Social Choice for Heterogeneous Fairness in Recommendation

Amanda Aird
amanda.aird@colorado.edu
Department of Information Science,
University of Colorado, Boulder
Boulder, Colorado, USA

Elena Štefancová
elena.stefancova@fmph.uniba.sk
Comenius University Bratislava
Bratislava, Slovakia

Cassidy All
cassidy.all@colorado.edu
Department of Information Science,
University of Colorado, Boulder
Boulder, Colorado, USA

Amy Voida
amy.voida@colorado.edu
Department of Information Science,
University of Colorado, Boulder
Boulder, Colorado, USA

Martin Homola
homola@fmph.uniba.sk
Comenius University Bratislava
Bratislava, Slovakia

Nicholas Mattei
nsmattei@tulane.edu
Department of Computer Science,
Tulane University
New Orleans, Louisiana, USA

Robin Burke
robin.burke@colorado.edu
Department of Information Science,
University of Colorado, Boulder
Boulder, Colorado, USA

ABSTRACT

Algorithmic fairness in recommender systems requires close attention to the needs of a diverse set of stakeholders that may have competing interests. Previous work in this area has often been limited by fixed, single-objective definitions of fairness, built into algorithms or optimization criteria that are applied to a single fairness dimension or, at most, applied identically across dimensions. These narrow conceptualizations limit the ability to adapt fairness-aware solutions to the wide range of stakeholder needs and fairness definitions that arise in practice. Our work approaches recommendation fairness from the standpoint of computational social choice, using a multi-agent framework. In this paper, we explore the properties of different social choice mechanisms and demonstrate the successful integration of multiple, heterogeneous fairness definitions across multiple data sets.

CCS CONCEPTS

• **Information systems → Recommender systems**; • **Computing methodologies → Multi-agent systems**; • **Social and professional topics → User characteristics**.

KEYWORDS

recommender systems, fairness, computational social choice

ACM Reference Format:
Amanda Aird, Elena Štefancová, Cassidy All, Amy Voida, Martin Homola, Nicholas Mattei, and Robin Burke. 2024. Social Choice for Heterogeneous Fairness in Recommendation. In *18th ACM Conference on Recommender Systems (RecSys '24), October 14–18, 2024, Bari, Italy*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3640457.3691706

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
RecSys '24, October 14–18, 2024, Bari, Italy
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0505-2/24/10
https://doi.org/10.1145/3640457.3691706

1 INTRODUCTION

Fairness in algorithmic systems is a complex and multi-faceted area that is of research and practical concern. Different definitions may apply in different contexts, to different stakeholders, and in different types of applications. This complexity is particularly evident in recommender systems [29] where there may be stakeholders of interest on the consumer or provider side of the recommendation interaction, or both [6]. In the literature on fair recommender systems, a plethora of fairness definitions have emerged in the literature [11, 27], each with its own logic and associated algorithmic techniques. What nearly all of this literature has in common is two simplifying assumptions that are severely limiting to the potential applications of these ideas, and a barrier to deploying fair recommender systems in practice [9].

The first limitation is an inability to capture the *multiplicity* of fairness issues. With few exceptions [2, 33, 34], fairness-aware recommender systems apply a single fairness definition to a single protected group, potentially one on each side of the interaction. This limitation is not realistic, as in most application contexts there will be multiple dimensions of fairness that need to be implemented. A second limitation is the assumption that a *single fairness definition* will be appropriate, regardless of the protected group or the area of application. An architecture that assumes fairness will always take a certain form and measure is limited in its applicability [26].

These limitations were addressed by the introduction of the Social Choice for Recommendation Fairness - Dynamic (SCRUF-D) architecture for fairness-aware re-ranking [1, 2, 7]. In this paper, we extend prior results with two agents with the same underlying fairness definition to three agents with differing fairness definitions, showing that the system can support multiple heterogeneous fairness definitions simultaneously. We examine how different social

choice mechanisms offer different trade-offs (and sometimes no trade-off) between fairness and accuracy.

2 RELATED WORK

Fairness in machine learning, especially in classification settings, is a popular topic, including formalizing definitions of fairness [8, 10, 13, 17, 21] and offering algorithmic techniques to mitigate unfairness [18, 25, 35, 36]. However, despite these recent research efforts, the state-of-the-art offers little concrete guidance to industry practitioners [9, 16]. The unique problems of fairness in recommender systems have also been studied; see Ekstrand et al. [11] for an overview. In recommendation, fairness concerns may arise on either the consumer side or on the provider side [6]. Other machine learning environments, such as classification, generally only need to consider the fairness properties of the system relative to the individuals being classified.

Both in the machine learning and the recommender systems formulation of fairness, there has been little recognition of the intersection of multiple fairness definitions and dimensions, although recent work has noted the benefits of combining multiple fairness definitions [3, 24]. Most existing research considers only a single protected class. Even in cases where multiple groups are considered [5, 15, 19, 37], fairness is defined the same way for all groups.

Many of the recent works on fair recommendation take a particular definition of fairness, and train a ranking model that includes fairness as a type of regularization on the subsequent loss function [23, 32, 33]. Such algorithms are brittle to changes in the definitions or measures of fairness. Post processing of the recommendations lists, often called re-ranking, is another popular method [12] and one that we employ here for its ability to allow multiple and changeable definitions of fairness.

3 THE SCRUF-D ARCHITECTURE

The Social Choice for Recommendation Under Fairness - Dynamic (SCRUF-D) platform was introduced in Aird et al. [2], as an extension to a simpler version found in Sonboli et al. [30]. At a high level, SCRUF-D embodies the fairness concerns of multiple stakeholders as a set of *fairness agents* that are *allocated* to a user when recommendations are being generated. Once the agents are allocated to a user, their ordering of items is aggregated with the user's preferences (here coming from a traditional recommendation algorithm) by using a *choice mechanism*, and the result is delivered to the user as their recommendations.

A fairness agent is defined by three functions: a fairness metric which is able to evaluate the history of the system and judge how fair the system has been to the agent over some time window, a compatibility metric which is the agent's evaluation of how much they want to be matched to a particular user, and a ranking function which expresses the agent's ordering/score of items.

When a user arrives at the system, the set of agents express their preference for being matched with this user and their evaluation of how fair the system has been. These scores are passed to an *allocation mechanism* which outputs a (randomized) allocation over the set of agents. Formally this is a probability distribution over the agents, and we examine different allocation functions. Once we have an allocation, the system generates an initial recommendation scoring of the items for the user using any standard recommendation algorithm. Each of the allocated fairness agents also express their ranking/scores of the items to give us a set of lists of items. These lists are then passed through a *choice function*, which aggregates these lists into a final recommendation.

4 FAIRNESS AGENTS

By necessity, our choice of fairness metrics is somewhat arbitrary. While our metrics are informed by research in particular application areas, we have not yet engaged in the full cycle of stakeholder consultation that would be required to formulate specific fairness metrics appropriate to each stakeholder group. Note also that we are examining only provider-side group fairness; consumer-side and individual fairness we leave to future work.

The literature of provider-side fairness measures in recommender systems is extensive; see Ekstrand et al. [11]. For the present study, we make use of fairness metrics that can be tied to a particular protected group. This excludes *global fairness* measures that provide a score for the overall performance of the system relative to a fairness target distribution [1].

In this paper, we explore the following types of fairness metrics. Note that we are not proposing these classes of metrics as appropriate or desirable for any particular application of fairness-aware recommendation. The goal in formulating these metrics is to implement common but very different types of fairness metrics and demonstrate the ability of SCRUF-D to accommodate this heterogeneity across agents:

Group Proportional Fairness (m_{GPF}): Under group proportional fairness, we assume that there is a fixed proportion of recommendation results associated with the protected group that counts as a fair result from that group's perspective. We count the number of items that protected items appear across some set of recommendation lists, divide by the total number of recommendations and normalize by the desired target proportion, truncating at 1 once the target proportion is reached to maintain the 0..1 range.

Group Utility Fairness (m_{GUF}): While m_{GPF} above captures the presence or absence of protected items in a list, it is indifferent to the item's position. However, highly-ranked positions may be of greater utility to providers. Also, m_{GPF} does not take into account the *number* of items that might be in each category. These considerations can be captured by modifying m_{GPF} to sum rank-discounted utilities (here we discount by log_2 of the rank) rather than just count occurrences, and to normalize the utility for protected and unprotected groups by their respective sizes.

Group MRR Fairness (m_{MRR}): The measures above look at all recommended items in a set of lists summing a total value for protected items. In some cases it might be desirable to focus on a minimal degree of representation across recommendation lists, and we capture this type of metric using *mean reciprocal rank*. We average the reciprocal rank of the highest ranking protected item across all lists and normalize by the target MRR value.

[1] An agent seeking fairness with respect to such measures could be implemented within SCRUF-D but it would be using a single definition of fairness on all groups, a different case from what we are considering here.

5 METHODOLOGY

We make use of two data sets MovieLens and Microlending. For each data set, we define three sensitive attributes and protected values and define an agent for each using metrics based on the three different definitions above. The MovieLens 1M dataset [14] contains user ratings for movies with has 3,900 movies, 6,040 users, and approximately 1 million ratings. We consider the sensitive attributes in this dataset to be (1) movies with at least one female writer and/or director, (2) movies with non-English scripts, and (3) older movies (released before 1990). Our second dataset is the Microlending 2017 dataset [28]. This dataset includes 2,673 pseudo-items[2], 4,005 lenders and 110,371 ratings. We consider the sensitive attributes in this dataset to be (1) loans from countries with low funding rates, (2) loans funding sectors that have low funding rates, and (3) loans larger than $5,000.

5.1 Agent Definitions

As noted in Section 3, fairness agents in SCRUF-D are defined by a fairness metric, a compatibility metric, and a ranking function. In this study, we concentrate on the fairness metric.

We define the compatibility between a user and an item feature as the frequency of occurrence of that feature among the items that the user likes. For the Microlending dataset, all funded loans are considered liked; in the Movies dataset, we categorize any movie with a rating over 3 as liked. Compatibility is calculated as $c_{u,f} = p_{u,f}/\bar{p}_f$ where $p_{u,i}$ = count of liked items with feature f/total count of items rated and $\bar{p}_f$ is the average number of items with feature f. The compatibility scores are normalized across features for each user.

For ranking, we use two approaches depending on the type of choice mechanism. For the *Rescoring Mechanism*, we use a simple binary partial order $\mathcal{V}_s > \mathcal{V}_{\neg s}$ where the agent prefers all items that it considers as protected. For our other choice mechanisms, this two-level partial ranking works poorly as it induces many ties, forcing the system to rely on an arbitrary tie-breaking rule. To address this issue, we implement a *cascaded* ranking function that integrates both the protected class preference and inherits, as a secondary ranking criterion, the ranking of the recommender system agent. This is effectively the same as moving all of the protected items to the front of the recommendation list, keeping the between-item preferences the same. This method removes ties, inducing a total order.

For each of our datasets we define three different fairness agents, with three different metrics, to explore the flexibility and trade-offs of SCRUF-D. For the Microlending experiments, the first agent is focused on loans in amounts greater than $5,000, which are considered to have proportionally stronger economic impact but also tend to be funded at a lower rate. This agent aims to have such loans represented across the recommendation lists with MRR target value of 0.5. The second agent is focused on loans from countries with low funding rates. The aim for this agent is to ensure higher utility for loans for countries with historically low funding rates. The third agent focuses on loans from sectors with low funding rates. This agent uses the proportional definition of fairness; the goal is that 20% of all loans recommended are from sectors with historically low funding rates. For the Movie data experiments, the first agent is focused on movies with women writers and directors. This agent is to ensure that a movie with a woman writer and/or director is in the top 1–2 positions of users' lists. The second agent is focused on non-English movies. This agent has a goal of ensuring non-English movies receive equal utility compared to movies in English. The third agent is focused on older movies, with the goal of making 25% of all recommended movies older movies.

[2]The Microlending 2017 dataset represents user preferences relative to clusters of items (pseudo-items) in place of individual loans since loans in this dataset have small numbers of associated users and the unclustered data is too sparse for collaborative recommendation. See [28] for additional details.

5.2 Mechanisms

Instantiating the SCRUF-D system requires choosing mechanisms that will be applied in the allocation and choice phases. There are a wide variety of options for each of these tasks. For the purposes of this study, we concentrate on three allocation mechanisms:

Least Fair: This mechanism allocates a single agent for each recommendation opportunity by comparing the fairness metric m_i for each agent and allocating the agent with the lowest value. If all agents have a fairness of 1.0 (the target), then no agents are allocated to the arriving user(s). This mechanism ignores the compatibility computation, which optimizes fairness but as we show, ignoring the user's preferences results in a greater loss of accuracy.

Lottery: This mechanism allocates only a single agent using a non-deterministic allocation [4] over the set of agents using a probability distribution. The distribution is computed by calculating the product of unfairness (1 - fairness) and compatibility, where each is raised to a small integer power. An agent is to a recommendation opportunity with high probability if fairness need and the compatibility are high, and the exponentiation allows the product to be tuned. In our experiments we set the power of compatibility to 2, which discounts it somewhat compared to fairness. These scores are computed over all agents and normalized to sum to 1.

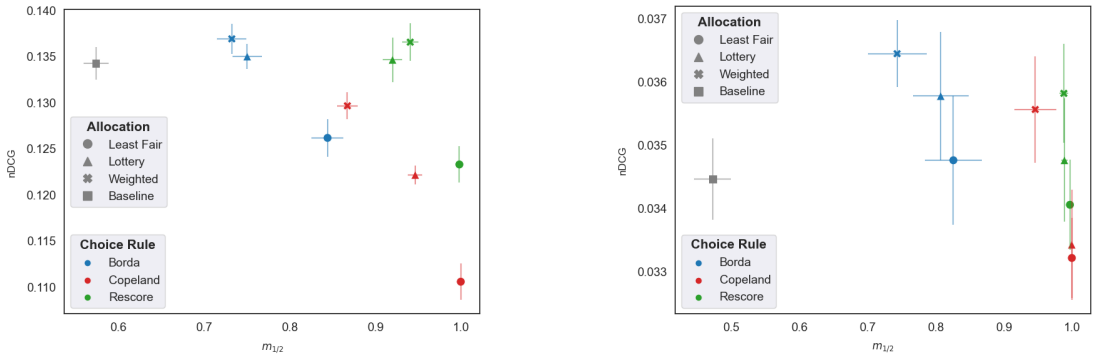
Weighted: The weighted mechanism uses the same distribution as calculated for the Lottery mechanism, but instead of selecting only a single agent, all agents with fairnness < 1.0 are allocated and weighted according to their value in the distribution.

These allocation mechanisms allow us to explore two key aspects of the SCRUF-D platform. First, by comparing Least Fair against the other mechanisms, we can see the value of considering user characteristics in agent allocation. Second, by comparing the Lottery vs Weighted schemes, we can see the difference between sending all of the agents to the choice phase as opposed to allocating a single agent. Since the scoring and re-ranking processes can be computationally intensive, there is an advantage to having only two agents in the choice phase if there is no cost in performance.

In the choice phase, we have again a wide variety of preference aggregation schemes to choose from. SCRUF-D integrates Whalrus[3], a well-known library implementing a variety of voting rules. From the available methods, this paper explores three:

Borda: The Borda method [38] assigns a score to each rank and sums these scores for every item to achieve a final scoring / ranking. To allow our implementation to be tuned for the best trade-off between fairness and accuracy, we use a weighted version

[3]https://github.com/francois-durand/whalrus



(a) Microlending data **(b) Movies data**

Figure 1: nDCG vs $l_{1/2}$ Fairness Norm for Microlending (Left) and Movies (Right).

of the method with the recommender weight set to 0.6. This gives the recommender 1.5x as much weight in the final outcome as compared to the allocated agents.

Copeland: The Copeland mechanism is a pair-wise method. It uses the win-loss record for each pair of items on each the ballot, and awards each item a point per win. These scores are then used to order the items [22]. We use a weighted version; as with Borda, this is realized by multiplying the number of ballots.

Rescoring: The Rescoring mechanism makes use of the scores from the recommendation function, rather than only rankings. A linear combination of the scores from each allocated agent (and the recommender) are computed and the items are reordered based on these values. In our scoring function, an agent contributes a fixed $\delta = 0.5$ to the score for protected items and 0 for unprotected items[4]. The values are weighted first by the agent's allocation weight and then by the inverse of the weight associated with the recommender.

These mechanisms allow us to explore several dimensions of preference aggregation. First, there is the pair-wise versus scoring-based distinction with Copeland, the pair-wise option. Second, there is the difference between ordinal social choice methods and the Rescoring method that makes use of the real-valued predicted ratings arising from the recommendation function.

Layering the allocation and choice mechanisms onto an existing recommender system introduces a small computational overhead (complexity), however, we see the risks of integrating a single fairness definition into the recommender (which would need to be completely retrained for a new metric) as far outweighing the (computational) cost. In terms of the allocation stage, each of the proposed mechanisms can be computed from the lottery history in $O(n)$ time where n is the length of the history to be considered. This computation is bounded by the window size and can scale depending on how reactive we want the system to be. For the choice mechanism, Borda and Rescoring can be computed in $O(n)$ while Copeland is $O(n^2)$ where n is the number of elements in the output list [38]. We expect in most practical applications for both of these to be smaller numbers which do not incur significant overhead.

[4]This δ value was found to provide a good fairness / accuracy trade-off in prior work.

5.3 Evaluation

We evaluate the results of these experiments examining fairness and accuracy. All results were computed over five training / test folds of the data and averaged. Fairness is computed relative to the metric associated with each agent. However, this summative fairness score is over the entire experiment, not the time-bounded window that agents compute over during experiment execution.

We summarize the fairness scores across all the agents using a function of the $l_{1/2}$ norm of over the fairness scores for each individual agent [20]. This metric is maximized (and equal to the mean) when all of the scores are equal, and it is below the mean when the scores are unequal. The value is normalized to $[0, 1]$.

Accuracy is computed using nDCG based on held-out test data in each data set. For the purposes of this study, we did not use any other algorithms as comparators; we are only comparing SCRUF-D's output against the non-re-ranked results. As noted in Section 2, there are very few algorithms that attempt to address multiple fairness concerns simultaneously and only one, OFair [31], is capable of supporting heterogeneous fairness definitions. Prior work found that OFair is not competitive against SCRUF-D [2].

6 RESULTS

Figure 1a shows the results for the Microlending data comparing accuracy (x-axis) with the $l_{1/2}$ fairness norm on the y-axis. (Individual fairness results for each agent are not included for reasons of space.) Although Least Fair has very strong fairness outcomes across different options for the choice mechanism, it suffers from low accuracy. This matches our expectation that taking user compatibility into account allows the system to make a better trade-off.

For the Borda and Rescore choice mechanisms, the Lottery and Weighted allocations look quite similar considering their 95% confidence intervals, suggesting that these allocation mechanisms can give similar results. This is encouraging because the Lottery has the potential for greater efficiency.

Across the choice mechanisms, Borda tends to have the smallest fairness improvement and does a somewhat better job of preserving accuracy. Copeland offers improved fairness, but these choice mechanisms are dominated by Rescore, which achieves better than 0.9 on the $l_{1/2}$ metric without a noticeable loss in accuracy.

Figure 1b shows the positioning of the different mechanism combinations in the accuracy / fairness space for the Movies dataset. We still see Rescoring as dominant and Borda in a lower fairness/higher accuracy position, with Copeland somewhere in the middle. We also see the Rescore and Copeland mechanisms clustered at the far right indicating that they are able to reach the fairness targets across all the agents almost fully.

One interesting difference is the positioning of the Lottery and Weighted mechanisms. In the Microlending data, these mechanisms gave similar results with Lottery a bit lower in accuracy. In the Movies results, the difference is more pronounced and there is greater cost for the Lottery mechanism. On the other hand, the accuracy values are closer to baseline in Movies, showing that high fairness does not have to come at the cost of accuracy loss.

7 CONCLUSION

This paper addresses two key limitations in existing fairness-aware recommendation research that have not been addressed in prior work. First, we formulate multiple definitions of provider-side fairness relevant to real-world datasets. Prior work has been limited to, at most two concerns, usually on different sides of the recommendation interaction. Second, the work posits heterogeneous fairness definitions, allowing different fairness issues to be represented by different metrics. We believe that allowing a multiplicity of concerns and allowing for varied fairness definitions and targets is essential in practical settings.

In the context of computational social choice, we show that SCRUF-D is compatible with a range of different allocation and choice mechanisms and, while some general patterns can be seen, that these mechanisms work differently across recommendation domains and datasets. We find that in some datasets a Lottery mechanism can be competitive with one that allocates multiple fairness agents at a time, suggesting potential efficiency in applying these techniques in practice.

Numerous additional challenges remain. This work has concentrated on provider-side group fairness under multiple definitions. Additional definitions exist and are worth exploring. There is also the question of scale: how many simultaneous agents can be supported? In addition, we believe that SCRUF-D is capable of supporting individual fairness and consumer-side fairness, but additional work needs to be done to demonstrate this capacity.

ACKNOWLEDGMENTS

Burke, Voida and Aird were supported by the National Science Foundation under grant awards IIS-1911025 and IIS-2107577. Mattei was supported by NSF Grant IIS-2107505. Stefancova was supported by Slovak Research and Development Agency under Contract no. APVV-20-0353 and the Fulbright program.

REFERENCES

[1] Amanda Aird, Cassidy All, Paresha Farastu, Elena Stefancova, Joshua Sun, Nicholas Mattei, and Robin Burke. 2023. Exploring Social Choice Mechanisms for Recommendation Fairness in SCRUF. Presented at the 2023 FAccTRec Workshop on Responsible Recommendation.. arXiv:2309.08621 [cs.IR] https://arxiv.org/abs/2309.08621

[2] Amanda Aird, Paresha Farastu, Joshua Sun, Elena Štefancová, Cassidy All, Amy Voida, Nicholas Mattei, and Robin Burke. 2024. Dynamic fairness-aware recommendation through multi-agent social choice. arXiv:2303.00968 [cs.AI] https://arxiv.org/abs/2303.00968

[3] Alex Beutel, Jilin Chen, Tulsee Doshi, Hai Qian, Li Wei, Yi Wu, Lukasz Heldt, Zhe Zhao, Lichan Hong, Ed H. Chi, and Cristos Goodrow. 2019. Fairness in Recommendation Ranking through Pairwise Comparisons. , 16 pages. arXiv:1903.00780 [cs.CY] https://arxiv.org/abs/1903.00780

[4] Felix Brandt. 2017. Rolling the dice: Recent results in probabilistic social choice. In *Trends in Computational Social Choice*, Ulle Endriss (Ed.). AI Access, Chapter 1, 3–26.

[5] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency*. ACM, 77–91.

[6] Robin Burke. 2017. Multisided Fairness for Recommendation. In *Workshop on Fairness, Accountability and Transparency in Machine Learning (FATML)*. Halifax, Nova Scotia, 5 pages. https://arxiv.org/abs/1707.00093

[7] Robin Burke, Nicholas Mattei, Vladislav Grozin, Amy Voida, and Nasim Sonboli. 2022. Multi-agent Social Choice for Dynamic Fairness-aware Recommendation. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. ACM, 234–244.

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

[9] Henriette Cramer, Kenneth Holstein, Jennifer W. Vaughan, Hal Daumé III, Miroslav Dudík, Hanna Wallach, Sravana Reddy, and Jean Garcia-Gathright. 2019. Challenges of incorporating algorithmic fairness into industry practice.

[10] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*. ACM, 214–226.

[11] Michael D. Ekstrand, Anubrata Das, Robin Burke, and Fernando Diaz. 2022. Fairness in Information Access Systems. arXiv:2105.05779 [cs.IR]

[12] Andres Ferraro, Xavier Serra, and Christine Bauer. 2021. Break the loop: Gender imbalance in music recommenders. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval*. 249–254.

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

[14] F Maxwell Harper and Joseph A Konstan. 2015. The MovieLens Datasets: History and Context. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 5, 4 (2015), 19.

[15] Úrsula Hébert-Johnson, Michael Kim, Omer Reingold, and Guy Rothblum. 2018. Multicalibration: Calibration for the (Computationally-identifiable) masses. In *International Conference on Machine Learning*. 1944–1953.

[16] Kenneth Holstein, Jennifer W. Vaughan, Hal Daumé III, Miroslav Dudík, and Hanna Wallach. 2019. Improving fairness in machine learning systems: What do industry practitioners need?. In *Proceedings of the Conference of Human Computer Interaction (CHI'19)*. ACM, 16 pages. arXiv:1812.05239v2[cs.HC]7Jan2019

[17] Ben Hutchinson and Margaret Mitchell. 2019. 50 years of test (un) fairness: Lessons for machine learning. In *Proceedings of the conference on fairness, accountability, and transparency*. ACM, 49–58.

[18] Faisal Kamiran, Toon Calders, and Mykola Pechenizkiy. 2010. Discrimination aware decision tree learning. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 869–874.

[19] Michael Kearns, Seth Neel, Aaron Roth, and Zhiwei Steven Wu. 2018. Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness. arXiv:1711.05144 [cs.LG]

[20] 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 *Proceedings of the Conference on Information and Knowledge Management*. 2243–2251.

[21] Arvind Narayanan. 2018. Translation tutorial: 21 fairness definitions and their politics. In *Proc. Conf. Fairness Accountability Transp., New York, USA*, Vol. 1170. ACM, 3.

[22] Eric Pacuit. 2019. Voting Methods. In *The Stanford Encyclopedia of Philosophy* (Fall 2019 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University.

[23] 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 *Proceedings of The Web Conference 2020*. 1194–1204.

[24] Gourab K Patro, Lorenzo Porcaro, Laura Mitchell, Qiuyue Zhang, Meike Zehlike, and Nikhil Garg. 2022. Fair ranking: a critical review, challenges, and future directions. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 1929–1942.

[25] Dino Pedreshi, Salvatore Ruggieri, and Franco Turini. 2008. Discrimination-aware data mining. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 560–568.

[26] Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency*. ACM, New York, 59–68.

[27] Jessie J Smith, Lex Beattie, and Henriette Cramer. 2023. Scoping Fairness Objectives and Identifying Fairness Metrics for Recommender Systems: The Practitioners' Perspective. In *Proceedings of the ACM Web Conference 2023*. 3648–3659.

[28] Nasim Sonboli, Amanda Aird, and Robin Burke. 2022. *Microlending 2017 Data Set*. University of Colorado Boulder. https://doi.org/10.25810/PGJK-RR19

[29] Nasim Sonboli, Robin Burke, Michael Ekstrand, and Rishabh Mehrotra. 2022. The multisided complexity of fairness in recommender systems. *AI magazine* 43, 2 (2022), 164–176.

[30] Nasim Sonboli, Robin Burke, Nicholas Mattei, Farzad Eskandanian, and Tian Gao. 2020. "And the Winner Is...": Dynamic Lotteries for Multi-group Fairness-Aware Recommendation. Presented at the 2020 FAccTRec Workshop on Responsible Recommendation. arXiv:2009.02590 [cs.IR]

[31] Nasim Sonboli, Farzad Eskandanian, Robin Burke, Weiwen Liu, and Bamshad Mobasher. 2020. Opportunistic Multi-Aspect Fairness through Personalized Re-Ranking. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) *(UMAP '20)*. Association for Computing Machinery, New York, NY, USA, 239–247. https://doi.org/10.1145/3340631.3394846

[32] 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 *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 3082–3092.

[33] Haolun Wu, Chen Ma, Bhaskar Mitra, Fernando Diaz, and Xue Liu. 2022. A multi-objective optimization framework for multi-stakeholder fairness-aware recommendation. *ACM Transactions on Information Systems* 41, 2 (2022), 1–29.

[34] Meike Zehlike, Tom Sühr, Ricardo Baeza-Yates, Francesco Bonchi, Carlos Castillo, and Sara Hajian. 2022. Fair Top-k Ranking with multiple protected groups. *Information Processing & Management* 59, 1 (2022), 102707.

[35] Rich Zemel, Yu Wu, Kevin Swersky, Toni Pitassi, and Cynthia Dwork. 2013. Learning fair representations. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*. 325–333.

[36] Lu Zhang and Xintao Wu. 2017. Anti-discrimination learning: a causal modeling-based framework. *International Journal of Data Science and Analytics* 4 (2017), 1–16.

[37] Ziwei Zhu, Xia Hu, and James Caverlee. 2018. Fairness-aware tensor-based recommendation. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 1153–1162.

[38] William S. Zwicker. 2016. Introduction to the Theory of Voting. In *Handbook of Computational Social Choice*, Felix Brandt, Vincent Conitzer, Ulle Endriss, Jérôme Lang, and Ariel D. Procaccia (Eds.). Cambridge University Press, 23–56. https://doi.org/10.1017/CBO9781107446984.003