the intents?
Utility	✓	✓	✓
User fairness	✓	×	✓
Item fairness	×	✓	×
Diversity	×	×	✓

Theorem 3. *For any non-degenerate ranking problem and user fairness function f, if a ranking policy π maximizes user fairness $UF(\pi|q)$, then π is Pareto efficient for the user groups.*

While the solution is utility-efficient for the user groups, the following Theorem 4 shows that this solution is not utility-efficient for the item groups.

Theorem 4. *There exists a ranking problem and a user fairness function f such that items are not ranked by the expected relevance within each item group under any ranking policy π that maximizes user fairness $UF(\pi|q)$.*

We summarize the results of our full utility efficiency analysis in Table 2 and provide the details in the appendix. Maximizing overall utility is the only criterion that ensures utility efficiency for all groups, but the solutions may be poor in terms of fairness or diversity as shown in the zero-utility analysis. Once we explicitly optimize for any one of the fairness or diversity goals, the ranking policy is generally not utility-efficient for the other goals (except that maximizing user fairness ensures utility efficiency for the intents). This implies that the utility efficiency of the three goals are in conflict with each other. Optimizing one of the three desiderata might cause harm or utility drop of the other two. A fair ranking algorithm should make sure the harm or the utility drop is fairly shared among different groups.

5 TSFD RANK: OPTIMIZING RANKINGS FOR FAIRNESS AND DIVERSITY

Driven by the theoretical analysis from the previous section, we now develop the first ranking algorithm — called TSFD Rank for Two-Sided Fairness and Diversity — that explicitly accounts for user fairness, item fairness, and diversity requirements. The algorithm proceeds in two steps. In the first step, it optimally satisfies user fairness and item fairness simultaneously through convex optimization. In the second step, the algorithm maximizes diversity subject to the fairness constraints from the first step.

5.1 Step 1: Convex Optimization for Item and User Fairness

In the first step, we optimize the marginal rank probability matrix representation Σ of the ranking policy to satisfy both user and item fairness. As already shown in Section 3, both user fairness and item fairness can be expressed in terms of Σ, which reduces the optimization problem from the exponential space of rankings to the polynomial space of marginal rank probability matrices. This leads to the following convex optimization problem

$$\begin{aligned} &\mathrm{argmax}_{\Sigma}\ UF(\Sigma|q) \\ &\text{s. t. } \mathbf{1}^\top\Sigma = \mathbf{1}^\top, \Sigma\mathbf{1} = \mathbf{1}, \forall i, j\ 0 \le \Sigma_{i,j} \le 1 \\ &\quad \Sigma \text{ satisfies item-fairness constraints} \end{aligned} \tag{13}$$

where UF is the user-fairness objective expressed in terms of the marginal rank probability matrix Σ and $\mathbf{1}$ is the vector of 1s. We enforce item-fairness in the constraints of the optimization problem, together with the linear constraints that ensure the marginal rank probability matrix Σ is doubly stochastic. As long as the user-fairness objective is concave and the item-fairness constraints are linear in Σ (e.g. the disparate treatment constraints in (6)), the problem can be solved efficiently and globally optimally with convex optimization algorithms [13].

5.2 Step 2: Sampling Diverse Rankings

Since we can not directly sample rankings from Σ, we still need to compute a ranking policy π that has Σ as its matrix of marginal rank probabilities, and thus the desired user and item fairness. For each matrix Σ, there are typically many different policies that produce these marginal rank probabilities. Among those policies, we aim to choose the one that provides maximum diversity. This can be formulated as the following optimization problem.

$$\begin{aligned} &\mathrm{argmax}_{\pi}\ D(\pi|q) \\ &\text{s. t. } \mathbb{E}_{\sigma\sim\pi(\cdot|q)}\left[\mathbb{1}_{\{\sigma(d_m)=n\}}\right] = \Sigma_{m,n} \quad \forall m, n \end{aligned} \tag{14}$$

The constraints in this optimization problem correspond to a Birkhoff-von Neumann (BvN) decomposition [11], for which efficient algorithms exist. We present a novel variant of Birkhoff's algorithm [11] to optimize diversity as illustrated in Algorithm 1. For each round of Birkhoff's algorithm, we find a permutation (ranking) σ that can be sampled from the marginal rank probability matrix Σ. This corresponds to finding a perfect matching σ of the bipartite graph generated from Σ. Then we add this σ to the ranking policy π with selection probability to be the smallest entry in the permutation σ. Then we subtract this selection probability from Σ for all the entries in the permutation. The algorithm is proved to be correct and we can always find a perfect matching from the bipartite graph generated from Σ in each round [11]. What is more, the resulting policy consists of no more than $(n-1)^2+1$ permutations [32] where n is dimension of Σ.

With the additional goal in the objective of constructing a policy that maximizes diversity, we choose the permutation matrices in each step greedily to maximize diversity as detailed in Algorithm 2. Since finding the permutation with the largest diversity that is satisfiable in Σ is NP-hard, we start with the permutation that maximizes utility among the ones that satisfy the conditions of the BvN decomposition. This can be solved by polynomial-time minimum-cost perfect matching algorithms [36]. We then adopt a local search strategy that switches two items if the switch increases diversity. We also tried more expensive search strategies that exhaustively search up to position 3 and found the difference to be small. We present the details of the other search strategy in the appendix.

Note that maximizing diversity reduces the utility variance to the users across the rankings drawn from π. This can be seen as a form of envy reduction [14, 48, 59], where envy measures the individual reduction in utility that a particular user experiences by not drawing the user's optimal-utility ranking from π. To show this, we derive an upper bound $D_{UB}^{\Sigma^{\pi,q}}$ of the diversity as a function of $\Sigma^{\pi,q}$

$$\begin{aligned} D(\pi|q) &= \mathbb{E}_{\sigma\sim\pi(\cdot|q), i\sim\mathcal{I}_{\mathcal{U}^q}}\left[g(U(\sigma|i,q))\right] \\ &\le \mathbb{E}_{i\sim\mathcal{I}_{\mathcal{U}^q}}\left[g\left(\mathbb{E}_{\sigma\sim\pi(\cdot|q)}\left[U(\sigma|i,q)\right]\right)\right] \\ &= \mathbb{E}_{i\sim\mathcal{I}_{\mathcal{U}^q}}\left[g\left(\left(\mathbf{r}^i\right)^\top\Sigma^{\pi,q}\mathbf{e}\right)\right] = D_{UB}^{\Sigma^{\pi,q}}, \end{aligned} \tag{15}$$

Algorithm 1: Greedy Algorithm for BvN Decomposition to Optimize Diversity

```
input  : A ranking problem RP, A diversity function g,
         A marginal rank probability matrix Σ
output : A ranking policy π
initialization: ∀σ π(σ|q) = 0
while Σ! = 0 do
    Construct a bipartite graph G with items and positions
     as vertices and with non-zero elements of Σ as edges.
     σ = Local-Search-Match(RP, g, G)
     π(σ|q) = min_m Σ_{m,σ(d_m)}
     for each item d_m do
        Σ_{m,σ(d_m)} = Σ_{m,σ(d_m)} − π(σ|q)
    end
end
return π
```

Algorithm 2: Local-Search-Match(RP, g, G)

```
input  : A ranking problem RP, a diversity function g
         A bipartite graph G
output : A ranking σ
σ* = find a perfect matching of G that maximizes utility.
 Improved = True
 while Improved do
    Improved = False
     for d_m, d_n in D^q × D^q do
        if (d_n, σ*(d_m)) and (d_m, σ*(d_n)) ∈ G then
            Construct σ′ by switching d_m and d_n in σ*
             if D(σ′|q) > D(σ*|q) then
                σ* = σ′ and Improved = True
            end
        end
    end
end
return σ*
```

where $\mathbf{r}^i$ is the relevance vector to the intent i with $\mathbf{r}^i_m = r(d_m, i)$. The equality holds when, for each user with a particular intent, the utility for that intent is the same across the rankings sampled from the ranking policy — which means that there is no envy of a user that receives a particular ranking to the other rankings that could have been sampled from the ranking policy.

Note that the upper bound is determined by $\Sigma^{\pi,q}$, which is optimized in the first step. The second step maximizes diversity to match this upper bound, which can be interpreted as reducing the envy of the users. This also illustrates a value judgment in the design of TSFD Rank, where we optimize user and item fairness as the primary criteria, and diversity as a secondary one. This is also reflected in Table 1, where TSFD Rank is shown to guarantee non-zero utility to the user and item groups, but not necessarily to cover the maximum amount of intent.

6 EMPIRICAL EVALUATION

In addition to the theoretical characterizations, we now evaluate empirically in how far different ranking algorithms affect utility, user fairness, item fairness and diversity on a movie recommendation dataset.

6.1 Dataset

We constructed the first benchmark dataset that provides intent, user group, and item group annotations. We collected 100 movies from different genres { Romance (20), Comedy (25), Action (25), Thriller (15), Sci-Fi (15) } that are lead by actors of different races {black-lead (20), white-lead (80)}. We treat the genres as the intent set and the leading-actor races as the item group set. The relevance of a movie to a genre is the average user rating on IMDB[1] if the movie belongs to that genre and 0 otherwise. To fully leverage the range of the ratings, we subtract the minimum rating 6 in the dataset from all the ratings to obtain the relevance. For the users, we regard male and female as two user groups and set the user proportion $\rho_{male} \in [0, 1]$ as 0.6 by default and $\rho_{female} = 1 - \rho_{male}$. To enable varying the intent similarity between the two user groups, we arbitrarily construct two dissimilar intent distributions $\mathcal{I}_1 = [0.5, 0.5, 0, 0, 0]$ and $\mathcal{I}_2 = [0, 0, 0.5, 0.25, 0.25]$ over the five genres. We use an intent similarity factor $s \in [0, 1]$ (0.5 by default) to control the intent similarity between the two user groups $\mathcal{I}_{male} = (1-0.5s)\mathcal{I}_1 + 0.5s\mathcal{I}_2$ and $\mathcal{I}_{female} = (1 - 0.5s)\mathcal{I}_2 + 0.5s\mathcal{I}_1$.

6.2 Experiment Setup

All results are averaged over 5,000 samples (50,000 samples for the results in Table 3), where each sample consists of 15 randomly selected movies to be ranked. To make sure the inputs to the merit and diversity functions are within their domains for all the algorithms while there is a clear trade-off between the policies, we set the user fairness function as $f(\cdot) = log(\cdot - 0.6)$ and the diversity function as $g(\cdot) = log(\cdot + 0.0001)$. To control the exposure steepness, we set the exposure function as $e(\cdot) = (\frac{1}{\cdot})^{\eta}$ where η controls the exposure steepness and we set $\eta = 1$ by default. For all the experiments, we use the default parameters introduced in this section unless explicitly stated otherwise.

[1] https://www.imdb.com/

Table 3: performance of different ranking algorithms[3]

The policy optimizing	Utility	Item unfairness	User fairness	Diversity	Diversity UB
Utility	**1.518**	0.186	1.447	1.016	1.016
Item fairness	1.509	**0.000**	1.437	1.010	1.013
User fairness	1.498	0.193	**1.476**	1.052	1.062
Diversity	1.428	0.185	1.390	**1.214**	1.214
TSFD Rank	1.489	**0.000**	1.466	1.045	1.055

We compare TSFD Rank with 4 other policies that maximize utility, item fairness, user fairness, and diversity respectively. For TSFD Rank and the policies that maximize item fairness and user fairness, we first satisfy the fairness goals through convex optimization[2], and then we optimize the diversity by running the greedy BvN decomposition algorithm. We use the greedy submodular optimization approximation algorithm with two matroid constraints to maximize diversity [38]. The algorithm is detailed in the appendix. To avoid cases where the item-fairness constraints can not be satisfied, we optimize the one-sided disparate treatment constraints proposed in [57] in the experiments: $\frac{E(\pi|DG_2)}{M(DG_2,q)} \leq \frac{E(\pi|DG_1)}{M(DG_1,q)}$ with $M(DG_1, q) \leq M(DG_2, q)$.

For clarity of presentation, we bring the user fairness, the diversity and the upper bound on the diversity calculated from the marginal rank probability matrix (diversity UB) on the same scale as the overall utility by applying the inverse of user fairness function and diversity function to each of them to get $f^{-1}(UF(\pi|q))$, $g^{-1}(D(\pi|q))$, and $g^{-1}(D_{UB}^{\Sigma^{\pi,q}})$, where $f^{-1}(\cdot) = e^{\cdot} + 0.6$ and $g^{-1}(\cdot) = e^{\cdot} - 0.0001$. Item unfairness is the amount of violation of the one-sided disparate treatment constraints $\max(0, \frac{E(\pi|DG_2)}{M(DG_2,q)} - \frac{E(\pi|DG_1)}{M(DG_1,q)})$ with $M(DG_1, q) \leq M(DG_2, q)$.

6.3 Empirical Results

How do different methods trade-off between user fairness, item fairness, diversity and utility? We show the empirical results with the default setup in Table 3. As expected, the ranking algorithms that consider only one of the measures excel at that measure but achieve sub-optimal performance on the other ones. In contrast, the proposed TSFD Rank explicitly controls all desiderata by sacrificing some utility to achieve perfect item fairness, second-best user fairness and third-best diversity (very close to the second-best). The diversity upper bound provides a skyline of how much diversity TSFD Rank can possibly achieve. So the small difference between the diversity achieved by the policy that maximizes item fairness, the policy that maximizes user fairness, the policy produced by TSFD Rank and their respective diversity upper bound shows that the greedy BvN decomposition algorithm achieves diversity very close to the upper bound.

How do user intent similarity, user group proportion, and exposure steepness affect user fairness? Figure 2 (a) (b) (c) show the effect of the three factors on the utility ratio between female and male user groups $U_{female}/U_{male} = U(\pi|female, q)/U(\pi|male, q)$, which measures the utility difference between the two user groups. For the policy that maximizes user fairness, the minority (female)

[2] We use MOSEK (https://www.mosek.com/) to solve the convex optimization problem.

[3] The standard error of each value presented in the table is smaller than 0.001.

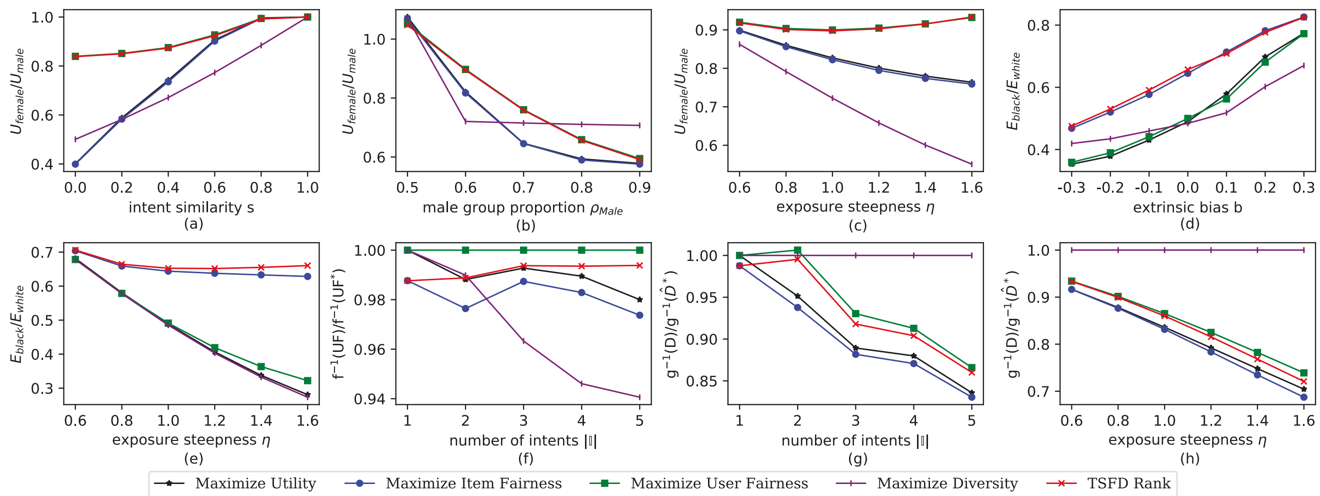

Figure 2: The effects of user intent similarity, user group proportion, extrinsic bias to an item group, exposure steepness and number of intents on the item groups, user groups and intents.

group gets a smaller ratio of utility as the intent similarity decreases. The ratio also decreases as the male group proportion increases, and it stays flat with varying exposure steepness. This is expected since the user fairness objective gives larger weight for the majority group but is oblivious to exposure steepness. The proposed TSFD Rank achieves almost the same ratio as the policy maximizing only user fairness, which shows its effectiveness on fairly distributing the utility drop due to the other desiderata between the two user groups. The policies that maximize item fairness or overall utility amplify the utility drop of the minority (female) user group more than TSFD Rank. The policies that maximize diversity sometimes amplify the utility drop while sometimes over-correcting it.

How do extrinsic bias and exposure steepness affect item fairness? Biased relevance estimates, which might come from biased data, can contribute to unfair exposure allocation to the items [56]. To simulate the bias, for each black-lead movie d of genre i, we set the biased relevance as $r_b(d,i) = (1+b)r(d,i)$ where b is the bias level to the black-lead movies. The results with varying biases are shown in Figure 2 (d). The policy maximizing item fairness ensures roughly a linear change in exposure ratio as the bias increases, which is expected since the exposure ratio is a linear function of the average relevance of black-lead movies, which in turn is a linear function of the bias level b. The proposed TSFD Rank achieves similar exposure ratio as the policy maximizing item fairness, while all the other methods lead to undesirable over-amplifications of the bias towards the less represented black-lead movies. Figure 2 (e) shows that when the exposure steepness increases, both TSFD Rank and the policy maximizing item fairness manage to control the winner-takes-all dynamics while all the other methods fail to ensure a more equitable amount of exposure to the less represented black-lead movies.

How do the number of intents and exposure steepness affect diversity? The diversity ratio $g^{-1}(D)/g^{-1}(\hat{D}^*) = \frac{g^{-1}(D(\pi|q))}{g^{-1}(\hat{D}^*)}$ and the user fairness ratio $f^{-1}(UF)/f^{-1}(UF^*) = \frac{f^{-1}(UF(\pi|q))}{f^{-1}(UF^*)}$ measure how far a policy deviates from the policies that optimize each desideratum where UF^* is the user fairness achieved by the policy maximizing user fairness and $\hat{D}^*$ is the diversity achieved by optimizing diversity by the greedy submodular approximation algorithm. Figure 2 (f) shows that as the number of intents gets larger, maximizing diversity gets further away from maximizing user fairness. Figure 2 (g) and (h) show that as the number of intents gets larger and as the exposure distribution gets steeper, the policies that satisfy other desiderata deviate further from the policy maximizing diversity. Combined with the other empirical findings, these results show that maximizing diversity fails to achieve user or item fairness and vice versa. That TSFD Rank achieves the third-best diversity is expected, since it prioritizes fairness over diversity and only considers diversity in the second step when the marginal rank probability matrix representation $\Sigma^{\pi,q}$ of the ranking policy with a sub-optimal diversity upper bound is already determined.

7 CONCLUSION

We analyzed the interplay between user fairness, item fairness and diversity for rankings in two-sided markets and found that they are three independent and conflicting goals. Driven by the analysis, we proposed TSFD Rank, the first ranking algorithm that explicitly enforces user fairness, item fairness and diversity. TSFD Rank can optimally satisfy user fairness and item fairness through convex optimization and then optimize diversity subject to the fairness constraints via a novel BvN decomposition algorithm. Empirical results on a movie recommendation dataset confirm that TSFD Rank can effectively and robustly control the three desiderata.

ACKNOWLEDGMENTS

This research was supported in part by NSF Awards IIS-1901168 and IIS-2008139. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

REFERENCES

[1] Himan Abdollahpouri, Gediminas Adomavicius, Robin Burke, Ido Guy, Dietmar Jannach, Toshihiro Kamishima, Jan Krasnodebski, and Luiz Pizzato. 2020. Multistakeholder recommendation: Survey and research directions. *User Modeling and User-Adapted Interaction* 30, 1 (2020), 127–158.

[2] Alekh Agarwal, Alina Beygelzimer, Miroslav Dudík, John Langford, and Hanna Wallach. 2018. A reductions approach to fair classification. In *International Conference on Machine Learning*. PMLR, 60–69.

[3] Rakesh Agrawal, Sreenivas Gollapudi, Alan Halverson, and Samuel Ieong. 2009. Diversifying search results. In *Proceedings of the second ACM international conference on web search and data mining*. 5–14.

[4] Mark Armstrong. 2006. Competition in two-sided markets. *The RAND Journal of Economics* 37, 3 (2006), 668–691.

[5] Abolfazl Asudeh, HV Jagadish, Julia Stoyanovich, and Gautam Das. 2019. Designing fair ranking schemes. In *Proceedings of the 2019 International Conference on Management of Data*. 1259–1276.

[6] Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.

[7] Kinjal Basu, Cyrus DiCiccio, Heloise Logan, and Noureddine El Karoui. 2020. A Framework for Fairness in Two-Sided Marketplaces. *arXiv preprint arXiv:2006.12756* (2020).

[8] David N Beede, Tiffany A Julian, David Langdon, George McKittrick, Beethika Khan, and Mark E Doms. 2011. Women in STEM: A gender gap to innovation. *Economics and Statistics Administration Issue Brief* 04-11 (2011).

[9] Alex Beutel, Jilin Chen, Tulsee Doshi, Hai Qian, Li Wei, Yi Wu, Lukasz Heldt, Zhe Zhao, Lichan Hong, Ed H Chi, et al. 2019. Fairness in recommendation ranking through pairwise comparisons. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining(KDD)*. 2212–2220.

[10] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of attention: Amortizing individual fairness in rankings. In *ACM Conference on Research and Development in Information Retrieval (SIGIR)*. 405–414.

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

[12] Léon Bottou, Jonas Peters, Joaquin Quiñonero-Candela, Denis X Charles, D Max Chickering, Elon Portugaly, Dipankar Ray, Patrice Simard, and Ed Snelson. 2013. Counterfactual reasoning and learning systems: The example of computational advertising. *The Journal of Machine Learning Research* 14, 1 (2013), 3207–3260.

[13] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. 2004. *Convex optimization*. Cambridge university press.

[14] Steven J Brams and Alan D Taylor. 1996. *Fair Division: From cake-cutting to dispute resolution*. Cambridge University Press.

[15] Robin Burke. 2017. Multisided fairness for recommendation. In *Workshop on Fairness, Accountability, and Transparency in Machine Learning*.

[16] Robin Burke, Nasim Sonboli, and Aldo Ordonez-Gauger. 2018. Balanced neighborhoods for multi-sided fairness in recommendation. In *Conference on Fairness, Accountability and Transparency*. 202–214.

[17] Bernard Caillaud and Bruno Jullien. 2003. Chicken & egg: Competition among intermediation service providers. *RAND journal of Economics* (2003), 309–328.

[18] Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *ACM SIGIR conference on Research and development in information retrieval*.

[19] L Elisa Celis, Anay Mehrotra, and Nisheeth K Vishnoi. 2020. Interventions for ranking in the presence of implicit bias. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 369–380.

[20] L Elisa Celis, Damian Straszak, and Nisheeth K Vishnoi. 2017. Ranking with fairness constraints. *arXiv preprint arXiv:1704.06840* (2017).

[21] Olivier Chapelle, Shihao Ji, Ciya Liao, Emre Velipasaoglu, Larry Lai, and Su-Lin Wu. 2011. Intent-based diversification of web search results: metrics and algorithms. *Information Retrieval* 14, 6 (2011), 572–592.

[22] Alexandra Chouldechova. 2017. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data* 5, 2 (2017), 153–163.

[23] Charles LA Clarke, Maheedhar Kolla, Gordon V Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. 2008. Novelty and diversity in information retrieval evaluation. In *ACM SIGIR conference on Research and development in information retrieval*.

[24] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An experimental comparison of click position-bias models. In *international conference on web search and data mining*.

[25] David S Evans and Richard Schmalensee. 2016. *Matchmakers: The new economics of multisided platforms*. Harvard Business Review Press.

[26] David Gale and Lloyd S Shapley. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly* 69, 1 (1962), 9–15.

[27] Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. 2019. Fairness-aware ranking in search & recommendation systems with application to linkedin talent search. In *International Conference on Knowledge Discovery & Data Mining*.

[28] James Grimmelmann. 2013. Speech engines. *Minn. L. Rev.* 98 (2013), 868.

[29] Moritz Hardt, Eric Price, and Nati Srebro. 2016. Equality of Opportunity in Supervised Learning. In *Advances in Neural Information Processing Systems*. 3315–3323.

[30] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé III, Miro Dudik, and Hanna Wallach. 2019. Improving fairness in machine learning systems: What do industry practitioners need?. In *CHI Conference on Human Factors in Computing Systems*.

[31] T. Joachims, A. Swaminathan, and T. Schnabel. 2017. Unbiased Learning-to-Rank with Biased Feedback. In *ACM Conference on Web Search and Data Mining (WSDM)*. 781–789.

[32] Diane M Johnson, AL Dulmage, and NS Mendelsohn. 1960. On an algorithm of G. Birkhoff concerning doubly stochastic matrices. *Canad. Math. Bull.* 3, 3 (1960), 237–242.

[33] Nathan Kallus and Angela Zhou. 2019. The fairness of risk scores beyond classification: Bipartite ranking and the xauc metric. In *Advances in Neural Information Processing Systems*. 3438–3448.

[34] Matthew Kay, Cynthia Matuszek, and Sean A Munson. 2015. Unequal representation and gender stereotypes in image search results for occupations. In *ACM Conference on Human Factors in Computing Systems*.

[35] Bettina Klaus and Flip Klijn. 2006. Procedurally fair and stable matching. *Economic Theory* 27, 2 (2006), 431–447.

[36] Jon Kleinberg and Eva Tardos. 2006. *Algorithm design*. Pearson Education India.

[37] Jon M. Kleinberg, Sendhil Mullainathan, and Manish Raghavan. 2017. Inherent Trade-Offs in the Fair Determination of Risk Scores. In *8th Innovations in Theoretical Computer Science Conference, ITCS (LIPIcs, Vol. 67)*. 43:1–43:23.

[38] Andreas Krause and Daniel Golovin. 2014. Submodular function maximization.

[39] Alex Kulesza and Ben Taskar. 2012. Determinantal point processes for machine learning. *arXiv preprint arXiv:1207.6083* (2012).

[40] F Masarani and Sadik S Gokturk. 1989. On the existence of fair matching algorithms. *Theory and Decision* 26, 3 (1989), 305–322.

[41] Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. 2018. Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In *acm international conference on information and knowledge management*.

[42] Marco Morik, Ashudeep Singh, Jessica Hong, and Thorsten Joachims. 2020. Controlling Fairness and Bias in Dynamic Learning-to-Rank. In *ACM Conference on Research and Development in Information Retrieval (SIGIR)*.

[43] Harikrishna Narasimhan, Andrew Cotter, Maya R Gupta, and Serena Wang. 2020. Pairwise Fairness for Ranking and Regression.. In *AAAI*. 5248–5255.

[44] Safiya Umoja Noble. 2018. *Algorithms of oppression: How search engines reinforce racism*. nyu Press.

[45] Gourab K Patro, Arpita Biswas, Niloy Ganguly, Krishna P Gummadi, and Abhijnan Chakraborty. 2020. FairRec: Two-Sided Fairness for Personalized Recommendations in Two-Sided Platforms. In *Web Conference*. 1194–1204.

[46] Gourab K Patro, Abhijnan Chakraborty, Niloy Ganguly, and Krishna Gummadi. 2020. Incremental Fairness in Two-Sided Market Platforms: On Smoothly Updating Recommendations. In *AAAI Conference on Artificial Intelligence*.

[47] Luiz Augusto Sangoi Pizzato, Tomek Rej, Thomas Chung, Irena Koprinska, and Judy Kay. 2010. RECON: a reciprocal recommender for online dating. In *ACM Conference on Recommender Systems(RecSys)*.

[48] Ariel D Procaccia. 2013. Cake cutting: not just child's play. *Commun. ACM* 56, 7 (2013), 78–87.

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

[50] Filip Radlinski, Robert Kleinberg, and Thorsten Joachims. 2008. Learning diverse rankings with multi-armed bandits. In *Proceedings of the 25th international conference on Machine learning*. 784–791.

[51] Bashir Rastegarpanah, Krishna P Gummadi, and Mark Crovella. 2019. Fighting fire with fire: Using antidote data to improve polarization and fairness of recommender systems. In *International Conference on Web Search and Data Mining*.

[52] Stephen E Robertson. 1977. The probability ranking principle in IR. *Journal of documentation* (1977).

[53] Jean-Charles Rochet and Jean Tirole. 2003. Platform competition in two-sided markets. *Journal of the european economic association* 1, 4 (2003), 990–1029.

[54] Piotr Sapiezynski, Wesley Zeng, Ronald E Robertson, Alan Mislove, and Christo Wilson. 2019. Quantifying the Impact of User Attentionon Fair Group Representation in Ranked Lists. In *World Wide Web Conference*.

[55] Mark Scott. 2017. Google Fined Record \$2.7 Billion in E.U. Antitrust Ruling. *New York Times* (2017).

[56] A. Singh and T. Joachims. 2018. Fairness of Exposure in Rankings. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*.

[57] A. Singh and T. Joachims. 2019. Policy Learning for Fairness in Rankings. In *Neural Information Processing Systems (NeurIPS)*.

[58] Andrew Stanton, Akhila Ananthram, Congzhe Su, and Liangjie Hong. 2019. Revenue, Relevance, Arbitrage and More: Joint Optimization Framework for Search Experiences in Two-Sided Marketplaces. *arXiv preprint arXiv:1905.06452* (2019).

[59] H Steihaus. 1948. The problem of fair division. *Econometrica* 16 (1948), 101–104.

[60] Julia Stoyanovich, Ke Yang, and HV Jagadish. 2018. Online set selection with fairness and diversity constraints. In *Proceedings of the EDBT Conference*.

[61] Tom Sühr, Asia J Biega, Meike Zehlike, Krishna P Gummadi, and Abhijnan Chakraborty. 2019. Two-sided fairness for repeated matchings in two-sided markets: A case study of a ride-hailing platform. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 3082–3092.

[62] Behzad Tabibian, Vicenç Gomez, Abir De, Bernhard Schölkopf, and Manuel Gomez Rodriguez. 2020. On the Design of Consequential Ranking Algorithms.

[63] Jan Vondrák. 2008. Optimal approximation for the submodular welfare problem in the value oracle model. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*. 67–74.

[64] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.

[65] Lin Xiao, Zhang Min, Zhang Yongfeng, Gu Zhaoquan, Liu Yiqun, and Ma Shaoping. 2017. Fairness-aware group recommendation with pareto-efficiency. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*. 107–115.

[66] Himank Yadav, Zhengxiao Du, and Thorsten Joachims. 2019. Fair learning-to-rank from implicit feedback. *arXiv preprint arXiv:1911.08054* (2019).

[67] Ke Yang and Julia Stoyanovich. 2017. Measuring fairness in ranked outputs. In *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*. 1–6.

[68] Longqi Yang, Yin Cui, Yuan Xuan, Chenyang Wang, Serge Belongie, and Deborah Estrin. 2018. Unbiased offline recommender evaluation for missing-not-at-random implicit feedback. In *ACM Conference on Recommender Systems*.

[69] Sirui Yao and Bert Huang. 2017. Beyond parity: Fairness objectives for collaborative filtering. In *Advances in Neural Information Processing Systems*. 2921–2930.

[70] Yisong Yue and Carlos Guestrin. 2011. Linear submodular bandits and their application to diversified retrieval. In *Advances in Neural Information Processing Systems*. 2483–2491.

[71] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. Fa* ir: A fair top-k ranking algorithm. In *ACM Conference on Information and Knowledge Management*.

[72] Meike Zehlike and Carlos Castillo. 2020. Reducing disparate exposure in ranking: A learning to rank approach. In *Proceedings of The Web Conference 2020*.

[73] ChengXiang Zhai, William W Cohen, and John Lafferty. 2015. Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In *ACM SIGIR Forum*, Vol. 49. ACM New York, NY, USA, 2–9.