Elizabeth Gómez, David Contreras, Maria Salamó, and Ludovico Boratto. Bringing Equity to Coarse and Fine-Grained Provider Groups in Recommender Systems. In Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP '24), July 1–4, 2024, Cagliari, Italy. ACM, New York, NY, USA, 2024.

http://dx.doi.org/10.1145/3627043.3659552

Bringing Equity to Coarse and Fine-Grained Provider Groups in Recommender Systems

ELIZABETH GÓMEZ, Facultat de Matemàtiques i Informàtica, Universitat de Barcelona, Spain
DAVID CONTRERAS, Facultad de Ingeniería y arquitectura, Universidad Arturo Prat, Chile
MARIA SALAMÓ, Facultat de Matemàtiques i Informàtica, Universitat de Barcelona, Spain
LUDOVICO BORATTO, Department of Mathematics and Computer Science, University of Cagliari, Italy

Provider fairness aims at regulating the recommendation lists, so that the items of different providers/provider groups are suggested by respecting notions of equity. When group fairness is among the goals of a system, a common way is to use coarse groups since the number of considered provider groups is usually small (e.g., two genders, or three/four age groups) and the number of items per group is large. From a practical point of view, having few groups makes it easier for a platform to manage the distribution of equity among them. Nevertheless, there are sensitive attributes, such as the age or the geographic provenance of the providers that can be characterized at a fine granularity (e.g., one might group providers at the country level, instead of the continent one), which increases the number of groups and decrements the number of items per group. In this study, we show that, in large demographic groups, when considering coarse-grained provider groups, the fine-grained provider groups are under-recommended by the state-of-the-art models. To overcome this issue, in this paper, we present an approach that brings equity to both coarse and fine-grained provider groups. Experiments on two real-world datasets show the effectiveness of our approach.

CCS Concepts: • **Information systems → Recommender systems**; **Personalization**; **Collaborative filtering**; **Provider Fairness**; • **Human-centered computing → Collaborative filtering**.

Additional Key Words and Phrases: Recommender systems, Bias, Provider Fairness, Geographic Groups, Disparate Impact.

ACM Reference Format:
Elizabeth Gómez, David Contreras, Maria Salamó, and Ludovico Boratto. 2024. Bringing Equity to Coarse and Fine-Grained Provider Groups in Recommender Systems. In *Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization (UMAP '24), July 1–4, 2024, Cagliari, Italy*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3627043.3659552

1 INTRODUCTION

Provider fairness in Recommender Systems aims at ensuring that (groups of) providers receive a visibility or an exposure in the recommendation lists that is proportional to the amount of interactions of the users with these providers. When group fairness is enabled by a system, usually a small number of groups is considered [2, 3, 15, 23]. Indeed, attributes such as gender are considered as binary, while age is divided into three or four groups. In the context of demographic groups, we refer to the terms *coarse* and *fine*-grained to describe the degree of specificity in the classification of populations. *Coarse*-grained involves a more generalized composition and broad categorization of demographic groups, whereas *fine*-grained involves a very specific and detailed categorization of demographic groups. Therefore, there are sensitive attributes that can be seen at different granularities, which can allow us to characterize demographic groups in different ways. An example of a sensitive attribute currently studied in the literature of provider fairness where large groups are considered is the *geographic provenance of the providers* [12–15]. Under this paradigm, item providers are usually considered as belonging to one of the seven continents. However, this attribute can be declined at finer granularities, and one might consider the country, the region, or the city of provenance, looking at this attribute as a sort of a Russian doll. Clearly, the finer the granularity at which we consider a sensitive attribute, the smaller the demographic groups we consider. The question that emerges, when considering coarse demographic groups (as in the current fairness literature), is *how the fine-grained provider groups inside each demographic group are treated in terms of fairness*.

The focus of this paper is two-fold: (i) to enhance that the use of coarse demographic groups under-recommend fine-grained provider groups; (ii) to show that a provider fairness algorithm that considers both coarse and fine-grained provider groups enhances equity and augments coverage. Specifically, in this paper, we present CONFIGRE (i.e., a COarse aNd FIne Grained RE-ranking) that not only focuses on providing fairness to large demographic groups but also enables provider fairness to each fine-grained group (in our study, the country) inside a coarse demographic group (in our study, the continent). Concretely, thanks to the use of buckets associating items to their continent and country of provenance, we ensure that the providers of a given country are given enough *visibility*, meaning that they are recommended a number of times proportional to the interaction of the users with their items. Note that our approach can be easily generalized to any granularity of the provider groups, but in the datasets we used in this work it would not make sense to consider finer granularities (e.g., region/city), due to the small representation these groups would have in the data.

2 RELATED WORK

In this paper, our focus is on *provider fairness* in recommender systems [4, 5], which explores how different providers, either individually or as members of protected groups, have their items included (or not) in the rankings generated by a recommender system [6]. In particular, a lot of work has been done in various scenarios [7, 8, 10, 11, 15, 21, 22, 29]. Different strategies can be employed when mitigating unfairness phenomena, including data pre-processing, in-processing modifications to the model, or post-processing of recommendation lists [6]. Recent advancements include the *CP-fairness* method [23], which integrates fairness constraints from both consumer and provider perspectives into an optimization-based re-ranking approach. Furthermore, Burke *et al.* [3] introduce the SCRUF-D framework, where providers and other stakeholders are presented as agents participating in the recommendation process through a two-stage social choice mechanism. Finally, Wu *et al.* [31] propose the Multi-FR framework, a multi-objective optimization approach for fairness-aware recommendations in multi-sided marketplaces, ensuring a Pareto optimal solution.

Concerning the mitigation strategy, various policies have been devised to examine the trade-offs between user relevance and fairness. Kamishima *et al.* [18] introduced the concept of recommendation independence and developed an objective function, which aims to minimize loss and maximize independence. Tahery *et al.* [30] expanded on this analysis by considering items belonging to more than two protected groups by an algorithm called FARGO. Assessing provider fairness typically involves metrics such as visibility and exposure. Visibility measures the frequency with which an item appears in the rankings [9, 34]. Exposure evaluates the ranking position of an item, specifically for users who receive recommendations for items from each provider [1, 35]. Mehrotra *et al.* [22] introduce a fairness metric that rewards diverse recommendation lists in terms of popularity bias. Other approaches, like the one presented by Karakolis *et al* [19], aim to provide fair recommendations across item providers by considering user diversity and coverage. Raj and Ekstrand [25] provide a comparative analysis of several recently introduced fairness metrics for measuring fair rankings, and Wu *et al.* [32] formulate a family of exposure fairness metrics based on the expected exposure metric, that address fairness concerns by considering group attributes of both users and items. Recently, Chen *et al.* [33] proposed a model called P-MMF, that aims to balance provider fairness and user preference.

Existing proposals in provider fairness primarily focused on ensuring sufficient visibility for providers or groups at a coarse granularity, without considering the impact for the fine-grained groups characterized by a given sensitive attribute. In contrast, our proposal –to the best of our knowledge– is the the first that provides guarantees that also the existing fine-grained of providers are not affected by disparities and receive a fair visibility in the recommendations.

3 PRELIMINARIES

Traditional recommendation scenarios are defined by a set of users, $U = \{u_1, u_2, ..., u_n\}$, that interact with a set of items, $I = \{i_1, i_2, ..., i_j\}$. A totally ordered set of values, V, can be used to express a preference together with a special symbol $\perp$. Considering a rating domain V, the set of ratings results from a map $r : U \times I \longrightarrow V$. If $r(u, i) = \perp$, then we say that user u did not rate item i. To easy notation, we denote $r(u, i)$ by r_{ui}. We define the set of ratings as $R = \{(u, i, r_{ui}) : u \in U, i \in I, r_{ui} \neq \perp\}$ and they can directly feed an algorithm in the form of triplets (point-wise approaches) or shape user-item observations (pair-wise approaches). We consider a random split of the data, where a fixed percentage of the ratings of the users goes to the training, and the rest goes to the test set. The recommendation goal is to learn a function f that estimates the relevance ($\hat{r}_{ui}$) of the user-item pairs that do not appear in the training data (i.e., $r_{ui} = \perp$). The term $\hat{R} = \{(u, i, \hat{r}_{ui}) : u \in U, i \in I\}$ refers to the set of recommendations.

Let $C = \{c_1, c_2, ..., c_g\}$ denote the set of g coarse-grained groups (i.e., in our case study, the geographic continents) associated with the items, and let $D = \{d_1, d_2, ..., d_h\}$ denote the set of h fine-grained groups (i.e., the set is the geographic countries) associated with items. We denote as $C_i \subseteq C$ and $D_i \subseteq D$ the set of coarse and fine-grained groups, respectively, associated with an item i. Specifically, for the geographic provenance domain, note that since an item could be produced by more than one provider, several geographic continents/countries may appear in a item, and thus, $|C_i| \geq 1$, and $|D_i| \geq 1$. In case two providers belong to the same geographic continent/country, that continent/country appears only once; this choice was made since we are dealing with group fairness, so when a group of providers is associated with an item (either once or multiple times), we account for the presence of that group. We use the geographic continents/countries to shape g/h demographic groups, which can be defined to group the ratings of the items produced in a continent/country (we denote the items in I produced in a continent $c \in C$ as I_c, and the ones produced in a country $d \in D$ as I_d). In our analysis and experiments, we use two metrics: *group representation* and *disparate visibility*.

Provider-group Representation. We compute the representation of a demographic group in the data as the number of ratings for items associated with that group in the data. We define with $\mathcal{R}$ the *representation* of a group $c \in C$ (i.e., continent of items) or $d \in D$ (i.e, country or items) as follows:

$$\mathcal{R}_* = \frac{|\{r_{ui} : u \in U, i \in I_*\}|}{|R|} \tag{1}$$

where the '*' symbol can be substituted by c or d, in case we are modeling the representation of a continent or a country, respectively. Eq. (1), which ranges in [0,1], accounts for the proportion of ratings given to the items of a demographic group associated with a continent or country. We compute the representation of a group only considering the training set. Trivially, the sum of the representations of all groups is equal to 1.

Disparate Visibility. Given a group $c \in C$ or $d \in D$, the *disparate visibility* returned by a recommender system for that group is measured as the difference between the share of recommendations for items of that group and the representation of that group in the input data:

$$\Delta\mathcal{V}_* = \left(\frac{1}{|U|} \sum_{u \in U} \frac{|\{\hat{r}_{ui} : i \in I_*\}|}{|\hat{R}|} \right) - \mathcal{R}_* \tag{2}$$

where, again, '*' can be either c or d. The range of values for this score is $[-\mathcal{R}_*, 1 - \mathcal{R}_*]$; specifically, it is 0 when the recommender system has no disparate visibility, while negative/positive values indicate that the group received a share of recommendations that is lower/higher than its representation. This metric is based on Fabbri *et al.* [9].

4 DISPARATE IMPACT ASSESSMENT

4.1 Experimental Setup

We consider four state-of-the-art Collaborative Filtering algorithms: **ItemKNN** [28], **UserKNN** [16]), **BPR** [26], and **SVD** [20]). In our experiments, we will report the results of the original non-fairness-aware recommendation algorithm, denoted as **OR** and two baselines a provider fairness algorithm, **P-Fair** [13], and a consumer-provider approach, **CP-Fair** [24]. We have used the P-Fair algorithm to control fairness for providers coming from different continents and countries; we denote them as **P-Fair**$_c$ and **P-Fair**$_d$, respectively. We follow a similar methodology to that described in [24] to evaluate our experiments' performance of **CP-Fair**, where we consider users as consumers and items (i.e., songs and movies) as producers. Since this algorithm does not allow a multi-group setting, as it addresses fairness as a binary setting, we used the most represented continent group (North America) versus the rest of the continents.

To run the recommendation models presented previously and generate consistently formatted lists that can be fed to CONFIGRE, P-Fair, and CP-Fair, we used the *Cornac* framework [27]. We used two datasets (publicly available at https://github.com/davidcontrerasaguilar/CONFIGRE.git): **(1)** the **MovieLens-1M (Movies)** extended to integrate the continent and the country of production of each movie, it provides 1M ratings (range 1-5), provided by 6,040 users, to 3,600 movies, 54 countries, and 6 continents; **(2)** a new created dataset called **DataSongs (Songs)**, which contains 1,777,981 ratings (range 1-5), provided by 30,759 users, to 16,380 songs, 62 countries, and 6 continents. Both datasets were randomly separated into a test (20%) and training (80%) sets. For each user, we generated the top-1000 recommendations (denoted in the paper as the top-n) to then re-rank the top-k (set up to 10) through the proposed CONFIGRE algorithm. To evaluate recommendation effectiveness, we measure the ranking quality of the lists by measuring the NDCG [17].

4.2 Group representation

Analyzing the training set in the **Songs** dataset, we observe that the North American (NA) continent has the highest representation, which is close to 63.64%. Note that the United States country represents nearly 89.90% of the continent, whereas only Canada, with a percentage of 7.49%, has a percentage above 1%; therefore, the remaining 2.61% is divided between 7 North American (NA) countries. Moreover, the United States has 57.21% of the representation of the whole dataset. Regarding the European (EU) continent, the United Kingdom has a 21.26% representation in the whole dataset and a 77.67% concerning the other suppliers from the EU continent, leaving only 22.33% for the other 24 countries. As it occurs in other continents, one country has more than half of the representation concerning other countries. Similarly, this is the case of South Korea in the Asia (AS) continent, South Africa in Africa (AF), Australia in Oceania (OC), and Brazil in South America (SA). Finally, it is also important to mention that, out of 54 countries, only 7 have a representation of more than 1%, equivalent to 91.11% of the total representation. On the other hand, the **Movies** dataset has a similar behavior, where the North American (NA) continent has 57.13% representation, and the United States has 53.71% of the whole dataset. Additionally, this country has 94.01% of representation concerning the NA continent, whereas only Canada, with 4.50%, has a representation higher than 1.48%, and the remaining 1% is distributed among the other 4 countries. Finally, it is essential to highlight that only 8 of 62 countries have a representation of more than 1%, equivalent to 89.59% of the total representation. **Observation 1.** *The countries inside a continent have very different representations, with a majority country that usually attracts most of the ratings. Hence, the question of how fairly the providers of the minority countries inside a continent are recommended emerges.*

5 CONFIGRE: A COARSE AND FINE GRAINED RE-RANKING PROVIDER FAIRNESS ALGORITHM

5.1 Algorithm

CONFIGRE's primary objective is to reach a provider group's target percentage observed in the training set, while minimizing the loss in user predictions. By continually adjusting the recommendation list, CONFIGRE, which focuses on re-ranking, ensures the provision of fairness for providers and guarantees equity. CONFIGRE works following three main steps: **Step 1**: We compute the representation $\mathcal{R}_*$ of each demographic group (i.e., the coarse groups, $c \in C$, and the fine-grained groups, $d \in D$) considering the ratings in the training set, as in Eq. (1). **Step 2**: Given the items predicted as relevant for a user by the recommender system, we create a bucket list considering each item-coarse group pair, (i.e., in the experiments, item-continent pair), which will store the predicted items. In the same manner, we create a bucket for each item-fine grained group pair (i.e., in the experiments item-country pair). Each bucket comes with an attribute, which is the representation of the coarse or fine-grained group. Specifically, the recommender system returns a list of top-n recommendations (where n is much larger than the cut-off value k, to be able to perform a re-ranking). Our starting point to fill a bucket is the relevance predicted for a user u and an item i, $\hat{r}_{ui}$. Each element in the bucket is a record that contains the user ID, the item ID, and the relevance predicted by the recommender system for that user, $\hat{r}_{ui}$. We sort each bucket by item relevance. **Step 3**: We perform the re-ranking on the basis of the created bucket lists. *The goal is to guarantee fairness for coarse-grained providers and to correctly distribute the recommendations among the different fine-grained groups.* This step includes three phases, the first one is the most constrained and the conditions for selection are relaxed in the second and third phases, so to complete the recommendation list of the user. First, in **Phase 1** we select items from the least represented fine-grained groups to the most represented ones in their corresponding buckets. The algorithm selects items that satisfy the following **conditions**: **(1)** the percentage of items in the recommendation list for a fine-grained group (i.e., country in our experiments) is lower or equal to its representation ($\mathcal{R}_*$); **(2)** the number of recommended items so far is lower than k. Second, in **Phase 2** we start this phase to include more items in the recommendation list when phase 1 finishes, but the top-k is incomplete. Specifically, the selection is made in the same way as in Phase 1. However, this time we do not care that the fine-grained percentage of items is exceeded, but that it belongs to the same coarse-grained group as the item without exceeding the percentage calculated in step 2, $\mathcal{R}_*$. In this phase, condition (2) is not applied. Again, the recommendation list, top-k, may be completed or not. If completed, the process finishes; otherwise, it is necessary to move to phase 3. Finally, in **Phase 3** we complete the recommendation list if the top-k recommendations cannot be reached due to the constraints in the previous phases. In this phase, we select the items that have the greater relevance for the user until we complete the top-k.

5.2 Assessment of the Impact of CONFIGRE

Table 1 shows the disparities of the algorithms in the Songs and Movies datasets. In the Songs dataset, the sum of absolute value of disparities in the OR models (i.e., negative represents lower visibility than expected and vice versa) show that BPR produces a country disparate visibility ($\Delta\mathcal{V}_d$) of 57.5%, whereas SVD, UserKNN and ItemKNN show lower

Table 1. Sum of the absolute value of the disparities $\Delta\mathcal{V}_d$ and $\Delta\mathcal{V}_c$ on the algorithms with respect to the training set in both datasets. Results are shown in percentages. In parentheses is the number of countries/continents recommended by each algorithm. At each row, in bold font the best $\Delta\mathcal{V}_*$ and the second one is underlined. Recommendations accuracy (NDCG) of algorithms is also shown.

	Algorithm		OR	P-Fair$_c$	P-Fair$_d$	CP-Fair	CONFIGRE
Songs	BPR	$\Delta\mathcal{V}_d$	57.5% (2)	56.8% (3)	56.8% (13)	43.0% (2)	**6.2% (16)**
		$\Delta\mathcal{V}_c$	45.3% (3)	44.6% (2)	38.0% (4)	18.0% (2)	**0.4% (5)**
		NDCG	0.0148	0.0134	0.0151	0.0121	0.0117
	SVD	$\Delta\mathcal{V}_d$	37.5% (3)	37.5% (3)	28.7% (20)	50.4% (3)	**3.2% (26)**
		$\Delta\mathcal{V}_c$	11.9% (3)	20.0% (3)	15.3% **(5)**	27.1% (2)	**0.2% (5)**
		NDCG	0.0128	0.0129	0.0109	0.0083	0.0110
	UserKNN	$\Delta\mathcal{V}_d$	13.2% (48)	10.4% (49)	9.5% (52)	33.8% (9)	**0.5% (53)**
		$\Delta\mathcal{V}_c$	10.8% **(6)**	7.9% **(6)**	7.5% **(6)**	17.1% (4)	**0.0% (6)**
		NDCG	0.0036	0.0035	0.0037	0.0024	0.0037
	ItemKNN	$\Delta\mathcal{V}_d$	10.2% **(54)**	8.2% (54)	7.6% **(54)**	23.4% (9)	**0.0% (54)**
		$\Delta\mathcal{V}_c$	8.1% **(6)**	5.9% **(6)**	5.9% **(6)**	5.7% (4)	**0.0% (6)**
		NDCG	0.0036	0.0035	0.0036	0.0029	0.0037
Movies	BPR	$\Delta\mathcal{V}_d$	47.4% (5)	54.7% (17)	42.1% (29)	63.9% (3)	**16.6% (44)**
		$\Delta\mathcal{V}_c$	15.0% (2)	14.6% (4)	13.7% (4)	57.2% (2)	**11.7% (5)**
		NDCG	0.1131	0.0733	0.1132	0.0811	0.1508
	SVD	$\Delta\mathcal{V}_d$	63.9% (4)	50.4% (7)	43.1% (34)	60.6% (5)	**18.9% (42)**
		$\Delta\mathcal{V}_c$	15.1% (2)	18.6% (4)	20.8% **(6)**	16.3% (2)	**9.8% (6)**
		NDCG	0.1872	0.1552	0.1400	0.1653	0.1795
	UserKNN	$\Delta\mathcal{V}_d$	41.9% (14)	46.1% (21)	40.1% (21)	63.1% (7)	**15.0% (43)**
		$\Delta\mathcal{V}_c$	14.1% (3)	16.5% (4)	13.0% (5)	30.5% (4)	**9.1% (6)**
		NDCG	0.0563	0.0618	0.0683	0.0046	0.0531
	ItemKNN	$\Delta\mathcal{V}_d$	36.9% (35)	35.4% (42)	36.6% (40)	31.2% (16)	**5.3% (61)**
		$\Delta\mathcal{V}_c$	15.0% **(6)**	17.3% **(6)**	15.7% **(6)**	11.6% (4)	**3.5% (6)**
		NDCG	0.0791	0.0761	0.0817	0.0015	0.0792

values, being 37.5%, 13.2%, and 10.2%, respectively. The recommendation lists of the OR models generate a significant geographical imbalance, especially in BPR and SVD algorithms, which only recommend 2 and 3 countries, respectively, out of the 54 countries in the dataset. The continent disparity, $\Delta\mathcal{V}_c$, in the OR models obtains better results than $\Delta\mathcal{V}_d$ in all recommendation algorithms.

Analyzing $\Delta\mathcal{V}_d$ in P-Fair$_c$, P-Fair$_d$, and CP-Fair, we can see that in the case of P-Fair$_c$, BPR and SVD produce a $\Delta\mathcal{V}_d$ almost equal to the OR. However, in UserKNN, it is reduced to 10.4%, and in ItemKNN to 8.2%. Although the P-Fair$_c$ method obtains a slight improvement, the $\Delta\mathcal{V}_d$ and $\Delta\mathcal{V}_c$ are still very high. P-Fair$_d$, presents results that are very similar to those obtained by P-Fair$_c$, with SVD being the algorithm that obtained the greatest reduction in both types of disparities. Similar results are obtained for the visibility of continents, where all results are less or equal to P-Fair$_c$. However, the number of countries recommended has been enlarged in this approach. In the case of CP-Fair, BPR produces a lower $\Delta\mathcal{V}_d$ of 43.0%, but the rest of the algorithms increase it. BPR continues to recommend items from 2 countries, but it is the only one that reduces the $\Delta\mathcal{V}_d$ with respect to the OR. SVD continues to recommend 3 countries, but the third country changes to one with the lowest representation in the dataset. In UserKNN and ItemKNN, the same number of countries are recommended, being much fewer than in OR, a total of 9 countries, which coincides with being the most representative ones. Note that CP-Fair is better than OR, but P-Fair$_d$ outperforms it, except for BPR.

Finally, analyzing the results of the CONFIGRE method, BPR produces a $\Delta\mathcal{V}_d$ of 6.2% (i.e., 51% lower than the OR, 57.5%), in SVD the $\Delta\mathcal{V}_d$ is reduced from 37.5% to 3.2% (-34%), UserKNN goes from 13.2% to 0.5% (-13%), and ItemKNN reduces the disparity from 10.2% to 0.0% (-10%). Moreover, our method recommends more countries for each original algorithm. For example, BPR recommends 16 countries, the SVD algorithm improves from 3 to 26 countries, UserKNN from 48 to 51, and ItemKNN continues to recommend all 54 countries. Note that CONFIGRE is the algorithm that reduces the most $\Delta\mathcal{V}_c$, being able to include more countries/continents in the recommendation lists.

In the Movies dataset, Table 1 shows that SVD has the highest $\Delta\mathcal{V}_d$ at 63.9%, BPR at 47.4%, UserKNN at 41.9%, and ItemKNN at 36.9%. Moreover, for these OR models, BPR recommended items from 5 different most represented countries in the dataset and, in the case of SVD, UserKNN, and ItemKNN 4, 14, and 35 countries, respectively. Using the P-Fair$_c$ approach, we obtain that SVD and ItemKNN drop the $\Delta\mathcal{V}_d$ to 50.4% and 35.4%, while BPR and UserKNN rise to 54.7% and 46.1%, respectively. Although this approach does not reduce $\Delta\mathcal{V}_d$, it increases the coverage of providers from different countries (e.g., in BPR to 17 countries). P-Fair$_d$, reduces $\Delta\mathcal{V}_d$ and increases the number of recommended countries. In CP-Fair, BPR increases $\Delta\mathcal{V}_d$ to 63.9% and UserKNN augments to 63.1%, whereas SVD reduces to 60.6% and ItemKNN to 31.2%. Moreover, considering the supplier's coverage, BPR reduces the recommended countries to 3, while SVD increases to 5.

Finally, the CONFIGRE method presents the best results, being the algorithm that better reduces the country/continent disparate visibility concerning the original algorithms. For example, in BPR, we obtain an improvement of $\Delta\mathcal{V}_d$ of 31%, reducing the country disparate visibility from 47.4% to 16.6%. In SVD, the $\Delta\mathcal{V}_d$ achieves an 18.9% (-45% w.r.t. the OR method), while in UserKNN, disparities are reduced from 41.9% to 15.0% (-27%), and in ItemKNN, from 36.9% to 5.3% (-32%). Therefore, the sum of the absolute values of the disparities shows a substantial improvement compared to the original results. Note that the coverage of countries in BPR increased from 5 to 44, SVD from 4 to 42, UserKNN from 14 to 43, and ItemKNN improved the coverage from 35 to 61 countries. Finally, it is important to note that similar to the Songs dataset, CONFIGRE is also the algorithm that reduces the most $\Delta\mathcal{V}_c$ compared to the OR and baselines models.

Additionally, we analyze how models impact the quality of recommendations using the NDCG metric. Analyzing the results obtained in the Songs dataset, we can observe that our method shows a better recommendation quality in contrast to the OR and P-Fair methods in UserKNN and ItemKNN. In the case of BRP and SVD, CONFIGRE is slightly smaller than the OR and P-Fair methods. Moreover, CONFIGRE is also better than CP-Fair in all methods except for the BPR algorithm. Regarding the Movies dataset, CONFIGRE obtains better recommendation quality concerning OR (in the case of BPR and ItemKNN) and P-Fair methods (in the case of BPR, SVD, and ItemKNN). In contrast, CONFIGRE obtained slightly lower quality recommendations than OR (UserKNN) and F-Pair (UserKNN). Finally, CONFIGRE obtained better quality recommendations than the CP-Fair method for all algorithms. **Observation 2.** *CONFIGRE can reduce both disparities (coarse and fine-grained) in the recommendations with a low impact on the recommendation quality while expanding the coverage for providers according to their country of origin.*

6 CONCLUSIONS

To facilitate the assessment of (un)fairness in recommender systems, a few, large, demographic groups are usually considered. In this study, we assessed the impact that this way of enabling fairness might have on the fine-grained groups. Specifically, we had provider fairness as a reference and studied the visibility of each coarse and fine-grained provider group in the recommendations. Our results show that the state-of-the-art approaches to regulating unfairness still bring disparities to the fine-grained groups within a demographic group. To overcome this issue, we presented an approach capable of regulating fairness for both coarse and fine-grained groups via a post-processing approach. Extensive experiments on novel datasets and against state-of-the-art baselines show that our solution can enable provider fairness at different granularities, with a negligible impact on the recommendation effectiveness.

ACKNOWLEDGMENTS

D. Contreras research was partially funded by grant No. 74200094 from ANID-Chile and by NLHPC ECM-02. M. Salamó belongs to the Associated unit to CSIC by IIIA and she was supported by MCIN-AEI-10.13039-501100011033-FEDER and by 2021 SGR 00313.

REFERENCES

[1] Asia J. Biega, Krishna P. Gummadi, and Gerhard Weikum. 2018. Equity of Attention: Amortizing Individual Fairness in Rankings. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018*. ACM, New York, NY, USA, 405–414. https://doi.org/10.1145/3209978.3210063

[2] Ludovico Boratto and Mirko Marras. 2021. Countering Bias in Personalized Rankings : From Data Engineering to Algorithm Development. In *37th IEEE International Conference on Data Engineering, ICDE 2021*. IEEE, 2362–2364. https://doi.org/10.1109/ICDE51399.2021.00266

[3] Robin Burke, Nicholas Mattei, Vladislav Grozin, Amy Voida, and Nasim Sonboli. 2022. Multi-agent Social Choice for Dynamic Fairness-aware Recommendation. In *UMAP '22: 30th ACM Conference on User Modeling, Adaptation and Personalization*. ACM, New York, NY, USA, 234–244. https://doi.org/10.1145/3511047.3538032

[4] Michael D. Ekstrand, Anubrata Das, Robin Burke, and Fernando Diaz. 2022. Fairness and Discrimination in Information Access Systems. *CoRR* abs/2105.05779 (2022). arXiv:2105.05779 https://arxiv.org/abs/2105.05779

[5] Michael D. Ekstrand, Anubrata Das, Robin Burke, and Fernando Diaz. 2022. *Fairness in Recommender Systems*. Springer US, New York, NY, 679–707. https://doi.org/10.1007/978-1-0716-2197-4_18

[6] Michael D. Ekstrand, Anubrata Das, Robin Burke, and Fernando Diaz. 2022. Fairness in Recommender Systems. In *Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, 679–707. https://doi.org/10.1007/978-1-0716-2197-4_18

[7] Michael D. Ekstrand and Daniel Kluver. 2021. Exploring author gender in book rating and recommendation. *User Modeling and User-Adapted Interaction* 31, 3 (2021), 377–420. https://doi.org/10.1007/s11257-020-09284-2

[8] Michael D. Ekstrand, Mucun Tian, Mohammed R. Imran Kazi, Hoda Mehrpouyan, and Daniel Kluver. 2018. Exploring author gender in book rating and recommendation. In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018*. ACM, 242–250. https://doi.org/10.1145/3240323.3240373

[9] Francesco Fabbri, Francesco Bonchi, Ludovico Boratto, and Carlos Castillo. 2020. The Effect of Homophily on Disparate Visibility of Minorities in People Recommender Systems. In *Proceedings of the Fourteenth International AAAI Conference on Web and Social Media, ICWSM 2020*. AAAI Press, USA, 165–175. https://ojs.aaai.org/index.php/ICWSM/issue/view/262

[10] Andres Ferraro, Xavier Serra, and Christine Bauer. 2021. What Is Fair? Exploring the Artists' Perspective on the Fairness of Music Streaming Platforms. In *Human-Computer Interaction - INTERACT 2021 - 18th IFIP TC 13 International Conference, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 12933)*. Springer, 562–584. https://doi.org/10.1007/978-3-030-85616-8_33

[11] Alireza Gharahighehi, Celine Vens, and Konstantinos Pliakos. 2021. Fair multi-stakeholder news recommender system with hypergraph ranking. *Information Processing & Management* 58, 5 (2021), 102663. https://doi.org/10.1016/j.ipm.2021.102663

[12] Elizabeth Gómez, Ludovico Boratto, and Maria Salamó. 2021. Disparate Impact in Item Recommendation: A Case of Geographic Imbalance. In *Advances in Information Retrieval - 43rd European Conference on IR Research, ECIR 2021, Virtual Event, March 28 - April 1, 2021, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 12656)*. Springer, 190–206. https://doi.org/10.1007/978-3-030-72113-8_13

[13] Elizabeth Gómez, Ludovico Boratto, and Maria Salamó. 2022. Provider fairness across continents in collaborative recommender systems. *Inf. Process. Manag.* 59, 1 (2022), 102719. https://doi.org/10.1016/j.ipm.2021.102719

[14] Elizabeth Gómez, Carlos Shui Zhang, Ludovico Boratto, Maria Salamó, and Mirko Marras. 2021. The Winner Takes it All: Geographic Imbalance and Provider (Un)fairness in Educational Recommender Systems. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 1808–1812. https://doi.org/10.1145/3404835.3463235

[15] Elizabeth Gómez, Carlos Shui Zhang, Ludovico Boratto, Maria Salamó, and Guilherme Ramos. 2022. Enabling cross-continent provider fairness in educational recommender systems. *Future Gener. Comput. Syst.* 127 (2022), 435–447. https://doi.org/10.1016/j.future.2021.08.025

[16] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2002. An Empirical Analysis of Design Choices in Neighborhood-Based Collaborative Filtering Algorithms. *Inf. Retr.* 5, 4 (2002), 287–310. https://doi.org/10.1023/A:1020443909834

[17] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.* 20, 4 (2002), 422–446. https://doi.org/10.1145/582415.582418

[18] Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Jun Sakuma. 2018. Recommendation Independence. In *Conference on Fairness, Accountability and Transparency, FAT 2018 (Proceedings of Machine Learning Research, Vol. 81)*. PMLR, 187–201. http://proceedings.mlr.press/v81/kamishima18a.html

[19] Evangelos Karakolis, Panagiotis Kokkinakos, and Dimitrios Askounis. 2022. Provider Fairness for Diversity and Coverage in Multi-Stakeholder Recommender Systems. *Applied Sciences* 12, 10 (2022). https://doi.org/10.3390/app12104984

[20] Yehuda Koren. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 426–434. https://doi.org/10.1145/1401890.1401944

[21] Mirko Marras, Ludovico Boratto, Guilherme Ramos, and Gianni Fenu. 2021. Equality of Learning Opportunity via Individual Fairness in Personalized Recommendations. *International Journal of Artificial Intelligence in Education* (2021). https://doi.org/10.1007/s40593-021-00271-1

[22] 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 27th ACM International Conference on Information and Knowledge Management, CIKM 2018*. ACM, New York, NY, USA, 2243–2251. https://doi.org/10.1145/3269206.3272027

[23] Mohammadmehdi Naghiaei, Hossein A. Rahmani, and Yashar Deldjoo. 2022. CPFair: Personalized Consumer and Producer Fairness Re-ranking for Recommender Systems. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 770–779. https://doi.org/10.1145/3477495.3531959

[24] Mohammadmehdi Naghiaei, Hossein A. Rahmani, and Yashar Deldjoo. 2022. CPFair: Personalized Consumer and Producer Fairness Re-Ranking for Recommender Systems. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Madrid, Spain) *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 770–779. https://doi.org/10.1145/3477495.3531959

[25] Amifa Raj and Michael D. Ekstrand. 2022. Measuring Fairness in Ranked Results: An Analytical and Empirical Comparison. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 726–736. https://doi.org/10.1145/3477495.3532018

[26] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence* (Montreal, Quebec, Canada) *(UAI '09)*. AUAI Press, Arlington, Virginia, USA, 452–461.

[27] Aghiles Salah, Quoc-Tuan Truong, and Hady W Lauw. 2020. Cornac: A comparative framework for multimodal recommender systems. *The Journal of Machine Learning Research* 21, 1 (2020), 3803–3807.

[28] Badrul Munir Sarwar, George Karypis, Joseph A. Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the Tenth International World Wide Web Conference, WWW 10*. ACM, 285–295. https://doi.org/10.1145/371920.372071

[29] Nasim Sonboli and Robin Burke. 2019. Localized Fairness in Recommender Systems. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, UMAP 2019*. ACM, New York, NY, USA, 295–300. https://doi.org/10.1145/3314183.3323845

[30] Saedeh Tahery, Seyyede Zahra Aftabi, and Saeed Farzi. 2021. A GA-based algorithm meets the fair ranking problem. *Information Processing & Management* 58, 6 (2021), 102711. https://doi.org/10.1016/j.ipm.2021.102711

[31] Haolun Wu, Chen Ma, Bhaskar Mitra, Fernando Diaz, and Xue Liu. 2022. A Multi-Objective Optimization Framework for Multi-Stakeholder Fairness-Aware Recommendation. *ACM Trans. Inf. Syst.* 41, 2, Article 47 (dec 2022), 29 pages. https://doi.org/10.1145/3564285

[32] Haolun Wu, Bhaskar Mitra, Chen Ma, Fernando Diaz, and Xue Liu. 2022. Joint Multisided Exposure Fairness for Recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Madrid, Spain) *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 703–714. https://doi.org/10.1145/3477495.3532007

[33] Chen Xu, Sirui Chen, Jun Xu, Weiran Shen, Xiao Zhang, Gang Wang, and Zhenhua Dong. 2023. P-MMF: Provider Max-min Fairness Re-ranking in Recommender System. In *Proceedings of the ACM Web Conference 2023* (Austin, TX, USA) *(WWW '23)*. Association for Computing Machinery, New York, NY, USA, 3701–3711. https://doi.org/10.1145/3543507.3583296

[34] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. FA*IR: A Fair Top-k Ranking Algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017*. ACM, New York, NY, USA, 1569–1578. https://doi.org/10.1145/3132847.3132938

[35] Meike Zehlike and Carlos Castillo. 2020. Reducing Disparate Exposure in Ranking: A Learning To Rank Approach. In *WWW '20: The Web Conference 2020*. ACM / IW3C2, New York, NY, USA, 2849–2855. https://doi.org/10.1145/3366424.3380048