Intra- and Inter-group Optimal Transport for User-Oriented Fairness in Recommender Systems

Zhongxuan Han[1], Chaochao Chen[1], Xiaolin Zheng[1]*, Meng Li[2], Weiming Liu[1], Binhui Yao[3], Yuyuan Li[1], Jianwei Yin[1]

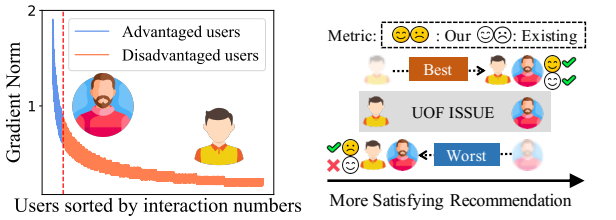
[1]College of Computer Science and Technology, Zhejiang University
[2]Harbin Institute of Technology (Shenzhen)
[3]Midea
{zxhan, zjuccc, xlzheng}@zju.edu.cn, mimas.li777@gmail.com, 21831010@zju.edu.cn, tony.yao@midea.com, 11821022@zju.edu.cn, zjuyjw@cs.zju.edu.cn

Abstract

Recommender systems are typically biased toward a small group of users, leading to severe unfairness in recommendation performance, i.e., User-Oriented Fairness (UOF) issue. Existing research on UOF exhibits notable limitations in two phases of recommendation models. **In the training phase**, current methods fail to tackle the root cause of the UOF issue, which lies in the unfair training process between advantaged and disadvantaged users. **In the evaluation phase**, the current UOF metric lacks the ability to comprehensively evaluate varying cases of unfairness. In this paper, we aim to address the aforementioned limitations and ensure recommendation models treat user groups of varying activity levels equally. *In the training phase*, we propose a novel **I**ntra- and **I**nter-**GrO**up **O**ptimal **T**ransport framework (II-GOOT) to alleviate the data sparsity problem for disadvantaged users and narrow the training gap between advantaged and disadvantaged users. *In the evaluation phase*, we introduce a novel metric called ξ-UOF, which enables the identification and assessment of various cases of UOF. This helps prevent recommendation models from leading to unfavorable fairness outcomes, where both advantaged and disadvantaged users experience subpar recommendation performance. We conduct extensive experiments on three real-world datasets based on four backbone recommendation models to prove the effectiveness of ξ-UOF and the efficiency of our proposed II-GOOT.

Introduction

Fairness is a critical research field in Machine Learning (ML) (Binns 2018; Dai et al. 2022; Mehrabi et al. 2021; Hutchinson and Mitchell 2019; Verma and Rubin 2018), and is also widely investigated in Recommender Systems (RSs) (Deldjoo et al. 2022; Chen et al. 2023; Han et al. 2023a). RS is a complex field involving frequent interactions between users and items (Su et al. 2023; Zheng et al. 2022b; Li et al. 2022, 2023). Fairness issues commonly arise from both the users' side (Li et al. 2021; Rahmani et al. 2022) and the items' side (Dash et al. 2021; Deldjoo et al. 2021a). In this paper, we focus on the fairness issue related to performance disparities among different user groups.



(a) Gradient Distribution (b) Different Levels of UOF

Figure 1: (a) visualizes the norm (i.e., $L_2 - norm$) of gradients coming from different users in a training epoch of LightGCN (He et al. 2020) in the Amazon Health dataset. (b) shows the best and worst cases of UOF that both share the same objective of facilitating equitable recommendation results for different user groups. Though the worst fairness will lead to dissatisfaction in both user groups, the existing metric treats these two cases as equally favorable.

RSs are always biased toward a small group of users, resulting in significant unfairness in the quality of recommendations (Li et al. 2021; Rahmani et al. 2022; Wen et al. 2022), i.e., the **U**ser-**O**riented **F**airness (UOF) issue. We define the users with more satisfied recommendation results as **advantaged users** and other users as **disadvantaged users**. Existing research has proved that advantaged users constitute only a small proportion of the total user base (Li et al. 2021) since many users suffer from the data sparsity (Han et al. 2023b; Zheng et al. 2022a) problem and fail to receive satisfying recommendation results. Therefore, addressing the UOF issue becomes crucial in RSs to enhance the overall quality of recommendation services.

To date, the relevant work of UOF is quite limited and exhibits notable limitations in both the training phase and evaluation phase. **In the training phase**, *the existing methods fail to tackle the root cause of the UOF issue.* The root of the UOF issue lies in the unfairness of the training process for recommendation models. Since recommendation models are always trained based on the interactions between users and items, we identify the advantaged users and disadvantaged users based on their interaction numbers (Li et al. 2021; Dai et al. 2022) and show the gradient distri-

*Xiaolin Zheng is the corresponding author.

bution in a training epoch of LightGCN in Figure 1(a). The majority of users (i.e., disadvantaged users) fail to provide sufficient training data for recommendation models due to the data sparsity problem. Consequently, recommendation models become dominated by advantaged users who contribute more to the model's updating process. Although existing research has proposed some re-ranking methods (Li et al. 2021; Dai et al. 2022) that adjust recommendation results after model training to achieve fairness, they cannot mitigate the unfair training process. Thus, the UOF issue cannot be solved well. Recently, researchers proposed a distributionally-robust optimization-based method (Wen et al. 2022) that aims to improve the worst-case user experience. Nevertheless, the limited availability of training samples from disadvantaged users restricts the performance of this method. **In the evaluation phase**, *The existing UOF metric fails to provide a comprehensive evaluation of recommendation models* As depicted in Figure 1(b), it is evident that the best case of UOF involves improving the quality of recommendation results for disadvantaged users to reach that of advantaged users, significantly surpassing the worst UOF. We argue that the UOF metric should be able to capture the differences among recommendation models with the worst and best fair. However, the existing metric (Li et al. 2021; Dai et al. 2022) solely compares whether different user groups receive nearly the same quality of recommendation services, treating both the best and worst fairness scenarios as equally favorable. This metric will encourage recommendation models to close the worst fair and lead to dissatisfaction in both user groups.

In this paper, we propose a comprehensive solution to the aforementioned limitations through the introduction of the **I**ntra- and **I**nter-**G**r**O**up **O**ptimal **T**ransport (II-GOOT) framework and a new metric ξ-UOF. In detail, **In the training phase**, we propose the II-GOOT framework to enhance the training process for disadvantaged users. Therefore, we can tackle the root cause of the UOF issue by reducing the training gap between advantaged users and disadvantaged users. The II-GOOT framework comprises two stages: the *intra-group* stage and the *inter-group* stage. (1) In the intra-group stage, our objective is to facilitate mutual assistance between pairs of similar disadvantaged users, thereby enhancing the training process. *Firstly*, we leverage the Optimal Transport (OT) (Liu et al. 2022b; Villani et al. 2009) mechanism to identify one-to-one similarities among disadvantaged users. *Secondly*, we facilitate the sharing of training samples between the two most similar disadvantaged users, thus alleviating the problem of data sparsity. Nevertheless, due to the significant data sparsity issue among disadvantaged users, the effectiveness of the intra-group stage might be constrained. Therefore, (2) in the inter-group stage, we further enable each disadvantaged user to learn from similar advantaged users who have been well-trained in the recommendation model. We propose the novel inter-group optimal clustering mechanism to explore the similarities between disadvantaged and advantaged users based on their shared interactions with items. Subsequently, we minimize the distance of embeddings between advantaged users and their similar disadvantaged users, aiming to narrow the training gap between these two user groups. **In the evaluation phase**, we define the *Best UOF* and the *Worst UOF* for a recommendation model, with different levels of recommendation accuracy. We emphasize that the optimal fair direction for a recommendation model is to achieve the Best UOF. However, attaining the ideal Best UOF in real-world scenarios is not feasible. Therefore, we introduce the ξ-UOF metric, which assesses the gap between the existing model and the model with Best UOF. ξ-UOF takes into account both the fairness between advantaged and disadvantaged users and the accuracy of the recommendation model. This metric aims to strike a balance between fairness and accuracy, providing a comprehensive evaluation of the UOF issue.

We have conducted extensive experiments based on four backbone models on three widely used real-world datasets. The experimental results demonstrate that II-GOOT outperforms State-Of-The-Art (SOTA) methods in addressing the UOF issue. Moreover, we substantiate that the proposed ξ-UOF has the ability to identify different cases of UOF, which are overlooked by the existing metric.

We summarize our contributions as follows: (1) We propose the II-GOOT framework to address the root cause of the UOF issue in the training phase. (2) We introduce the novel ξ-UOF metric, providing a comprehensive evaluation of the UOF issue in recommendation models in the evaluation phase. (3) We conduct extensive experiments to demonstrate the efficiency of II-GOOT and the effectiveness of ξ-UOF.

Related Work

Fair Recommendation

Fairness among different stakeholders in recommendation systems has attracted considerable attention in recent years. Considering the subject of fairness, fairness in recommender systems can be decoupled into user fairness, item fairness, and provider fairness (Deldjoo et al. 2023).

The ultimate goal of fair recommendation is to mitigate disparities among different subject groups. *For user fairness*, many works strive to provide similar users with similar recommendation results, e.g., ranking accuracy (Deldjoo, Bellogin, and Di Noia 2021), diversity, coverage (Melchiorre et al. 2021), under-ranking (Gorantla, Deshpande, and Louis 2021), and selection rate (Sühr, Hilgard, and Lakkaraju 2021). *For item fairness*, similar items should receive equal exposure regardless of sensitive attributes (Rastegarpanah, Gummadi, and Crovella 2019; Deldjoo et al. 2021b; Dash et al. 2021) or past exposure (Biega, Gummadi, and Weikum 2018), like the typical cold-start scene. *For provider fairness*, providers with more history interactions may be recommended more often than the rest (Ferraro 2019; Gharahighehi, Vens, and Pliakos 2021), leading to the superstar effect. Exposure disparity caused by the correspondence between providers and items (Sühr, Hilgard, and Lakkaraju 2021) and private characteristics (Shakespeare et al. 2020) should also be mitigated to create an equal market.

In this paper, we focus on the rarely explored fairness issue among users with different activity levels, i.e., the UOF

problem. Different from existing work (Li et al. 2021; Rahmani et al. 2022; Wen et al. 2022), we dive into the training process to mitigate the learning gap between advantaged and disadvantaged user groups and propose a novel metric.

Optimal Transport

Optimal transport has garnered significant attention due to its excellent ability to match between two distributions or spaces. Concerning OT as a field of mathematics, a broad range of literature is available (Villani et al. 2009; Santambrogio 2015; Figalli and Glaudo 2021). Notably, (Santambrogio 2015) unified the two classical formulations of OT: Monge formulation and Kantorovich formulation. Recent advances in accelerating OT computation have unveiled its potential in Machine Learning. Computation of Wasserstein distances and Wasserstein Barycenters was greatly sped up by (Cuturi 2013; Cuturi and Doucet 2014).

Many attempts have been made to utilize OT to improve some downstream tasks in natural language processing (Asano, Rupprecht, and Vedaldi 2019; Chen et al. 2019), transfer learning (Flamary et al. 2016; Courty et al. 2017; Damodaran et al. 2018; Xu et al. 2020), adversarial learning (Arjovsky, Chintala, and Bottou 2017), neural architecture search (Yang, Liu, and Xu 2023), and recommendation systems (Liu et al. 2021, 2022a; Liu, Fang, and Wu 2023).

As the user embeddings in collaborative filtering models can be seen as a kind of latent space, in this paper, we apply OT to match between users in the latent space. The matching results are then utilized to enhance the training process.

Methodology

In this section, we introduce the proposed II-GOOT framework to solve the UOF issue **in the training phase** of recommendation models.

Problem Formulation

We use $\mathcal{U}$ and $\mathcal{I}$ to represent the user set and the item set. We divide users into disadvantaged user group $\mathcal{D}$ and advantaged user group $\mathcal{A}$ based on their interaction numbers according to (Li et al. 2021; Rahmani et al. 2022). Users with more interactions are more likely to be advantaged. We denote the initial average recommendation performance (e.g., HitRatio, NDCG) of these two groups of users as $P_{\mathcal{D}}$ and $P_{\mathcal{A}}$ with $P_{\mathcal{A}} > P_{\mathcal{D}}$ in most cases. In this paper, we aim to narrow the gap in the recommendation performance between $\mathcal{D}$ and $\mathcal{A}$ to achieve UOF and maintain the overall recommendation performance simultaneously.

Overview

In this section, we proposed a novel **I**ntra- and **I**nter-**GrO**up **O**ptimal **T**ransport framework, namely II-GOOT to solve the UOF issue in the training phase. II-GOOT is a general framework that can be integrated with any recommendation models (i.e., backbone models) to achieve UOF. The overall architecture of the framework is depicted in Figure 2, and it is divided into two key stages: *the intra-group stage* and *the inter-group stage*. (1) The intra-group stage aims to address the data sparsity problem encountered by disadvantaged users, thereby enhancing the modeling process for this group. To achieve this, we divide the disadvantaged users into two distinct groups and introduce the intra-group Optimal Transport (OT) to explore one-to-one similarities between these groups. Consequently, each pair of disadvantaged users can share their training samples, effectively mitigating the data sparsity issue. (2) In the inter-group stage, we introduce the novel inter-group optimal clustering mechanism to explore the similarities between advantaged and disadvantaged users. This step enables disadvantaged users to learn from their similar advantaged counterparts, further enhancing the training process for the disadvantaged group. By employing these two stages, we successfully reduce the training gap between advantaged and disadvantaged users, thereby mitigating the root cause of the UOF issue.

Intra-Group Stage

In this stage, we aim to alleviate the data sparsity problem of disadvantaged users. As depicted in Figure 2, firstly, we utilize the intra-group optimal transport mechanism to explore one-to-one similarities among disadvantaged users. Then, we enable disadvantaged users to share their training samples with their most similar users.

Intra-Group Optimal Transport. To ensure users with limited training samples can benefit from those with more extensive training data, we sort the disadvantaged users based on their interactions with items and divide them into two subgroups $\mathcal{G}_1$ and $\mathcal{G}_2$. Each subgroup comprises half of the disadvantaged users (i.e., $|\mathcal{G}_1| = |\mathcal{G}_2|$) with users in $\mathcal{G}_1$ having fewer interactions with items compared to those in $\mathcal{G}_2$. The primary goal of the intra-group optimal transport is to ascertain users' similarities based on their interactions with items. We achieve this objective in several steps:

Firstly, we construct one-hot interaction embeddings $H \in \{0,1\}^{|\mathcal{U}|\times|\mathcal{I}|}$ of users, where $H_{ij} = 1$ indicates that $\mathcal{U}_i$ has interacted with $\mathcal{I}_j$, and $H_{ij} = 0$ otherwise. We denote one-hot embeddings of $\mathcal{G}_1$ and $\mathcal{G}_2$ as $H^{\mathcal{G}_1}$ and $H^{\mathcal{G}_2}$, respectively.

Secondly, we extract the optimal transport matrix by solving the Monge-Kantorovich Problem (Bogachev and Kolesnikov 2012) to explore similar user pairs between $\mathcal{G}_1$ and $\mathcal{G}_2$. Considering $h^{\mathcal{G}_1}$ and $h^{\mathcal{G}_2}$ are two variables respectively sampled from $H^{\mathcal{G}_1}$ and $H^{\mathcal{G}_2}$. Then, the Monge-Kantorovich Problem is defined as follows:

Problem 1 (Monge-Kantorovich Problem) *Given the transport cost matrix $C \in \mathbb{R}_+^{|\mathcal{G}_1|\times|\mathcal{G}_2|}$, the objective of the Monge-Kantorovich Problem is to find the joint probability $W \in \mathbb{R}_+^{|\mathcal{G}_1|\times|\mathcal{G}_2|}$ that minimizes the total transport cost:*

$$d_C(H^{\mathcal{G}_1}, H^{\mathcal{G}_2}) = \min_W \int_{H^{\mathcal{G}_1}\times H^{\mathcal{G}_2}} C(h^{\mathcal{G}_1}, h^{\mathcal{G}_2}) dW(h^{\mathcal{G}_1}, h^{\mathcal{G}_2}). \tag{1}$$

Here, W_{ij} indicates the possibility of transporting $\mathcal{U}_i$ in $\mathcal{G}_1$ to $\mathcal{U}_j$ in $\mathcal{G}_2$, which reflects the similarity between $\mathcal{U}_i$ and $\mathcal{U}_j$. To identify the most similar user in another group for a

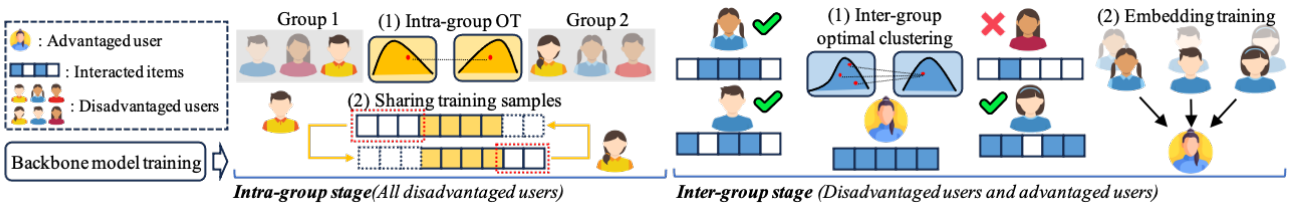


Figure 2: The overall framework of II-GOOT. In the intra-group stage, we enable each disadvantaged user to share training samples with his/her most similar disadvantaged user to mitigate the data sparsity problem. In the inter-group stage, we let disadvantaged users learn from advantaged users to further narrow the training gap between them.

given user, we introduce a constraint on W:

$$\sum_{i=1}^{|\mathcal{G}_1|}\sum_{j=1}^{|\mathcal{G}_2|} W_{ij} = 1,$$
$$W_{ij} \in \left(0, \frac{1}{|\mathcal{G}_1|}\right), \sum_{i=1}^{|\mathcal{G}_1|} W_{ij} = \frac{1}{|\mathcal{G}_1|}, \sum_{j=1}^{|\mathcal{G}_1|} W_{ij} = \frac{1}{|\mathcal{G}_1|}, \tag{2}$$

where $W_{ij} = \frac{1}{|\mathcal{G}_1|}$ indicates that $\mathcal{U}_j$ is the most similar user in $\mathcal{G}_2$ for $\mathcal{U}_i$, and vice versa.

However, solving the Monge-Kantorovich Problem can be time-consuming, with a worst-case time complexity of $O(|\mathcal{G}_1|^3)$. To overcome this, we introduce the sinkhorn divergence (Cuturi 2013) to smooth the objective with an entropic regularization:

$$d_C^{\epsilon}(H^{\mathcal{G}_1}, H^{\mathcal{G}_2}) = \min_W \int_{H^{\mathcal{G}_1}\times H^{\mathcal{G}_2}} C(h^{\mathcal{G}_1}, h^{\mathcal{G}_2}) dW(h^{\mathcal{G}_1}, h^{\mathcal{G}_2}) + \epsilon \cdot \sum_{i=1}^{|\mathcal{G}_1|}\sum_{j=1}^{|\mathcal{G}_2|} W_{ij}(\log(W_{ij}) - 1). \tag{3}$$

The derived new objective can be efficiently solved through Sinkhorn's matrix scaling algorithm with a complexity of $O(|\mathcal{G}_1| \cdot |\mathcal{G}_2|)$ (Cuturi 2013). We introduce the detailed optimization process for Equation (3) in Appendix A.

Thirdly, we construct the cost matrix C based on cosine similarities between $H^{\mathcal{G}_1}$ and $H^{\mathcal{G}_2}$:

$$C_{ij} = \frac{H_i^{\mathcal{G}_1} \cdot H_j^{\mathcal{G}_2}}{|H_i^{\mathcal{G}_1}| \times |H_j^{\mathcal{G}_2}|}. \tag{4}$$

Therefore, C can reflect the initial similarities between each user and give additional constraints to the probability measure. By calculating W in Equation(3) based on C, we can explore the one-to-one similar user pairs between $\mathcal{G}_1$ and $\mathcal{G}_2$.

Sharing Training Samples. During the training process of the backbone recommendation model, we enable similar users in $\mathcal{G}_1$ and $\mathcal{G}_2$ to share their training samples to mitigate the data sparsity problem of disadvantaged users based on the result of W. For example, if $W_{ij} = \frac{1}{|\mathcal{G}_1|}$, then $\mathcal{U}_i$ in $\mathcal{G}_1$ and $\mathcal{U}_j$ in $\mathcal{G}_2$ will train together.

Inter-Group Stage

In this stage, we enable disadvantaged users to learn from their similar advantaged users, thereby reducing the training gap between these two user groups. While the intra-group stage helps mitigate data sparsity among disadvantaged users, it alone may be insufficient to address the UOF issue due to the limited training samples available for disadvantaged users. Therefore, as shown in Figure 2, firstly, we propose a novel inter-group optimal clustering mechanism to explore similar disadvantaged users for each advantaged user. Then, disadvantaged users will learn from the corresponding similar advantaged user to receive better recommendation results.

Inter-Group Optimal Clustering. To explore n-to-one similarities among disadvantaged users and disadvantaged users, we propose the novel inter-group optimal clustering mechanism. In this approach, each advantaged user acts as a cluster center, while disadvantaged users are considered nodes to be clustered around these centers. To achieve this goal, we need to solve the following Monge-Kantorovich problem smoothed with the sinkhorn divergence:

$$d_C^{\epsilon}(H^{\mathcal{D}}, H^{\mathcal{A}}) = \min_X \int_{H^{\mathcal{D}}\times H^{\mathcal{A}}} M(h^{\mathcal{D}}, h^{\mathcal{A}}) dX(h^{\mathcal{D}}, h^{\mathcal{A}}) + \epsilon \sum_{i=1}^{|\mathcal{D}|}\sum_{j=1}^{|\mathcal{A}|} X_{ij}(\log(X_{ij}) - 1), \tag{5}$$

where $M \in \mathbb{R}_+^{|D|\times|A|}$ is the cost matrix, similar to C. $h^{\mathcal{D}}$ and $h^{\mathcal{A}}$ are two embeddings sampled from $H^{\mathcal{D}}$ and $H^{\mathcal{A}}$, respectively. The above problem aims to find similarities between disadvantaged and advantaged users. To achieve our objective of clustering each disadvantaged user into a specific advantaged user, we design a restriction of X as follows:

$$\sum_{i=1}^{|\mathcal{D}|}\sum_{j=1}^{|\mathcal{A}|} X_{ij} = 1,$$
$$X_{ij} \in \{0, \frac{1}{|\mathcal{D}|}\}, \sum_{i=1}^{|\mathcal{D}|} X_{ij} = \frac{1}{|\mathcal{A}|}, \sum_{j=1}^{|\mathcal{A}|} X_{ij} = \frac{1}{|\mathcal{D}|}. \tag{6}$$

By restricting $\sum_{j=1}^{|\mathcal{A}|} X_{ij} = \frac{1}{|\mathcal{D}|}$, we ensure balanced clusters and avoid too many disadvantaged users being clustered together, which will cause over-smoothing of features. After

solving the above optimization problem, we can get X which represents the clustering result. For example, $X_{ij} = \frac{1}{|\mathcal{D}|}$ indicates that $\mathcal{D}_i$ belongs to the cluster centered with $\mathcal{A}_j$.

Embedding Training. During the training of the backbone recommendation model, we enable disadvantaged users to learn from advantaged users by enhancing the cohesion of each cluster. We calculate the inter-group loss as follows:

$$L_{inter} = \sum_{i}^{|\mathcal{D}|} ||E_{\mathcal{D}_i} - E_{\mathcal{T}_i}||^2, \tag{7}$$

where $\mathcal{T}_i$ represents the clustering center of $\mathcal{D}_i$, E represents the user embedding. By minimizing L_{inter}, disadvantaged users can learn from their similar advantaged users, and the distributions of disadvantaged users and advantaged users will be closer, which will improve the recommendation performance of disadvantaged users.

Theorem 1 *Let $\mathcal{H}$ be a hypothesis space of recommendation models with $h \in \mathcal{H}$. Let $R_{\mathcal{D}}(h)$ and $R_{\mathcal{A}}(h)$ be the expected error in user group $\mathcal{D}$ and $\mathcal{A}$, $\hat{R}_{\mathcal{A}}(h)$ be the empirical estimate of $R_{\mathcal{A}}(h)$:*

$$R_{\mathcal{D}}(h) \leqslant \hat{R}_{\mathcal{A}}(h) + \hat{d}_{\mathcal{H}}(\mathcal{D}, \mathcal{A}) + \gamma, \tag{8}$$

where γ is a constant.

Theorem 1 tells us that, to obtain a recommendation model with a small $R_{\mathcal{D}}(h)$, it is necessary to minimize the $\mathcal{H}$-divergence $\hat{d}_{\mathcal{H}}(\mathcal{D}, \mathcal{A})$ together with $\hat{R}_{\mathcal{A}}(h)$. As pointed-out by (Ben-David et al. 2006), a strategy to control $\mathcal{H}$-divergence is to find two user groups that are as indistinguishable as possible. Therefore, by aligning distributions of disadvantaged and advantaged users to be similar, we can improve the recommendation performance of disadvantaged users and narrow the performance gap. The proof of Theorem 1 can be found in Appendix B.

We combine L_{inter} together with the recommendation loss $L_{utility}$ of the backbone model to achieve fairness and maintain the overall recommendation performance simultaneously:

$$L = L_{utility} + L_{inter}. \tag{9}$$

Through the intra-group stage and the inter-group stage, we enhance the training process for disadvantaged users and effectively address the root cause of the UOF issue. It's important to highlight that both the intra-group optimal transport and the inter-group optimal clustering processes can be executed ahead of the actual model training. Therefore, the II-GOOT framework is time-efficient.

A Novel Metric: ξ-UOF

In this section, we solve the limitations of UOF research **in the evaluation phase** by introducing a novel metric.

Existing UOF

User-Oriented Fairness (UOF) is a kind of group fairness (Dwork et al. 2012; Hardt, Price, and Srebro 2016), that strives to establish equitable treatment for both advantaged and disadvantaged users within a recommendation model. Given $\mathcal{M}$ that indicates a metric (e.g., NDCG and HitRatio) that can evaluate the recommendation performance, UOF is defined as follows (Li et al. 2021; Rahmani et al. 2022):

Definition 1 (User-Oriented Fairness (UOF))

$$\mathbb{E}[\mathcal{M}(\mathcal{A})] = \mathbb{E}[\mathcal{M}(\mathcal{D})]. \tag{10}$$

UOF aims to offer users with different activity levels the same recommendation performance, which is usually impossible in real-world RS. Therefore, researchers (Li et al. 2021; Rahmani et al. 2022) always calculate the difference in average recommendation performance for different user groups to evaluate the fairness of a model:

Definition 2 (The UOF metric)

$$\mathcal{M}_{UOF}(\mathcal{A}, \mathcal{D}) = \left| \frac{1}{|\mathcal{A}|} \sum_{i=1}^{|\mathcal{A}|} \mathcal{M}(A_i) - \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} \mathcal{M}(D_i) \right|. \tag{11}$$

However, the above $\mathcal{M}_{UOF}$ only evaluates the recommendation performance gap between advantaged and disadvantaged users, overlooking whether recommendation results are satisfying. For instance, if both groups receive equally poor recommendations, the model would still be considered favorable according to $\mathcal{M}_{UOF}$. Such a metric may encourage recommendation models to achieve fairness at a low accuracy level.

Our Proposed ξ-UOF

To address the limitations and provide a comprehensive evaluation, we propose our metric, referred to as ξ-UOF, which takes both fairness and accuracy into account. To begin with, we define the Best UOF and the Worst UOF of a recommendation model as Figure 1(b) shows:

Definition 3 (The Best UOF)

$$\mathbb{E}[\mathcal{M}(\mathcal{A})] = \mathbb{E}[\mathcal{M}(\mathcal{D})] = P_{\mathcal{A}}. \tag{12}$$

Definition 4 (The Worst UOF)

$$\mathbb{E}[\mathcal{M}(\mathcal{A})] = \mathbb{E}[\mathcal{M}(\mathcal{D})] = P_{\mathcal{D}}. \tag{13}$$

Clearly, if a recommendation model achieves the Best UOF, it can provide a fair recommendation result with high accuracy. Genuine fairness entails enhancing the recommendation outcomes for disadvantaged users in order to narrow the recommendation gap between them and advantaged users, i.e., the Best UOF. It does not involve reducing the recommendation quality of advantaged users to match the lower level experienced by disadvantaged users, i.e., the Worst UOF. Nevertheless, attaining the ideal Best UOF is impractical within real-world recommender systems due to limitations in training samples. Therefore, we define ξ-UOF, which evaluates the gap between a recommendation model and the model with Best UOF.

Definition 5 (ξ-UOF)

$$\begin{aligned} \xi \geqslant \mathcal{M}_{UOF}(\mathcal{A}, \mathcal{D}) = & \frac{|\mathcal{A}|}{|\mathcal{U}|} max\left\{0, P_{\mathcal{A}} - \frac{1}{|\mathcal{A}|} \sum_{i=1}^{|\mathcal{A}|} \mathcal{M}(A_i)\right\} \\ & + \frac{|\mathcal{D}|}{|\mathcal{U}|} max\left\{0, P_{\mathcal{A}} - \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} \mathcal{M}(D_i)\right\}, \end{aligned} \tag{14}$$

where the majority user group (i.e., disadvantaged users) contributes more to the value of ξ-UOF. A smaller value of ξ indicates a fairer model, and when $\xi = 0$, it signifies that a model has achieved the Best UOF and even surpassed it. Through ξ-UOF, we encourage recommendation models to achieve the Best UOF and provide satisfying recommendation results for both advantaged and disadvantaged users. We conduct experiments in section to further analyze the limitations of the existing metric and advantages of ξ-UOF.

Experiments and Analysis

We conduct extensive experiments to answer the following questions: **Q1**: Does II-GOOT outperform the existing methods in effectively addressing the UOF issue and enhancing recommendation performance? **Q2**: Can ξ-UOF provide a robust evaluation of the UOF issue? **Q3**: What is the respective impact of the intra-group and inter-group stages on the performance of II-GOOT? **Q4**: How robust is the generalizability of the II-GOOT framework when subjected to variations in the categorization of advantaged and disadvantaged users? **Q5**: Can II-GOOT narrow the training gap between advantaged users and disadvantaged users?

Datasets and Experimental Settings

Dataset. We conduct our experiments on three public Amazon datasets *Beauty*, *Grocery & Gourmet Food (Grocery)*, and *Health & Personal Care (Health)*, which are widely used to evaluate the UOF issue (Li et al. 2021). We give a detailed description of the datasets in Appendix C.1.

Baselines and Backbone Models. We compare II-GOOT with the SOTA methods UFR (Li et al. 2021) and S-DRO (Wen et al. 2022). Besides, we choose four backbone models, including MF (Koren, Bell, and Volinsky 2009), NeuMF (He et al. 2017), VAECF (Liang et al. 2018), and LightGCN (He et al. 2020) to evaluate the performance. We give a detailed introduction of baseline models and backbone models in Appendix C.2.

Evaluation Protocols and Parameter Settings. We extract the top 5% users as advantaged users according to their interaction numbers, leaving others as disadvantaged users. Besides, we adopt Normalized Discounted Cumulative Gain (NDCG) (Wang et al. 2013) and Hit Ratio (HR) (Waters 1976) to evaluate the recommendation performance (Li et al. 2021; Dai et al. 2022). Then, we utilize our proposed ξ-UOF to evaluate the UOF level of a recommendation model, with a lower value of ξ-UOF means a fairer performance. We give detailed evaluation protocols and parameter settings in Appendix C.3.

Overall Comparison (Q1, Q2)

We conduct extensive experiments on three public datasets. The results are reported in Table 1.

To Answer Q1. The experimental results demonstrate that II-GOOT outperforms all baselines with fairer recommendation results and higher overall performance. *Compared with original backbone models*, II-GOOT particularly enhances the training process for disadvantaged users. Since model training involves both advantaged and disadvantaged users, the accuracy of recommendations for advantaged users also experiences an improvement. With both user groups being more satisfied with recommendation results, II-GOOT effectively fosters fairness in recommendation models, moving closer to the Best Fair and enhancing the overall recommendation performance. *Compared with UFR*, II-GOOT has the ability to solve the root cause of UOF, i.e., the training bias between advantaged and disadvantaged users. UFR aims to narrow the recommendation gap between these two groups of users by re-ranking recommendation results. Its effectiveness is hindered by inadequate training of disadvantaged users. *Compared with S-DRO*, II-GOOT solves the data sparsity problem of disadvantaged users by expanding the pool of training samples. While S-DRO focuses solely on minimizing the loss function for disadvantaged users alongside advantaged users, its potential is curtailed by the limitation of training data for the former group.

To Answer Q2. The experimental results prove that ξ-UOF has the ability to identify different levels of fairness and comprehensively evaluate the UOF issue. As shown in Table 1, recommendation models closer to the Best UOF have a lower value of ξ-UOF. This trend is reasonable since by optimizing recommendation models to the Best UOF, disadvantaged users can receive more satisfying recommendation results and the recommendation gap between advantaged and disadvantaged users can be narrowed. Among baseline models, UFR aims to simply ensure that both advantaged and disadvantaged users experience similar recommendation performance, as shown in Equation (11). However, since the re-ranking method, UFR, cannot mitigate the training bias during model training, it reduces the recommendation quality of advantaged users to match the low level experienced by disadvantaged users. For instance, in the Beauty dataset, UFR reduces the NDCG value for advantaged users from 0.3046 to 0.1956 in the NeuMF model, while the value for disadvantaged users shows a marginal increase from 0.1861 to 0.1863. Such performance will lead to dissatisfaction in both user groups and should not be encouraged. ξ-UOF recognizes it to be unfavorable with the high metric value of 0.1178.

Ablation Study (Q3)

In this section, we choose LightGCN as the backbone model to demonstrate the effectiveness of the intra-group stage and the inter-group stage. The experimental results are reported in Table 2, with Intra and Inter indicating the model with only the intra-group stage and the inter-group stage, respectively. Both the intra-group stage and the inter-group stage yield a fairer model, accompanied by a better overall recommendation performance. These outcomes underscore the significance of enhancing the training process for disadvantaged users. Compared with the inter-group stage, the intra-group stage offers a more substantial improvement. The reason is that the key issue of insufficient training for disadvantaged users is the data sparsity problem. The intra-group stage expands the pool of training samples, giving a more efficient solution. II-GOOT has the best performance, serving as evidence that both stages play integral roles in achieving

			Beauty				Grocery				Health			
			Ove.	Adv.	Dis.	ξ.	Ove.	Adv.	Dis.	ξ.	Ove.	Adv.	Dis.	ξ.
MF	NDCG	Original	0.152	0.266	0.146	0.114	0.176	0.320	0.168	0.144	0.171	0.341	0.162	0.170
		UFR	0.158	0.154	0.158	0.109	0.171	0.191	0.170	0.149	0.162	0.163	0.162	0.179
		S-DRO	0.159	0.266	0.153	0.108	0.180	0.316	0.172	0.141	0.170	0.342	0.161	0.171
		II-GOOT	**0.194***	**0.271***	**0.189***	**0.073***	**0.228***	**0.350***	**0.222***	**0.093***	**0.199***	**0.347***	**0.191***	**0.143***
	HR	Original	0.256	0.439	0.247	0.182	0.316	0.478	0.308	0.162	0.293	0.463	0.284	0.170
		UFR	0.251	0.277	0.250	0.187	0.305	0.313	0.304	0.174	0.288	0.310	0.287	0.175
		S-DRO	0.260	0.426	0.251	0.179	0.320	0.479	0.312	0.158	0.289	0.447	0.281	0.174
		II-GOOT	**0.286***	**0.445***	**0.278***	**0.153***	**0.364***	**0.483***	**0.358***	**0.115***	**0.322***	**0.469***	**0.315***	**0.141***
NeuMF	NDCG	Original	0.192	0.305	0.186	0.113	0.200	0.344	0.193	0.144	0.196	0.359	0.188	0.162
		UFR	0.187	0.196	0.186	0.118	0.200	0.292	0.196	0.144	0.192	0.231	0.190	0.167
		S-DRO	0.196	0.311	0.190	0.109	0.200	0.331	0.193	0.144	0.207	0.352	0.199	0.152
		II-GOOT	**0.219***	**0.323***	**0.213***	**0.087***	**0.219***	**0.351***	**0.212***	**0.125***	**0.234***	**0.391***	**0.225***	**0.127***
	HR	Original	0.282	0.481	0.272	0.199	0.328	0.504	0.319	0.176	0.298	0.520	0.287	0.221
		UFR	0.262	0.259	0.262	0.219	0.331	0.357	0.329	0.173	0.294	0.301	0.293	0.226
		S-DRO	0.289	0.467	0.280	0.191	0.336	0.487	0.328	0.168	0.298	0.510	0.286	0.222
		II-GOOT	**0.321***	**0.509***	**0.311***	**0.162***	**0.356***	**0.510***	**0.348***	**0.148***	**0.349***	**0.530***	**0.340***	**0.171***
VAECF	NDCG	Original	0.208	0.343	0.201	0.135	0.206	0.351	0.199	0.145	0.242	0.410	0.233	0.168
		UFR	0.205	0.212	0.205	0.139	0.208	0.216	0.208	0.143	0.233	0.260	0.231	0.177
		S-DRO	0.216	0.332	0.210	0.127	0.208	0.352	0.200	0.144	0.240	0.401	0.231	0.171
		II-GOOT	**0.228***	**0.357***	**0.221***	**0.116***	**0.235***	**0.364***	**0.228***	**0.117***	**0.261***	**0.427***	**0.252***	**0.150***
	HR	Original	0.324	0.528	0.313	0.204	0.350	0.529	0.340	0.179	0.323	0.555	0.311	0.232
		UFR	0.320	0.322	0.320	0.208	0.345	0.368	0.343	0.185	0.314	0.356	0.312	0.241
		S-DRO	0.334	0.529	0.323	0.194	0.357	0.527	0.348	0.173	0.328	0.560	0.316	0.227
		II-GOOT	**0.353***	**0.537***	**0.343***	**0.175***	**0.361***	**0.541***	**0.352***	**0.169***	**0.349***	**0.562***	**0.338***	**0.206***
LightGCN	NDCG	Original	0.245	0.435	0.235	0.190	0.257	0.401	0.249	0.144	0.263	0.501	0.250	0.238
		UFR	0.243	0.309	0.239	0.193	0.251	0.307	0.248	0.150	0.264	0.355	0.259	0.237
		S-DRO	0.248	0.421	0.239	0.187	0.260	0.405	0.252	0.141	0.272	0.492	0.260	0.229
		II-GOOT	**0.286***	**0.443***	**0.278***	**0.150***	**0.289***	**0.407***	**0.283***	**0.112***	**0.309***	**0.502***	**0.299***	**0.192***
	HR	Original	0.375	0.625	0.362	0.250	0.393	0.622	0.380	0.230	0.376	0.633	0.362	0.257
		UFR	0.368	0.404	0.366	0.257	0.396	0.441	0.394	0.226	0.383	0.419	0.381	0.250
		S-DRO	0.376	0.602	0.365	0.249	0.414	0.620	0.403	0.208	0.387	0.639	0.374	0.246
		II-GOOT	**0.433***	**0.641***	**0.422***	**0.193***	**0.443***	**0.628***	**0.434***	**0.180***	**0.448***	**0.637***	**0.438***	**0.185***

Table 1: Experimental result. Ove. indicates the overall recommendation performance. ξ. indicates the value of ξ-UOF. Adv. indicates advantaged users. Dis. indicates disadvantaged users. The results of II-GOOT are highlighted in bold. The best results are marked with *. The second-best results are underlined.

	Beauty		Grocery		Health	
	Ove.	ξ.	Ove.	ξ.	Ove.	ξ.
NDCG						
Original	0.245	0.190	0.257	0.144	0.263	0.238
Intra	0.274	0.160	0.279	0.124	0.290	0.207
Inter	0.269	0.159	0.277	0.128	0.280	0.219
II-GOOT	**0.286**	**0.149**	**0.289**	**0.112**	**0.309**	**0.192**
HR						
Original	0.375	0.250	0.393	0.230	0.376	0.257
Intra	0.420	0.214	0.424	0.191	0.425	0.205
Inter	0.410	0.220	0.420	0.193	0.417	0.212
II-GOOT	**0.433**	**0.193**	**0.443**	**0.180**	**0.448**	**0.185**

Table 2: Ablation study. Ove. indicates the overall recommendation performance. ξ. indicates the value of ξ-UOF.

optimal outcomes.

Generalizability of II-GOOT (Q4)

We conduct experiments in Appendix D.1 to prove that II-GOOT has strong generalizability in narrowing the recommendation gap across various user distributions.

The Change in L_{inter} (Q5)

We conduct experiments in Appendix D.2 to prove that II-GOOT has the ability to narrow the training gap between advantaged and disadvantaged users.

Conclusion

This paper focuses on rarely studied User-Oriented Fairness (UOF) in recommender systems, with the objective of reducing the recommendation performance gap between advantaged and disadvantaged users. We address the UOF issue in two phases of recommendation models. *In the training phase*, we propose an **I**ntra- and **I**nter-**G**r**O**up **O**ptimal **T**ransport (II-GOOT) framework. This framework effectively narrows the training gap between advantaged and disadvantaged users through the intra-group stage and the inter-group stage. *In the evaluation phase*, we propose a novel ξ-UOF metric to give a comprehensive evaluation of the UOF issue. We conduct extensive experiments on three real-world datasets based on four backbone models, demonstrating the efficiency of II-GOOT and the effectiveness of ξ-UOF.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (No. 72192823), the "Ten Thousand Talents Program" of Zhejiang Province for Leading Experts (No. 2021R52001).

References

Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein generative adversarial networks. In *International conference on machine learning*, 214–223. PMLR.

Asano, Y. M.; Rupprecht, C.; and Vedaldi, A. 2019. Self-labelling via simultaneous clustering and representation learning. *arXiv preprint arXiv:1911.05371*.

Ben-David, S.; Blitzer, J.; Crammer, K.; and Pereira, F. 2006. Analysis of representations for domain adaptation. *Advances in neural information processing systems*, 19.

Biega, A. J.; Gummadi, K. P.; and Weikum, G. 2018. Equity of attention: Amortizing individual fairness in rankings. In *The 41st international acm sigir conference on research & development in information retrieval*, 405–414.

Binns, R. 2018. Fairness in machine learning: Lessons from political philosophy. In *Conference on fairness, accountability and transparency*, 149–159. PMLR.

Bogachev, V. I.; and Kolesnikov, A. V. 2012. The Monge-Kantorovich problem: achievements, connections, and perspectives. *Russian Mathematical Surveys*, 67(5): 785.

Chen, J.; Dong, H.; Wang, X.; Feng, F.; Wang, M.; and He, X. 2023. Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems*, 41(3): 1–39.

Chen, L.; Zhang, Y.; Zhang, R.; Tao, C.; Gan, Z.; Zhang, H.; Li, B.; Shen, D.; Chen, C.; and Carin, L. 2019. Improving sequence-to-sequence learning via optimal transport. *arXiv preprint arXiv:1901.06283*.

Courty, N.; Flamary, R.; Habrard, A.; and Rakotomamonjy, A. 2017. Joint distribution optimal transportation for domain adaptation. *Advances in neural information processing systems*, 30.

Cuturi, M. 2013. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26.

Cuturi, M.; and Doucet, A. 2014. Fast computation of Wasserstein barycenters. In *International conference on machine learning*, 685–693. PMLR.

Dai, E.; Zhao, T.; Zhu, H.; Xu, J.; Guo, Z.; Liu, H.; Tang, J.; and Wang, S. 2022. A comprehensive survey on trustworthy graph neural networks: Privacy, robustness, fairness, and explainability. *arXiv preprint arXiv:2204.08570*.

Damodaran, B. B.; Kellenberger, B.; Flamary, R.; Tuia, D.; and Courty, N. 2018. Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation. In *Proceedings of the European conference on computer vision (ECCV)*, 447–463.

Dash, A.; Chakraborty, A.; Ghosh, S.; Mukherjee, A.; and Gummadi, K. P. 2021. When the umpire is also a player: Bias in private label product recommendations on e-commerce marketplaces. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 873–884.

Deldjoo, Y.; Anelli, V. W.; Zamani, H.; Bellogin, A.; and Di Noia, T. 2021a. A flexible framework for evaluating user and item fairness in recommender systems. *User Modeling and User-Adapted Interaction*, 1–55.

Deldjoo, Y.; Anelli, V. W.; Zamani, H.; Bellogin, A.; and Di Noia, T. 2021b. A flexible framework for evaluating user and item fairness in recommender systems. *User Modeling and User-Adapted Interaction*, 1–55.

Deldjoo, Y.; Bellogin, A.; and Di Noia, T. 2021. Explaining recommender systems fairness and accuracy through the lens of data characteristics. *Information Processing & Management*, 58(5): 102662.

Deldjoo, Y.; Jannach, D.; Bellogin, A.; Difonzo, A.; and Zanzonelli, D. 2022. A survey of research on fair recommender systems. *arXiv preprint arXiv:2205.11127*.

Deldjoo, Y.; Jannach, D.; Bellogin, A.; Difonzo, A.; and Zanzonelli, D. 2023. Fairness in recommender systems: research landscape and future directions. *User Modeling and User-Adapted Interaction*, 1–50.

Dwork, C.; Hardt, M.; Pitassi, T.; Reingold, O.; and Zemel, R. 2012. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, 214–226.

Ferraro, A. 2019. Music cold-start and long-tail recommendation: bias in deep representations. In *Proceedings of the 13th ACM Conference on Recommender Systems*, 586–590.

Figalli, A.; and Glaudo, F. 2021. *An Invitation to Optimal Transport, Wasserstein Distances, and Gradient Flows*. European Mathematical Society/AMS. ISBN 978-3-98547-010-5.

Flamary, R.; Courty, N.; Tuia, D.; and Rakotomamonjy, A. 2016. Optimal transport for domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell*, 1: 1–40.

Gharahighehi, A.; Vens, C.; and Pliakos, K. 2021. Fair multi-stakeholder news recommender system with hypergraph ranking. *Information Processing & Management*, 58(5): 102663.

Gorantla, S.; Deshpande, A.; and Louis, A. 2021. On the problem of underranking in group-fair ranking. In *International Conference on Machine Learning*, 3777–3787. PMLR.

Han, Z.; Chen, C.; Zheng, X.; Liu, W.; Wang, J.; Cheng, W.; and Li, Y. 2023a. In-processing User Constrained Dominant Sets for User-Oriented Fairness in Recommender Systems. In *Proceedings of the 31st ACM International Conference on Multimedia*, 6190–6201.

Han, Z.; Zheng, X.; Chen, C.; Cheng, W.; and Yao, Y. 2023b. Intra and Inter Domain HyperGraph Convolutional Network for Cross-Domain Recommendation. In *Proceedings of the ACM Web Conference 2023*, 449–459.

Hardt, M.; Price, E.; and Srebro, N. 2016. Equality of opportunity in supervised learning. *Advances in neural information processing systems*, 29.

He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; and Wang, M. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, 639–648.

He, X.; Liao, L.; Zhang, H.; Nie, L.; Hu, X.; and Chua, T.-S. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, 173–182.

Hutchinson, B.; and Mitchell, M. 2019. 50 years of test (un) fairness: Lessons for machine learning. In *Proceedings of the conference on fairness, accountability, and transparency*, 49–58.

Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer*, 42(8): 30–37.

Li, Y.; Chen, C.; Zheng, X.; Zhang, Y.; Han, Z.; Meng, D.; and Wang, J. 2023. Making Users Indistinguishable: Attribute-wise Unlearning in Recommender Systems. In *Proceedings of the 31st ACM International Conference on Multimedia*, MM '23, 984–994. New York, NY, USA: Association for Computing Machinery. ISBN 9798400701085.

Li, Y.; Chen, H.; Fu, Z.; Ge, Y.; and Zhang, Y. 2021. User-oriented fairness in recommendation. In *Proceedings of the Web Conference 2021*, 624–632.

Li, Y.; Zheng, X.; Chen, C.; and Liu, J. 2022. Making recommender systems forget: Learning and unlearning for erasable recommendation. *arXiv preprint arXiv:2203.11491*.

Liang, D.; Krishnan, R. G.; Hoffman, M. D.; and Jebara, T. 2018. Variational autoencoders for collaborative filtering. In *Proceedings of the 2018 world wide web conference*, 689–698.

Liu, W.; Su, J.; Chen, C.; and Zheng, X. 2021. Leveraging distribution alignment via stein path for cross-domain cold-start recommendation. *Advances in Neural Information Processing Systems*, 34: 19223–19234.

Liu, W.; Zheng, X.; Hu, M.; and Chen, C. 2022a. Collaborative filtering with attribution alignment for review-based non-overlapped cross domain recommendation. In *Proceedings of the ACM Web Conference 2022*, 1181–1190.

Liu, W.; Zheng, X.; Su, J.; Hu, M.; Tan, Y.; and Chen, C. 2022b. Exploiting variational domain-invariant user embedding for partially overlapped cross domain recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 312–321.

Liu, Z.; Fang, Y.; and Wu, M. 2023. Mitigating popularity bias for users and items with fairness-centric adaptive recommendation. *ACM Transactions on Information Systems*, 41(3): 1–27.

Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; and Galstyan, A. 2021. A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR)*, 54(6): 1–35.

Melchiorre, A. B.; Rekabsaz, N.; Parada-Cabaleiro, E.; Brandl, S.; Lesota, O.; and Schedl, M. 2021. Investigating gender fairness of recommendation algorithms in the music domain. *Information Processing & Management*, 58(5): 102666.

Rahmani, H. A.; Naghiaei, M.; Dehghan, M.; and Aliannejadi, M. 2022. Experiments on generalizability of user-oriented fairness in recommender systems. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2755–2764.

Rastegarpanah, B.; Gummadi, K. P.; and Crovella, M. 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.

Santambrogio, F. 2015. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63): 94.

Shakespeare, D.; Porcaro, L.; Gómez, E.; and Castillo, C. 2020. Exploring artist gender bias in music recommendation. *arXiv preprint arXiv:2009.01715*.

Su, J.; Chen, C.; Liu, W.; Wu, F.; Zheng, X.; and Lyu, H. 2023. Enhancing Hierarchy-Aware Graph Networks with Deep Dual Clustering for Session-based Recommendation. In *Proceedings of the ACM Web Conference 2023*, 165–176.

Sühr, T.; Hilgard, S.; and Lakkaraju, H. 2021. Does fair ranking improve minority outcomes? understanding the interplay of human and algorithmic biases in online hiring. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 989–999.

Verma, S.; and Rubin, J. 2018. Fairness definitions explained. In *Proceedings of the international workshop on software fairness*, 1–7.

Villani, C.; et al. 2009. *Optimal transport: old and new*, volume 338. Springer.

Wang, Y.; Wang, L.; Li, Y.; He, D.; and Liu, T.-Y. 2013. A theoretical analysis of NDCG type ranking measures. In *Conference on learning theory*, 25–54. PMLR.

Waters, S. 1976. Hit ratios. *The Computer Journal*, 19(1): 21–24.

Wen, H.; Yi, X.; Yao, T.; Tang, J.; Hong, L.; and Chi, E. H. 2022. Distributionally-robust Recommendations for Improving Worst-case User Experience. In *Proceedings of the ACM Web Conference 2022*, 3606–3610.

Xu, R.; Liu, P.; Zhang, Y.; Cai, F.; Wang, J.; Liang, S.; Ying, H.; and Yin, J. 2020. Joint Partial Optimal Transport for Open Set Domain Adaptation. In *IJCAI*, 2540–2546.

Yang, J.; Liu, Y.; and Xu, H. 2023. HOTNAS: Hierarchical Optimal Transport for Neural Architecture Search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11990–12000.

Zheng, X.; Su, J.; Liu, W.; and Chen, C. 2022a. DDGHM: dual dynamic graph with hybrid metric training for cross-domain sequential recommendation. In *Proceedings of the 30th ACM International Conference on Multimedia*, 471–481.

Zheng, X.; Wu, R.; Han, Z.; Chen, C.; Chen, L.; and Han, B. 2022b. Heterogeneous Information Crossing on Graphs for Session-based Recommender Systems. *ACM Transactions on the Web*.