Optimizing Neighborhoods for Fair Top-N Recommendation

Stavroula Eleftherakis, Georgia Koutrika, Sihem Amer-Yahia

Optimizing Neighborhoods for Fair Top-N Recommendation (Author's Copy)

STAVROULA ELEFTHERAKIS, Athena Research Center, Greece and Grenoble Alpes University, France
GEORGIA KOUTRIKA, Athena Research Center, Greece
SIHEM AMER-YAHIA, French National Center for Scientific Research, Grenoble Alpes University, France

We address demographic bias in *neighborhood-learning* models for collaborative filtering recommendations. Despite their superior ranking performance, these methods can learn neighborhoods that inadvertently foster discriminatory patterns. Little work exists in this area, highlighting an important research gap. A notable yet solitary effort, *Balanced Neighborhood Sparse LInear Method (BNSLIM)* aims at balancing neighborhood influence across different demographic groups. Yet, BNSLIM is hampered by computational inefficiency, and its rigid balancing approach often impacts accuracy. In that vein, we introduce two novel algorithms. The first, an enhancement of BNSLIM, incorporates the *Alternating Direction Method of Multipliers (ADMM)* to optimize all similarities concurrently, greatly reducing training time. The second, *Fairly Sparse Linear Regression (FSLR)*, induces controlled sparsity in neighborhoods to reveal correlations among different demographic groups, achieving comparable efficiency while being more accurate. Their performance is evaluated using standard exposure metrics alongside a new metric for user coverage disparities. Our experiments cover various applications, including a novel exploration of bias in course recommendations by teachers' country development status. Our results show the effectiveness of our algorithms in imposing fairness compared to BNSLIM and other well-known fairness approaches.

CCS Concepts: • **Information systems → Recommender systems**; • **Computing methodologies → Learning linear models**.

Additional Key Words and Phrases: balanced neighborhoods, collaborative filtering, fairness, neighborhood learning, recommender systems, sparse linear models

ACM Reference Format:
Stavroula Eleftherakis, Georgia Koutrika, and Sihem Amer-Yahia. 2024. Optimizing Neighborhoods for Fair Top-N Recommendation (Author's Copy). 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, 16 pages. https://doi.org/10.1145/3627043.3659539

1 INTRODUCTION

Recommenders have become indispensable decision-making tools across diverse domains, from e-commerce to healthcare [17, 18]. As a result, their unbiased functioning has become a key concern for the scientific community. To mitigate bias, various dimensions of fairness have been explored in the literature [30, 38]. *Individual fairness* aims for similar individuals to be treated similarly, whereas *group fairness* aims for different groups to be treated similarly. These groups are often divided into *protected* and *non-protected*, with the former being subject to bias. Fairness in recommenders also requires an understanding of their *multi-stakeholder* nature [5]. *Consumer fairness (C-fairness)* looks at fairness from the side of the users, whereas *Provider fairness (P-fairness)* looks at fairness from the side of the items. Fairness benefits may also vary: *exposure* refers to the uniformity with which items or item groups are presented across all users or user

This work was supported by the European Union's Horizon 2020 program under the FAIRCORE4EOSC project, grant agreement No. 101057264.

Manuscript submitted to ACM

groups, and *effectiveness* refers to the degree to which an exposure is effective [11]. Based on the stage at which fairness considerations take place, approaches can be categorized into: *pre-processing*, *in-processing*, and *post-processing*.

Collaborative Filtering (CF) recommends items based on past user-item interactions, such as ratings or clicks [24]. While traditional CF methods use predefined similarity measures to compute user-user or item-item similarities, neighborhood-learning CF models [34], considered state-of-the-art (SOTA) [1, 15], dynamically derive these similarities from the interaction data. The SOTA models, *Sparse LInear Methods (SLIM)* [35] and *Embarrassingly Shallow AutoEncoders (EASE)* [41], differ in approach. SLIM estimates sparse similarities to preclude less important neighbors, while EASE estimates dense similarities to include correlated neighbors. Given that similarities may be derived from imbalanced interactions, a critical question arises: *do the computed similarities unwittingly sustain or exacerbate existing biases?*

Imbalanced interactions between different user or item demographics make these methods vulnerable to demographic bias, in which certain user or item demographic groups are unfairly treated over others [7]. We will use two motivating examples that highlight demographic bias in user-based and item-based methods.

User-neighborhood example. On an academic platform where students can be grouped based on their gender as either females or males, assume a preference imbalance: females mainly engage with the humanities, while males are more inclined towards STEM (Science, Technology, Engineering, and Mathematics). Consider a female student who has a balanced interest in both the humanities and STEM. If her neighborhood mostly includes female students, her recommendations are likely to favor courses in the humanities, potentially overlooking her interests in STEM.

Item-neighborhood example. On an academic platform where courses can be grouped based on the country of the teacher, assume a preference imbalance: courses by teachers from developed countries receive more interactions than those from developing countries. Consider a user engaging with a course from the developed country group. If its neighborhood mostly includes courses by teachers from developed countries, the user's recommendations are likely to favor such courses, potentially overlooking related courses by teachers from developing countries.

In that vein, the *balanced neighborhoods* concept was introduced, aimed at balanced group representation in user or item neighborhoods for C- or P-fairness, respectively, as realized by *Balanced Neighborhood SLIM (BNSLIM)* [7] that adds the *balanced regularization term* into SLIM. BNSLIM learns similarities such that, within each neighborhood, the cumulative similarity across neighbors of the protected group is equal to that of the non-protected group. For our user-neighborhood example, this means having a balance between the influence of female and male neighbors. For our item-neighborhood example, this means having a balance between the influence of courses by teachers from developed and developing countries. BNSLIM employs *Coordinate Descent (CD)* for optimizing each similarity individually, which is impractical for large datasets. To partly address this, during each CD step, the algorithm focuses solely on the closest neighbors, defined at the beginning and fixed throughout the algorithm [6]. Still, the computational burden is significant [3]. More crucially, such static neighbor selection potentially harms the intended balancing effect due to a lack of group diversity; thus, the true potential of balancing neighborhoods for fairness remains unverified to its full extent. Lastly, this rigid balancing approach may adversely affect personalization by increasing (decreasing) the similarity scores with irrelevant (relevant) neighbors just for the sake of influence balance.

Overcoming these limitations, *our work introduces two novel in-processing algorithms for group fairness to mitigate demographic bias in neighborhood-learning models*. We particularly focus on the benefit of exposure[1], reducing exposure disparities in top-N lists: C-fairness involves balancing the exposure of item groups across user groups, while P-fairness involves balancing the exposure of item groups within the top-N lists of all users. For this purpose, our

[1]We use *exposure* and *visibility* interchangeably to describe how frequently items appear in top-N recommendation lists.

algorithms optimize user or item neighborhoods for C- or P-fairness, respectively. The first algorithm, $BNSLIM_{ADMM}$, tackles the efficiency bottleneck of BNSLIM by allowing simultaneous optimization of all similarities. The second algorithm, *Fairly Sparse Linear Regression (FSLR)*, induces controlled sparsity in neighborhoods, ensuring representation by demographically varied, yet *correlated*, counterparts so as to preserve personalization. This is achieved by integrating a novel regularization term into EASE. For our user-neighborhood example, this means that the recommendations of the female student will largely be influenced by the preferences of similar male students. For our item-neighborhood example, this means that the recommendations of the user will largely be influenced by similar courses from the developing countries group. Both algorithms employ the *Alternating Direction Method of Multipliers (ADMM)* for its ability to decompose complex optimization problems and convergence properties [4].

We ran extensive experiments across various recommendation domains: movies (Movielens 1M [19]), music (LastFM 1K [8]), courses (COCO [12]), and books (Goodreads [44]). We explore how teachers' country development status affects course recommendations, a P-fairness scenario that (to the best of our knowledge) has not previously been studied. Furthermore, we study fairness not only by measuring exposure disparities of item groups across all users or user groups (our primary fairness goal), but also by measuring disparities in the coverage of user groups (in terms of the item categories they access through recommendations) or item groups. For the latter, we present *User-coverage Parity (u-Parity)*, a novel exposure metric for C-fairness that measures the difference in percentages of protected and non-protected users covered by each item category. Results show that under the given fairness goals, our algorithms outperform other leading fairness-aware models [7, 9, 46] in balancing personalization and fairness, with FSLR showing superior accuracy than $\text{BNSLIM}_{\text{ADMM}}$. Notably, $\text{BNSLIM}_{\text{ADMM}}$ significantly reduces training time by around 99%, compared to BNSLIM. Overall, our algorithms are uniquely capable of achieving satisfactory results for both C- and P-fairness.

Our contributions are summarized as follows:

(1) We introduce $\text{BNSLIM}_{\text{ADMM}}$, a faster implementation of BNSLIM that updates all similarities concurrently.
(2) We introduce FSLR, a novel model that induces controlled sparsity in neighborhoods, ensuring representation by demographically varied, yet correlated, counterparts so as to preserve personalization.
(3) We present u-Parity, a novel exposure metric for C-fairness, which measures the disparity in coverage percentages between protected and non-protected users within each item category.
(4) We study the influence of teachers' country development status in course recommendations, offering insights into a previously unexplored P-fairness application.
(5) We provide empirical evidence across diverse datasets, comparing $\text{BNSLIM}_{\text{ADMM}}$ and FSLR to other established fair recommendation models [7, 9, 46], showing their superiority in balancing accuracy and fairness.

2 RELATED WORK

Fairness definitions. In recommenders, fairness definitions are shaped by task (i.e., rating prediction or ranking) and fairness type (i.e., individual or group, C- or P-fairness). In the rating-prediction task, C- and P-fairness aim to balance average ratings or prediction errors across user or item groups [22, 46], while individual C-fairness aims to minimize prediction error variance across all pairs of users (individual C-fairness) [39]. In the ranking task, P-fairness aims to balance the exposure of item groups [7, 16, 48], or their coverage in top-N lists [29], whereas C-fairness aims to balance the exposure of items [9] or item groups [7], or ensure comparable relevance levels [13, 26], across user groups. *We target group fairness in top-N lists (rankings), focusing on either C-fairness (balancing the exposure of item groups across user groups) or P-fairness (balancing the exposure of item groups within the top-N lists of all users).*

Table 1. Notation and Definitions.

Symbol	Description	Symbol	Description
$\mathcal{U}$, $\mathcal{I}$	Sets of user and item identifiers	$\oslash$	Element-wise division
$\lvert\cdot\rvert$	The size of a set	$\odot$	Element-wise multiplication (Hadamard product)
$\boldsymbol{R}$	The $\lvert\mathcal{I}\rvert \times \lvert\mathcal{U}\rvert$ interaction matrix	$\lVert\boldsymbol{W}\rVert_F^2$	Squared Frobenius norm of $\boldsymbol{W}$
$\boldsymbol{r}_{i.}$, $\boldsymbol{r}_{.u}$	Row and column vectors of $\boldsymbol{R}$	$\lVert\boldsymbol{W}\rVert_1$	Sum of the absolute elements of $\boldsymbol{W}$
$\mathrm{Diag}(\boldsymbol{\gamma})$	Diagonal matrix with elements of vector $\boldsymbol{\gamma}$	$\lVert\boldsymbol{W}\rVert_\infty$	Maximum absolute value in $\boldsymbol{W}$
$\mathrm{diag}(\boldsymbol{W})$	Vector of diagonal elements of matrix $\boldsymbol{W}$	$\mathbf{1}_{\text{condition}}$	Indicator function, 1 if condition is true, otherwise 0
$\langle\cdot,\cdot\rangle$	Frobenius inner product	$\lVert\boldsymbol{a}\rVert_2^2$	Squared Euclidean norm of vector $\boldsymbol{a}$
$s_\tau(\cdot)$	Soft-thresholding operator with threshold τ	$(\cdot)_+$	Positive part operator

Fairness-aware CF recommenders can be grouped into the following three categories:

Fairness-aware embedding methods. These methods learn bias-free embeddings. [20–22] introduced regularizers to decouple predicted ratings from sensitive attributes. [46] focused on the dependency of those attributes with prediction errors. [47] proposed the use of a sensitive embedding matrix along with an orthogonality regularizer. [48] proposed a ranking model based on adversarial learning to handle item under-recommendation. [45] introduced a user model comprising a bias-aware embedding built through sensitive attribute prediction and a bias-free embedding built on adversarial learning. [27] enabled consumers to define their sensitive attributes, using adversarial learning for rankings that are personalized yet free from these attributes' influence. *Instead of purifying the learned user or item embeddings from sensitive attributes, our work regulates the neighborhood distribution of users or items to promote fairness.*

Data-driven fairness methods. Data augmentation approaches mitigate unfairness using a dual-optimization strategy, optimizing the model and fake data, such as user profiles [39] or interactions [9], concurrently. However, each method follows its own approach to fake data synthesis: [39] tries to meet predefined fairness objectives, whereas [9], independent of fairness metrics, employs two hypotheses to balance interactions between user groups. Instead of increasing the size of the dataset, a different approach is to modify the user or item relationships. BNSLIM balances influences within user or item neighborhoods, aiming for equal representation of protected and non-protected members within a neighborhood [7]. *Like BNSLIM, our algorithms target C- or P-fairness by optimizing neighborhoods.*

Rebalancing methods. These methods mitigate unfairness by rebalancing the input or output of recommenders. [14, 31] studied data resampling techniques to adjust the representation of protected and non-protected user groups, mitigating demographic bias. [40] proposed a probabilistic re-ranker for balancing item groups' exposure. [29] proposed a re-ranker for balancing users' ranking accuracy and provider diversity. [26] introduced a re-ranking framework for balancing ranking accuracy across user groups with different activity levels. [33] introduced a framework that extends the concept of [26] to also balance the exposure of long-tail and short-head items. *Our algorithms address fairness directly during the processing stage, eliminating the need for additional re-samping or re-ranking steps.*

3 PRELIMINARIES

This section provides an overview of the neighborhood-learning models SLIM and EASE, alongside ADMM. Originally designed for item-item similarity learning, SLIM and EASE can adapt to learn user-user similarities by transposing the user-item interaction matrix $\boldsymbol{R}$. *For our purposes, focusing on user-user similarities, we directly define $\boldsymbol{R}$ as an item-user interaction matrix to streamline notation and eliminate the need for constant transposing.* Table 1 outlines our notation with sets denoted by calligraphic letters or {.}, vectors by bold lowercase, and matrices by bold uppercase.

3.1 Sparse LInear Methods (SLIM)

SLIM [35] optimizes a least-squares regression model with elastic net regularization, yielding sparse similarity estimates for each user that represent its neighborhood. The acquired user-user similarities are then utilized as weights for the item vectors: for an arbitrary user $u \in \mathcal{U}$ and item $i \in \mathcal{I}$, the ranking estimate is $\hat{r}_{iu} = \boldsymbol{r}_{i.}\boldsymbol{w}_{.u}$, where $\boldsymbol{w}_{.u} \in \mathbb{R}^{|\mathcal{U}|}$ the similarity weights of user u relative to all users. Ranking estimates for unobserved items are ranked in descending order, recommending the top-N. The similarities are computed by solving the following optimization problem:

$$\boldsymbol{W}^* = \arg\min_{\boldsymbol{W}} \frac{1}{2}\|\boldsymbol{R} - \boldsymbol{R}\boldsymbol{W}\|_F^2 + \lambda_1\|\boldsymbol{W}\|_1 + \frac{\lambda_2}{2}\|\boldsymbol{W}\|_F^2, \text{ s.t. } \operatorname{diag}(\boldsymbol{W}) = \boldsymbol{0} \text{ and } \boldsymbol{W} \geq 0, \tag{1}$$

where $\boldsymbol{W} = [\boldsymbol{w}_{.1}, \ldots, \boldsymbol{w}_{.|\mathcal{U}|}] \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{U}|}$ the user-user similarity matrix and $\lambda_1, \lambda_2 > 0$ the regularization parameters. The first term minimizes reconstruction error, the second term shapes the neighborhoods, and the third term handles overfitting. The first constraint, *self-similarity constraint*, precludes self-similarity, where $\boldsymbol{W}^* = \boldsymbol{I}_{|\mathcal{U}|}$; and the second constraint, *non-negativity constraint*, permits only positive relations between users for interpretability purposes.

3.2 Embarrassingly Shallow AutoEncoders (EASE)

Steck [41] omitted the sparsity-enforcing regularization and the non-negativity constraint of SLIM, resulting in a new model called EASE. The similarities are now computed by solving the following optimization problem:

$$\boldsymbol{W}^* = \arg\min_{\boldsymbol{W}} \frac{1}{2}\|\boldsymbol{R} - \boldsymbol{R}\boldsymbol{W}\|_F^2 + \frac{\lambda_2}{2}\|\boldsymbol{W}\|_F^2, \text{ s.t. } \operatorname{diag}(\boldsymbol{W}) = \boldsymbol{0}. \tag{2}$$

This simplification allows EASE to benefit from a closed-form solution, a feature SLIM lacks. EASE leads to increased ranking accuracy and is faster due to the direct computation of similarities enabled by the closed-form solution [42].

3.3 Alternating Direction Method of Multipliers (ADMM)

ADMM, popular for its ability to decompose complex optimization problems, is widely used in various domains [4], including recommenders [10, 42]. It is a first-order optimization method for optimizing problems of the form:

$$\min_{\boldsymbol{W},\boldsymbol{Z}} f(\boldsymbol{W}) + g(\boldsymbol{Z}) \text{ s.t. } \boldsymbol{A}\boldsymbol{W} + \boldsymbol{B}\boldsymbol{Z} = \boldsymbol{C}, \tag{3}$$

where f and g are two convex closed functions; $\boldsymbol{W}$ and $\boldsymbol{Z}$ are the unknown matrices of variables to be optimized; and $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are the coefficient matrices in the constraint of the optimization problem.

The solution procedure can be simplified by minimizing the augmented Lagrangian of the above-given problem. This function merges the primal and dual problems, i.e., the original objective function and its constraint, into one:

$$\mathcal{L}_\rho(\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{Y}) = f(\boldsymbol{W}) + g(\boldsymbol{Z}) + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{W} + \boldsymbol{B}\boldsymbol{Z} - \boldsymbol{C}\|_F^2 \; + \langle \boldsymbol{Y}, (\boldsymbol{A}\boldsymbol{W} + \boldsymbol{B}\boldsymbol{Z} - \boldsymbol{C})\rangle\,, \tag{4}$$

where $\rho > 0$ is a parameter that controls the trade-off between solution accuracy and convergence rate. $\boldsymbol{Y}$ denotes the matrix of Lagrangian multipliers, representing the magnitude of the constraint violation.

In ADMM, the augmented Lagrangian is minimized through an iterative process that progressively converges towards the solution. This procedure entails three sequential updates, with each iteration denoted by k:

$$\textcircled{1}\; \boldsymbol{W}^{k+1} = \arg\min_{\boldsymbol{W}} \mathcal{L}_\rho(\boldsymbol{W}, \boldsymbol{Z}^k, \boldsymbol{Y}^k)\; \textcircled{2}\; \boldsymbol{Z}^{k+1} = \arg\min_{\boldsymbol{Z}} \mathcal{L}_\rho(\boldsymbol{W}^{k+1}, \boldsymbol{Z}, \boldsymbol{Y}^k)\; \textcircled{3}\; \boldsymbol{Y}^{k+1} = \boldsymbol{Y}^k + \rho(\boldsymbol{A}\boldsymbol{W}^{k+1} + \boldsymbol{B}\boldsymbol{Z}^{k+1} - \boldsymbol{C}) \tag{5}$$

The steps outlined in Eq. (5) are repeated until convergence or a maximum iteration limit is reached.

4 BALANCED NEIGHBORHOODS

We first revisit BNSLIM's objective and then reformulate it to enable the simultaneous optimization of all similarities through the ADMM framework. While presented for *C-fairness through balanced user neighborhoods*, BNSLIM is also able to address *P-fairness through balanced item neighborhoods* by transposing the item-user interaction matrix $\boldsymbol{R}$.

4.1 Balanced Neighborhood SLIM (BNSLIM)

BNSLIM attempts to learn a matrix of user-user relationships, with the goal that each user can be represented equally by users from two distinct demographic groups: the protected $\mathcal{G}_{\text{p}}$, which is subject to bias, and the non-protected $\mathcal{G}_{\text{np}}$. Specifically, it aims for each user to be equally influenced by members from both of these groups in the resultant similarity matrix. The optimization problem of BNSLIM for synthesizing this balanced matrix $\boldsymbol{W}^*$ is formulated as:

$$\boldsymbol{W}^* = \arg\min_{\boldsymbol{W}} \frac{1}{2}\|\boldsymbol{R} - \boldsymbol{R}\boldsymbol{W}\|_F^2 + \lambda_1\|\boldsymbol{W}\|_1 + \frac{\lambda_2}{2}\|\boldsymbol{W}\|_F^2 + \frac{\lambda_3}{2}\sum_{u\in\mathcal{U}}(\boldsymbol{p}^\top \boldsymbol{w}_{\cdot u})^2 \text{ s.t. } \operatorname{diag}(\boldsymbol{W}) = \boldsymbol{0} \text{ and } \boldsymbol{W} \geq 0, \tag{6}$$

where $\boldsymbol{W} = [\boldsymbol{w}_{\cdot 1}, \ldots, \boldsymbol{w}_{\cdot|\mathcal{U}|}] \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{U}|}$ is the user-user similarity matrix; $\boldsymbol{p} \in \mathbb{R}^{|\mathcal{U}|}$ is a vector constructed with elements set to 1 for members of $\mathcal{G}_{\text{p}}$ and -1 for members of $\mathcal{G}_{\text{np}}$; and $\lambda_1, \lambda_2, \lambda_3 > 0$ are the regularization parameters. The inclusion of the non-negativity constraint in BNSLIM is crucial as it ensures similarities within $\mathcal{G}_{\text{p}}$ and $\mathcal{G}_{\text{np}}$ are not negated. Note that omitting the fourth term from Eq. (6) effectively transforms the model into SLIM [35]. This fourth term is known as the balanced regularization term and is responsible for learning balanced neighborhoods.

The optimization problem delineated in Eq. (6) is conventionally resolved using CD [7]. Within this framework, for each distinct pair $(u, v) \in \mathcal{U} \times \mathcal{U}$, where $u \neq v$, the similarity is iteratively updated as per the rule:

$$w_{uv} = \frac{(s_{\lambda_1}(\sum_{i=1}^{|\mathcal{I}|}(r_{iu} - \sum_{l=1,\, l\neq u,v}^{|\mathcal{U}|} r_{il}w_{lu}) + \lambda_3 p_v \sum_{l=1,\, l\neq u,v}^{|\mathcal{U}|} p_l w_{lu}))_+}{\sum_{i=1}^{|\mathcal{I}|} r_{iv}^2 + \lambda_2 + \lambda_3}. \tag{7}$$

From Eq. (7), it is apparent that optimizing each similarity individually poses a computational challenge, especially for large user bases. In an effort to improve performance, BNSLIM considers updating similarities between the closest neighbors only, defined at the beginning and fixed throughout the algorithm [6]. Still, this heuristic does not lead to notable performance gains and results in suboptimal solutions [3]. To address this, we propose to re-write Eq. (6) so that the balanced regularization term is expressed in a compact matrix form, i.e.,

$$\boldsymbol{W}^* = \arg\min_{\boldsymbol{W}} \frac{1}{2}\|\boldsymbol{R} - \boldsymbol{R}\boldsymbol{W}\|_F^2 + \lambda_1\|\boldsymbol{W}\|_1 + \frac{\lambda_2}{2}\|\boldsymbol{W}\|_F^2 + \frac{\lambda_3}{2}\|\boldsymbol{p}^\top\boldsymbol{W}\|_2^2 \text{ s.t. } \operatorname{diag}(\boldsymbol{W}) = \boldsymbol{0} \text{ and } \boldsymbol{W} \geq 0. \tag{8}$$

We may now solve this optimization problem for the entire $\boldsymbol{W}$ matrix, addressing the efficiency bottleneck of BNSLIM.

4.2 Optimization via ADMM

Under the ADMM framework, the augmented Lagrangian for this problem is given by

$$\mathcal{L}_\rho(\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{Y}) = \overbrace{\frac{1}{2}\|\boldsymbol{R} - \boldsymbol{R}\boldsymbol{W}\|_F^2 + \frac{\lambda_2}{2}\|\boldsymbol{W}\|_F^2 + \frac{\lambda_3}{2}\|\boldsymbol{p}^\top\boldsymbol{W}\|_2^2}^{f(\boldsymbol{W})} + \underbrace{\lambda_1\|\boldsymbol{Z}\|_1}_{g(\boldsymbol{Z})} + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{W} + \boldsymbol{B}\boldsymbol{Z} - \boldsymbol{C}\|_F^2 + \langle \boldsymbol{Y}, \boldsymbol{A}\boldsymbol{W} + \boldsymbol{B}\boldsymbol{Z} - \boldsymbol{C}\rangle, \tag{9}$$

where $\boldsymbol{A} = \boldsymbol{I}_{|\mathcal{U}|}$ (the $|\mathcal{U}| \times |\mathcal{U}|$ identity matrix), $\boldsymbol{B} = -\boldsymbol{I}_{|\mathcal{U}|}$, and $\boldsymbol{C} = \boldsymbol{O}_{|\mathcal{U}|}$ (the $|\mathcal{U}| \times |\mathcal{U}|$ matrix of zeroes).

The updating steps of ADMM, as outlined in Section 3, can be adapted to fit the form and constraints of the BNSLIM model. In what follows, we provide detailed descriptions of each step within the ADMM framework.

Algorithm 1 BNSLIM via ADMM

Input: $\boldsymbol{R}, \boldsymbol{p}, \lambda_1, \lambda_2, \lambda_3, \rho$ **Output:** $\boldsymbol{W}^* = \boldsymbol{W}^{k+1}$
Initialize: $k = 0$, $\boldsymbol{W}^0 = \boldsymbol{Z}^0$, and $\boldsymbol{Y}^0$
$\boldsymbol{G} = \boldsymbol{R}^\top \boldsymbol{R}$, $\boldsymbol{P} = (\boldsymbol{G} + (\lambda_2 + \rho)\boldsymbol{I}_{|\mathcal{U}|} + \lambda_3 \boldsymbol{p}\boldsymbol{p}^\top)^{-1}$
repeat
$\boldsymbol{Q}^k = \boldsymbol{P}(\boldsymbol{G} + \rho \boldsymbol{Z}^k - \boldsymbol{Y}^k)$
$\boldsymbol{W}^{k+1} = \boldsymbol{Q}^k - \boldsymbol{P}\,\mathrm{Diag}(\mathrm{diag}\,(\boldsymbol{Q}^k) \oslash \mathrm{diag}\,(\boldsymbol{P}))$
$\boldsymbol{Z}^{k+1} = (s_{\lambda_1/\rho}(\boldsymbol{W}^{k+1} + \frac{1}{\rho}\boldsymbol{Y}^k))_+$
$\boldsymbol{Y}^{k+1} = \boldsymbol{Y}^k + \rho \cdot (\boldsymbol{W}^{k+1} - \boldsymbol{Z}^{k+1})$
$k = k + 1$
until convergence

Algorithm 2 FSLR via ADMM

Input: $\boldsymbol{R}, \boldsymbol{M}, \lambda_1, \lambda_2, \rho$ **Output:** $\boldsymbol{W}^* = \boldsymbol{W}^{k+1}$
Initialize: $k = 0$, $\boldsymbol{W}^0 = \boldsymbol{Z}^0$, and $\boldsymbol{Y}^0$
$\boldsymbol{G} = \boldsymbol{R}^\top \boldsymbol{R}$, $\boldsymbol{P} = (\boldsymbol{G} + (\lambda_2 + \rho)\boldsymbol{I}_{|\mathcal{U}|})^{-1}$
repeat
$\boldsymbol{Q}^k = \boldsymbol{P}(\boldsymbol{G} + \rho \boldsymbol{Z}^k - \boldsymbol{Y}^k)$
$\boldsymbol{W}^{k+1} = \boldsymbol{Q}^k - \boldsymbol{P}\,\mathrm{Diag}(\mathrm{diag}\,(\boldsymbol{Q}^k) \oslash \mathrm{diag}\,(\boldsymbol{P}))$
$\boldsymbol{Z}^{k+1} = s_{\lambda_1/\rho}(\boldsymbol{W}^{k+1} + \frac{1}{\rho}\boldsymbol{Y}^k) \odot \boldsymbol{M} + (\boldsymbol{W}^{k+1} + \frac{1}{\rho}\boldsymbol{Y}^k) \odot \tilde{\boldsymbol{M}}$
$\boldsymbol{Y}^{k+1} = \boldsymbol{Y}^k + \rho \cdot (\boldsymbol{W}^{k+1} - \boldsymbol{Z}^{k+1})$
$k = k + 1$
until convergence

Step 1: W update. The self-similarity constraint is incorporated into the update process using Lagrange multipliers, penalizing the loss function when it is violated. Therefore, the new loss for updating $\boldsymbol{W}$ is given by

$$\mathcal{L}_\rho^{(c)}(\boldsymbol{W}, \boldsymbol{Z}^k, \boldsymbol{Y}^k) = f(\boldsymbol{W}) + g(\boldsymbol{Z}^k) + \frac{\rho}{2}\|\boldsymbol{W} - \boldsymbol{Z}^k\|_F^2 + \left\langle \boldsymbol{Y}^k, \boldsymbol{W} - \boldsymbol{Z}^k \right\rangle + \boldsymbol{\gamma}^\top \mathrm{diag}(\boldsymbol{W}), \tag{10}$$

where $f(\cdot)$ includes all the involving differentiable norms[2] and $\boldsymbol{\gamma} \in \mathbb{R}^{|\mathcal{U}|}$ is the Lagrangian multiplier vector.
By taking the partial derivative of $\mathcal{L}_\rho^{(c)}(\boldsymbol{W}, \boldsymbol{Z}^k, \boldsymbol{Y}^k)$ w.r.t. $\boldsymbol{W}$ and setting it equal to zero, we find that

$$\boldsymbol{W}^{k+1} = \boldsymbol{Q}^k - \boldsymbol{P}\,\mathrm{Diag}\,(\boldsymbol{\gamma}), \text{ where } \boldsymbol{Q}^k = \boldsymbol{P}\boldsymbol{R}^\top \boldsymbol{R} + \boldsymbol{P}(\rho \boldsymbol{Z}^k - \boldsymbol{Y}^k) \text{ and } \boldsymbol{P} = (\boldsymbol{R}^\top \boldsymbol{R} + (\lambda_2 + \rho)\boldsymbol{I}_{|\mathcal{U}|} + \lambda_3 \boldsymbol{p}\boldsymbol{p}^\top)^{-1}. \tag{11}$$

Applying the constraint $\mathrm{diag}(\boldsymbol{W}^{k+1}) = \boldsymbol{0}$, it holds that $\boldsymbol{\gamma} = \mathrm{diag}\,(\boldsymbol{Q}^k) \oslash \mathrm{diag}\,(\boldsymbol{P})$. Thus, Eq. (11) can be reformulated as:

$$\boldsymbol{W}^{k+1} = \boldsymbol{Q}^k - \boldsymbol{P}\,\mathrm{Diag}(\mathrm{diag}\,(\boldsymbol{Q}^k) \oslash \mathrm{diag}\,(\boldsymbol{P})). \tag{12}$$

Step 2: Z update. Setting the partial derivative of $\mathcal{L}_\rho(\boldsymbol{W}^{k+1}, \boldsymbol{Z}, \boldsymbol{Y}^k)$ w.r.t. $\boldsymbol{Z}$ to zero involves handling the $\|\cdot\|_1$ norm, which is non-differentiable at zero. For this purpose, the soft-thresholding operator, $s_{\lambda_1/\rho}(\cdot)$, is applied to act as a proximal operator for the norm. We then project the output onto the non-negative orthant to preserve the non-negativity constraint. The update formula for $\boldsymbol{Z}$ is thus expressed as:

$$\boldsymbol{Z}^{k+1} = \left(s_{\lambda_1/\rho}(\boldsymbol{W}^{k+1} + \frac{1}{\rho}\boldsymbol{Y}^k)\right)_+. \tag{13}$$

Step 3: Y update. The final step involves updating $\boldsymbol{Y} \in \mathbb{R}^{|\mathcal{U}| \times |\mathcal{U}|}$ using the last formula given in Eq. (5). This update is crucial for ensuring that the variables $\boldsymbol{W}$ and $\boldsymbol{Z}$, updated in the earlier steps, converge towards equality.

The update steps of the ADMM solution for Eq. (8) are summarized in Alg. 1.

5 FAIRLY SPARSE NEIGHBORHOODS

This section presents the FSLR model, aimed at *C-fairness through fairly sparse user neighborhoods*. FSLR is also able to address *P-fairness through fairly sparse item neighborhoods* by simply transposing the item-user interaction matrix $\boldsymbol{R}$.

5.1 Fairly Sparse Linear Regression (FSLR)

BNSLIM's strategy of balancing neighborhood influence across different demographic groups may hurt accuracy as it may lead to increasing (decreasing) the similarity scores with irrelevant (relevant) neighbors. Acknowledging the

[2] f embodies EASE with the balanced regularization term. Offline experiments revealed that without sparsity-enforcing regularization, the model struggles to balance neighborhoods, likely due to EASE's tendency to estimate dense similarities.

importance of achieving fairness with minimal impact on personalization, we also propose an alternative approach: induce controlled sparsity in neighborhoods to reveal *correlations* among users of different demographic groups. In particular, we introduce FSLR, a model that attempts to learn a partially, i.e., *fairly*, sparse matrix of user-user relationships, with the goal that each user is primarily represented as a linear combination of users belonging to different groups. The optimization problem of FSLR for synthesizing this fairly sparse matrix $\boldsymbol{W}^*$ is formulated as:

$$\boldsymbol{W}^* = \arg\min_{\boldsymbol{W}} \frac{1}{2}\|\boldsymbol{R} - \boldsymbol{R}\boldsymbol{W}\|_F^2 + \lambda_1\|\boldsymbol{M} \odot \boldsymbol{W}\|_1 + \frac{\lambda_2}{2}\|\boldsymbol{W}\|_F^2, \text{ s.t. } \operatorname{diag}(\boldsymbol{W}) = \boldsymbol{0}, \tag{14}$$

where $\boldsymbol{M} = [\boldsymbol{\mu}_{\cdot,1}, \ldots, \boldsymbol{\mu}_{\cdot,|\mathcal{U}|}] \in \{0,1\}^{|\mathcal{U}|\times|\mathcal{U}|}$ is the membership matrix with $\mu_{uv} = 1$ if user u belongs to the same group as user v and $\mu_{uv} = 0$ otherwise, $\forall u, v \in \mathcal{U}$; and $\lambda_1, \lambda_2 > 0$ are the regularization parameters.

Observe that applying the non-negativity constraint, $\boldsymbol{W} \geq 0$, into Eq. (14) yields to a modified version of SLIM that enforces elastic-net regularization within demographic groups. Omitting the second term of Eq. (14) reverts the model to its EASE form. The novelty lies on incorporating $\boldsymbol{M}$ into the $\|\cdot\|_1$ norm as λ_1 selectively reduces the influence of neighbors from the same group on each individual user, facilitating the discovery of important relationships with users from different groups. Adding the balanced regularization term introduces a conflict between prioritizing cross-group similarities (FSLR) and balancing group similarities (BNSLIM), justifying the proposal of two separate models.

5.2 Optimization via ADMM

Under the ADMM framework, the augmented Lagrangian for FSLR is given by:

$$\mathcal{L}_\rho(\boldsymbol{W}, \boldsymbol{Z}, \boldsymbol{Y}) = \overbrace{\frac{1}{2}\|\boldsymbol{R} - \boldsymbol{R}\boldsymbol{W}\|_F^2 + \frac{\lambda_2}{2}\|\boldsymbol{W}\|_F^2}^{f(\boldsymbol{W})} + \underbrace{\lambda_1\|\boldsymbol{M} \odot \boldsymbol{Z}\|_1}_{g(\boldsymbol{Z})} + \frac{\rho}{2}\|\boldsymbol{A}\boldsymbol{W} + \boldsymbol{B}\boldsymbol{Z} - \boldsymbol{C}\|_F^2 + \langle \boldsymbol{Y}, \boldsymbol{A}\boldsymbol{W} + \boldsymbol{B}\boldsymbol{Z} - \boldsymbol{C}\rangle, \tag{15}$$

where $\boldsymbol{A} = \boldsymbol{I}_{|\mathcal{U}|}$ (the $|\mathcal{U}| \times |\mathcal{U}|$ identity matrix), $\boldsymbol{B} = -\boldsymbol{I}_{|\mathcal{U}|}$, and $\boldsymbol{C} = \boldsymbol{O}_{|\mathcal{U}|}$ (the $|\mathcal{U}| \times |\mathcal{U}|$ matrix of zeroes).

This formulation enables concurrent optimization of EASE's objective and our controlled sparsity objective. The last two terms of the loss balance the discrepancy between the solutions of these two subproblems.

As before, the updating steps of ADMM are adapted to accommodate the specificities of this model.

Step 1: $\boldsymbol{W}$ update. The self-similarity constraint is incorporated into the update process using Lagrange multipliers, penalizing the loss function when it is violated. Therefore, the new loss for updating $\boldsymbol{W}$ is given by

$$\mathcal{L}_\rho^{(c)}(\boldsymbol{W}, \boldsymbol{Z}^k, \boldsymbol{Y}^k) = f(\boldsymbol{W}) + g(\boldsymbol{Z}^k) + \frac{\rho}{2}\|\boldsymbol{W} - \boldsymbol{Z}^k\|_F^2 + \left\langle \boldsymbol{Y}^k, \boldsymbol{W} - \boldsymbol{Z}^k \right\rangle + \boldsymbol{\gamma}^\top \operatorname{diag}(\boldsymbol{W}). \tag{16}$$

To determine the optimal solution, we take the partial derivative of $\mathcal{L}_\rho^{(c)}(\boldsymbol{W}, \boldsymbol{Z}^k, \boldsymbol{Y}^k)$ w.r.t. $\boldsymbol{W}$ and set it to zero, i.e.,

$$\boldsymbol{W}^{k+1} = \boldsymbol{Q}^k - \boldsymbol{P}\operatorname{Diag}(\boldsymbol{\gamma}), \text{ where } \boldsymbol{Q}^k = \boldsymbol{P}\boldsymbol{R}^\top\boldsymbol{R} + \boldsymbol{P}(\rho\boldsymbol{Z}^k - \boldsymbol{Y}^k) \text{ and } \boldsymbol{P} = (\boldsymbol{R}^\top\boldsymbol{R} + (\lambda_2 + \rho)\boldsymbol{I}_{|\mathcal{U}|})^{-1}. \tag{17}$$

Applying the constraint $\operatorname{diag}(\boldsymbol{W}^{k+1}) = \boldsymbol{0}$, it holds that $\boldsymbol{\gamma} = \operatorname{diag}(\boldsymbol{Q}^k) \oslash \operatorname{diag}(\boldsymbol{P})$. Thus, Eq. (17) can be reformulated as:

$$\boldsymbol{W}^{k+1} = \boldsymbol{Q}^k - \boldsymbol{P}\operatorname{Diag}(\operatorname{diag}(\boldsymbol{Q}^k) \oslash \operatorname{diag}(\boldsymbol{P})). \tag{18}$$

Step 2: $\boldsymbol{Z}$ update. By setting the partial derivative of $\mathcal{L}_\rho(\boldsymbol{W}^{k+1}, \boldsymbol{Z}, \boldsymbol{Y}^k)$ w.r.t. $\boldsymbol{Z}$ to zero, it follows that

$$\boldsymbol{Z}^{k+1} = s_{\lambda_1/\rho}(\boldsymbol{W}^{k+1} + \frac{1}{\rho}\boldsymbol{Y}^k) \odot \boldsymbol{M} + (\boldsymbol{W}^{k+1} + \frac{1}{\rho}\boldsymbol{Y}^k) \odot \tilde{\boldsymbol{M}}, \tag{19}$$

where $\tilde{\boldsymbol{M}} = \boldsymbol{J}_{|\mathcal{U}|} - \boldsymbol{M}$, with $\boldsymbol{J}_{|\mathcal{U}|}$ being a $|\mathcal{U}| \times |\mathcal{U}|$ matrix of ones, indicating $\tilde{\boldsymbol{M}}$ as the binary complement of $\boldsymbol{M}$.

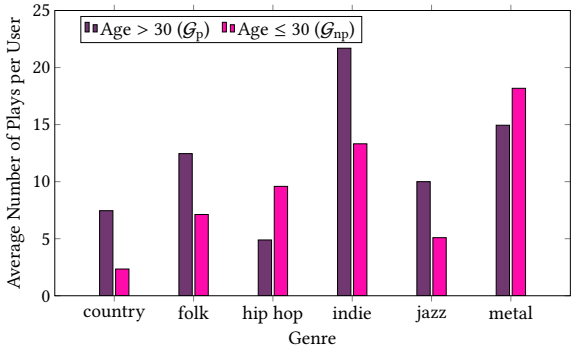

Fig. 1. Imbalances in Music Genre Preferences Across Different Age Groups.

In essence, this step applies a penalized update to those elements of the similarity matrix associated with users within the same group, while the remaining elements undergo an update process that is unaffected by the L1 regularizer.

Step 3: Y update. Finally, $Y \in \mathbb{R}^{|\mathcal{U}|\times|\mathcal{U}|}$ is updated using the last formula given in Eq. (5). This update is crucial for ensuring that the variables W and Z, updated in the earlier steps, converge towards equality.

The update steps of the ADMM solution for Eq. (14) are summarized in Alg. 2.

Complexities. BNSLIM exhibits a computational complexity of $O(k|\mathcal{U}|^3|\mathcal{I}|)$, necessitated by k iterations to update $|\mathcal{U}|^2$ similarities, each involving computations across all users and items, see Eq. (7). Conversely, the overall computational complexity for $\text{BNSLIM}_{\text{ADMM}}$ and FSLR is $O(k|\mathcal{U}|^3)$. Our improved performance is also attributed to being able to conduct efficient operations on square matrices, unlike the less flexible cyclic CD of BNSLIM that limits matrix operations. For item neighborhoods, complexities adjust by interchanging $|\mathcal{U}|$ and $|\mathcal{I}|$.

6 DATASETS AND METRICS

Datasets. For C-fairness, we cover two distinct recommendation applications: movies and music. For movies, we used the Movielens 1M (ML1M) dataset [19]. We transformed ratings (1 to 5) into binary format by interpreting ratings of 4 or higher as positive feedback [9, 41]. Users were grouped by *gender*, with females as the protected group and males as the non-protected group. For music, we focused on artist recommendations using the LastFM 1K (LFM1K) dataset [8]. For each user, we aggregated song plays per artist, and if the count exceeded the average play count of that user, it was marked as positive feedback for this artist by the user. Users were grouped by *age*, with those over 30 as the protected group and those 30 or younger as the non-protected group [31]. As LFM1K lacks music genre information, we used the genres extracted from MusicBrainz by [2]. In our analysis, we targeted genres where notable age imbalances among listeners were observed, specifically: *country*, *folk*, *hip hop*, *indie*, *jazz*, and *metal* (see Fig. 1).

For P-fairness, we also cover two distinct recommendation applications: courses and books. For courses, we used the COCO dataset [12], binarizing ratings (1 to 5) such that ratings of 5 are interpreted as positive feedback [2]. COCO features courses taught by teachers from various *geographic provinces*. We focused on single-teacher courses, grouping them as protected for teachers from developing countries and non-protected for those from developed countries, based on the OECD's classification [36]. *To the best of our knowledge, this is the first exploration of bias in course recommendations based on teachers' country development status.* For books, we used the *young adult* subset of the Goodreads dataset [44], a genre where users exhibit bias based on *author gender* [43]. We retained books by single authors and binarized ratings (1 to 5) with ratings of 4 or higher considered as positive feedback. Subsequently, we fetched author gender (female or

Table 2. Summary of Processed Data ($\mathcal{G}_{\text{p}}$: protected group, $\mathcal{G}_{\text{np}}$: non-protected group, *inter*$(\cdot)$: interaction counts).

Dataset	Fairness	Attribute	Sparsity	$\lvert\mathcal{U}\rvert$	$\lvert\mathcal{I}\rvert$	$\lvert\mathcal{G}_{\text{p}}\rvert$	$\lvert\mathcal{G}_{\text{np}}\rvert$	*inter*$(\mathcal{G}_{\text{p}})$	*inter*$(\mathcal{G}_{\text{np}})$
ML1M	Consumer	Gender	97.27	5,950	3,532	1,682	4,268	145,372	429,247
LFM1K	Consumer	Age	98.01	236	2,233	37	199	2,032	8,003
COCO	Provider	Province	99.81	3,803	7,972	2,146	5,826	13,349	45,159
Goodreads	Provider	Gender	99.88	186,919	26,576	6,574	20,002	1,170,604	4,802,037

male) using the Wikidata API. Books were grouped by author gender, with male-written books viewed as protected and female-written books as non-protected.

Table 2 provides data statistics. All datasets underwent 10-core user filtering, i.e., keeping users with at least 10 interactions, to secure robust training data, and they were split: an 80-20 training-testing split and an additional 80-20 split within the training set for validation. For C-fairness, to ensure consistent group sizes across all splits, focusing on the study of our mitigation strategies' impact, the datasets underwent *weak generalization*. This involves the same users appearing in all sets, a necessary condition for user-neighborhood models, which cannot accommodate unseen users. For P-fairness, we applied *strong generalization*, where users in each set are different, chosen to simulate real-world conditions, and because such a split would not notably alter the results, unlike in C-fairness.

Accuracy metrics. Accuracy is evaluated using *Recall* and *Normalized Discounted Cumulative Gain (NDCG)* [28, 31]. These metrics measure predictive performance, with higher scores indicating better accuracy. Recall measures the presence of held-out items in top-N lists, whereas NDCG additionally accounts for their ranking.

C-fairness metrics. Recall that C-fairness is defined as balancing the exposure of item groups across user groups. In our datasets, these item groups represent item categories: movie genres (ML1M dataset) and music genres (LFM1K dataset). To assess different aspects of this fairness objective, we utilize the following two metrics:

Consumer-side Equity (c-Equity) [7] measures the ratio of the observed probability of recommending an item category to protected users ($\mathcal{G}_{\text{p}}$) relative to that of non-protected users ($\mathcal{G}_{\text{np}}$). Due to its formulation, this metric is computed and reported separately for each item category. For convenience, we reformulated it as the absolute differences between these probabilities, averaged across all item categories, i.e.,

$$\textit{c-Equity@}N = \frac{1}{|C|}\sum_{c\in C}\left|\frac{\sum_{u\in\mathcal{G}_{\text{p}}}\sum_{i\in\mathcal{R}_u}\mathbf{1}_{i\in C_c}}{|\mathcal{G}_{\text{p}}|\cdot N} - \frac{\sum_{u\in\mathcal{G}_{\text{np}}}\sum_{i\in\mathcal{R}_u}\mathbf{1}_{i\in C_c}}{|\mathcal{G}_{\text{np}}|\cdot N}\right|, \tag{20}$$

where C is the set of item category identifiers, $\mathcal{R}_u$ is the top-N recommendation list for user u, and C_c includes all item identifiers within category c. The c-Equity metric ranges from 0 (balanced visibility) to 1 (imbalanced visibility), indicating the average difference in the visibility of item categories between two user groups. Attaining a c-Equity score of 1 indicates a rare scenario where $\mathcal{G}_{\text{p}}$ exclusively gets recommendations from one category and $\mathcal{G}_{\text{np}}$ from a different category with no overlap. A c-Equity score, multiplied by N, represents the average additional number of positions occupied by any item category in the top-N recommendation lists of a user group.

We also introduce the User-coverage Parity (u-Parity) metric, which measures the disparity between the proportions of protected and non-protected users receiving recommendations from a particular category, averaged across all item categories. u-Parity is formally defined as follows:

$$\textit{u-Parity@}N = \frac{1}{|C|}\sum_{c\in C}\left|\frac{\sum_{u\in\mathcal{G}_{\text{p}}}\mathbf{1}_{\exists i\in\mathcal{R}_u\,:\,i\in C_c}}{|\mathcal{G}_{\text{p}}|} - \frac{\sum_{u\in\mathcal{G}_{\text{np}}}\mathbf{1}_{\exists i\in\mathcal{R}_u\,:\,i\in C_c}}{|\mathcal{G}_{\text{np}}|}\right|. \tag{21}$$

Table 3. Results on the ML1M and LFM1K Datasets (top-10 lists). Best metric scores are highlighted in gold, and the slowest time performance in red. Rows corresponding to our algorithms are shaded in gray for easy identification.

Algorithm	ML1M					LFM1K				
	Training (secs)	Accuracy Metrics		Fairness Metrics		Training (secs)	Accuracy Metrics		Fairness Metrics	
		Recall↑	NDCG↑	c-Equity↓	u-Parity↓		Recall↑	NDCG↑	c-Equity↓	u-Parity↓
EASE	6.854±2.786	0.323±0.001	0.332	0.050±0.001	0.097±0.001	0.041±0.017	0.222±0.006	0.208±0.008	0.060±0.004	0.108±0.014
BNSLIM	11,160.704±3,597.703	0.239±0.007	0.246±0.006	0.038±0.001	0.066±0.002	478.099±25.274	0.142±0.027	0.127±0.023	0.031±0.012	0.060±0.013
$BNSLIM_{ADMM}$	120.580±4.490	0.226±0.013	0.240±0.016	0.015±0.004	0.027±0.005	0.085±0.026	0.116±0.004	0.098±0.007	0.010±0.004	0.026±0.008
FSLR	60.893±39.747	0.255±0.010	0.270±0.010	0.026±0.007	0.059±0.007	0.055±0.018	0.164±0.010	0.151±0.009	0.039±0.007	0.114±0.034
FairMF	5.878±3.042	0.161±0.015	0.164±0.019	0.013±0.002	0.028±0.003	0.201±0.169	0.103±0.049	0.092±0.045	0.016±0.004	0.039±0.007
FDA	3,156.129±807.204	0.250±0.023	0.251±0.025	0.057±0.003	0.118±0.007	45.212±10.284	0.159±0.011	0.145±0.016	0.060±0.007	0.116±0.011

u-Parity ranges from 0 to 1, with 0 indicating that equal percentages of users (from both groups) receive recommendations from each category and 1 indicating a rare scenario where each category covers users from only one group.

c-Equity helps identify if there is an imbalance in how item categories are recommended to distinct user groups, u-Parity helps identify if there is an imbalance in how different user groups are covered by recommendations across categories.

P-fairness metrics. Recall that P-fairness is defined as balancing the exposure of item groups within the top-N lists of all users. In our datasets, these item groups represent providers: teachers (COCO dataset) and authors (Goodreads dataset). To assess different aspects of this fairness objective, we utilize the following two metrics:

Bilateral Disparate Visibility (BDV) is a modification of the c-Equity metric for P-fairness. It measures the difference between the observed probability of recommending protected items ($\mathcal{G}_\text{p}$) and that of non-protected items ($\mathcal{G}_\text{np}$), i.e.,

$$BDV@N = \left| \frac{\sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{R}_u} \mathbf{1}_{i \in \mathcal{G}_\text{p}}}{|\mathcal{U}| \cdot N} - \frac{\sum_{u \in \mathcal{U}} \sum_{i \in \mathcal{R}_u} \mathbf{1}_{i \in \mathcal{G}_\text{np}}}{|\mathcal{U}| \cdot N} \right|. \tag{22}$$

The BDV metric ranges from 0 to 1, indicating the difference in visibility between protected and non-protected items in top-N lists. A score of 0 indicates balanced visibility, while a score of 1 indicates completely imbalanced visibility. A BDV score, multiplied by N, represents the additional number of positions occupied by one group over another.

Average Provider Coverage Rate (APCR) [29] measures the degree of provider coverage within the recommendation lists of all users in the test set. APCR is formally defined as follows:

$$APCR@N = \frac{\sum_{u \in \mathcal{U}_\text{test}} prov(\mathcal{R}_u)}{|\mathcal{P}||\mathcal{U}_\text{test}|}, \tag{23}$$

where $\mathcal{U}_\text{test}$ is the user set in the test dataset, $prov(\mathcal{R}_u)$ counts distinct providers in $\mathcal{R}_u$, and $\mathcal{P}$ is the set of providers. In our datasets, we have two providers: the protected and non-protected teachers or authors, i.e., $|\mathcal{P}| = 2$. The APCR metric ranges from 0 to 1. A score of 0 indicates a lack of coverage, with recommendations dominated by a single provider group, while a score of 1 indicates perfect coverage, with all providers appearing across all top-N lists.

u-Parity and APCR metrics emphasize coverage fairness in top-N lists: u-Parity evaluates balanced item category access for different consumer groups, while APCR evaluates balanced provider appearance for all consumers.

7 BENCHMARKING AND RESULTS

The evaluation of our algorithms regarding C- and P-fairness is guided by the following research questions:

RQ1. Can our algorithms achieve C- and P-fairness? What is the impact on accuracy?
RQ2. How do they compare to other fairness-aware algorithms in terms of fairness, accuracy, and time efficiency?
RQ3. How do the different characteristics and sensitive attributes of datasets affect our algorithms' behavior?

Table 4. Results on the COCO and Goodreads Datasets (top-10 lists). Best metric scores are highlighted in gold, and the slowest time performance in red. Rows corresponding to our algorithms are shaded in gray for easy identification.

Algorithm	COCO					Goodreads				
	Training (secs)	Accuracy Metrics		Fairness Metrics		Training (secs)	Accuracy Metrics		Fairness Metrics	
		Recall↑	NDCG↑	BDV↓	APCR↑		Recall↑	NDCG↑	BDV↓	APCR↑
EASE	7.863±0.945	0.275±0.005	0.214±0.005	0.586±0.013	0.870±0.010	117.301±15.450	0.554±0.001	0.529±0.001	0.527±0.001	0.898±0.001
BNSLIM	33,237.556±5,116.273	0.031±0.006	0.020±0.007	0.680±0.094	0.873±0.051	- did not complete within the 24-hour limit -				
BNSLIM$_{\text{ADMM}}$	160.103±69.683	0.153±0.047	0.129±0.038	0.207±0.265	0.831±0.038	3,435.894±21.644	0.498±0.009	0.470±0.008	0.482±0.010	0.916±0.004
FSLR	182.845±50.534	0.197±0.021	0.157±0.014	0.079±0.040	0.898±0.012	2,009.628±1,540.76	0.499±0.005	0.472±0.004	0.476±0.004	0.917±0.002
FairMF	24.040±16.832	0.002±0.001	0.001	0.437±0.021	0.983±0.005	1,424.151±812.837	0.000	0.000	0.449±0.035	0.980±0.006

Competitors. We consider two baselines. EASE, as a SOTA lacking fairness capabilities, benchmarks our fairness improvements. BNSLIM, the sole existing fairness-aware neighborhood-learning algorithm, emphasizes the demand for our more efficient alternatives. We utilize both the user- and item-neighborhood versions of EASE for C- and P-fairness experiments, respectively. Two advanced fairness-aware algorithms are also considered. *Fair Matrix Factorization (FairMF)* [46] is the classical MF model enhanced with the *non-parity regularizer* of [22], chosen for its interpretability when binary data are present and its frequent use in related work [25, 45]. *Fairness-aware Data Augmentation (FDA)* using *Bayesian Personalized Ranking (BPR)* as its backbone [9], chosen as the most recent data-driven fairness strategy, akin to the proposed algorithms. BNSLIM and FairMF are applicable for both C- and P-fairness. FDA, originally designed for C-fairness, presents challenges in applying for P-fairness; hence, we use it solely for C-fairness.

Implementations. Neighborhood-learning algorithms were developed using *Recpack* [32]. Our Python implementation for BNSLIM adheres to its original Java version [6]. For FDA, we used the implementation available by its authors[3]. For FairMF, we developed an implementation using *PyTorch* [37], due to the lack of existing code.

Fine-tuning. We optimized for $N = 10$ using *Hyperopt* [23], with parameter ranges derived from the corresponding papers. For BNSLIM$_{\text{ADMM}}$, the parameter space included λ_1 ranging from 10^{-3} to 50, λ_2 from 1 to 10^4, and λ_3 from 10^{-3} to 10^3. For FSLR, the parameter space was set for λ_1 from 10^{-3} to 10 and λ_2 from 1 to 10^4. The *Tree-structured Parzen Estimator (TPE)* algorithm guided this process, limited to either 50 trials or 24 hours. Convergence for BNSLIM was set at $\|\boldsymbol{W}^k - \boldsymbol{W}^{k+1}\|\infty < 10^{-4}$ or 50 iterations, while for BNSLIM$_{\text{ADMM}}$ and FSLR, it was set at $\|\boldsymbol{W}^{k+1} - \boldsymbol{Z}^{k+1}\|\infty < 10^{-4}$ or 50 iterations. To enhance the computational efficiency of BNSLIM, we used a cap of 100 neighbors.

Addressing the lack of standard fairness optimization and the frequent omission of such details in existing studies and code repositories, we explain our own approach. Given our interest in both accuracy and fairness, it was deemed reasonable to tune all fairness-aware algorithms by combining NDCG (accuracy) with c-Equity or BDV for C- or P-fairness, respectively. This combination was formulated as $\alpha \times (1 - accuracy@N) + (1 - \alpha) \times fairness@N$, where α is a weighting parameter that allows system designers to tailor the balance between accuracy and fairness according to marketplace needs. Prioritizing fairness in this work, we chose $\alpha = 0.2$. EASE was tuned solely on NDCG.

Experiments were conducted on a server featuring an Intel® Xeon® Gold 5318Y processor at 2.10GHz with 48 cores and 386GB of RAM, running Ubuntu 22. The source code for the experiments is available on GitHub[4].

7.1 C-fairness Results

Each experiment was repeated five times to ensure reliability. Table 3 details the mean performance and standard deviation for each metric. Note that the small fairness scores are the result of balancing the exposure of multiple item categories within the top-10 lists amidst the inherent diversity of user interests within each group.

[3]https://github.com/newlei/FDA
[4]https://github.com/Selefth/fair_neighborhood

RQ1. In both datasets, our $\text{BNSLIM}_{\text{ADMM}}$ considerably improves fairness compared to EASE over its predecessor, BNSLIM, both in terms of item group exposure (c-Equity) and user coverage (u-Parity), while preserving accuracy losses comparable to those observed between BNSLIM and EASE. In the LFM1K dataset, it reduced the c-Equity score from 6% to 1% compared to EASE, implying a nearly negligible disparity in the representation of each item category across top-N lists for different user groups. This favorably impacted the u-Parity metric, decreasing from 10.8% to 2.6%, which shows more even user coverage by each category. Similar reductions are observed in the ML1M dataset. However, in the LFM1K dataset, these improvements lead to a recall and NDCG reduction of about 48-53% from EASE.

In contrast to $\text{BNSLIM}_{\text{ADMM}}$, FSLR strikes a balance, improving fairness while achieving the least reduction in ranking accuracy from EASE in both datasets. FSLR's superior accuracy over $\text{BNSLIM}_{\text{ADMM}}$ is attributed to the latter's stringent approach to balancing neighborhoods: by increasing (decreasing) the similarity scores with irrelevant (relevant) neighbors for the sake of influence balance, i.e., it may increase (potentially artificially) fairness but adversely affect personalization, as shown by the significant decrease in recall and NDCG of $\text{BNSLIM}_{\text{ADMM}}$.

Answer to RQ1. *Consequently, our algorithms effectively enhance C-fairness in neighborhood learning, with FSLR compromising less accuracy than $\text{BNSLIM}_{\text{ADMM}}$, following a more reasonable approach to imposing fairness.*

RQ2. While EASE leads in accuracy, $\text{BNSLIM}_{\text{ADMM}}$ and FairMF are more fair, albeit at the cost of accuracy. $\text{BNSLIM}_{\text{ADMM}}$ achieves the best fairness scores compared to all fairness-aware competitors in both datasets, followed by FairMF, which slightly surpasses $\text{BNSLIM}_{\text{ADMM}}$ in ML1M, achieving the best c-Equity score. FairMF is also the fastest but ranks last in terms of personalization, making it less appealing for imposing fairness due to its high accuracy loss.

Apart from improved fairness, $\text{BNSLIM}_{\text{ADMM}}$ has another notable advantage over BNSLIM: it is orders of magnitude faster. For instance, for the ML1M dataset, it achieves a reduction of about 99% in training time, where $5,950^2$ similarities must be learned. BNSLIM, in contrast, only computes 100 neighbors per user to manage its computational load, highlighting $\text{BNSLIM}_{\text{ADMM}}$'s superior efficiency in processing the full set of similarities more rapidly. Among the two algorithms, $\text{BNSLIM}_{\text{ADMM}}$ is a more efficient choice for attaining fairness through balanced neighborhoods.

FDA is also time-intensive due to its data-augmentation approach. This model underperformed in our C-fairness context, as even if the estimated interactions are balanced, some item categories may still be underrepresented or overrepresented in the recommendations provided to different user groups. Finally, our FSLR is faster than $\text{BNSLIM}_{\text{ADMM}}$, and, while not as strong in fairness as $\text{BNSLIM}_{\text{ADMM}}$, leads in accuracy among fairness-aware methods.

Answer to RQ2. *Compared to other fairness-aware methods, our algorithms demonstrate optimal performance in balancing accuracy, fairness, and reasonable training time.*

RQ3. FSLR surpasses BNSLIM in fairness metrics in ML1M but not in LFM1K, while securing the best ranking accuracy among fairness-aware methods. This variation could be due to the protected group's size in LFM1K, affecting FSLR's ability to form strong neighbor relations with users from different groups. On the other hand, $\text{BNSLIM}_{\text{ADMM}}$ consistently outperforms BNSLIM in fairness in both datasets, as, in contrast to BNSLIM, it considers all possible neighbors to achieve balance. Lastly, FSLR consistently achieves accuracy closer to that of EASE.

Answer to RQ3. *Our comparison with EASE revealed that our algorithms are consistent in reducing unfairness in the ML1M and LFM1K datasets, focusing on distinct sensitive attributes: users' genders and users' ages, respectively.*

7.2 P-fairness Results

Each experiment was repeated five times to ensure reliability. Table 4 details the mean performance outcomes for each metric, along with their corresponding standard deviations. P-fairness scores are larger compared to their C-fairness counterparts due to our focus on only two item groups, making any disparities more noticeable.

RQ1. In the COCO dataset, EASE's 58.6% overrepresentation of courses from developed countries (BDV), translating to roughly six more slots for developed countries in users' top-10 lists, is notably reduced by our algorithms: BNSLIM$_{\text{ADMM}}$ reduces this to 20.7% and FSLR further decreases it to 7.9%. The high standard deviation in BNSLIM$_{\text{ADMM}}$'s BDV score suggests a sensitivity to strong generalization, unlike FSLR. This also contributes to its slightly poorer (compared to EASE) APCR. Despite the fairness gains, there are accuracy trade-offs: Compared to EASE, BNSLIM$_{\text{ADMM}}$ shows a drop in accuracy by nearly 40-44%, whereas FSLR's accuracy decreases by about 27-28%. In the Goodreads dataset, EASE exhibits a 52.7% overrepresentation of women-authored books. While our mitigation methods' impact is not as pronounced as in COCO, both of them managed to reduce this imbalance to 48%, with a 2% increase in APCR. Accuracy for both algorithms falls by 10-11% compared to EASE, a reasonable trade-off for our modest fairness improvements.

Answer to RQ1. *Our algorithms effectively enhance P-fairness, with FSLR compromising less accuracy than BNSLIM$_{ADMM}$ in the COCO dataset, following the trends in C-fairness. Strong generalization impacts BNSLIM$_{\text{ADMM}}$'s performance.*

RQ2. BNSLIM exhibits poor performance, ranking as the slowest among all evaluated algorithms and failing to complete within the 24-hour limit on Goodreads. BNSLIM$_{\text{ADMM}}$ outperforms BNSLIM with an impressive 99% reduction in training time for COCO and completes training on Goodreads, where BNSLIM could not. It also surpasses BNSLIM in both accuracy and fairness, establishing itself as a feasible solution for balancing neighborhoods. FSLR achieves the best BDV score in COCO, notably improving over EASE. In both datasets, it shows strong APCR performance and surpasses all other fairness-aware methods in accuracy. BNSLIM$_{\text{ADMM}}$ ranks near FSLR in fairness, but for the COCO dataset, it offers less personalized recommendations. FairMF leads in APCR for both datasets and BDV for Goodreads, yet it underperforms in personalization, attributed to datasets' sparsity (see Table 2). Experiments on denser COCO subsets showed accuracy and fairness gains for FairMF, yet our algorithms preserved a superior balance of both (results omitted due to space limitations). FairMF is slower in Goodreads than in COCO, mainly due to its dependence on the number of users, while the item-neighborhood versions of our algorithms scale solely with the number of items.

Answer to RQ2. *Compared to other fairness-aware methods, our algorithms optimally balance accuracy and fairness while scaling well to large datasets in P-fairness scenarios, with scalability driven solely by the number of items. Additionally, they demonstrate robust performance on sparse datasets, where other algorithms struggle.*

RQ3. BNSLIM$_{\text{ADMM}}$ and FLSR reduce unfairness, more notably in the COCO dataset, possibly due to Goodreads' extensive volume of interactions (nearly 6 million). Extending training beyond 50 iterations would enhance fairness further. BNSLIM$_{\text{ADMM}}$ is faster than FLSR in COCO, but it needs more time to converge in the larger Goodreads. Finally, in COCO, FSLR achieves better visibility (BDV score), whereas in Goodreads, both algorithms go hand-in-hand.

Answer to RQ3. *Our algorithms are consistent in reducing unfairness in the COCO and Goodreads datasets, focusing on distinct sensitive attributes: teachers' geographic provinces and authors' genders, respectively.*

8 CONCLUSIONS

We introduced two algorithms, BNSLIM$_{\text{ADMM}}$ and FSLR, to address demographic bias in neighborhood-learning models, targeting group fairness for consumers or providers. BNSLIM$_{\text{ADMM}}$ improves upon BNSLIM by employing the ADMM so as to balance the influences across demographic groups faster. FSLR induces controlled sparsity in neighborhoods, ensuring representation by demographically varied, yet correlated, counterparts so as to preserve personalization. We verified our algorithms empirically on several real-world datasets and showed that they outperformed existing solutions while achieving an optimal balance of accuracy, fairness, and time efficiency. An interesting research direction for future work involves developing algorithms for multi-group neighborhood balancing and exploring how these neighborhoods can preserve fairness while dynamically adapting to temporal changes.

REFERENCES

[1] Vito Walter Anelli, Alejandro Bellogín, Tommaso Di Noia, Dietmar Jannach, and Claudio Pomo. 2022. Top-n recommendation algorithms: A quest for the state-of-the-art. In *Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*. 121–131.

[2] Ludovico Boratto, Gianni Fenu, and Mirko Marras. 2021. Interplay between upsampling and regularization for provider fairness in recommender systems. *User Modeling and User-Adapted Interaction* 31, 3 (2021), 421–455.

[3] Ludovico Boratto, Gianni Fenu, Mirko Marras, and Giacomo Medda. 2023. Practical perspectives of consumer fairness in recommendation. *Information Processing & Management* 60, 2 (2023), 103208.

[4] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning* 3, 1 (2011), 1–122.

[5] Robin Burke. 2017. Multisided fairness for recommendation. *arXiv preprint arXiv:1707.00093* (2017).

[6] Robin Burke, Masoud Mansoury, and Nasim Sonboli. 2020. Experimentation with fairness-aware recommendation using librec-auto: hands-on tutorial. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*.

[7] Robin Burke, Nasim Sonboli, and Aldo Ordonez-Gauger. 2018. Balanced neighborhoods for multi-sided fairness in recommendation. In *Conference on fairness, accountability and transparency*. PMLR, 202–214.

[8] O. Celma. 2010. *Music Recommendation and Discovery in the Long Tail*. Springer.

[9] Lei Chen, Le Wu, Kun Zhang, Richang Hong, Defu Lian, Zhiqiang Zhang, Jun Zhou, and Meng Wang. 2023. Improving Recommendation Fairness via Data Augmentation. *arXiv preprint arXiv:2302.06333* (2023).

[10] Yao Cheng, Liang Yin, and Yong Yu. 2014. LorSLIM: low rank sparse linear methods for top-n recommendations. In *2014 IEEE International Conference on Data Mining*. IEEE, 90–99.

[11] Yashar Deldjoo, Dietmar Jannach, Alejandro Bellogin, Alessandro Difonzo, and Dario Zanzonelli. 2023. Fairness in recommender systems: research landscape and future directions. *User Modeling and User-Adapted Interaction* (2023), 1–50.

[12] Danilo Dessì, Gianni Fenu, Mirko Marras, and Diego Reforgiato Recupero. 2018. Coco: Semantic-enriched collection of online courses at scale with experimental use cases. In *Trends and Advances in Information Systems and Technologies: Volume 2 6*. Springer, 1386–1396.

[13] Michael D Ekstrand and Maria Soledad Pera. 2017. The demographics of cool. *Poster Proceedings at ACM RecSys. ACM, Como, Italy* (2017).

[14] Michael D Ekstrand, Mucun Tian, Ion Madrazo Azpiazu, Jennifer D Ekstrand, Oghenemaro Anuyah, David McNeill, and Maria Soledad Pera. 2018. All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In *Conference on fairness, accountability and transparency*. PMLR, 172–186.

[15] Maurizio Ferrari Dacrema, Simone Boglio, Paolo Cremonesi, and Dietmar Jannach. 2021. A troubling analysis of reproducibility and progress in recommender systems research. *ACM Transactions on Information Systems (TOIS)* 39, 2 (2021), 1–49.

[16] Elizabeth Gómez, Ludovico Boratto, and Maria Salamó. 2022. Provider fairness across continents in collaborative recommender systems. *Information Processing & Management* 59, 1 (2022), 102719.

[17] Felix Gräßer, Stefanie Beckert, Denise Küster, Susanne Abraham, Hagen Malberg, Jochen Schmitt, and Sebastian Zaunseder. 2017. Neighborhood-based Collaborative Filtering for Therapy Decision Support.. In *HealthRecSys@ RecSys*. 22–26.

[18] Mihajlo Grbovic, Vladan Radosavljevic, Nemanja Djuric, Narayan Bhamidipati, Jaikit Savla, Varun Bhagwan, and Doug Sharp. 2015. E-commerce in your inbox: Product recommendations at scale. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 1809–1818.

[19] F Maxwell Harper and Joseph A Konstan. 2015. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)* 5, 4 (2015), 1–19.

[20] Toshihiro Kamishima and Shotaro Akaho. 2017. Considerations on recommendation independence for a find-good-items task. (2017).

[21] Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Jun Sakuma. 2012. Enhancement of the Neutrality in Recommendation.. In *Decisions@ RecSys*. 8–14.

[22] Toshihiro Kamishima, Shotaro Akaho, Hideki Asoh, and Jun Sakuma. 2013. Efficiency Improvement of Neutrality-Enhanced Recommendation. In *Decisions@ RecSys*. Citeseer, 1–8.

[23] Brent Komer, James Bergstra, and Chris Eliasmith. 2019. Hyperopt-sklearn. *Automated Machine Learning: Methods, Systems, Challenges* (2019), 97–111.

[24] Yehuda Koren, Steffen Rendle, and Robert Bell. 2021. Advances in collaborative filtering. *Recommender systems handbook* (2021), 91–142.

[25] Pigi Kouki, Shobeir Fakhraei, James Foulds, Magdalini Eirinaki, and Lise Getoor. 2015. Hyper: A flexible and extensible probabilistic framework for hybrid recommender systems. In *Proceedings of the 9th ACM Conference on Recommender Systems*. 99–106.

[26] Yunqi Li, Hanxiong Chen, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. 2021. User-oriented fairness in recommendation. In *Proceedings of the Web Conference 2021*. 624–632.

[27] Yunqi Li, Hanxiong Chen, Shuyuan Xu, Yingqiang Ge, and Yongfeng Zhang. 2021. Towards personalized fairness based on causal notion. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1054–1063.

[28] Dawen Liang, Rahul G Krishnan, Matthew D Hoffman, and Tony Jebara. 2018. Variational autoencoders for collaborative filtering. In *Proceedings of the 2018 world wide web conference*. 689–698.

[29] Weiwen Liu and Robin Burke. 2018. Personalizing fairness-aware re-ranking. *arXiv preprint arXiv:1809.02921* (2018).

[30] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)* 54, 6 (2021), 1–35.

[31] Alessandro B Melchiorre, Navid Rekabsaz, Emilia Parada-Cabaleiro, Stefan Brandl, Oleg Lesota, and Markus Schedl. 2021. Investigating gender fairness of recommendation algorithms in the music domain. *Information Processing & Management* 58, 5 (2021), 102666.

[32] Lien Michiels, Robin Verachtert, and Bart Goethals. 2022. Recpack: An (other) experimentation toolkit for top-n recommendation using implicit feedback data. In *Proceedings of the 16th ACM Conference on Recommender Systems*. 648–651.

[33] 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*. 770–779.

[34] Athanasios N Nikolakopoulos, Xia Ning, Christian Desrosiers, and George Karypis. 2021. Trust your neighbors: a comprehensive survey of neighborhood-based methods for recommender systems. *Recommender Systems Handbook* (2021), 39–89.

[35] Xia Ning and George Karypis. 2011. Slim: Sparse linear methods for top-n recommender systems. In *2011 IEEE 11th international conference on data mining*. IEEE, 497–506.

[36] Organisation for Economic Co-operation and Development (OECD). 2021. Guidance: Countries defined as developing by the OECD. https://www.gov.uk/government/publications/countries-defined-as-developing-by-the-oecd/countries-defined-as-developing-by-the-oecd.

[37] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* 32 (2019).

[38] Evaggelia Pitoura, Kostas Stefanidis, and Georgia Koutrika. 2022. Fairness in rankings and recommendations: an overview. *The VLDB Journal* (2022), 1–28.

[39] Bashir Rastegarpanah, Krishna P Gummadi, and Mark Crovella. 2019. Fighting fire with fire: Using antidote data to improve polarization and fairness of recommender systems. In *Proceedings of the twelfth ACM international conference on web search and data mining*. 231–239.

[40] Ashudeep Singh and Thorsten Joachims. 2018. Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 2219–2228.

[41] Harald Steck. 2019. Embarrassingly shallow autoencoders for sparse data. In *The World Wide Web Conference*. 3251–3257.

[42] Harald Steck, Maria Dimakopoulou, Nickolai Riabov, and Tony Jebara. 2020. Admm slim: Sparse recommendations for many users. In *Proceedings of the 13th International Conference on Web Search and Data Mining*. 555–563.

[43] Mike Thelwall. 2019. Reader and author gender and genre in Goodreads. *Journal of Librarianship and Information Science* 51, 2 (2019), 403–430.

[44] Mengting Wan and Julian J. McAuley. 2018. Item recommendation on monotonic behavior chains. In *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018*, Sole Pera, Michael D. Ekstrand, Xavier Amatriain, and John O'Donovan (Eds.). ACM, 86–94. https://doi.org/10.1145/3240323.3240369

[45] Chuhan Wu, Fangzhao Wu, Xiting Wang, Yongfeng Huang, and Xing Xie. 2021. Fairness-aware news recommendation with decomposed adversarial learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 4462–4469.

[46] Sirui Yao and Bert Huang. 2017. Beyond parity: Fairness objectives for collaborative filtering. *Advances in neural information processing systems* 30 (2017).

[47] Ziwei Zhu, Xia Hu, and James Caverlee. 2018. Fairness-aware tensor-based recommendation. In *Proceedings of the 27th ACM international conference on information and knowledge management*. 1153–1162.

[48] Ziwei Zhu, Jianling Wang, and James Caverlee. 2020. Measuring and mitigating item under-recommendation bias in personalized ranking systems. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. 449–458.