Enabling Cross-continent Provider Fairness in Educational Recommender Systems

Elizabeth Gómez[a], Carlos Shui Zhang[a], Ludovico Boratto[b], Maria Salamó[a], Guilherme Ramos[c]

[a] Facultat de Matemàtiques i Informàtica, Universitat de Barcelona, Barcelona, Spain
[b] Department of Mathematics and Computer Science, University of Cagliari, Italy
[c] Department of Electrical and Computer Engineering, Faculty of Engineering, University of Porto, Portugal.

Abstract

With the widespread diffusion of Massive Online Open Courses (MOOCs), educational recommender systems have become central tools to support students in their learning process. While most of the literature has focused on students and the learning opportunities that are offered to them, the teachers behind the recommended courses get a certain *exposure* when they appear in the final ranking. Underexposed teachers might have reduced opportunities to offer their services, so accounting for this perspective is of central importance to generate equity in the recommendation process. In this paper, we consider groups of teachers based on their geographic provenience and assess provider (un)fairness based on the continent they belong to. We consider measures of visibility and exposure, to account (*i*) in how many recommendations and (*ii*) wherein the ranking of the teachers belonging to different groups appear. We observe disparities that favor the most represented groups, and we overcome these phenomena with a re-ranking approach that provides each group with the expected visibility and exposure, thus controlling fairness of providers coming from different continents (*cross-continent provider fairness*). Experiments performed on data coming from a real-world MOOC platform show that our approach can provide fairness without affecting recommendation effectiveness.

Keywords: Educational recommender systems, provider fairness, geographic groups.

1. Introduction

Historically, recommender systems have been used to promote the consumption of items [1]. Their recent employment in domains such as tourism [2, 3], health [4, 5], and education [6, 7], has shown that this class of algorithms can support users in their decision-making processes, beyond pure sales and streams.

Educational recommender systems have particularly flourished, due to the widespread use of Massive Online Open Courses (MOOCs) [8]. In MOOC platforms, recommender systems learn users' learning needs and preferences, and direct them towards possible resources of interest [6]. With the recent pandemics, the subscription to MOOC platforms has increased by 25-30%[1], which makes the research on recommender systems in these platforms more and more relevant. Among the many types of entities that can be recommended in MOOC platforms, we focus on the main one, i.e., *course recommendation*.

Producing effective recommendations is not the sole goal in a domain such as education. Indeed, the emergence of biases, such as course popularity, can push the recommendation of only popular courses [6] or affect users' learning opportunities [9]. If we go beyond the learners' perspective and of how recommendations can affect them, to consider a multi-stakeholder perspective [10, 11], we can observe that teachers are also directly affected by how recommendations are produced. Indeed, when their courses are recommended by an algorithm, they receive a certain *exposure* in the final ranking. Under- or over-exposing, certain providers might generate or exacerbate disparities and affect the opportunities that are given to teachers to offer their services. When these disparities are associated with sensitive attributes, a recommender system unfairly *discriminates* teachers (*provider unfairness*) [10, 12].

In this paper, we focus on possible unfairness emerging from the provenience of the teachers offering the courses. Specifically, we tackle a *continent-based perspective*, considering demographic groups formed by the continent of provenience of the teachers[2]. Previous studies have shown that geographic perspectives can impact the way users consume items [13]. Delving into the context of our study, considering a geographic perspective to provider fairness is a problem of central relevance in the context of course recommendation to (*i*) avoid affecting teachers belonging to geographic areas that have low representation in the data, by under-recommending their courses, and (*ii*) increase cultural diversity in the recommendation process, by putting learners in touch with courses coming from different parts of the world. Hence, equity for providers from a geographical perspective can provide benefits to both teachers and learners.

Our study begins by assessing unfairness, considering the share of recommendations associated with a demographic group,

Email addresses: egomezye13@alumnes.ub.edu (Elizabeth Gómez), cshuizha7@alumnes.ub.edu (Carlos Shui Zhang), ludovico.boratto@acm.org (Ludovico Boratto), maria.salamo@ub.edu (Maria Salamó), guilhermeramos21@gmail.com (Guilherme Ramos)

[1] https://www.classcentral.com/report/mooc-stats-pandemic/

[2] In the context of this work, we will refer to a group of teachers belonging to a certain continent simply as a "demographic group".

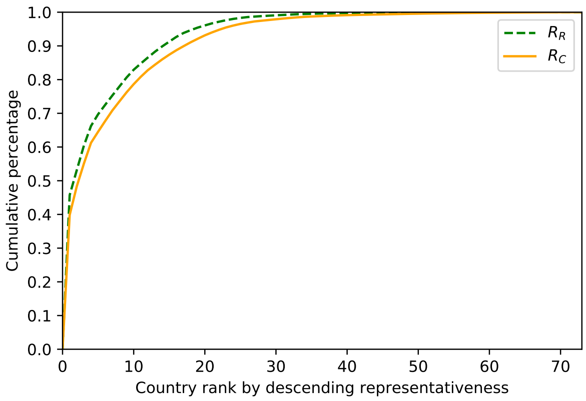


Figure 1: **Country imbalance**. Cumulative percentage of learners' feedback (in blue) and online courses (in green) for each country in COCO [15].

and contextualizing it to the *representation* of the group in the data. We will consider two forms of representation, based on (*i*) the number of courses the teachers in a group offer and (*ii*) the number of interactions between learners and the courses offered by that demographic group. Specifically, we assess unfairness by considering both the *visibility* received by the teachers in a group (i.e., the percentage of recommendations having them as teachers) and their *exposure*, which accounts for the position in which courses are recommended [14]. Hence, with these two metrics, we measure, respectively, (*i*) the share of recommendations of a group and (*ii*) the relevance that is given to that group. Both metrics are relevant to assess disparate impact in this context. Visibility alone might lead a group of teachers not being reached by learners in case they appear only at the bottom of the list, and exposure alone might not guarantee that the courses of a group are being offered to enough learners (indeed, if we optimized only for exposure, then a single course at the top of the recommendation list for one learner would lead that group to get high exposure, but might mean that the opportunities for that group to get recommended to other learners are strongly reduced). We do this assessment on state-of-the-art collaborative recommendation approaches, covering both model- and memory-based approaches and point- and pair-wise algorithms.

Our choice to shape demographic groups based on their continent of provenience was made because a country-based perspective led to a too fine-grained granularity. Considering the data we work with (presented in detail in Section 3), the teachers come from 74 different countries. Figure 1 presents the imbalance in the rating and course distributions, considering the countries in descending order, based on our two notions of representation. We can observe that the top-20 countries respectively attract and cover around 90% of the ratings and courses. This severe imbalance shows that mitigating unfairness at the country level would be unfeasible, due to the very high number of countries we deal with and the low representation of the great majority of countries. We discuss in Section 6 how to deal with fairness at the country level.

We mitigate disparities emerging from our previous assessment with a novel multi-class re-ranking strategy, which optimizes both the visibility and exposure given to teachers, based on their representation in the data. Thanks to our approach, we can regulate how recommendations are distributed along with the different demographic groups (*cross-continent provider fairness*), following a distributive norm based on *equity* [16].

Our contributions can be summarized as follows:

- We consider, for the first time in the literature of educational recommendation, provider fairness for demographic groups based on their geographic provenience;
- We assess unfairness on real-world data coming for a MOOC platform;
- We mitigate unfairness with a novel approach and evaluate its effectiveness.

The rest of the paper is structured as follows: in Section 2 we cover related work, and in Section 3 we provide the foundation to our study. We assess unfairness in Section 4 and mitigate disparities in Section 5. Finally, we conclude our paper in Section 6.

2. Related Work

This section presents literature related to our work. We divided it into different sections, according to the topics we analyze. First of all, we start with education recommender systems. Next, we overview related work on visibility and exposure in rankings. We continue by analyzing provider fairness in recommender systems literature and then focus on the specific topic of our work, fairness in education Artificial Intelligence. Finally, we conclude this section contextualizing our work with respect to the existing literature.

2.1. Educational Recommender Systems

Recommender systems in educational platforms can involve the suggestion of different entities, such as courses [6, 17, 18], threads [19, 20], peers with whom to connect [21, 22, 23], and learning elements [24, 23]. In this section, we focus on course recommendation, which is the main focus of this paper. When designing course recommender systems, several sources of data are considered, such as previous user preferences [18, 25, 26] the combination between user preferences, demographic data, and pre-requisites [27], or the learning style of learners [28]. The classic recommendation models are employed to process the recommendations, namely collaborative filtering [6, 27, 18, 25], content-based filtering [17, 18], and hybrid approaches [29]. Specifically, in this work, we focus on collaborative filtering algorithms.

2.2. Visibility and Exposure in Rankings

Given a ranking, visibility, and exposure metrics respectively assess the number of times an item is present in the rankings [30, 31] and *where* an item is ranked [32, 33]. They were introduced in the context of non-personalized rankings, where the objects being ranked are individual users (e.g., job candidates). These metrics can operate at the *individual* level, thus guaranteeing that similar individuals are treated similarly [32, 34], or at *group* level, by making sure that users belonging to different groups are given adequate visibility or exposure [31, 33]. Under the group setting, the visibility/exposure of a group is proportional to its representation in the data [35, 36, 37].

2.3. Provider Fairness in Recommender Systems

The concepts of visibility and exposure have a direct impact on the providers behind the recommended items. When a system does not discriminate providers based on sensitive attributes, it is known to offer *provider fairness (P-fairness)*. P-fairness guarantees that the providers of the recommended objects that belong to different groups or are similar at the individual level, will get recommended according to their representation in the data. In this domain, Ekstrand et al. [38] assessed that collaborative filtering methods recommend books of authors of a given gender with a distribution that differs from that of the original user profiles. Liu and Burke [39] propose a re-ranking function, which balances recommendation accuracy and fairness, by dynamically adding a bonus to the items of the uncovered providers. Sonboli and Burke [40] define the concept of local fairness, to equalize access to capital across all types of businesses. Mehrotra et al. [41] assess unfairness based on the popularity of the providers. Several policies are defined to study the trade-offs between user-relevance and fairness. Kamishima et al. [42] introduce recommendation independence, which leads to recommendations that are statistically independent of sensitive features.

2.4. Fairness in Educational Artificial Intelligence

Defining when a user or a group of users gets discriminated by an Artificial Intelligence (AI) system highly depends on the context that is being studied [43, 44, 45, 46]. Yu et al. [47] assessed that a fair prediction, for the under-represented groups, of long- and short-term students' success is only possible if institutional data is integrated with the learning management system data. In the context of adaptive learning technologies, Doroudi and Brunskill [48] have shown that the existing algorithms can be inequitable when they rely on inaccurate models; the integration of the additive factor model, usually employed to perform knowledge tracing, can improve fairness in these systems. Hu and Rangwala [49] have focused on models that ensure individual fairness when predicting students at risk of underperforming. Individual fairness was also guaranteed to learners in course recommender systems, by ensuring equal learning opportunities [9].

2.5. Contextualizing our Work

As our analysis of the existing literature shows, our work provides novelty in the intersection of the four areas we have analyzed. Specifically, the concepts of visibility and exposure were never analyzed for demographic groups based on their provenience. None of the educational AI systems has dealt with our notion of fairness. Specifically, our work is the first to provide fairness guarantees to teachers based on their provenience, thus enabling recommender systems to tackle equity in the learning process from a novel perspective.

3. Preliminaries

Here, we present the preliminaries to provide foundations for our work. First of all, Section 3.1 details the recommendation scenario. Next, the metrics are described in Section 3.2. In Section 3.3, we present the recommendation algorithms. Finally, we describe the dataset used in this study in Section 3.4.

3.1. Recommendation scenario

Let $U = \{u_1, u_2, ..., u_n\}$ be a set of learners, $C = \{c_1, c_2, ..., c_j\}$ be a set of courses, and V be a totally ordered set of values that can be used to express a preference together with a special symbol $\perp$. The set of ratings result from a map $r : U \times C \rightarrow V$, where V is the ratings' domain. If $r(u, c) = \perp$ then we say that u did not rate c. To easy notation, we denote $r(u, c)$ by r_{uc}. Now, we can define the set of ratings as $R = \{(u, c, r_{uc}) : u \in U, c \in C, r_{uc} \neq \perp\}$. These ratings can directly feed an algorithm in the form of triplets (point-wise approaches) or shape learner-course observations (pair-wise approaches).

To assess the real impact of the recommendations, we consider a temporal split of the data, where a fixed percentage of the ratings of the learners (ordered by timestamp) goes to the training and the rest goes to the test set [50].

The recommendation goal is to learn a function f that estimates the relevance ($\hat{r}_{uc}$) of the learner-course pairs that do not appear in the training data (i.e., $r_{uc} = \perp$). We denote as $\hat{R}$ the set of recommendations, and as $\hat{R}_G$ those involving courses of a group G, i.e., $\hat{R}_G = \{\hat{r}_{uc} : u \in U, c \in G \subseteq C\}$.

Let $A = \{a_1, a_2, ..., a_g\}$ denote the set of g geographic areas in which courses are organized. Specifically, we consider a geographic area as the continent of provenience of each teacher providing a course. We denote as A_c the set of geographic areas associated with a course c. Note that, since teachers of a course could be from different geographical areas, several geographic areas may appear in a course, and thus, $|A_c| \geq 1$. In case two teachers belong to the same geographic area, it appears only once. We use the geographic areas to shape g demographic groups, where the ith demographic group is defined as $G_i = \{c \in C : a_i \in A_c\}$, for $i = 1, \ldots, g$.

3.2. Metrics

In this section, we describe the metrics used in our analysis and experiments, i.e., the representation of a group, disparate visibility, and disparate exposure.

Representation. The representation of a group is the number of times in which that group appears in the data. We consider two forms of representation, based on (*i*) the number of courses offered by a group and (*ii*) the number of ratings collected for that group. We define with $\mathcal{R}$ the *representation* of a group G

($\mathcal{R}_C$ denotes a course-based representation, while $\mathcal{R}_R$ a rating-based representation):

$$\mathcal{R}_C(G) = |G|/|C| \tag{1}$$

$$\mathcal{R}_R(G) = |\{r_{uc} \,:\, u \in U,\, c \in G \subseteq C\}| \,/|R| \tag{2}$$

Eq. (1) accounts for the proportion of courses of a group, while Eq. (2) for the proportion of ratings associated with a group. Both metrics are between 0 and 1.

The representation of a group is measured by considering only the training set.

Disparate Impact. We assess unfairness with notions of *disparate impact* generated by a recommender system. Specifically, we assess disparate impact with two metrics.

Definition 1 (Disparate visibility)**.** The *disparate visibility* of a group is computed as the difference between the share of recommendations for items of that group and the representation of that group:

$$\Delta\mathcal{V}(G) = \left(\frac{1}{|U|} \sum_{u \in U} \frac{|\{\hat{r}_{uc} : \hat{r}_{uc} \in \hat{R}_G,\, c \in G \subseteq C\}|}{|\hat{R}|} \right) - \mathcal{R}_*(G) \tag{3}$$

where '*' refers to C or R. Its range is in $[-\mathcal{R}_*(G), 1-\mathcal{R}_*(G)]$; it is 0 when there is no disparate visibility, while negative/positive values indicate that the group received a share of recommendations lower/higher than its representation. This metric is based on that considered by Fabbri et al. [30].

Definition 2 (Disparate exposure)**.** The *disparate exposure* of a group is the difference between the exposure obtained by the group in the recommendation lists [14] and the representation of that group:

$$\Delta\mathcal{E}(G) = \left(\frac{1}{|U|} \sum_{u \in U} \frac{\sum_{pos=1}^{k} \frac{1}{log_2(\hat{r}^u_G(pos)+1)}}{\sum_{pos=1}^{k} \frac{1}{log_2(\hat{r}^u_C(pos)+1)}} \right) - \mathcal{R}_*(G), \tag{4}$$

where $\hat{r}^u_G(pos)$ denotes the rating $\hat{r}_{uc}$ that takes position pos in the list $\hat{R}^u_G = \{\hat{r}_{vc} \,:\, v = u,\, c \in G \subseteq C\}$, $u \in U$, sorted by a decreasing order.

This metric also ranges in $[-\mathcal{R}_*(G), 1-\mathcal{R}_*(G)]$; it is 0 when there is no disparate exposure, while negative/positive values indicate that the exposure given to the group in the recommendations is lower/higher than its representation.

> **Remark**. *We do not define a unique "disparate impact" metric, to control both visibility and exposure, so that* teachers are recommended enough times and with enough exposure. *A unique metric would not allow us to balance both, by compressing everything in a unique number. Later in this paper, we show why both metrics are relevant to enable provider fairness in this context.*

3.3. Recommendation algorithms

In this work, we consider five state-of-the-art Collaborative Filtering approaches, which are known to be the most employed class of algorithms for course recommendation [6]. We cover both classes of point-wise and pair-wise approaches and memory-based and model-based algorithms. In addition, we consider two baseline algorithms.

Our baselines are non-personalized algorithms, which will allow us to contextualize the results obtained with different classes of approaches.

- **MostPopular** recommends items based on their popularity in the dataset, by counting the number of items an item was rated. In this way, the algorithm considers only the item perspective, without associating the ratings to the individual users and their preferences.
- **RandomGuess** establishes the maximum and minimum ratings in the data and returns a random rating for each user-item pair to predict.

For the class of memory-based approaches, we consider the following neighborhood-based algorithms:

- **UserKNN** [51] selects the K neighbors closest to the target user, and recommends the elements that other users more similar to him liked.
- **ItemKNN** [52] works in a similar way to the previous one, but in this case the target user is recommended the items that are more similar to other items that they liked before.

Matrix Factorization algorithms divide the data into matrices, representing them in latent factors to determine the degree of affinity that users and items have with those factors. For this class of approaches, we consider the following algorithms:

- **BPR.** [53] Bayesian Personalized Ranking is a state-of-the-art algorithm, optimized to generate recommendation lists, creating a probability function from the Bayesian probability function. The preference function is based on the ratings of pairs of items.
- **BiasedMF.** [54] Basic factorization of the matrix that includes the global mean, user bias, and item bias.
- **SVD++** [55] takes into account the implicit interactions, as well as the user's latent factors and the item's latent factors.

3.4. Dataset

We analyze data from the educational context, exploring the role of the geographic provenience of teachers in the recommendation process. We remark that the experimentation is made difficult because there are very few large-scale educational datasets coming from this specific field of online education. To the best of our knowledge, COCO [15] is the only educational dataset that contains the geographic provenience of the users. The dataset was collected from an online course platform, and includes 43,045 courses and 4,123,127 learners who gave 6,564,870 ratings. Each course is associated with one or more teachers, belonging to 74 different countries.

We pre-processed the dataset to remove all users with less than 3 ratings. Our final dataset contains 12,472 courses and 298,644 learners, which provided 1,296,598 ratings. Out of these courses, 379 are associated with two or more continents, while the rest to only to one.

We shape demographic groups considering the following continents: Africa, Asia, Europe, North America, Oceania, and South America. No course from the seventh continent (Antarctica) was available in the dataset.

Other educational datasets, proposed by [56, 57, 58], generally include (*learner*, *course*, *rating*) triplets only, as needed in traditional recommendation scenarios, thus not fitting the problem tackled in this study (no information about the teachers' sensitive attributes is available).

4. Disparate Impact Assessment

In this section, we run the algorithms presented in Section 3.3 to assess their effectiveness and the disparate impact they generate. Before doing so, we present the experimental setting and analyze the training data, to get insights into the representation of the different groups.

4.1. Experimental setting

For the dataset presented in Section 3.4, the test set was composed of the most recent 20% of the ratings of each learner. To run the recommendation algorithms presented in Section 3.3, we considered the LibRec library (version 2). For each user, we generate 100 recommendations (denoted in the paper as the top-n) so that we can mitigate the disparate impact through a re-ranking algorithm. The final recommendation list for each learner is composed of 20 courses (denoted as top-k).

We performed a grid search to optimize the hyper-parameters of each algorithm and we chose the ones that achieved the best NDCG. Intending to facilitate the reproducibility of our experiments, we detail the hyper-parameters used to run each algorithm:

- **UserKNN.** similarity: Pearson; neighbors: 50; similarity shrinkage: 10;
- **ItemKNN.** similarity: Cosine; neighbors: 200; similarity shrinkage: 10;
- **BPR.** iterator learnrate: 0.01; iterator learnrate maximum: 0.01; iterator maximum: 100; user regularization: 0.01; item regularization: 0.01; factor number: 10; learnrate bolddriver: false; learnrate decay=1.0;
- **BiasedMF.** iterator learnrate: 0.01; iterator learnrate maximum: 0.01; iterator maximum: 10; user regularization: 0.01; item regularization: 0.01; bias regularization: 0.01; number of factors: 10; learnrate bolddriver: false; learnrate decay: 1.0;

Table 1: **Group representation.** Course-based ($\mathcal{R}_C$) and rating-based ($\mathcal{R}_R$) representations of each group. Groups appear in alphabetical order by the name of the continent.

	$\mathcal{R}_C$	$\mathcal{R}_R$
Africa	0.0569	0.0492
Asia	0.1043	0.0526
Europe	0.1974	0.1812
North America	0.5268	0.5796
Oceania	0.0443	0.0694
South America	0.0702	0.0680

- **SVD++.** iterator learnrate: 0.01; iterator learnrate maximum: 0.01; iterator maximum: 13; user regularization: 0.01; item regularization: 0.01; impItem regularization: 0.001; number of factors: 10; learnrate bolddriver: false; learnrate decay: 1.0.

To evaluate recommendation effectiveness, we measure the ranking quality of the lists by measuring the *Normalized Discounted Cumulative Gain* (NDCG) [59].

$$DCG@k = \sum_{u \in U} \hat{r}^u_G(pos) + \sum_{pos=2}^{k} \frac{\hat{r}^u_G(pos)}{log_2(pos)}$$

$$NDCG@k = \frac{DCG@k}{IDCG@k},$$

The ideal *DCG* (*IDCG*) is computed by sorting courses based on decreasing true relevance (true relevance is 1 if the learner interacted with the course in the test set, 0 otherwise). The higher the better.

4.2. Characterizing Representation

The first step towards the assessment of disparate impact is to characterize the representation of the different groups in the data, which we present in Table 1.

The first phenomenon we can observe is that the ranking of the groups is the same, regardless of the form of representation we consider. Most of the courses are taught by North American teachers, covering almost 52.7% of the courses. Europe follows with 19.7% of the courses, and Asia takes a 10.4% share. The remainder of the groups (Africa, Oceania, and South America) have less than 10% representation. This imbalance associated with North America is exacerbated when considering the rating-based representation, where the group covers around 60% of the ratings. This leads the rest of the groups to have a lower representation w.r.t. the course-based one, regardless of Oceania, which accounts for 6.9% of the ratings. We conjecture that learners might interact with courses from Oceania because its main language is English. We performed an additional analysis of the language of the courses, which confirmed that the vast majority of the courses where teachers are from Oceania are taught in English. This analysis connects the vast number of interactions between learners and courses from North America with their interactions with courses from Oceania.

We cannot draw similar conclusions for the two spoken languages both in Europe and South America, Spanish and Portuguese. Indeed, we observed that Spanish learners following courses in Spanish, mainly do from courses that are also organized in Spain. The same holds for South American learners and the courses in Spanish they interact with, which are mainly organized in South America. For the courses in Portuguese, learners from Portugal and Brazil mainly interact with courses provided in their own country. Hence, the representations of Europe and South America are not directly affected by the fact that the continents share two languages.

Observation 1. *North America represents the majority group, with over 50% of the offered courses. These courses attract even more interactions by the learners, thus increasing the group's rating-based representation. All the other groups have a rating-based representation that is lower than the course-based one, minus Oceania. Hence, when courses are offered in English, a group attracts a share of ratings higher than the rate of courses it offers. The same does not hold for courses in Spanish and Portuguese, where learners mainly follow courses in these languages organized in their own country.*

4.3. Assessing Effectiveness and Disparate Impact

In this section, we report the results in terms of effectiveness (NDCG) obtained by each algorithm, and the disparate visibility and exposure associated with each demographic group, based on the two forms of representation. Table 2 summarizes the results.

The first aspect that emerges is that the most effective algorithm in terms of NDCG is ItemKNN. Interestingly, this leads the algorithm to return, for several groups, visibility or exposure proportional to the number of ratings. This scenario is the case for the exposure in Europe and North America, obtaining the lowest $\Delta\mathcal{E}_R$, and for South America in terms of visibility ($\Delta\mathcal{V}_R$). The second most performing algorithm in terms of NDCG is BPR; we can connect this result to the analysis of the dataset made in [6], where it was observed that most of the ratings were equal to 5. Hence, most of these interactions can be treated as binary observations, leading to the capability of the algorithm to produce a good ranking in this context. For the remaining groups, this is the approach that better adjusts to the rating-based representation, in terms of visibility ($\Delta\mathcal{V}_R$) for Europe, North America, and Oceania, and of exposure ($\Delta\mathcal{E}_R$) for Africa and South America. North America and South America are also, respectively, the two groups receiving the best visibility and exposure, given to them by BPR.

Observation 2. *Ranking effectiveness is associated with good visibility and exposure when considering a rating-based representation of the groups. The ratings given by learners help to produce good recommendations and to adapt to the preferences (in terms of ratings) that each demographic group had received.*

Focusing on the course-based representation, two interesting phenomena can be observed. The first is that Random Guess is the one adapting best to the offer in terms of courses. This phenomena is the case for the visibility, $\Delta\mathcal{V}_C$, in all the groups, and for the exposure, $\Delta\mathcal{E}_C$, in Europe, Oceania, and South America. South America is also the place where the best (and almost perfect) visibility and exposure are given to a group, also thanks to Random Guess. Nevertheless, this is also the algorithm that achieves the worst NDCG. Hence, a random choice of the courses to recommend adapts well to the offer of each group but is not effective. The other algorithm offering a good course-based visibility exposure is SVD++. What we can observe here is the presence of exposure equity for both the majority group (North America) and one of the smallest ones (Africa). This means that the factors built by the algorithm capture well the original distribution of the data, thus adapting well to the course offer. Also, in this case, the NDCG of the algorithm is very low, leading to the following observation.

Observation 3. *If an algorithm can provide a group with equitable visibility and exposure, when considering its representation in terms of offered courses, then its effectiveness is very low.*

Finally, we can analyze the scenarios in which the most severe disparities can be observed. Trivially, Most Popular is the algorithm associated with the highest disparate impact values, which can be observed for North America. This result connects to previous studies on popularity bias in educational recommendation [6, 7], and extends them to the unfairness provided by an algorithm.

Observation 4. *Popularity-based recommendation exacerbates disparities, favoring the largest group and at the expense of the smallest ones.*

5. Mitigating Disparate Impact

The previous section allowed us to observe that groups are receiving disproportional visibility and exposure concerning their representation in the data. In this section, we propose a re-ranking algorithm to mitigate disparities. The algorithm introduces courses of the disadvantaged groups in the recommendation list, to reach visibility and exposure proportional to their representation.

A re-ranking algorithm is the only option when optimizing ranking-based metrics, such as visibility and exposure. An in-processing regularization, such as those that have been presented in [42, 60], would not be possible, since at the prediction stage the algorithm does not predict *if and where* an item will be ranked in a recommendation list; hence, no direct comparison with these approaches is possible. This is not due to the specific choice of algorithms, since this consideration would also hold for list-wise approaches. Re-rankings have been introduced to reduce disparities, both in the context

Table 2: **Results of state-of-the-art recommender systems before mitigation.** Each column reports the results of an algorithm, with the first line containing the global Normalized Discounted Cumulative Gain (NDCG). The table continues with one block per demographic group, reporting (*i*) the Disparate Visibility when considering the rating-based representation as a reference ($\Delta\mathcal{V}_R$), (*ii*) Disparate Exposure when considering the rating-based representation as a reference ($\Delta\mathcal{E}_R$), (*iii*) Disparate Visibility when considering the course-based representation as a reference ($\Delta\mathcal{V}_C$), and (*iv*) Disparate Exposure when considering the course-based representation as a reference ($\Delta\mathcal{E}_C$). The underlined values indicate the best ones for each metric and demographic group, while those in bold indicate the overall best result for each metric.

		AF	AS	EU	NA	OC	SA
MostPop	$\Delta\mathcal{V}_R$	-0.0492	-0.0054	-0.0393	0.1364	0.0254	-0.0680
	$\Delta\mathcal{E}_R$	-0.0152	-0.0145	-0.0538	0.1057	-0.0028	-0.0194
	$\Delta\mathcal{V}_C$	-0.0569	-0.0571	-0.0556	0.1893	0.0505	-0.0702
	$\Delta\mathcal{E}_C$	-0.0226	-0.0260	-0.0575	0.1206	0.0082	-0.0228
RandomG	$\Delta\mathcal{V}_R$	0.0066	0.0504	0.0159	-0.0497	-0.0257	0.0024
	$\Delta\mathcal{E}_R$	0.0028	0.0125	0.0064	-0.0148	-0.0077	0.0009
	$\Delta\mathcal{V}_C$	-0.0011	-0.0013	-0.0003	0.0031	-0.0006	0.0001
	$\Delta\mathcal{E}_C$	-0.0004	-0.0005	-0.0002	0.0012	-0.0002	0.0000
UserKNN	$\Delta\mathcal{V}_R$	0.0259	0.0043	0.0378	0.0367	0.0505	0.0456
	$\Delta\mathcal{E}_R$	-0.0105	-0.0069	-0.0180	0.0397	0.0012	-0.0056
	$\Delta\mathcal{V}_C$	-0.0065	-0.0527	0.0081	0.0367	-0.0008	0.0153
	$\Delta\mathcal{E}_C$	-0.0128	-0.0102	-0.0196	0.0411	0.0016	-0.0001
ItemKNN	$\Delta\mathcal{V}_R$	-0.0120	-0.0122	0.0058	0.0045	0.0120	0.0019
	$\Delta\mathcal{E}_R$	-0.0052	-0.0043	-0.0016	0.0050	0.0070	-0.0009
	$\Delta\mathcal{V}_C$	-0.0197	-0.0639	-0.0105	0.0573	0.0371	-0.0016
	$\Delta\mathcal{E}_C$	-0.0066	-0.0197	-0.0066	0.0248	0.0098	-0.0003
BPR	$\Delta\mathcal{V}_R$	-0.0053	-0.0124	0.0025	0.0009	0.0096	0.0047
	$\Delta\mathcal{E}_R$	-0.0016	-0.0044	-0.0054	0.0061	0.0049	0.0004
	$\Delta\mathcal{V}_C$	-0.0130	-0.0642	-0.0138	0.0537	0.0347	0.0025
	$\Delta\mathcal{E}_C$	-0.0038	-0.0213	-0.0106	0.0275	0.0083	-0.0001
BiasedMF	$\Delta\mathcal{V}_R$	0.0331	0.0021	0.0332	0.0060	-0.0672	-0.0073
	$\Delta\mathcal{E}_R$	0.0269	-0.0072	0.0038	-0.0138	-0.0219	0.0123
	$\Delta\mathcal{V}_C$	0.0254	-0.0496	0.0169	0.0589	-0.0421	-0.0095
	$\Delta\mathcal{E}_C$	0.0271	0.0000	-0.0033	0.0049	-0.0151	0.0108
SVD++	$\Delta\mathcal{V}_R$	0.0012	-0.0010	0.0244	-0.0161	-0.0259	0.0175
	$\Delta\mathcal{E}_R$	0.0024	0.0028	0.0040	-0.0189	-0.0065	0.0162
	$\Delta\mathcal{V}_C$	-0.0130	-0.0642	-0.0138	0.0537	0.0347	0.0025
	$\Delta\mathcal{E}_C$	-0.0001	-0.0150	-0.0014	-0.0008	0.0017	0.0157

of non-personalized rankings [31, 14, 32, 61, 33, 35] and of recommender systems [41, 62, 39], with approaches such as Maximal Marginal Relevance [63]. However, all these algorithms optimize only one property (either visibility or exposure). As we will show later in our ablation study, optimizing for one metric is not enough. Nevertheless, we studied the impact of the approach by Liu and Burke [39] in our context, which aims at introducing provider fairness via a re-ranking approach. Concretely, the predicted relevance is increased if a provider has not appeared yet in the top-k of a user. Since we are dealing with a provider fairness setting, we increase the predicted rating if a geographic area has not appeared yet in the ranking of a user. We remind readers to [39] for the technical details of the re-ranking approach we compare with. Hyperparameter λ of the original algorithm proposed in [39] was set to 2.

5.1. Algorithm

Our mitigation algorithm is based on the idea to *move up in the recommendation list the course that causes the minimum loss in prediction for all the learners, until the target visibility or exposure is reached.* Our approach at introducing fairness via a re-ranking is the only one providing guarantees that equity of visibility and exposure is possible since we keep changing the recommendation list until equity from both perspectives is reached. The approaches at the state of the art, based on Maximal Marginal Relevance, make interventions on the predicted relevance for the items, thus not optimizing and not offering guarantees for the final visibility and exposure goals.

The mitigation algorithm is described in Algorithm 1, while Algorithm 2 describes the its support methods. The input is a recommendation list for all the learners (the top-n items) and the target proportions to reach of each continent. The output is the re-ranked list of courses.

The first method, called *optimizeVisibilityExposure* (lines 1-6), calls our mitigation function twice, to have the first intervention in terms of visibility and the second one in terms of exposure. The first *mitigation* call (line 3) is devoted to targeting the desired visibility, to make sure the courses of the disadvantaged groups are recommended enough times. This mitigation step adds the courses of the disadvantaged groups to the top-k.

```
Input: recList: ranked list (records contain user, item, prediction, exposure, continent, position)
       targetProportions: list with the target proportions of each continent
Output: reRankedList: ranked list adjusted by visibility and exposure
1  define optimizeVisibilityExposure (recList, targetProportions)
2  begin
3      reRankedList ← mitigation(recList, “visibility”, targetProportions) ; // mitigation to target the desired visibility
4      reRankedList ← mitigation(reRankedList, “exposure”, targetProportions) ; // mitigation to regulate the exposure
5      return reRankedList ; // re-ranked list adjusted by visibility and exposure
6  end
7  define mitigation (list, reRankingType, targetProportions)
8  begin
9      itemsIn, itemsOut, possibleSwaps, continentList ← list(), list(), list(), list() ; // initializes 4 empty lists to store
        candidate items to add to the list, candidate items to remove, all possible swaps of items, and the
        disparities per continent, respectively
10     proportions ← initialProportions(list, reRankingType); // compute continents’ proportions in the ranked list
11     continentList ← updateDisparity(proportions, targetProportions) ; // updates disparity of each continent
12     foreach user ∈ list do // for each user
13         foreach list.item ∈ top-n do // we loop over all items that belong to this user
14             if checkPosition(list.item, itemsOut, reRankingType)==True and
                checkDisadvantagedGroup(list.continent,continentList)==False then
15                 itemsOut.add(list.item) ; // adds the item as possible candidate to move out if it belongs to an
                    advantaged group and belongs to the top-k
16             else if checkPosition(list.item, itemsOut, reRankingType)==False and
                checkDisadvantagedGroup(list.continent,continentList)==True then
17                 itemsIn.add(list.item) ; // adds the item as possible candidate to move in if it belongs to a
                    disadvantaged group and it is not in the top-k
18             end
19         end
20         while !itemsIn.empty() and !itemsOut.empty() do
21             itemIn ← itemsIn.pop(first); // item ranked higher in the top-n, outside the top-k
22             itemOut ← itemsOut.pop(last); // item ranked lower in the top-k
23             loss ← itemOut.prediction − itemIn.prediction ; // compute the loss if swapped the elements in the list
24             possibleSwaps.add(id, user, itemOut, itemIn, loss); // add the possible swap
25         end
26     end
27     sortByLoss(possibleSwaps); // sort the possible swaps by loss, from minor to major
28     i ← 0;
       // do swaps until the target proportions are reached
29     while proportions < targetProportions and i < len(possibleSwaps) do
30         elem ← possibleSwaps.get(i) ; // gets candidate swap of items with the minor loss
31         if checkPosition(elem.id, elem.itemOut, reRankingType)==True and
            checkDisadvantagedGroup(elem.itemIn.continent,continentList)==False then
32             list ← swap(list, elem.itemOut, elem.itemIn); // makes the swap of the candidate with the minor loss
33             exp ← itemOut.exposure − itemIn.exposure; ; // computes the exposure difference of the swap performed
34             proportions ← updateProportions(elem.itemOut, reRankingType, exp, −1); // reduces continents’ proportions
35             proportions ← updateProportions(elem.itemIn, reRankingType, exp, 1); // adds continents’ proportions
36             continentList ← updateDisparity(proportions, targetProportions) ; // updates disparity of each continent
37         i ← i + 1 ; // advances to the next possible swap with minor loss
38     end
39     return list ; // re-ranked list
40 end
```

Algorithm 1: Visibility and exposure mitigation algorithm

The second mitigation call (line 4) is devoted to regulating the exposure, by moving courses up in the top-k inside the recommendation list, to reach the target exposure.

In lines 7-40, the *mitigation* method regulates the visibility and exposure inside the recommendation list. First of all, several lists are initialized (line 9). Next, in lines 10 and 11, the continent's proportions and their disparities are computed. Following, from line 12 to 26, the algorithm computes for each user all possible swaps of disadvantaged groups that can be done in their recommendations list. Note that it loops over all items (i.e., courses) that belong to each learner and it checks two situations, (*i*) the course's position in the list and (*ii*) if the course is

```
1  define checkPosition(item, itemsOut, reRankingType) // check the position of an item
2  begin
3      if reRankingType == "visibility" then return item.position < top-k ;
4      else if reRankingType == "exposure" then return item.position < itemsOut.last.position ;
5  end
6  define checkDisadvantagedGroup (continent, continentList) // check disadvantaged continent
7  begin
8      for cont ∈ continent do sumDeltas += continentList.get(cont) ; // adds the disparity of the continent
9      return (sumDeltas > 0);
10 end
11 define initialProportions(list, reRankingType) // check initial continents' proportions
12 begin
13     proportions ← 0; // set up each continent' proportion to 0
14     foreach user ∈ list do // for each user
15         foreach list.item ∈ top-k do // we loop over the top-k items that belong to this user
16             if reRankingType == "visibility" then
17                 for cont ∈ list.continent do proportions[cont] += 1 ;
18             else if reRankingType == "exposure" then
19                 for cont ∈ list.continent do proportions[cont] += list.exposure ;
20             end
21         end
22     end
23     return proportions
24 end
25 define updateProportions(item, reRankingType, exp, value) // update proportions after a swap
26 begin
27     if reRankingType == "visibility" then
28         for cont ∈ item.continent do proportions[cont] += (1 × value) ;
29     else if reRankingType == "exposure" then
30         for cont ∈ item.continent do proportions[cont] += ( exp × value) ;
31     end
32     return proportions
33 end
34 define updateDisparity(proportions, targetProportions) // update disparities after a swap
35 begin
36     continentList ← proportions − targetProportions
37     return continentList
38 end
```

Algorithm 2: Support methods for the main mitigation algorithm

in a disadvantaged group or not. So, in the end, *possibleSwaps* contains a set of swaps, where each swap contains the user, the item to extract, the item to add in the recommendation list (top-k) of that user, and the loss we would observe if the swap was done. After that, we sort the possible swaps by loss (line 27). Next, a while loop deals with all the swaps (lines 29-38). We iterate through all possible swaps until the target proportions are reached or there are no more swaps available. Before the *swap* method is called, we check that the candidate swap still makes sense. That is, the candidate course to move up still belongs to a disadvantaged group and the candidate to move down is still in an advantaged group. If the conditions are satisfied by the candidate swap, we proceed to make the swap and update both the group proportions and the disparities. Finally, the method returns the re-ranked list (line 39).

Algorithm 2 details the support methods called in Algorithm 1. The *checkPosition* method (lines 1-5) is responsible for checking the position of an item in the list, taking into account if we perform a visibility or exposure mitigation. In lines 6-10, the method *checkDisadvantagedGroup* verifies whether the item belongs to a disadvantaged continent or not. Note that the method contains a for loop since multiple continents may occur in a course. In that case, we compute the total sum of disparities to define a global disparity of the course. The method returns true when the disparity is positive, false otherwise. The method *initialProportions* (lines 11-24) computes the proportion of each continent. In case of mitigating visibility it accounts the number of courses per continent and, when it mitigates exposure, it computes the sum of exposure per continent. Specifically, the *updateProportions* method (see lines 25-33) updates the proportions per group, based on the ranking type. In case of mitigating visibility, it updates the number of courses per continent and, when it mitigates exposure, modifies the sum of exposure of each continent. Finally, the method *updateDisparity* (lines 34-38) computes the differences between the current proportions per continent and the target proportions.

5.2. Impact of mitigation

In this section, we analyze the impact of our mitigation algorithm, analyzing both the recommendation effectiveness and the visibility and exposure given to the different groups.

> **Remark**. *Since our study is based on a temporal split of the data, we could not run any statistical test to assess the difference in the results between the original algorithm and our re-ranking.*

In Table 3, we report the results obtained by our algorithm after mitigating to regulate both visibility and exposure, having as target the rating- and course-based representations of the group.[3] Readers should note that we are reporting only the NDCG values, because we successfully mitigated both disparate visibility and exposure for all groups; all the values were exactly 0, with some minor deviations at the third or fourth decimal in very few cases. What we can observe is that the effectiveness of the algorithm shows negligible losses in both cases.

> **Observation 5**. *Cross-continent provider fairness for demographic groups of teachers can be achieved without having a negative impact in terms of recommendation effectiveness. Thanks to our approach, we can distribute the recommendation in equitable ways between the different groups, without affecting the learners.*

In Figure 2, we visually show the benefits of moving from the original models to our mitigation in terms of disparate visibility and exposure, considering both a rating- and a course-based representation of the groups. The results confirm that we can provide consistent benefits and introduce equity, regardless of the algorithm, the metric, and the form of representation we consider.

To validate our mitigation strategy, which optimizes for both the target visibility and exposure, we run an ablation study, where we mitigate only for visibility. Results are reported in Table 4. The disparate visibility is mitigated by design. What we can observe is that in all of the groups and all the representations, disparate exposure is never fully mitigated. Referring to the phenomena we previously highlighted, Most Popular still over-exposes North America by 10%, at the expense of other groups, such as Europe (-5%). More broadly, we can observe that the disparate exposure values remain more or less the same as those of Table 2.

> **Observation 6**. *Regulating the visibility given to a group does not provide the group with enough exposure. Disparities in terms of exposure are attenuated, but not fully mitigated. Specific interventions to regulate the given exposure are needed.*

To sum up, the ablation study shows that it is not enough to mitigate unfairness for demographic groups only considering the visibility received by the teachers in a group. Thus, our proposal of mitigating both visibility and exposure is an imperative need. The novelty of our approach comes from the idea of considering both metrics, visibility, and exposure, to address provider unfairness. It is important to remark that our results show that the proposed algorithm (see Algorithm 1) can reach the target proportions with a minimal loss in NDCG.

5.3. Contextualization with the State of the Art

In this section, we compare the results of our mitigation with that proposed in [39]. Table 5 reports the obtained results.

While our approach is capable of introducing equity by mitigating both disparate visibility and exposure, as we have previously observed, this is not the case for the baseline approach in our context. Indeed, disparities are reduced by little concerning those returned by the original models, and, in some cases, they are even slightly worse. This effect is because the baseline approach favors the introduction in the top-k of courses produced in more than one continent (in other words, belonging to more than one geographic group). This observation means that, while a disadvantaged group might gain visibility and/or exposure, the accompanying group also receives the same treatment, even though it might be advantaged.

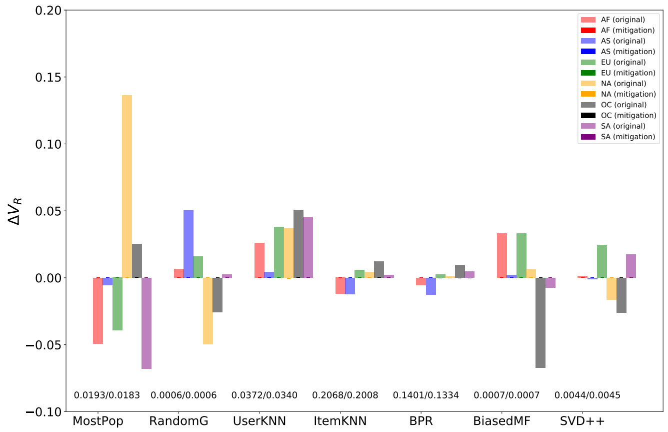
The reason why the original approach can only partially mitigate disparity is since an item of the group becomes more relevant than what it was predicted, whenever that group is not yet in the top-k. Once the group is included in the recommendation list, the items stop getting a boost. However, there is no guarantee that disparities are fully mitigated. On the contrary, our approach keeps injecting items in the top-k as long as disparities are fully mitigated.

> **Observation 7**. *Introducing provider fairness requires interventions at recommendation-list level. Mitigating by boosting predicted relevance for the disadvantaged groups does not provide guarantees of equity of visibility and exposure are fully mitigated. Disparities are only partially mitigated.*

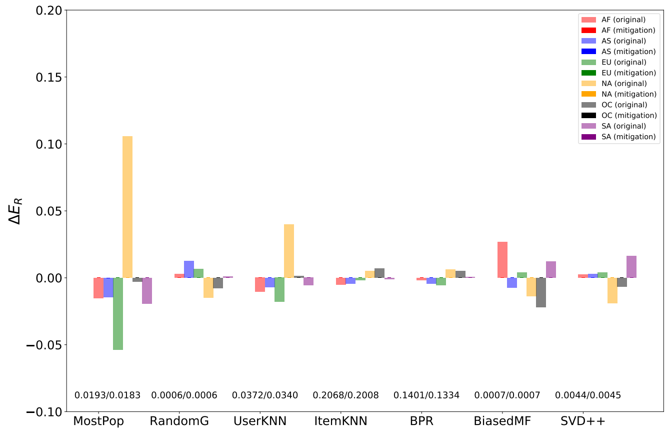
6. Conclusions and Future Work

Accounting for provider fairness in the recommendation process is a central aspect to account for equity in the way recommendations are produced. In this paper, we considered a course recommendation scenario and assessed unfairness for demographic groups based on the continent of provenience of the teachers. We run state-of-the-art collaborative filtering approaches on real-world data coming from a MOOC platform, and observed disparities in the visibility and exposure at the expense of the smaller demographic groups. We mitigated these disparities with a novel re-ranking multi-class approach, which adjusted the final ranking based on the target visibility and exposure, thus enabling *cross-continent provider fairness* to teachers. Results have shown that the disparities in visibility and ex-

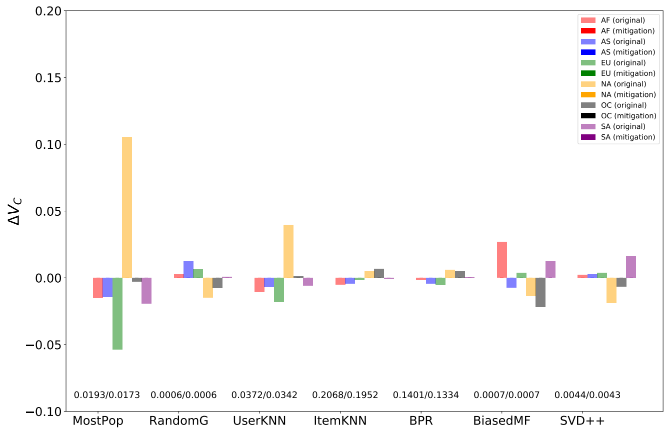
[3]The last line, indicating the NDCG values returned after running the mitigation with the baseline approach, will be analyzed in the context of Section 5.3.



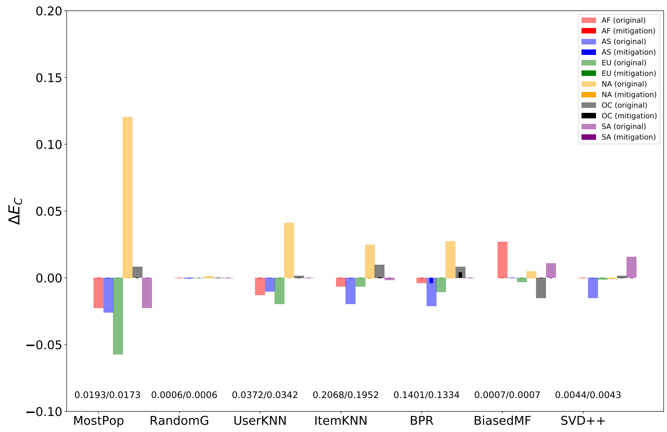
(a) Disparate visibility (rating-based representation)



(b) Disparate exposure (rating-based representation)



(c) Disparate visibility (course-based representation)



(d) Disparate exposure (course-based representation)

Figure 2: **Disparate impact**. Disparate impact returned by the state-of-the-art models (thick bars) and by the mitigation proposed in [64] (thin bars). Each figure contains one section for each algorithm and a color for each continent. The text at the bottom of each figure contains the NDCG returned by the original model and after the mitigation, separated by a "/". In (a) and (b), we report the disparate visibility and disparate exposure obtained when considering a rating-based representation, while in (c) and (d), the disparate visibility and disparate exposure obtained when considering a course-based representation representation.

Table 3: **Results of state-of-the-art recommender systems after full mitigation (both visibility and exposure).** Normalized Discounted Cumulative Gain (NDCG) of the original algorithm, after mitigating based on the rating-based representation ($\mathcal{V}_R \rightarrow \mathcal{E}_R$), after mitigating based on the course-based representation ($\mathcal{V}_C \rightarrow \mathcal{E}_C$), and after mitigating with the baseline.

NDCG	MostPop	RandomG	UserKNN	ItemKNN	BPR	BiasedMF	SVD++
Original	0.0193	0.0006	0.0372	**0.2068**	0.1401	0.0007	0.0044
$\mathcal{V}_R$	0.0195	0.0006	0.0368	0.2066	0.1398	0.0007	0.0045
$\mathcal{V}_C$	0.0187	0.0006	0.0367	0.2039	0.1373	0.0007	0.0043
$\mathcal{V}_R \rightarrow \mathcal{E}_R$	0.0183	0.0006	0.0340	0.2008	0.1334	0.0007	0.0045
$\mathcal{V}_C \rightarrow \mathcal{E}_C$	0.0173	0.0006	0.0342	0.1952	0.1334	0.0007	0.0043
Baseline	0.0193	0.0002	0.0376	**0.2075**	0.1400	0.0005	0.0036

posure can be overcome without affecting the recommendation effectiveness for learners.

While we have highlighted that mitigating disparities at the level of individual countries can be very challenging, it is still relevant to generate equity also at this granularity. Indeed, highly represented countries inside a continent (e.g., the United States in North America) can be over-exposed, thus maintaining unfairness. In future work, we plan to introduce a two-stage process to regulate the distribution of recommendations inside a continent and guarantee fairness for teachers also at this level.

At the moment, only the dataset we considered in this study is available to study these phenomena. In the future, we plan to enrich other existing educational datasets with synthetic demographic groups to validate our approach under different scenarios.

Finally, we plan to study our multi-class mitigation in different application scenarios, such as movies or books, to study the impact of recommender systems in the context of pure consumption items.

References

[1] F. Ricci, L. Rokach, B. Shapira (Eds.), Recommender Systems Handbook, Springer, 2015. doi:10.1007/978-1-4899-7637-6. URL https://doi.org/10.1007/978-1-4899-7637-6

[2] P. Sánchez, A. Bellogín, Applying reranking strategies to route recommendation using sequence-aware evaluation, User Model. User Adapt. Interact. 30 (4) (2020) 659–725. doi:10.1007/s11257-020-09258-4. URL https://doi.org/10.1007/s11257-020-09258-4

[3] S. Torrijos, A. Bellogín, P. Sánchez, Discovering related users in location-based social networks, in: T. Kuflik, I. Torre, R. Burke, C. Gena (Eds.), Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, UMAP 2020, Genoa, Italy, July 12-18, 2020, ACM, 2020, pp. 353–357. doi:10.1145/3340631.3394882. URL https://doi.org/10.1145/3340631.3394882

[4] L. Boratto, S. Carta, W. Iguider, F. Mulas, P. Pilloni, Predicting workout quality to help coaches support sportspeople, in: D. Elsweiler, B. Ludwig, A. Said, H. Schäfer, H. Torkamaan, C. Trattner (Eds.), Proceedings of the 3rd International Workshop on Health Recommender Systems, HealthRecSys 2018, co-located with the 12th ACM Conference on Recommender Systems (ACM RecSys 2018), Vancouver, BC, Canada, October 6, 2018, Vol. 2216 of CEUR Workshop Proceedings, CEUR-WS.org, 2018, pp. 8–12. URL http://ceur-ws.org/Vol-2216/healthRecSys18_paper_2.pdf

[5] T. N. T. Tran, A. Felfernig, C. Trattner, A. Holzinger, Recommender systems in the healthcare domain: state-of-the-art and research issues, Journal of Intelligent Information Systems (2020) 1–31doi:10.1007/s10844-020-00633-6.

[6] L. Boratto, G. Fenu, M. Marras, The effect of algorithmic bias on recommender systems for massive open online courses, in: L. Azzopardi, B. Stein, N. Fuhr, P. Mayr, C. Hauff, D. Hiemstra (Eds.), Advances in Information Retrieval - 41st European Conference on IR Research, ECIR 2019, Cologne, Germany, April 14-18, 2019, Proceedings, Part I, Vol. 11437 of Lecture Notes in Computer Science, Springer, 2019, pp. 457–472. doi:10.1007/978-3-030-15712-8_30. URL https://doi.org/10.1007/978-3-030-15712-8_30

[7] L. Boratto, G. Fenu, M. Marras, Connecting user and item perspectives in popularity debiasing for collaborative recommendation, Inf. Process. Manag. 58 (1) (2021) 102387. doi:10.1016/j.ipm.2020.102387. URL https://doi.org/10.1016/j.ipm.2020.102387

[8] G. Fenu, M. Marras, M. Meles, A learning analytics tool for usability assessment in moodle environments, Journal of e-Learning and Knowledge Society 13 (3) (September 2017). URL https://www.learntechlib.org/p/180986

[9] M. Marras, L. Boratto, G. Ramos, G. Fenu, Equality of learning opportunity via individual fairness in personalized recommendations, CoRR abs/2006.04282 (2020). arXiv:2006.04282. URL https://arxiv.org/abs/2006.04282

[10] H. Abdollahpouri, G. Adomavicius, R. Burke, I. Guy, D. Jannach, T. Kamishima, J. Krasnodebski, L. A. Pizzato, Multistakeholder recommendation: Survey and research directions, User Model. User Adapt. Interact. 30 (1) (2020) 127–158. doi:10.1007/s11257-019-09256-1. URL https://doi.org/10.1007/s11257-019-09256-1

[11] C. Bauer, E. Zangerle, Leveraging multi-method evaluation for multi-stakeholder settings, CoRR abs/2001.04348 (2020). arXiv:2001.04348.

[12] L. Boratto, G. Fenu, M. Marras, Interplay between upsampling and regularization for provider fairness in recommender systems, CoRR abs/2006.04279 (2020). arXiv:2006.04279. URL https://arxiv.org/abs/2006.04279

[13] C. Bauer, M. Schedl, Global and country-specific mainstreaminess measures: Definitions, analysis, and usage for improving personalized music recommendation systems, PLOS ONE 14 (6) (2019) 1–36. doi:10.1371/journal.pone.0217389.

[14] A. Singh, T. Joachims, Fairness of exposure in rankings, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, ACM, 2018, pp. 2219–2228. doi:10.1145/3219819.3220088. URL https://doi.org/10.1145/3219819.3220088

[15] D. Dessì, G. Fenu, M. Marras, D. Reforgiato Recupero, COCO: semantic-enriched collection of online courses at scale with experimental use cases, in: Á. Rocha, H. Adeli, L. P. Reis, S. Costanzo (Eds.), Trends and Advances in Information Systems and Technologies - Volume 2 [WorldCIST'18, Naples, Italy, March 27-29, 2018], Vol. 746 of Advances in Intelligent Systems and Computing, Springer, 2018, pp. 1386–1396. doi:10.1007/978-3-319-77712-2_133. URL https://doi.org/10.1007/978-3-319-77712-2_133

[16] E. Walster, E. Berscheid, G. W. Walster, New directions in equity research., Journal of personality and social psychology 25 (2) (1973) 151.

[17] D. Yanhui, W. Dequan, Z. Yongxin, L. Lin, A group recommender system for online course study, in: Proceedings of the 2015 7th International Conference on Information Technology in Medicine and Education (ITME), ITME '15, IEEE Computer Society, USA, 2015, p. 318–320. doi:10.1109/ITME.2015.99. URL https://doi.org/10.1109/ITME.2015.99

[18] H. Zhang, H. Yang, T. Huang, G. Zhan, Dbncf: Personalized courses

Table 4: **Results of state-of-the-art recommender systems after mitigating only for visibility.** Each column reports the results of an algorithm, with the first two line containing the global Normalized Discounted Cumulative Gain (NDCG) obtained after the two mitigations. The table continues with one block per demographic group, reporting (*i*) the Disparate Visibility when considering the rating-based representation as a reference ($\Delta\mathcal{V}_R$), (*ii*) Disparate Exposure when considering the rating-based representation as a reference ($\Delta\mathcal{E}_R$), (*iii*) Disparate Visibility when considering the course-based representation as a reference ($\Delta\mathcal{V}_C$), and (*iv*) Disparate Exposure when considering the course-based representation as a reference ($\Delta\mathcal{E}_C$).

		AF	AS	EU	NA	OC	SA
MostPop	$\Delta\mathcal{V}_R$	-0.0004	0.0000	0.0000	0.0000	0.0004	0.0000
	$\Delta\mathcal{E}_R$	0.0152	0.0145	0.0538	-0.1057	0.0028	0.0194
	$\Delta\mathcal{V}_C$	-0.0081	0.0000	0.0000	0.0000	0.0081	0.0000
	$\Delta\mathcal{E}_C$	0.0226	0.0260	0.0575	-0.1206	-0.0082	0.0228
RandomG	$\Delta\mathcal{V}_R$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$\Delta\mathcal{E}_R$	-0.0028	-0.0125	-0.0064	0.0148	0.0077	-0.0009
	$\Delta\mathcal{V}_C$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$\Delta\mathcal{E}_C$	0.0004	0.0005	0.0002	-0.0012	0.0002	0.0000
UserKNN	$\Delta\mathcal{V}_R$	0.0000	0.0000	0.0000	-0.0006	0.0007	0.0000
	$\Delta\mathcal{E}_R$	0.0105	0.0069	0.0180	-0.0397	-0.0012	0.0056
	$\Delta\mathcal{V}_C$	0.0000	0.0001	0.0000	-0.0001	0.0000	0.0000
	$\Delta\mathcal{E}_C$	0.0128	0.0102	0.0196	-0.0411	-0.0016	0.0001
ItemKNN	$\Delta\mathcal{V}_R$	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
	$\Delta\mathcal{E}_R$	0.0052	0.0043	0.0016	-0.0050	-0.0070	0.0009
	$\Delta\mathcal{V}_C$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$\Delta\mathcal{E}_C$	0.0066	0.0197	0.0066	-0.0248	-0.0098	0.0016
BPR	$\Delta\mathcal{V}_R$	0.0003	0.0001	-0.0001	-0.0002	-0.0001	0.0000
	$\Delta\mathcal{E}_R$	0.0016	0.0044	0.0054	-0.0061	-0.0049	-0.0004
	$\Delta\mathcal{V}_C$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$\Delta\mathcal{E}_C$	0.0038	0.0213	0.0106	-0.0275	-0.0083	0.0001
BiasedMF	$\Delta\mathcal{V}_R$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$\Delta\mathcal{E}_R$	-0.0269	0.0072	-0.0038	0.0138	0.0219	-0.0123
	$\Delta\mathcal{V}_C$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$\Delta\mathcal{E}_C$	-0.0271	0.0000	0.0033	-0.0049	0.0151	-0.0108
SVD++	$\Delta\mathcal{V}_R$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$\Delta\mathcal{E}_R$	-0.0024	-0.0028	-0.0040	0.0189	0.0065	-0.0162
	$\Delta\mathcal{V}_C$	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	$\Delta\mathcal{E}_C$	0.0004	0.0005	0.0002	-0.0012	0.0002	0.0000

recommendation system based on dbn in mooc environment, in: 2017 International Symposium on Educational Technology (ISET), 2017, pp. 106–108. doi:10.1109/ISET.2017.33.

[19] D. Yang, D. Adamson, C. P. Rosé, Question recommendation with constraints for massive open online courses, in: A. Kobsa, M. X. Zhou, M. Ester, Y. Koren (Eds.), Eighth ACM Conference on Recommender Systems, RecSys '14, Foster City, Silicon Valley, CA, USA - October 06 - 10, 2014, ACM, 2014, pp. 49–56. doi:10.1145/2645710.2645748.
URL https://doi.org/10.1145/2645710.2645748

[20] Z. Chen, C. D. Epp, Csclrec: Personalized recommendation of forum posts to support socio-collaborative learning, in: A. N. Rafferty, J. Whitehill, C. Romero, V. Cavalli-Sforza (Eds.), Proceedings of the 13th International Conference on Educational Data Mining, EDM 2020, Fully virtual conference, July 10-13, 2020, International Educational Data Mining Society, 2020.
URL https://educationaldatamining.org/files/conferences/EDM2020/papers/paper_64.pdf

[21] H. Labarthe, F. Bouchet, R. Bachelet, K. Yacef, Does a peer recommender foster students' engagement in moocs?, in: T. Barnes, M. Chi, M. Feng (Eds.), Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016, Raleigh, North Carolina, USA, June 29 - July 2, 2016, International Educational Data Mining Society (IEDMS), 2016, pp. 418–423.
URL http://www.educationaldatamining.org/EDM2016/proceedings/paper_171.pdf

[22] B. A. Potts, H. Khosravi, C. Reidsema, A. Bakharia, M. Belonogoff, M. Fleming, Reciprocal peer recommendation for learning purposes, in: A. Pardo, K. Bartimote-Aufflick, G. Lynch, S. B. Shum, R. Ferguson, A. Merceron, X. Ochoa (Eds.), Proceedings of the 8th International Conference on Learning Analytics and Knowledge, LAK 2018, Sydney, NSW, Australia, March 07-09, 2018, ACM, 2018, pp. 226–235. doi:10.1145/3170358.3170400.
URL https://doi.org/10.1145/3170358.3170400

[23] J. Gong, S. Wang, J. Wang, W. Feng, H. Peng, J. Tang, P. S. Yu, Attentional graph convolutional networks for knowledge concept recommendation in moocs in a heterogeneous view, in: J. Huang, Y. Chang, X. Cheng, J. Kamps, V. Murdock, J. Wen, Y. Liu (Eds.), Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020, ACM, 2020, pp. 79–88. doi:10.1145/3397271.3401057.
URL https://doi.org/10.1145/3397271.3401057

[24] A. Elbadrawy, A. Polyzou, Z. Ren, M. Sweeney, G. Karypis, H. Rangwala, Predicting student performance using personalized analytics, Computer 49 (4) (2016) 61–69. doi:10.1109/MC.2016.119.
URL https://doi.org/10.1109/MC.2016.119

[25] H. Zhang, T. Huang, Z. Lv, S. Liu, Z. Zhou, MCRS: A course recommendation system for moocs, Multim. Tools Appl. 77 (6) (2018) 7051–7069. doi:10.1007/s11042-017-4620-2.
URL https://doi.org/10.1007/s11042-017-4620-2

[26] F. Fauzan, D. Nurjanah, R. Rismala, Apriori association rule for course recommender system, Indonesia Journal on Computing (Indo-JC) 5 (2) (2020) 1–16. doi:10.34818/INDOJC.2020.5.2.434.

Table 5: **Disparate impact with different mitigation strategies.** Disparate impact metrics returned by the different models for each continent (AF: Africa, AS: Asia, EU: Europe, NA: North America, OC: Oceania, SA: South America). For each algorithm we report the results obtained by the baseline and by our multiclass mitigation, in terms of disparate visibility and exposure when considering the rating-based representation as a reference ($\Delta\mathcal{V}_R$ and $\Delta\mathcal{E}_R$ lines) and with the course-based representation ($\Delta\mathcal{V}_C$ and $\Delta\mathcal{E}_C$ lines).

		AF		**AS**		**EU**		**NA**		**OC**		**SA**	
		mitigation	**baseline**	**mitigation**	**baseline**	**mitigation**	**baseline**	**mitigation**	**baseline**	**mitigation**	**baseline**	**mitigation**	**baseline**
MostPop	$\Delta\mathcal{V}_R$	-0.0004	-0.0428	0.0000	-0.0039	0.0000	-0.0353	0.0000	0.1230	0.0004	0.0268	0.0000	-0.0680
	$\Delta\mathcal{E}_R$	0.0000	-0.0112	0.0002	-0.0137	-0.0002	-0.0513	0.0000	0.0974	0.0000	-0.0019	0.0000	-0.0194
	$\Delta\mathcal{V}_C$	-0.0081	-0.0505	0.0000	-0.0556	0.0000	-0.0516	0.0000	0.1759	0.0081	0.0519	0.0000	-0.0702
	$\Delta\mathcal{E}_C$	0.0000	-0.0186	0.0000	-0.0252	0.0000	-0.0550	0.0000	0.1123	0.0000	0.0091	0.0000	-0.0228
RandomG	$\Delta\mathcal{V}_R$	0.0000	0.0196	0.0000	0.0728	0.0000	0.0200	0.0000	-0.1267	0.0000	-0.0009	0.0000	0.0152
	$\Delta\mathcal{E}_R$	0.0000	0.0094	0.0000	0.0246	0.0000	0.0085	0.0000	-0.0549	0.0000	0.0053	0.0000	0.0074
	$\Delta\mathcal{V}_C$	0.0000	0.0119	0.0000	0.0211	0.0000	0.0038	0.0000	-0.0739	0.0000	0.0242	0.0000	0.0129
	$\Delta\mathcal{E}_C$	0.0000	0.0062	0.0000	0.0116	0.0000	0.0019	0.0000	-0.0389	0.0000	0.0128	0.0000	0.0065
UserKNN	$\Delta\mathcal{V}_R$	0.0000	0.0303	0.0000	0.0060	0.0000	0.0416	-0.0006	0.0230	0.0007	0.0543	0.0000	0.0456
	$\Delta\mathcal{E}_R$	0.0000	-0.0079	0.0000	-0.0059	0.0000	-0.0157	0.0000	0.0315	0.0000	0.0035	0.0000	-0.0056
	$\Delta\mathcal{V}_C$	0.0000	-0.0021	0.0001	-0.0510	0.0000	0.0119	-0.0001	0.0230	0.0000	0.0030	0.0000	0.0153
	$\Delta\mathcal{E}_C$	0.0000	-0.0102	0.0000	-0.0092	0.0000	-0.0173	0.0000	0.0329	0.0000	0.0039	0.0000	-0.0001
ItemKNN	$\Delta\mathcal{V}_R$	0.0001	-0.0066	0.0000	-0.0114	0.0000	0.0101	0.0000	-0.0076	0.0000	0.0137	0.0000	0.0019
	$\Delta\mathcal{E}_R$	0.0000	-0.0021	0.0000	-0.0039	0.0000	0.0010	0.0000	-0.0021	0.0000	0.0080	0.0000	-0.0009
	$\Delta\mathcal{V}_C$	0.0000	-0.0143	0.0000	-0.0631	0.0000	-0.0062	0.0000	0.0452	0.0000	0.0388	0.0000	-0.0016
	$\Delta\mathcal{E}_C$	0.0000	-0.0035	0.0000	-0.0193	0.0000	-0.0040	0.0005	0.0177	-0.0005	0.0108	0.0000	-0.0003
BPR	$\Delta\mathcal{V}_R$	0.0003	-0.0029	0.0001	-0.0075	-0.0001	0.0059	-0.0002	-0.0135	-0.0001	0.0130	0.0000	0.0049
	$\Delta\mathcal{E}_R$	0.0000	-0.0002	0.0000	-0.0015	0.0000	-0.0034	0.0000	-0.0025	0.0000	0.0070	0.0000	0.0005
	$\Delta\mathcal{V}_C$	0.0000	-0.0106	0.0000	-0.0593	0.0000	-0.0104	0.0000	0.0393	0.0000	0.0381	0.0000	0.0027
	$\Delta\mathcal{E}_C$	0.0000	-0.0024	0.0040	-0.0184	0.0000	-0.0086	0.0000	0.0189	-0.0040	0.0104	0.0000	0.0000
BiasedMF	$\Delta\mathcal{V}_R$	0.0000	0.0431	0.0000	0.0006	0.0000	0.0366	0.0000	-0.0206	0.0000	-0.0630	0.0000	0.0031
	$\Delta\mathcal{E}_R$	0.0000	0.0325	0.0000	-0.0079	0.0000	0.0055	0.0000	-0.0280	0.0000	-0.0196	0.0000	0.0177
	$\Delta\mathcal{V}_C$	0.0000	0.0354	0.0000	-0.0511	0.0000	0.0203	0.0000	0.0323	0.0000	-0.0379	0.0000	0.0009
	$\Delta\mathcal{E}_C$	0.0000	0.0327	0.0000	-0.0007	0.0000	-0.0016	0.0000	-0.0093	0.0000	-0.0128	0.0000	0.0162
SVD++	$\Delta\mathcal{V}_R$	0.0000	0.0136	0.0000	-0.0041	0.0000	0.0323	0.0000	-0.0566	0.0000	-0.0212	0.0000	0.0361
	$\Delta\mathcal{E}_R$	0.0000	0.0093	0.0000	0.0012	0.0000	0.0081	0.0000	-0.0406	0.0000	-0.0040	0.0000	0.0260
	$\Delta\mathcal{V}_C$	0.0000	-0.0006	0.0000	-0.0673	0.0000	-0.0059	0.0000	0.0132	0.0000	0.0394	0.0000	0.0211
	$\Delta\mathcal{E}_C$	0.0000	0.0068	0.0000	-0.0166	0.0000	0.0027	0.0000	-0.0225	0.0000	0.0042	0.0000	0.0255

[27] Y. Pang, N. Wang, Y. Zhang, Y. Jin, W. Ji, W. Tan, Prerequisite-related mooc recommendation on learning path locating, Computational Social Networks 6 (1) (2019) 1–16.

[28] J. Gope, S. K. Jain, A learning styles based recommender system prototype for edx courses, in: 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), IEEE, 2017, pp. 414–419.

[29] V. Garg, R. Tiwari, Hybrid massive open online course (mooc) recommendation system using machine learning, in: International Conference on Recent Trends in Engineering, Science Technology - (ICRTEST 2016), 2016, pp. 1–5. doi:10.1049/cp.2016.1479.

[30] F. Fabbri, F. Bonchi, L. Boratto, C. Castillo, 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, 2020, pp. 165–175.
URL https://aaai.org/ojs/index.php/ICWSM/article/view/7288

[31] M. Zehlike, F. Bonchi, C. Castillo, S. Hajian, M. Megahed, R. Baeza-Yates, Fa*ir: A fair top-k ranking algorithm, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, ACM, 2017, pp. 1569–1578. doi:10.1145/3132847.3132938.
URL https://doi.org/10.1145/3132847.3132938

[32] A. J. Biega, K. P. Gummadi, G. Weikum, Equity of attention: Amortizing individual fairness in rankings, in: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, ACM, 2018, pp. 405–414. doi:10.1145/3209978.3210063.
URL https://doi.org/10.1145/3209978.3210063

[33] M. Zehlike, C. Castillo, Reducing disparate exposure in ranking: A learning to rank approach, in: WWW '20: The Web Conference 2020, ACM / IW3C2, 2020, pp. 2849–2855. doi:10.1145/3366424.3380048.
URL https://doi.org/10.1145/3366424.3380048

[34] F. Diaz, B. Mitra, M. D. Ekstrand, A. J. Biega, B. Carterette, Evaluating stochastic rankings with expected exposure, CoRR abs/2004.13157 (2020). arXiv:2004.13157.
URL https://arxiv.org/abs/2004.13157

[35] G. K. Patro, A. Biswas, N. Ganguly, K. P. Gummadi, A. Chakraborty, Fairrec: Two-sided fairness for personalized recommendations in two-sided platforms, in: WWW '20: The Web Conference 2020, ACM / IW3C2, 2020, pp. 1194–1204. doi:10.1145/3366423.3380196.
URL https://doi.org/10.1145/3366423.3380196

[36] P. Sapiezynski, W. Zeng, R. E. Robertson, A. Mislove, C. Wilson, Quantifying the impact of user attentionon fair group representation in ranked lists, in: Companion of The 2019 World Wide Web Conference, WWW 2019, ACM, 2019, pp. 553–562. doi:10.1145/3308560.3317595.
URL https://doi.org/10.1145/3308560.3317595

[37] K. Yang, J. Stoyanovich, Measuring fairness in ranked outputs, in: Proceedings of the 29th International Conference on Scientific and Statistical Database Management, ACM, 2017, pp. 22:1–22:6. doi:10.1145/3085504.3085526.
URL https://doi.org/10.1145/3085504.3085526

[38] M. D. Ekstrand, M. Tian, M. R. I. Kazi, H. Mehrpouyan, D. Kluver, Exploring author gender in book rating and recommendation, in: Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, ACM, 2018, pp. 242–250. doi:10.1145/3240323.3240373.
URL https://doi.org/10.1145/3240323.3240373

[39] W. Liu, R. Burke, Personalizing fairness-aware re-ranking, CoRR abs/1809.02921 (2018). arXiv:1809.02921.
URL http://arxiv.org/abs/1809.02921

[40] N. Sonboli, R. Burke, Localized fairness in recommender systems, in: Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, UMAP 2019, ACM, 2019, pp. 295–300. doi:10.1145/3314183.3323845.
URL https://doi.org/10.1145/3314183.3323845

[41] R. Mehrotra, J. McInerney, H. Bouchard, M. Lalmas, F. Diaz, 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, 2018, pp. 2243–2251. doi:10.1145/3269206.3272027.
URL https://doi.org/10.1145/3269206.3272027

[42] T. Kamishima, S. Akaho, H. Asoh, J. Sakuma, Recommendation independence, in: Conference on Fairness, Accountability and Transparency, FAT 2018, Vol. 81 of Proceedings of Machine Learning Research, PMLR, 2018, pp. 187–201.
URL http://proceedings.mlr.press/v81/kamishima18a.html

[43] K. Holstein, S. Doroudi, Fairness and equity in learning analytics systems

(fairlak), in: Companion Proceedings of the Ninth International Learning Analytics & Knowledge Conference (LAK 2019), 2019.

[44] B. Green, L. Hu, The myth in the methodology: Towards a recontextualization of fairness in machine learning, in: Proceedings of the machine learning: the debates workshop, 2018.

[45] D. Dessì, G. Fenu, M. Marras, D. Reforgiato Recupero, Leveraging cognitive computing for multi-class classification of e-learning videos, in: E. Blomqvist, K. Hose, H. Paulheim, A. Lawrynowicz, F. Ciravegna, O. Hartig (Eds.), The Semantic Web: ESWC 2017 Satellite Events - ESWC 2017 Satellite Events, Portorož, Slovenia, May 28 - June 1, 2017, Revised Selected Papers, Vol. 10577 of Lecture Notes in Computer Science, Springer, 2017, pp. 21–25. doi:10.1007/978-3-319-70407-4_5.
URL https://doi.org/10.1007/978-3-319-70407-4_5

[46] S. Barra, M. Marras, G. Fenu, Continuous authentication on smartphone by means of periocular and virtual keystroke, in: M. H. Au, S. Yiu, J. Li, X. Luo, C. Wang, A. Castiglione, K. Kluczniak (Eds.), Network and System Security - 12th International Conference, NSS 2018, Hong Kong, China, August 27-29, 2018, Proceedings, Vol. 11058 of Lecture Notes in Computer Science, Springer, 2018, pp. 212–220. doi:10.1007/978-3-030-02744-5_16.
URL https://doi.org/10.1007/978-3-030-02744-5_16

[47] R. Yu, Q. Li, C. Fischer, S. Doroudi, D. Xu, Towards accurate and fair prediction of college success: Evaluating different sources of student data, in: A. N. Rafferty, J. Whitehill, C. Romero, V. Cavalli-Sforza (Eds.), Proceedings of the 13th International Conference on Educational Data Mining, EDM 2020, Fully virtual conference, July 10-13, 2020, International Educational Data Mining Society, 2020.
URL https://educationaldatamining.org/files/conferences/EDM2020/papers/paper_194.pdf

[48] S. Doroudi, E. Brunskill, Fairer but not fair enough on the equitability of knowledge tracing, in: Proceedings of the 9th International Conference on Learning Analytics & Knowledge, LAK 2019, Tempe, AZ, USA, March 4-8, 2019, ACM, 2019, pp. 335–339. doi:10.1145/3303772.3303838.
URL https://doi.org/10.1145/3303772.3303838

[49] Q. Hu, H. Rangwala, Towards fair educational data mining: A case study on detecting at-risk students, in: A. N. Rafferty, J. Whitehill, C. Romero, V. Cavalli-Sforza (Eds.), Proceedings of the 13th International Conference on Educational Data Mining, EDM 2020, Fully virtual conference, July 10-13, 2020, International Educational Data Mining Society, 2020.
URL https://educationaldatamining.org/files/conferences/EDM2020/papers/paper_157.pdf

[50] A. Bellogín, P. Castells, I. Cantador, Statistical biases in information retrieval metrics for recommender systems, Inf. Retr. Journal 20 (6) (2017) 606–634.

[51] J. L. Herlocker, J. A. Konstan, J. Riedl, An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms, Inf. Retr. 5 (4) (2002) 287–310. doi:10.1023/A:1020443909834.
URL https://doi.org/10.1023/A:1020443909834

[52] B. M. Sarwar, G. Karypis, J. A. Konstan, J. Riedl, Item-based collaborative filtering recommendation algorithms, in: Proceedings of the Tenth International World Wide Web Conference, WWW 10, ACM, 2001, pp. 285–295. doi:10.1145/371920.372071.
URL https://doi.org/10.1145/371920.372071

[53] S. Rendle, C. Freudenthaler, Z. Gantner, L. Schmidt-Thieme, BPR: bayesian personalized ranking from implicit feedback, in: UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, AUAI Press, 2009, pp. 452–461.
URL https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=1630&proceeding_id=25

[54] Y. Koren, R. M. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, IEEE Computer 42 (8) (2009) 30–37. doi:10.1109/MC.2009.263.
URL https://doi.org/10.1109/MC.2009.263

[55] Y. Koren, 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, 2008, pp. 426–434. doi:10.1145/1401890.1401944.
URL https://doi.org/10.1145/1401890.1401944

[56] W. Feng, J. Tang, T. X. Liu, Understanding dropouts in moocs, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019, AAAI Press, 2019, pp. 517–524. doi:10.1609/aaai.v33i01.3301517.
URL https://doi.org/10.1609/aaai.v33i01.3301517

[57] J. Zhang, B. Hao, B. Chen, C. Li, H. Chen, J. Sun, Hierarchical reinforcement learning for course recommendation in moocs, in: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019, AAAI Press, 2019, pp. 435–442. doi:10.1609/aaai.v33i01.3301435.
URL https://doi.org/10.1609/aaai.v33i01.3301435

[58] J. Qiu, J. Tang, T. X. Liu, J. Gong, C. Zhang, Q. Zhang, Y. Xue, Modeling and predicting learning behavior in moocs, in: P. N. Bennett, V. Josifovski, J. Neville, F. Radlinski (Eds.), Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, San Francisco, CA, USA, February 22-25, 2016, ACM, 2016, pp. 93–102. doi:10.1145/2835776.2835842.
URL https://doi.org/10.1145/2835776.2835842

[59] K. Järvelin, J. Kekäläinen, Cumulated gain-based evaluation of IR techniques, ACM Trans. Inf. Syst. 20 (4) (2002) 422–446. doi:10.1145/582415.582418.

[60] A. Beutel, J. Chen, T. Doshi, H. Qian, L. Wei, Y. Wu, L. Heldt, Z. Zhao, L. Hong, E. H. Chi, C. Goodrow, Fairness in recommendation ranking through pairwise comparisons, in: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019., ACM, 2019, pp. 2212–2220. doi:10.1145/3292500.3330745.
URL https://doi.org/10.1145/3292500.3330745

[61] L. E. Celis, D. Straszak, N. K. Vishnoi, Ranking with fairness constraints, in: 45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, Vol. 107 of LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018, pp. 28:1–28:15. doi:10.4230/LIPIcs.ICALP.2018.28.
URL https://doi.org/10.4230/LIPIcs.ICALP.2018.28

[62] R. Burke, N. Sonboli, A. Ordonez-Gauger, Balanced neighborhoods for multi-sided fairness in recommendation, in: Conference on Fairness, Accountability and Transparency, FAT 2018, Vol. 81 of Proceedings of Machine Learning Research, PMLR, 2018, pp. 202–214.
URL http://proceedings.mlr.press/v81/burke18a.html

[63] J. G. Carbonell, J. Goldstein, The use of mmr, diversity-based reranking for reordering documents and producing summaries, in: SIGIR '98: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 1998, pp. 335–336. doi:10.1145/290941.291025.
URL https://doi.org/10.1145/290941.291025

[64] E. Gómez, L. Boratto, M. Salamó, Disparate impact in item recommendation: A case of geographic imbalance, in: D. Hiemstra, M.-F. Moens, J. Mothe, R. Perego, M. Potthast, F. Sebastiani (Eds.), Advances in Information Retrieval, Springer International Publishing, Cham, 2021, pp. 190–206. doi:10.1007/978-3-030-72113-8_13.