

# A novel approach for measuring demographic parity fairness in group recommendation

Raciel Yera<sup>1\*</sup>, Manuel J. Barranco<sup>1</sup> and Luis Martínez<sup>1</sup>

<sup>1</sup>Computer Science Department, University of Jaén, Campus Las Lagunillas s/n, Jaén, 23071, Spain.

\*Corresponding author(s). E-mail(s): [ryera@ujaen.es](mailto:ryera@ujaen.es);  
Contributing authors: [barranco@ujaen.es](mailto:barranco@ujaen.es); [martin@ujaen.es](mailto:martin@ujaen.es);

## Abstract

Fairness is currently becoming a necessary dimension to consider in contemporary artificial intelligence(AI)-based systems, according to the recent Ethics Guidelines for a Trustworthy AI. Being recommender systems popular applications that incorporate AI in a larger or lesser extent, the literature analysis identifies a research gap related to the exploration of demographic parity fairness in the group recommendation scenario. The current work is focused on this gap, developing a group recommendation framework that has as main novelty the measuring of the consumer fairness taking into account the presence of advantaged and disadvantaged class of users. Experimental studies are developed for measuring the performance of the proposal in a real recommendations scenario, illustrating that it is able to distinguish different fairness levels across the delivered recommendations.

**Keywords:** group recommender systems, demographic parity fairness, experiments

This paper is the Accepted Version of the following published contribution:

Yera, R., Barranco, M. J., & Martínez, L. (2024). A Novel Approach for Measuring Demographic Parity Fairness in Group Recommendation. In *Intelligent Management of Data and Information in Decision Making (FLINS-ISKE 2024)* (pp. 195-202). [https://doi.org/10.1142/9789811294631\\_0025](https://doi.org/10.1142/9789811294631_0025)

# 1 Introduction

The Ethics Guidelines for a Trustworthy AI <sup>1</sup>, published by the European Commission in 2019, identified that the development, deployment, and use of artificial intelligence (AI) systems should adhere to the ethical principles of respect for human autonomy, prevention of harm, explainability, and fairness. Particularly, fairness concerns a commitment to ensure that individuals and groups in the system, are free from unfair bias and discrimination in AI, assuring an equal and fair distribution of benefits and costs. Considering its social impact in gender and diversity-related issues, the studies on measuring and improving the systems' fairness are becoming a relevant and necessary topic in AI research.

Recently, the concept of fairness has been brought to the recommender systems (RS) research area. As popular AI-based systems, RSs are focused on providing users with items that best fit their preferences and needs, in an overloaded search space of possible options [1]. Since 90s, they have been applied in scenarios such as e-commerce, e-learning, e-health, and so on [1].

Formally, fairness is associated with the recognition of RS as a multi-stakeholder scenario, in which the main stakeholders are consumers (C) and providers (P) [2]. Specifically C-fairness, the focus of this work, is formerly centered on the disparate impact of recommendations on protected classes of consumers, associated with sensitive features, e.g., gender, race, and age.

In other direction, recommender systems have been recently brought to scenarios where users consume items in groups and not individually, such as TV programs, restaurants, or tourism packages [3, 4]. For these scenarios, the concept of group recommender systems has been developed for suggesting items focused on satisfying the overall preference of the group beyond the particularities of each individual preference of the users.

However, research on fairness has been mainly centered on individual RS [2, 5, 6], being on the other hand very limited the research on recommendation fairness in group recommender systems (GRS). Current GRS proposals only consider fairness by optimising the exposure of items in the final recommendation lists (e.g. the recommendation of items that are preferred by the whole group, as far as possible [7, 8]), without taking into account the user's demographic attributes, which are considered as key information for characterizing algorithmic fairness [9]. Here the item exposure makes reference to the position of the item in the list of generated recommendations, having a top position being considered as a higher exposure.

At this stage, we consider that it is necessary to boost the study of fairness in the traditional GRS scenario. With this goal in mind, the current paper extensively brings to the GRS context the **C-fairness** concept proposed by Burke [2] which is supported by the demographic attributes of the consumers.

The paper is structured as follows. Section 2 presents some antecedents on fairness in group recommendation. Section 3 explores how to measuring demographic attributes-related fairness in GRS. Section 4 develops the evaluation of the presented approach. Section 5 concludes the contribution.

---

<sup>1</sup><https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

## 2 Fairness in group recommender systems

In contrast to fairness in individual RS [5, 6], fairness in GRS has been hardly covered by the specialized literature, and with a much more limited scope. Xiao et al. [10] proposed several definitions of fairness, encouraging the group members to minimize the gap between least and highest recommendation utilities. Serbos et al. [11] explore fairness in the package-to-group recommendation problem, by formalizing the concepts of m-proportional fairness and m-envy-freeness, modelling greedy algorithms to find appropriate recommendations. Sacharidis [7] also focused on fairness in GRS, defining the member utility of a group recommendation list as the similarity value of the list and the member ground truth. The fairness of the list is then defined as the lowest member utility of the group, and is reached through the items probability to be in the top-n pareto-optimal items for the group.

A similar goal is followed by Stratigi et al. [12] focused on an iterative process to find a top-n list for the group that maximizes global utility. Also, Felfernig et al. [13] present FAI, a method where group’s individuals incrementally build the recommendation list, in a fair turn-based way.

Finally, Kaya et al. [8] have focused on a rank-sensitive balancing of relevance across the group. Herein, the group recommendation list is built in a greedy way, according to a score that depends on the probability that the item is relevant to each member and that there is at least one relevant item in the list for each group’s member.

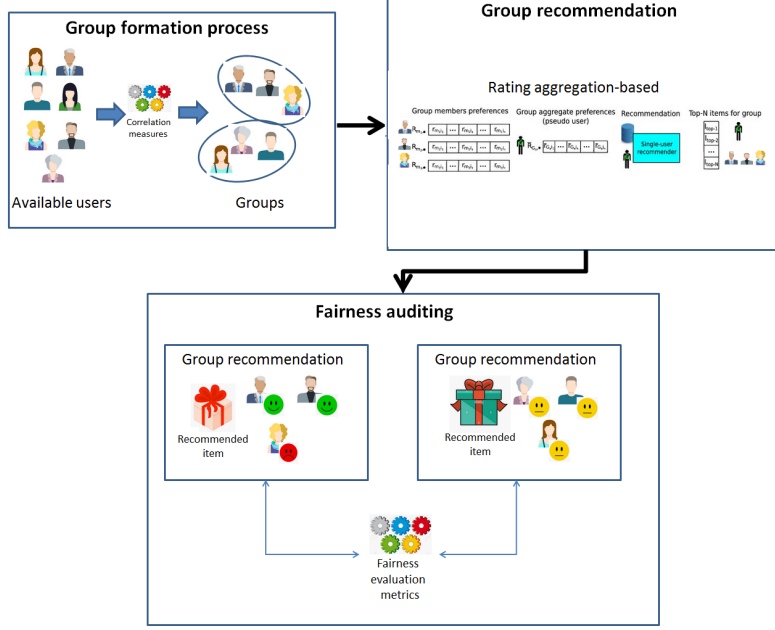
In short, current proposals for GRSs consider fairness by optimizing user fairness in item exposure at the final recommendation list, a particular C-fairness scenario in individual RS. Furthermore, Pitoura et al. [9] have recently pointed out that in GRS the few fairness-aware approaches have been centered only on individual and consumer fairness. It is then necessary to take into account fairness from a more holistic viewpoint than in individual RS, by evaluating C-fairness from the demographic parity viewpoint (performance in protected classes, e.g. age, gender), in the GRS. This is the goal of the current work.

## 3 Formulating consumer fairness in the GRS scenario

This section is focused on presenting the framework that will be used in the current work, for exploring fairness in group recommendation. This architecture is composed of three stages (see Figure 1): 1) group formation, 2) group recommendation process, and 3) fairness measuring in group recommendation.

*Group formation process.* The typical group formation process in GRS, usually predefines the associated groups in a previous stage before the recommendation generation. The current work will consider the selection of groups with users having a large similarity in their rating profiles as group formation strategy. This approach has been explored by several research works focused on this issue [3]

*Group recommendation process.* In this stage it will be used a rating aggregation-based GRS approach, as one of the basic approaches in group recommendation [4]. It is based on the creation, for each group, of a pseudo-user that contains the preferences



**Fig. 1** Overall architecture for measuring fairness in group recommendation

of the underlying users. This pseudo-user is then used for generating recommendations as it was an individual, typical user [4]. In this work it will be considered the average approach for aggregating the user’s preferences in the composition of the group profile.

*Fairness measuring in GRS.* Here we will present the novel strategy for characterizing fairness in group recommendation. As novel feature, it is focused on bringing the demographic attributes-related fairness, into the framework of group recommendation. Specifically, it is centered on exploring whether the group members belonging to the disadvantaged classes (regarding users’ demographic features), are adversely affected by the generated recommendations.

With this purpose in mind, at first we identify the level of deviation of the actual user preferences  $r_{ui}$ , in relation to the predictions  $r_{Gi}$  done for the group using a GRS algorithm (Equation 1). In this case we consider a signed deviation regarding we are interested in exploring whether the users’ ratings are over or under the predicted group preferences.

$$User\_Group\_Dev(u) = \frac{1}{|R_u|} \sum_{r_{ui} \in R_u} (r_{ui} - r_{Gi}) \quad (1)$$

In a second stage, we characterize the recommendation unfairness  $Unfairn$  of the user  $u$  in a group recommendation framework, as how close are his/her  $User\_Group\_Dev(u)$  value, in relation to the  $User\_Group\_Dev(v)$  values of the other users  $v \in G$  (Equation 2). If this difference is far enough, we consider that the underlying GRS model performed different for this user, being associated this behavior with a larger unfairness.

Average unfairness values according to Equation 2, for disadvantaged and advantaged classes in group members

	Gender	Age	Age
Disadvantaged users	<i>Gender = Woman</i> 0.2454±0.1893	<i>Age &lt; 25</i> 0.1572±0.1202	<i>Age &gt; 55</i> 0.1866±0.1476
Advantaged users	<i>Gender = Man</i> 0.2070±0.1469	<i>Age &gt;= 25</i> 0.2007±0.1516	<i>Age &lt;= 55</i> 0.2312±0.1573

Average unfairness values for the identified groups and the disadvantaged classes of users, according to Equation 3

	Gender	Age	Age
Group unfairness	<i>Gender = Woman</i> 0.1790±0.1029	<i>Age &lt; 25</i> 0.1455±0.0741	<i>Age &gt; 55</i> 0.2037±0.1291

$$Unfairn(u) = \min(|User\_Group\_Dev(u) - User\_Group\_Dev(v)|); v \in G \quad (2)$$

Finally, the unfairness level associated to a group  $G$  (Equation 3), is represented as the absolute difference between the average unfairness of the users respectively belonging to the set  $A$  and  $D$  of the advantaged and disadvantaged users in the group  $G$ .

$$Unfairn(G) = \left| \frac{1}{|A|} \sum_{u \in A} Unfairn(u) - \frac{1}{|D|} \sum_{v \in D} Unfairn(v) \right| \quad (3)$$

The next section will be focused on exploring this unfairness criteria in a traditional RS dataset.

## 4 Experiments

In this study we will use the traditional Movielens 100K dataset [14]. It contains the ratings provided by 943 users over a set of 1682 movies, having 100000 ratings values in the range [1, 5], and a sparsity around 94% [14]. In addition to the preference data, there is available some demographic information for each user, specifically age, gender, and occupation.

The Pearson correlation coefficient is used as similarity metric, building groups with users with a pairwise similarity  $\alpha \geq 0$ . Particularly, we will build groups of size 4; 3 out of these 4 users belonging to the *advantaged* demographic user class, and the fourth user belonging to the *disadvantaged* user class. In this context, the goal of the current experiment is to characterize the Unfair value for each of the 4 members of the group (Equation 2), as well as the overall Unfair value for the whole group (Equation 3). The used GRS approach was the rating aggregation-based GRS approach with the average aggregation function [4].

Table 4 illustrates the unfairness degree associated to the mentioned dataset, according to three possible disadvantaged classes of users, which are *Gender =*

*Woman*,  $Age < 25$ , and  $Age > 55$ . The selection of these three classes follows a common sense connected to the user gender ( $F$ ), to young people ( $Age < 25$ ), and old people ( $Age > 55$ ). In addition, it is also presented the average unfairness associated to the users in the advantaged class ( $M$ ,  $Age \geq 25$ , and  $Age \leq 55$ ).

According to Table 1, the larger unfairness values were associated to the gender attribute. In this case, the selected disadvantaged class was associated to the larger unfairness values, lying around 0.2454. However, in the case of the age criteria, it was obtained a higher unfairness value for the advantaged class in relation to the disadvantaged class. Beyond this result, it was obtained a larger unfairness value for the criteria  $Age > 55$  in relation to  $Age < 25$ .

Table 2 illustrates the average unfairness values associated to the groups. Here it is interesting to point out that even though the disadvantaged class of users with the larger unfairness according to Table 1 was  $Gender = Woman$ , in this second table the larger group unfairness was associated to  $Age > 55$ . This behavior can be explained by the fact that in the case of gender it was obtained a high individual unfairness according to Equation 2 for both disadvantaged and advantaged classes, leading to a lower unfairness for the groups according to Equation 3 that is focused on measuring the difference between unfairness at advantaged and disadvantaged classes. In a different direction, the lower unfairness value was obtained for " $Age < 25$ ".

Overall, the experimental results lead to the following findings:

- The approach exposed at Section 3 for characterizing unfairness in group recommendation, is able to identify different unfairness levels in GRS.

- The difference in the results obtained in Tables 1 and 2 regarding the larger unfairness value ( $Gender = Woman$  in Table 1;  $Age > 55$  in Table 2), illustrate that it would be interesting to manage two different unfairness notions in this GRS scenario, which are 1) a different prediction nature (over or under the actual preferences) in relation to the other group members, and 2) a different prediction nature in relation to the users belonging to other demographic attributes.

- The criterion  $Age < 25$  leads to the lower unfairness value regarding the considered disadvantaged classes, being interesting the exploration of other ages based on the same criteria.

## 5 Conclusions

The current research work has been centered on bringing a novel framework for characterizing demographic parity fairness in group recommendation. The performed experiments illustrate that the proposal is able to distinguish different fairness levels in demographic attributes such as gender and age. Future works will be focused on studying other kinds of attributes and other fairness dimensions in the GRS scenario.

**Acknowledgements:** This research is supported by the European Union's Horizon Europe research and innovation program under the Marie Skłodowska-Curie grant agreement number 101106164. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

## References

- [1] Lu, J., Wu, D., Mao, M., Wang, W., Zhang, G.: Recommender system application developments: a survey. *Decision Support Systems* **74**, 12–32 (2015)
- [2] Burke, R.: Multisided fairness for recommendation. In: *Workshop on Fairness, Accountability, and Transparency in Machine Learning*, pp. 1–5 (2017)
- [3] Pérez-Almaguer, Y., Yera, R., Alzahrani, A.A., Martínez, L.: Content-based group recommender systems: a general taxonomy and further improvements. *Expert Systems with Applications* **184**, 115444 (2021)
- [4] De Pessemier, T., Dooms, S., Martens, L.: Comparison of group recommendation algorithms. *Multimedia Tools and Applications* **72**(3), 2497–2541 (2014) <https://doi.org/10.1007/s11042-013-1563-0>
- [5] Bobadilla, J., Lara-Cabrera, R., Gonzalez-Prieto, A., Ortega, F.: Deepfair: Deep learning for improving fairness in recommender systems. *International Journal of Interactive Multimedia and Artificial Intelligence* **6**(6), 86–95 (2021)
- [6] Deldjoo, Y., Jannach, D., Bellogin, A., Difonzo, A., Zanzonelli, D.: Fairness in recommender systems: research landscape and future directions. *User Modeling and User-Adapted Interaction*, 1–50 (2023)
- [7] Sacharidis, D.: Top-n group recommendations with fairness. In: *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, pp. 1663–1670 (2019)
- [8] Kaya, M., Bridge, D., Tintarev, N.: Ensuring fairness in group recommendations by rank-sensitive balancing of relevance. In: *Proceedings of the 14th ACM Conference on Recommender Systems*, pp. 101–110 (2020)
- [9] Pitoura, E., Stefanidis, K., Koutrika, G.: Fairness in rankings and recommendations: An overview. *The VLDB Journal* **31**(3), 431–458 (2022)
- [10] Xiao, L., Min, Z., Yongfeng, Z., Zhaoquan, G., Yiqun, L., Shaoping, M.: Fairness-aware group recommendation with pareto-efficiency. In: *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pp. 107–115 (2017)
- [11] Serbos, D., Qi, S., Mamoulis, N., Pitoura, E., Tsaparas, P.: Fairness in package-to-group recommendations. In: *Proceedings of the 26th International Conference on World Wide Web*, pp. 371–379 (2017)
- [12] Stratigi, M., Nummenmaa, J., Pitoura, E., Stefanidis, K.: Fair sequential group recommendations. In: *Proceedings of the 35th Annual ACM Symposium on Applied Computing*, pp. 1443–1452 (2020)
- [13] Felfernig, A., Boratto, L., Stettinger, M., Tkalčič, M., *et al.*: *Group Recommender Systems: An Introduction*. Springer, ??? (2018)

- [14] Harper, F.M., Konstan, J.A.: The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TIIS)* **5**(4), 1–19 (2015)