Individual Fairness in Algorithmic Hiring

STAVROS DAVIDOPOULOS, University of the Aegean, Greece
PANAGIOTIS SYMEONIDIS, University of the Aegean, Greece
DIMITRIS SACHARIDIS, Université Libre de Bruxelles, Brussels

In this paper, we study individual fairness in job recommendations and make two concrete contributions. To the best of our knowledge, we are the first to introduce the concept of ε-individual fairness to job recommendations; the smaller the value of ε, the stronger the guarantee of individual fairness is. Compared to existing definitions of individual fairness, e.g., for classification tasks, the output of a recommender is a ranked list of items, here jobs. Therefore, the novel aspect is that we propose the use of Kendall's τ distance as a measure of similarity between recommendation lists. To ensure individual fairness, we introduce a novel post-processing approach. Initially, we cluster individuals with similar non-protected attributes. For each cluster, we construct a Kemeny-optimal aggregate recommendation list, which will serve as a template to generate individual recommendations. As a result, similar individuals will receive similar recommendations. Our experiments show that our method can effectively control the individual fairness guarantee, i.e., the value of ε.

Keywords: Individual fairness; Job recommendations; Effectiveness

Reference Format:
Stavros Davidopoulos, Panagiotis Symeonidis, and Dimitris Sacharidis. 2025. Individual Fairness in Algorithmic Hiring. In *Proceedings of Fourth European Workshop on Algorithmic Fairness (EWAF'25).* Proceedings of Machine Learning Research, 6 pages.

1 Introduction

Job recommendation systems are widely used in online platforms to match job seekers with job postings. They use machine learning algorithms to analyze job seekers' profiles and recommend job postings that match their qualifications and preferences. However, these algorithms may introduce biases that can lead to unfair outcomes for job seekers. For example, a system may consistently recommend high-paying job postings to men rather than women. Therefore, algorithmic fairness is important when designing job recommendation systems.

Two main categories of algorithmic fairness exist [11, 18]. *Individual fairness* requires that similar individuals should be treated similarly, while *group fairness* requires that groups of people, defined by certain *protected* attributes (such as gender, race, religion) should be treated similarly. While group fairness has been studied in depth [12, 14, 16], there are few studies for individual fairness because it is hard to operationalize, as it requires domain knowledge to define meaningful, task-specific similarity metrics for individuals [5]. Individual fairness offers advantages over group fairness, as it can provide more fine-grained fairness guarantees and is robust to

Authors' Contact Information: Stavros Davidopoulos, University of the Aegean, Greece, icsdd23012@icsd.aegean.gr; Panagiotis Symeonidis, University of the Aegean, Samos, Greece, psymeon@aegean.gr; Dimitris Sacharidis, Université Libre de Bruxelles, Brussels, dimitris.sacharidis@ulb.be.

EWAF'25, June 30–July 02, 2025, Eindhoven, NL

fairness gerrymandering [9], which refers to a situation where an algorithm appears to be fair when considering overall metrics but is unfair or discriminatory when examined at a more granular level, e.g., across subpopulations.

We consider a job recommendation system where the users are the job seekers, the items are the job postings, and the system produces a ranked list of job postings for each job seeker. We aim to ensure that the recommender is *accurate*, and *individually fair*. *Accuracy* means that the system should match job seekers with job postings that they are likely to be interested in. *Individual fairness* means that job seekers with similar qualifications (i.e., the task-specific and non-protected attributes) should receive similar job recommendations. For example, a male and a female with similar qualifications should receive similar job recommendations, ensuring that the system does not introduce biases based on the, irrelevant for the task, gender attribute.

Contributions. In this work, we introduce ε-individual fairness for the recommendation task. Briefly, our definition requires that similar, in terms of qualifications, individuals will receive similar recommendation lists, where the latter similarity is measured in terms of Kendall's τ distance. The ε parameter dictates how much the recommendation similarity can differ with respect to the individuals' similarity, which is in line with the corresponding definition for classification [5]. Moreover, we introduce the notion of *individual fairness loss* as the smallest value of ε that holds for a given set of recommendation lists.

To ensure individual fairness, we propose the iFair algorithm that tries to minimize the individual fairness loss while preserving the recommendation effectiveness. iFair first clusters job seekers setting a distance threshold θ on the similarity of their qualifications. Then it derives the Kemeny's aggregated ranking [10], which minimizes the Kendall τ distance among pairs of ranked lists within a cluster. This aggregate ranking acts as a *template* describing the qualifications that best achieve a consensus among the users within each cluster. To return recommendations to job seekers, iFair selects the best job postings that match the template of the cluster they belong to. The contributions of our work are summarized as follows: (i) We define ε-individual fairness for the recommendation task, providing an algorithmic guarantee of equitable treatment of alike job seekers. (ii) We introduce the iFair algorithm that seeks to maximize recommendation effectiveness while guaranteeing individual fairness. (iii) We conduct experiments on two real-life data sets (XING16 and CAREERBUILDER12), showing that iFair can effectively control the individual fairness loss with a minimal drop in recommendation effectiveness.

Related Work. Past work at the intersection of algorithmic fairness and job recommendation is scarce. Mashayekhi et al. [14] propose ReCon to reduce congestion in job recommendations by using optimal transport theory. Congestion exists when there is an unequal distribution of the jobs in how often they are recommended (i.e., some jobs may be recommended much more than others). Their approach aims at a more fair job market, looking to ensure a more equal spread of jobs over job seekers. More related to ours, is the work of Markert et al. [13] that explore the existence of individual fairness for a job candidates ranking system. They follow the framework of Ruoss et al. [15], which involves training an encoder with both classification loss and fairness loss to create informative and fair representations of the data. They seek to achieve individual fairness *indirectly*, through fair representation learning, whereas we directly target individual fairness and provide guarantees.

2 Defining Individual Fairness for Recommendation Lists

Individual fairness dictates that *similar individuals* should be *treated similarly*. Crucially, determining similarity among individuals should be task-specific [5]. To operationalize this principle, one needs to define two task-specific distance metrics: the distance between two inputs (in our setting, users), and the distance between two outputs (in our setting, recommendation lists). Then, individual fairness enforces a Lipschitz continuity constraint on the ML model [5]. Concretely, let f denote the model that takes an input x and produces output $y = f(x)$. The model is *individually fair* if for any pair of inputs x, x' their outputs are not far off from each other: $d_Y(y, y') \leq \varepsilon \cdot d_X(x, x')$, where d_X, d_Y are distance metrics (satisfying triangle inequality) in the space of inputs and outputs respectively, and ε is a non-negative real constant. Please note that in our job recommendation scenario, we denote the task-specific *user distance* between u, u' over their job qualifications as $d_{\mathbb{U}}(u, u')$.

Before we define a distance metric for recommendation lists, we first discuss the rationale assuming that the output is a single item (job posting). To allow for personalization, we consider two items to be identical as long as they have the same *type*, meaning identical values in the item attributes that are pertinent for the task at hand. For example, two job postings t and t' are considered identical if they are in the same *industry* and *department*, require the same level of *experience*, and offer the same level of *salary*, even though they are for two different companies at two distinct locations, say *city* and *city'*, respectively. This way, two identical users u and u' (w.r.t. the job-pertinent attributes) who live in *city* and *city'*, respectively, are treated fairly when they are recommended job postings of the same type t and t', respectively.

The output of a job recommender to an individual is a ranked list of job postings. A natural way to compare two ranked lists is to compute their Kendall τ distance [4] of the job requirements (i.e., items' attributes), which measures how many pairs of elements (i.e., attributes of the job postings) are in *discordance*, i.e., are ranked differently in the two lists; two identical lists have zero discordant pairs. Concretely, consider two ranked lists ρ, σ containing the same $n = |\rho| = |\sigma|$ elements. For an element i, let $\rho[i]$ (resp. $\sigma[i]$) denote its position in the list ρ (resp. σ). A pair of elements (i, j) is *discordant* if: $\rho[i] > \rho[j],\ \sigma[i] < \sigma[j]$ or $\rho[i] < \rho[j],\ \sigma[i] > \sigma[j]$. The Kendall τ distance is the number of discordant pairs, and can be normalized to lie in the $[0, 1]$ range by dividing with the number of unordered pairs $\binom{n}{2}$. Please note that we do not define the Kendal τ correlation coefficient. Henceforth, we define the *ranking distance* between two users u, u', denoted as $d_{\mathbb{R}}(R_u, R_{u'})$ with the Kendall τ distance.

Having defined the two task-specific distance metrics, the user $d_{\mathbb{U}}$ and the ranking $d_{\mathbb{R}}$ distances of job requirements (i.e., item's attributes), we can now formally define individual fairness for the task of job recommendation.

Definition 2.1 (ε-Individual Fairness). Given a set $\mathbb{R}$ of recommendation lists generated by a recommender for a set $\mathbb{U}$ of users, the recommender is *ε-individually fair* if for any pair of users $u, u' \in \mathbb{U}$ and their corresponding lists $R_u, R_{u'} \in \mathbb{R}$ it holds that

$$\forall u, u' \in \mathbb{U} \quad d_{\mathbb{R}}(R_u, R_{u'}) \leq \varepsilon \cdot d_{\mathbb{U}}(u, u') \tag{1}$$

Definition 2.1 guarantees that for any two users u and u' who are similar in their non-protected attributes (e.g., level of experience, certificate) the recommender's outputs will not differ much, ensuring similar treatment.

3 The iFair Method

In this section, we present the iFair method that re-ranks a given set of job recommendation lists provided to the users so that they are accurate and individually fair. The method consists of three steps (refer to Algorithm 1).

(1) We ensure that the job recommender is calibrated [17], requiring that recommendation lists contain job postings in proportion to the job seekers' preferences, as observed in their previous interactions. For this purpose, we extend the DetCons [6] algorithm to consider multiple job attributes.

Algorithm 1: The iFair Algorithm

Input: $\mathbb{U}$ set of users; $\mathbb{T}$ set of items; $\mathbb{A}$ set of attributes; $\mathbb{R}$ set of recommendation lists for all users; $\hat{\text{rel}}_u$ predicted relevance of items for each u; p_u^A desired proportion of values of each attribute A for each u; θ user distance threshold
Output: $\mathbb{R}''$ set of individually fair and calibrated reranked recommendation lists

▷ *Step 1: Fairness Re-ranking*
foreach $R_u \in \mathbb{R}$ **do**
  $R'_u \leftarrow \text{DetCons}(R_u, \hat{\text{rel}}_u, p_u^A)$ ▷ *re-ranked list R'_u for user u*

▷ *Step 2: Group Formation*
$\mathscr{G} \leftarrow \varnothing$ ▷ *set of groups*
while $\mathbb{U}$ is not empty **do**
  $u \leftarrow$ random-choice($\mathbb{U}$) ▷ *select a random user u*
  $G \leftarrow \{u' \in U \mid d_{\mathbb{U}}(u,u') \leq \theta/2\}$ ▷ *put all users that are close to random user u in a group*
  $\mathscr{G} \leftarrow \mathscr{G} \cup \{G\}$ ▷ *add group to the set of groups*
  $\mathbb{U} \leftarrow \mathbb{U} \setminus G$ ▷ *remove grouped users from the set of users*

▷ *Step 3: Individual Fairness Ensurance*
$\mathbb{R}'' \leftarrow \varnothing$ ▷ *the set of individually fair and calibrated lists*
foreach $G \in \mathscr{G}$ **do**
  $\mathbb{R}_G \leftarrow \{R'_u \mid u \in G\}$ ▷ *the set of calibrated rankings in the group*
  $R_G^{\mathbb{A}} \leftarrow \text{kemenize}(\mathbb{R}_G)$ ▷ *compute the Kemeny aggregate ranking*
  foreach $R'_u \in \mathbb{R}_G$ **do** ▷ *rerank each ranking in the group*
    $R''_u \leftarrow []$ ▷ *initialize list R''_u for user u*
    foreach $k \in [1, |R_G^{\mathbb{A}}|]$ **do**
      $t^* \leftarrow \arg\max\{\hat{\text{rel}}_u(t) \mid t \in R'_u \setminus R''_u \wedge t$ satisfies the attibute combination in $R_G^{\mathbb{A}}[k]\}$ ▷ *the item with the highest predicted relevance among those that have the required attributes' value at position k*
      $R''_u[k] \leftarrow t^*$ ▷ *insert the item in the list*
    $\mathbb{R}'' \leftarrow \mathbb{R}'' \cup \{R''_u\}$
return $\mathbb{R}''$

(2) We define a distance metric $d_{\mathbb{U}}$ between two users' job qualifications considering the non-protected attributes (e.g., *career level*, and *discipline*) and their corresponding values (e.g., *junior, intermediate, senior*). We set a threshold θ, and form clusters of job seekers that are similar within θ.

(3) In each cluster, we abstract the recommendation lists at the level of attribute values, rather than job postings. Then we find the Kemeny's [10] optimal aggregated ranked list, which best expresses a consensus among the attribute-level ranking of all similar job seekers within the cluster. Finally, for each cluster member, we recommend the most relevant job postings that follow the consensus template ranking.

4 Experimental Evaluation

Data Sets. The first dataset was originally used in RecSys Challenge 2016 [2, 3], with data were provided by XING. Henceforth, we will refer to this as the XING16 dataset. XING16 dataset contains 8,826,678 interactions of 1,367,057 users with 1,358,098 job postings. It also contains content features about the users (*career_level, discipline_id, region*, etc.) and job postings (*career_level, discipline_id, region*, etc.). The second dataset, Careerbuilder12 was originally used in an open Kaggle competition [1], called Job Recommendation Challenge, and was provided by the online employment Web site CareerBuilder. The CAREERBUILDER12 dataset contains 661,910 interactions of 120,147 users with 197,590 job postings. It is also contains content features about users (*MajTopic, Total Years Experience, State*, etc.) and job postings (*State, RecTopic, DescTopic*, etc.)

Evaluation Metrics. To evaluate accuracy we use the classic metrics, such as Utility ($\mathscr{U}$) [6], normalized Discounted Cumulative Gain (nDCG) [8]. To evaluate Group Fairness, we use Normalized Discounted KL-divergence (nDKL) [6, 17]. To assess the individual fairness for a pair of lists $R_u, R_{u'}$, we define the *fairness loss* as: $\ell^f(R_u, R_{u'}) = \frac{d_{\mathbb{R}}(R_u, R_{u'})}{d_{\mathbb{U}}(u, u')}$. Clearly, if for all pairs of users the fairness loss is below some constant ε, the recommender is ε-individually fair. To assess the overall individual fairness of a recommender, we consider the set $\mathbb{R}$ of all recommendation lists and measure the *average fairness loss* avg ℓ^f. The smaller the average fairness loss is, the more fair the recommender is.

Results. We evaluate the performance of iFair in terms of recommendation effectiveness, calibration, and individual fairness. We compare iFair with calibration techniques (DetGreedy, DetCons, and DetRelax from [6]), which however are not individual fairness-aware. All methods use the same base recommender, NeuMF [7].

Table 1 presents the results. The base recommender NeuMF outperforms all algorithms in recommendation effectiveness showcasing its strength in delivering highly relevant job recommendations to job seekers. It achieves perfect utility $\mathscr{U}@10 = 1$ (for top-10 recommended job postings) by design, since all other methods re-rank the lists generated by NeuMF. However, NeuMF is not fairness-aware and thus performs the worst in fairness metrics.

Table 1. Comparison of Algorithms' Performance.

		XING16					**CAREERBUILDER12**				
		NeuMF	DetGreedy	DetCons	DetRelax	iFair	NeuMF	DetGreedy	DetCons	DetRelax	iFair
Effectiv.	nDGG@10 $\uparrow$	**0.2137**	0.2051	0.1945	0.1945	0.1987	**0.080**	0.078	0.054	0.054	0.074
	$\mathscr{U}$@10 $\uparrow$	**1.0**	0.96	0.91	0.91	0.93	**1.0**	0.97	0.68	0.68	0.93
Calibr.	nDKL@10 $\downarrow$	0.52	0.31	0.33	0.33	**0.11**	0.57	0.33	0.33	0.11	**0.11**
	nDKL@10 $\downarrow$	0.52	0.26	0.33	0.33	**0.12**	0.57	0.23	0.33	0.33	**0.13**
	nDKL@10 $\downarrow$	1.32	0.72	0.92	0.92	**0.51**	1.48	0.67	0.92	0.92	**0.41**
Fairness	avg ℓ^f@10 $\downarrow$	2553.19	6472.55	0.05	718.40	**0.04**	426.0	547.45	0.04	412.86	**0.03**
	avg $d_{\mathbb{R}}$@10 $\downarrow$	0.21	0.10	0.16	0.16	**0.09**	0.58	0.16	0.34	0.24	**0.08**

In terms of fairness, iFair demonstrates significant improvement over all competitors. iFair outperforms others in the fairness metric nDKL on both datasets, achieves the lowest scores in nDKL@10, underlining its capacity to offer balanced and fair job recommendations. At the same time, iFair maintains high nDCG@10 ensuring that its enhanced fairness does not compromise the relevance of its suggestions.

In terms of individual fairness metrics, iFair significantly outperforms the other algorithms, as seen in the last two rows of Table 1. The main reason is that iFair re-ranks job postings based on the Kemeny's optimal aggregated list of attributes, which ensures that all job seekers will get jobs with same non-protected attributes (i.e., salary, discipline, region), which is a prerequisite for equitable treatment of job seekers with same qualifications.

5 Conclusion

In this paper, we introduced ε-individual fairness for job recommendations, providing a robust method to ensure equitable treatment of job seekers. Our iFair algorithm re-ranks job postings based on Kemeny's optimal ranking, which ensures fairness without significant loss of effectiveness. For future work, we plan to integrate multi-objective optimization for balancing fairness, accuracy, and explainability.

References

[1] 2012. *Kaggle Job Recommendation Challenge*. Kaggle. https://kaggle.com/competitions/job-recommendation

[2] Fabian Abel, András Benczúr, Daniel Kohlsdorf, Martha Larson, and Róbert Pálovics. 2016. Recsys challenge 2016: Job recommendations. In *Proceedings of the 10th ACM conference on recommender systems*. 425–426.

[3] Fabian Abel, Yashar Deldjoo, Mehdi Elahi, and Daniel Kohlsdorf. 2017. Recsys challenge 2017: Offline and online evaluation. In *Proceedings of the eleventh acm conference on recommender systems*. 372–373.

[4] Stephan Arndt, Carolyn Turvey, and Nancy C Andreasen. 1999. Correlating and predicting psychiatric symptom ratings: Spearmans r versus Kendalls tau correlation. *Journal of psychiatric research* 33, 2 (1999), 97–104.

[5] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*. 214–226.

[6] Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. 2019. Fairness-aware ranking in search & recommendation systems with application to linkedin talent search. In *Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining*. 2221–2231.

[7] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*. 173–182.

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

[9] Michael Kearns, Seth Neel, Aaron Roth, and Zhiwei Steven Wu. 2018. Preventing fairness gerrymandering: Auditing and learning for subgroup fairness. In *Proceedings of the 35th International Conference on Machine Learning (ICML 2018)* (Stockholm, Sweden). JMLR.org, 2569–2577.

[10] John G Kemeny. 1959. Mathematics without numbers. *Daedalus* 88, 4 (1959), 577–591.

[11] Yunqi Li, Hanxiong Chen, Shuyuan Xu, Yingqiang Ge, Juntao Tan, Shuchang Liu, and Yongfeng Zhang. 2023. Fairness in recommendation: Foundations, methods, and applications. *ACM Transactions on Intelligent Systems and Technology* 14, 5 (2023), 1–48.

[12] Yunqi Li, Michiharu Yamashita, H. Chen, Dongwon Lee, and Yongfeng Zhang. 2023. Fairness in Job Recommendation under Quantity Constraints. https://api.semanticscholar.org/CorpusID:259336772

[13] Kilian Markert, Adam Ahouzi, and Philipp Debus. 2022. Fairness in Regression – Analysing a Job Candidates Ranking System. In *INFORMATIK 2022*. https://doi.org/10.18420/inf2022_109

[14] Yoosof Mashayekhi, Bo Kang, Jefrey Lijffijt, and Tijl De Bie. 2023. ReCon: Reducing Congestion in Job Recommendation using Optimal Transport. In *Proceedings of the 17th ACM Conference on Recommender Systems*. 696–701.

[15] Anian Ruoss, Mislav Balunovic, Martin Fischer, and Martin Vechev. 2020. Learning Certified Individually Fair Representations. In *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*.

[16] Clara Rus, Jeffrey Luppes, Harrie Oosterhuis, and Gido H Schoenmacker. 2022. Closing the gender wage gap: Adversarial fairness in job recommendation. *arXiv preprint arXiv:2209.09592* (2022).

[17] Harald Steck. 2018. Calibrated recommendations. In *Proceedings of the 12th ACM Conference on Recommender Systems* (Vancouver, British Columbia, Canada) *(RecSys '18)*. Association for Computing Machinery, New York, NY, USA, 154–162. https://doi.org/10.1145/3240323.3240372

[18] Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. 2023. A survey on the fairness of recommender systems. *ACM Transactions on Information Systems* 41, 3 (2023), 1–43.