Incentives for Item Duplication under Fair Ranking Policies

Giorgio Maria Di Nunzio, Alessandro Fabris ✉, Gianmaria Silvello, and Gian Antonio Susto

University of Paqua, Padua, Italy
{dinunzio, fabrisal, silvello, sustogia}@dei.unipd.it

Abstract. Ranking is a fundamental operation in information access systems, to filter information and direct user attention towards items deemed most relevant to them. Due to position bias, items of similar relevance may receive significantly different exposure, raising fairness concerns for item providers and motivating recent research into fair ranking. While the area has progressed dramatically over recent years, no study to date has investigated the potential problem posed by duplicated items. Duplicates and near-duplicates are common in several domains, including marketplaces and document collections available to search engines. In this work, we study the behaviour of different fair ranking policies in the presence of duplicates, quantifying the extra-exposure gained by redundant items. We find that fairness-aware ranking policies may conflict with diversity, due to their potential to incentivize duplication more than policies solely focused on relevance. This fact poses a problem for system owners who, as a result of this incentive, may have to deal with increased redundancy, which is at odds with user satisfaction. Finally, we argue that this aspect represents a blind spot in the normative reasoning underlying common fair ranking metrics, as rewarding providers who duplicate their items with increased exposure seems unfair for the remaining providers.

Keywords: Algorithmic Fairness · Duplicates · Fair Ranking.

1 Introduction

Ranking is a central component in search engines, two-sided markets, recommender and match-making systems. These platforms act as intermediaries between providers and consumers of items of diverse nature, facilitating access to information, entertainment, accommodation, products, services, jobs and workers. The rank of an item in a result page is a strong predictor of the attention it will receive, as users devote most of their attention to the top positions in a list, and are less likely to view low-ranking items [15]. This *position bias* is at the root of fairness concerns for providers of ranked items, as comparably relevant results may receive remarkably different exposure. In the absence of countermeasures, unfair exposure can affect item providers on e-commerce websites, job-search

platforms and commercial search engines, such as Amazon sellers, Airbnb hosts, job candidates on LinkedIn and owners of contents ranked by Google [6,14].

Unfair exposure can compound and increase over time, as the same query, issued multiple times to a system, is met with the same ranking. Each time the query is processed by the system, items gain a fixed, potentially unfair, level of exposure; this is a severe problem with static ranking policies, which map relevance scores to rankings in a deterministic fashion. Non-static policies, on the other hand, can respond to identical queries with different rankings, and they are more suited to equitably distribute exposure among relevant items. In recent years, fair ranking policies and measures have been proposed, which consider repetitions of the same query and encourage rotation of relevant items in the top-ranked positions [5,6,10,21,23].

While these measures and approaches are surely a sign of solid progress in the area of fair ranking, in this paper we highlight a potential blind spot in their normative reasoning: duplicates. Item duplicates and near-duplicates are not uncommon in online domains such as e-commerce websites [2] and online document collections [13]. Anecdotal evidence for this phenomenon can be found in official forums for item providers of popular marketplaces[1][2][3] and in Figure 1. Based on the reasoning brought forth in recent works, requiring equal exposure for equally relevant items [6,10], two copies of the same item deserve more exposure (in sum) than a single copy. On the contrary, in some situations it is reasonable to postulate that multiple copies of an item deserve the same attention the item would be entitled to on its own, especially if the copies benefit the same provider.

The key contribution of this work is to analyze the tension between fairness and diversity when duplicates are not properly considered and factored into the fair ranking objective. More in detail, we show that, under different settings, common fair ranking policies reward duplicates more than static policies solely focused on relevance. We argue that this phenomenon is unfair to the providers of unique (non-duplicate) items and problematic for system owners and users as it introduces an incentive for redundancy. The rest of this paper is organized as follows. Section 2 introduces related work, covering fairness and diversity in ranking. Section 3 is the core of our study, where we formalize the problem and analyze the benefits obtained by providers through item duplication. We quantify the extra-attention earnt by duplicated items in a controlled setting, as we vary the relevance of the items available to a system, its ranking policy and the cost of duplication. Section 4 contains closing remarks and outlines directions for future work.

---

[1]https://community.withairbnb.com/t5/Hosting/Unfair-duplication-of-same-listing-to-gain-more-exposure/td-p/850319, all links accessed on 02-03-21

[2]https://community.withairbnb.com/t5/Help/Duplicate-photos-in-listings-and-terms-of-service/td-p/1081009

[3]https://sellercentral.amazon.com/forums/t/duplicate-search-results/445552
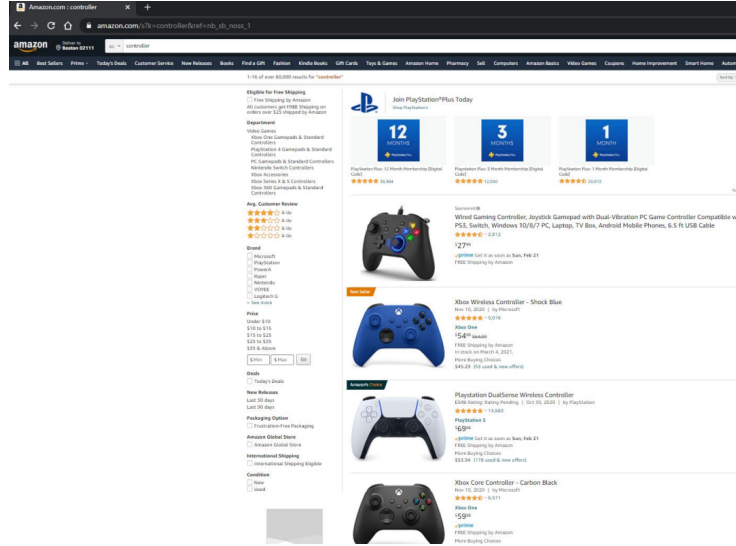
Fig. 1: Amazon result page for query `controller` issued on February 19, 2021 by a Boston-based unregistered user in *incognito* browser mode. Top 4 results comprise near-duplicates in positions 1 and 3 (0-based indexing).

2 Related Work

2.1 Fairness in Ranking

Fairness in ranking requires that the items ranked by a system receive a suitable share of exposure, so that the overall allocation of user attention is considered fair according to a criterion of choice [6,21]. Fair ranking criteria depend on the specific context and normative reasoning, often inheriting and adapting notions from the machine learning fairness literature, such as independence and separation [3].

Position bias and *task repetition* are peculiar aspects of many fair ranking problems. *Position bias* refers to the propensity of users of ranking systems to concentrate on the first positions in a list of ranked items, while devoting less attention to search results presented in lower positions [15]. Common measures of accuracy in ranking, such as Expected Reciprocal Rank (ERR) [8], hinge on this property: they reward rankings where items are presented in decreasing order of relevance, so that the positions which attract most user attention are occupied by the most relevant items. These are static ranking measures, which summarize the performance of a system with respect to an information need by modeling a single user-system interaction. However, users can issue the same query multiple times, requiring a search engine to repeatedly attend to the same

task (*task repetition*). Repeated queries, stemming from the same information need, are sometimes called *query impressions*.[4]

Recently, several measures of fairness in rankings have been proposed, which take into account the peculiarities of ranking problems [6,10,21]. These measures incorporate *position bias*, by suitably modeling user browsing behaviour when estimating item exposure, and consider *task repetition* by evaluating systems over multiple impressions of the same query, thus encouraging rotation of relevant items in top ranks. For example, *equity of amortized fairness* [6] considers cumulative attention and relevance of items over multiple query repetitions, and is defined as follows. "A sequence of rankings $\rho^1, \ldots, \rho^J$ offers equity of amortized attention if each subject receives cumulative attention proportional to her cumulative relevance", where the accumulation and amortization process are intended over multiple queries and impressions.

Depending on its amortization policy, measures of fairness in rankings can be (i) *cross-query* or (ii) *within-query*. (i) Cross-query measures are aimed at matching cumulative attention and relevance across different information needs [6]; this approach has the advantage of naturally weighing information needs based on their frequency and to enforce fairness over a realistic query load. On the downside, these fairness measures may end up rewarding systems that display irrelevant items in high-ranking positions. (ii) Within-query measures, on the other hand, enforce fairness over impressions of the same query [10]; this amortization policy results in one measurement for each information need and does not run the risk of rewarding systems that compensate an item's exposure across different information needs, which may result in balancing false negatives (missed exposure when relevant) with false positives (undue exposure when irrelevant).

Different approaches have been proposed to optimize ranking systems against a given fairness measure. Most of them make use of the *task repetition* property by employing stochastic ranking policies. These systems are non-deterministic since, given a set of estimated relevance scores, the resulting rankings are not necessarily fixed. A key advantage of stochastic ranking policies over deterministic ones lies in the finer granularity with which they can distribute exposure over multiple impressions of a query. Depending on whether they keep track of the unfairness accumulated by items, policies can be *stateless* or *stateful*. Stateless systems are based on drawing rankings from an ideal distribution independently [10,21,22]. This family of approaches can yield high-variance exposure for items, especially over few impressions, due to rankings being independent from one another. Moreover, they are not suitable to target cross-query measures as they would require estimating the future query load. Stateful solutions, on the other hand, keep track of the unfairness accumulated by items [6,18,23], and exploit it to build controllers or heuristics that can actively drive rankings toward solutions that increase the average cumulative fairness.

---

[4]`https://fair-trec.github.io/2020/doc/guidelines-2020.pdf`

2.2   Diversity in Ranking

The diversity of items in search results is important for users of products and services that feature a ranking component [7,9,11]. In the absence of ad-hoc approaches measuring and favouring diversity, very similar items may be present in a result page [7,19]. Duplicates and near-duplicates are the most severe example of redundant results [4,9].

Redundant items are present in collections for web search [4], with repercussions on search engines that need to handle duplicates at different stages of their life cycle, including training, testing and deployment [7,12,13]. Moreover, duplicate or near-duplicate listings of items can be present in online marketplaces, such as Airbnb and Amazon [2]. Multiple listings for the same item can derive from a legitimate need to highlight slightly different conditions under which a product or service is provided, or from an adversarial attempt to increase visibility on the platform through redundancy. Especially in the latter case, duplicates are viewed unfavourably by system owners for their negative impact on user experience [1].

3   Duplicates and Fair Ranking

In this section, we illustrate the mechanism through which fairness in ranking may reward duplicates by granting multiple copies of the same item more exposure than it would obtain as a single item. Section 3.1 introduces the problem, explains its root cause and presents a basic model of duplication and its cost. Section 3.2 details the synthetic experimental setup, summarizing the key parameters of the problem and the values chosen for the present analysis. Results are reported in Section 3.3. The relevant notation is summarized in Table 1.

3.1   Rewarding Duplicates

If a ranking approach is purely based on the relevance of single items and unaware of their potential redundancy, two copies of the same item will receive more attention than a single copy. For example, let us consider an item u_i ranked in position $\rho_i^j = n$, receiving in turn a share of user attention a_i^j. If a copy $u_{\tilde{i}}$ is created, in the absence of a ranking signal that rewards diversity, the item pair will rank in positions $\rho_i^j = n$, $\rho_{\tilde{i}}^j = n + 1$ (or viceversa). As a result, under a non-singular attention model, the sum of attentions received by the item pair is greater than the original a_i^j.

The above consideration holds true regardless of notions of fairness in rankings. However, in the presence of fairness constraints, there may be a further advantage for duplicates. For example, under *equity of amortized fairness* [6], which requires cumulative exposure of items proportional to their cumulative relevance, two items of identical relevance deserve twice as much exposure as a single item.

In reality, there are several factors that make duplication "expensive" in terms of the features commonly exploited by systems for item retrieval and

Table 1: Notation employed in this work.

Symbol	Meaning
u_i	items to be ranked by system, $i \in \{1, \ldots, I\}$
$u_{\tilde{i}}$	duplicate of item u_i
q_j	query impressions issued to system, $j \in \{1, \ldots, J\}$
ρ^j	ranking of items in response to query q_j
ρ_i^j	rank of item u_i in ranking ρ^j
π	a ranking policy
ρ_π	$\{\rho^1, \ldots, \rho^J\}$ sequence of rankings obtained via policy π
$u(\rho_\pi)$	utility function rewarding ranking sequence based on user satisfaction
a_i^j	attention received by item u_i in ranking ρ^j
r_i^j	relevance of item u_i for the information need expressed by q_j
k	cost of duplication, such that $r_{\tilde{i}}^j = k r_i^j$, $k \in (0,1)$
$\delta_{i+1,i}^j$	difference in relevance for adjacently ranked items (simplified to δ)
A_i	$\sum_{j=1}^{J} a_i^j$, i.e. cumulative attention received by item u_i
R_i	$\sum_{j=1}^{J} r_i^j$, i.e. cumulative relevance of item u_i over queries $\{q_1, \ldots, q_J\}$
$f(A, R)$	fairness function, defining the ideal relationship between A_i and R_i

ranking. Firstly, some of these features, such as user ratings, stem from the interaction of users with items; if an item is duplicated, its interactions with users will be distributed across its copies, presumably reducing their relevance score and lowering their rank in result pages. Moreover, a ranking system may explicitly measure the diversity of retrieved items [9] and favour rankings with low redundancy accordingly [19]. Finally, some platforms forbid duplication, in the interest of user experience, and enforce this ban with algorithms for duplicate detection and suppression procedures [1].

Therefore, we treat duplication as an expensive procedure, with a negative impact on items' relevance scores. We assume that the cost of duplicating an item only affects the new copy, while leaving the relevance of the original copy intact. We model duplication cost as a multiplicative factor $k \in (0,1)$, reducing the relevance score of new copies of an item. In other words, if a copy $u_{\tilde{i}}$ of item u_i is created, then r_i remains constant, while $r_{\tilde{i}} = k r_i$. Richer models of duplication cost are surely possible (e.g. also reducing the relevance r_i of the original copy), and should be specialized depending on the application at hand.

3.2   Experimental Setup

We cast fair ranking as an unconstrained optimization problem over a ranking policy π, whose objective function is a linear combination of a utility measure and a fairness measure.

$$\mathcal{Q}(\pi) = \lambda u(\rho_\pi) + (1 - \lambda) f(R_i, A_i(\rho_\pi)), \quad \lambda \in (0,1) \tag{1}$$

Here $u(\cdot)$ is a function computing the utility of a sequence of rankings ρ_π for item consumers, produced by a policy π. For example, in an IR setting, $u(\cdot)$ is a proxy

for the information gained by users from ρ_π. We measure the utility of a ranking (for a single impression) via normalized ERR [8], where the normalization ensures that a ranking where items are perfectly sorted by relevance has utility equal to 1, regardless of the items available for ranking and their relevance. ERR is based on a cascade browsing model, where users view search results from top to bottom, with a probability of abandoning at each position which increases with rank and relevance of examined items.[5] The overall utility $u(\rho_\pi)$ is computed as the average utility over all impressions. More in general, utility can be broadly characterized as user satisfaction, potentially including notions of diversity in rankings [9].

The objective function $\mathcal{Q}$ also comprises a function $f(\cdot)$ which combines the cumulative relevance R_i and exposure A_i of items to compute the fairness of ranking policy π toward item providers. To this end, we follow [5] by requiring that items receive a share of cumulative exposure that matches their share of cumulative relevance. More precisely, let us define the attention and relevance accumulated by item u_i over J queries as $A_i = \sum_{j=1}^{J} a_i^j$, $R_i = \sum_{j=1}^{J} r_i^j$, and let us denote as $\bar{A}_i$ and $\bar{R}_i$ their normalized versions

$$\bar{A}_i = \frac{A_i}{\sum_{i=1}^{I} A_i}; \quad \bar{R}_i = \frac{R_i}{\sum_{i=1}^{I} R_i}. \tag{2}$$

Cumulative unfairness is then quantified by the ℓ_2 norm of vector

$$\bar{A} - \bar{R} = \left[\bar{A}_1 - \bar{R}_1, \ldots, \bar{A}_I - \bar{R}_I\right] \tag{3}$$

and fairness by its negation:

$$f(A, R) = -||\bar{A} - \bar{R}||_2. \tag{4}$$

To quantify the attention a_i^j received by an item ranked in position ρ_i^j we use, again, the browsing model of ERR [8], so that the same user model underlies the estimates of item exposure and user utility.

We adopt a *within-query* policy for fairness amortization and focus on a single information need. Five items $(u_0, u_1, u_2, u_3, u_4)$ of relevance $r = (r_0, r_1, r_2, r_3, r_4)$ compete for attention in rankings over multiple impression of the same query. While in practice systems are required to rank large sets of items, this reduced cardinality allows us to enumerate all solutions and compute the perfect greedy solution without resorting to approximations or heuristics. We consider three relevance distributions, corresponding to different availability of relevant items and, consequently, different targets of ideal exposure. Specifically, item relevance decreases linearly from $r_0 = 1$, for the most relevant item, to $r_4 = \min(r_i^j)$ for the least relevant one, so that $r = (1, 1-\delta, 1-2\delta, 1-3\delta, 1-4\delta)$. Three values are tested for parameter δ, namely $\delta = 0.25$ (large relevance difference), $\delta = 0.125$ (intermediate relevance difference) and $\delta = 0.05$ (small relevance difference).

---

[5]Following [5], the probability of user stopping at position p, after viewing items $\{u_1, \ldots, u_p\}$ is set to $P(\text{stop}|u_1, \ldots, u_p) = \gamma^{p-1} c r_p \prod_{i=1}^{p-1}(1 - c r_i), \quad c = 0.7, \gamma = 0.5$

To solve the optimization problem, we test a *stateful* policy that explicitly targets the objective function (Equation 1) by keeping track of the relevance and exposure accumulated by items up to some impression t. At step $t+1$, we exploit this information to compute the best greedy solution via enumeration. This ranking policy is compared against a *stateless* approach that exploits Placket-Luce (PL) sampling [17,20]. PL sampling is based on drawing the top-ranked item at random from a categorical distribution where the probability of drawing item u_i is equal to $r_i^j / \sum_{i=1}^{I} r_i^j$. The chosen item is then removed from the pool of candidate items, from which the second ranking item is drawn in the same way, based on ℓ_1-normalization of the relevance scores of remaining candidates. The procedure is repeated until all items are drawn.

To evaluate the effects of item duplication in systems where copies incur different relevance penalties, we let the cost of duplication take values $k = 1$ (free copy) and $k = 0.5$ (relevance of duplicate item is halved). For each combination of relevance distribution, summarized by parameter δ, and cost of duplication k, we test each policy in six different settings. In each setting a different item u_i is duplicated, and an additional setting accounts for a scenario without any duplicates.

3.3   Results

As a first step, we ensure that the stateful policy can effectively trade off relevance and fairness in the basic setting where no duplicates are present. In Figure 2, we evaluate the impact of parameter λ on the utility and unfairness of rankings produced by the stateful policy, where unfairness is defined as the negation of function $f(\cdot)$ in Equation 4. As a baseline, we test the PL-based policy π_{PL}, reporting median values for utility and unfairness over 1,000 repetitions, along with the 5th and 95th percentile. Each panel in Figure 2 corresponds to a different combination of relevance difference, parametrized by δ, and number of impressions J. The top row corresponds to a less frequent query ($J = 20$) and the bottom row to a more frequent one ($J = 100$). Panels on the left depict results for a large relevance difference ($\delta = 0.25$), middle panels correspond to an intermediate relevance difference ($\delta = 0.125$) and left panels to a small one ($\delta = 0.05$).

We find that, over large relevance differences (left panels), a value $\lambda \geq 0.3$ is required to approach zero unfairness, while, for small relevance differences (right panels), $\lambda = 0.1$ is sufficient. This is expected: as relevance becomes uniform across items, even a policy marginally focused on fairness ($\lambda = 0.1$) can bring about major improvements in the distribution of attention. Moreover, for a small relevance difference, the trade-off between fairness and utility is less severe, which is also expected. When items have a similar relevance, a policy can more easily grant them a share of exposure proportional to their share of relevance, while only suffering a minor loss in terms of utility. Furthermore, the unfairness of exposure brought about by a solution purely based on relevance ($\lambda = 0$) increases as the difference in relevance for the available items become smaller. This is a desirable property of unfairness function $u(\cdot)$. Indeed, if a small difference in relevance
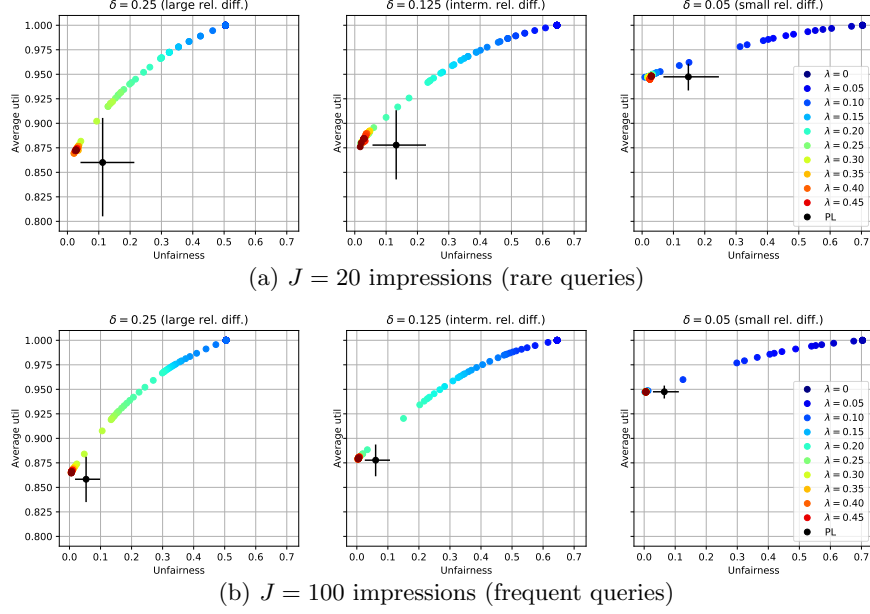
(a) $J = 20$ impressions (rare queries)

(b) $J = 100$ impressions (frequent queries)

Fig. 2: Unfairness (x axis) vs average utility (y axis) for *stateful* greedy solutions over different values of λ (downsampled and color-coded in legend). In black, summary of 1,000 Plackett-Luce repetitions, reporting median, 5th and 95th percentile for utility and unfairness. Each column corresponds to a different relevance profile for the available items, namely large relevance difference (left – $\delta = 0.25$), intermediate difference (middle – $\delta = 0.125$) and small difference (right – $\delta = 0.05$). Solutions with $\lambda > 0.5$ are omitted for better color-coding as they are all in a close neighbourhood of $\lambda = 0.5$.

($\frac{r_0}{r_4} = 1.25$) corresponds to a large difference in attention ($\frac{A_0}{A_4} > 1,000$), then the distribution of exposure stands out as particularly unfair for item u_4.

Although dominated by the stateful approach we just analyzed, the baseline PL-based policy π_{PL} consistently provides low-unfairness solutions. Unfairness is especially low for frequent queries, while for a rare query PL sampling is less likely to successfully distribute the cumulative exposure of items so that it matches their cumulative relevance. For frequent queries, the intervals spanned by the 5th and 95th percentile are narrower, signalling lower variance. PL sampling has been found to be a good baseline approach to obtain a fair policy from relevance scores under top-heavy browsing models such as the one underlying ERR [10]. Overall, our experiment confirms this finding.

To study the interaction between item duplication and equity of ranking, we firstly concentrate on a stateful greedy solution with $\lambda = 0.5$, representing a policy strongly focused on fairness. In Figure 3, solid lines represent the distribution of cumulative attention A_i for each item u_i under a fairness-aware policy
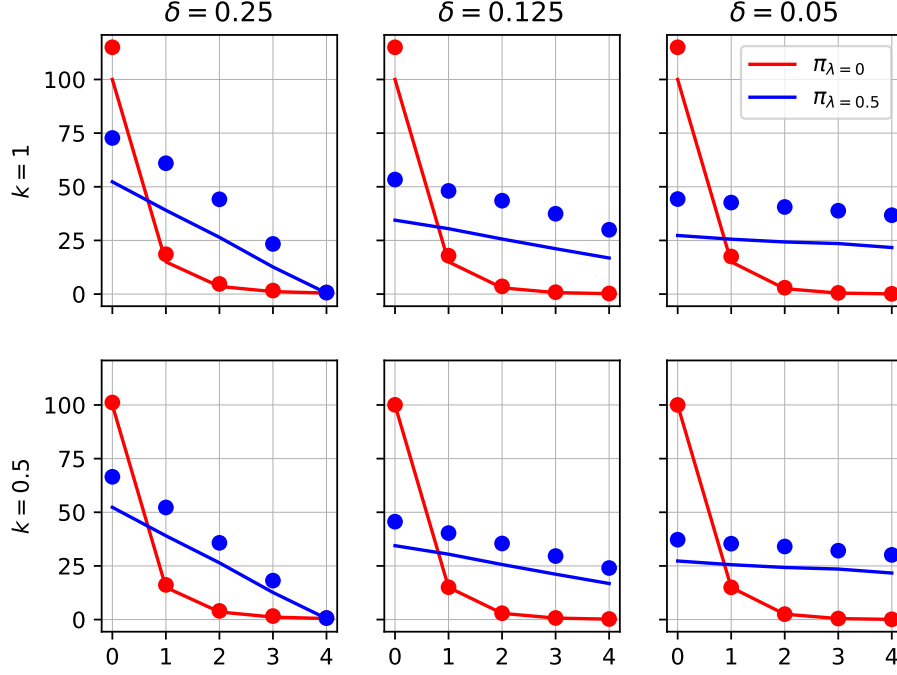
Fig. 3: Solid lines represent the attention A_i accumulated by each item (y axis) under a fairness-aware policy $\pi_{\lambda=0.5}$ (blue) or a policy solely focused on (ERR-based) utility $\pi_{\lambda=0}$ (red) in the absence of duplicates, over $J = 100$ impressions of the same query. Item indices $i \in \{0, \ldots, 4\}$ vary along the x axis. Round markers summarize the extra-attention one item would obtain if duplicated. Each column corresponds to a different relevance profile for the available items, namely large relevance difference (left – $\delta = 0.25$), intermediate difference (middle – $\delta = 0.125$) and small difference (right – $\delta = 0.05$). Each row corresponds to a different relevance multiplier for duplicates, namely $k = 1$ (top) and $k = 0.5$ (bottom).

($\lambda = 0.5$ - blue) and a policy solely focused on utility ($\lambda = 0$ - red), in the absence of duplicates. The query of interest is repeated for $J = 100$ impressions. Each column corresponds to a different value of δ, which determines large difference (left panels), intermediate difference (middle panels) and small difference (right panels) in relevance for items u_i. Interestingly, cumulative attention under policy $\pi_{\lambda=0.5}$ ends up resembling the distribution of relevance r_i^j for items u_i, i.e. a linear distribution with variable steepness. Policy $\pi_{\lambda=0}$, on the other hand, is not affected by the distribution of relevance.

Each round marker at position $x = i$ represents the sum of the attentions $(A_i + A_{\tilde{i}})$ received by item u_i and its copy $u_{\tilde{i}}$, if said item was duplicated (while remaining items are not). In other words, compared against solid lines of the same color, round markers summarize the extra-attention one item would ob-
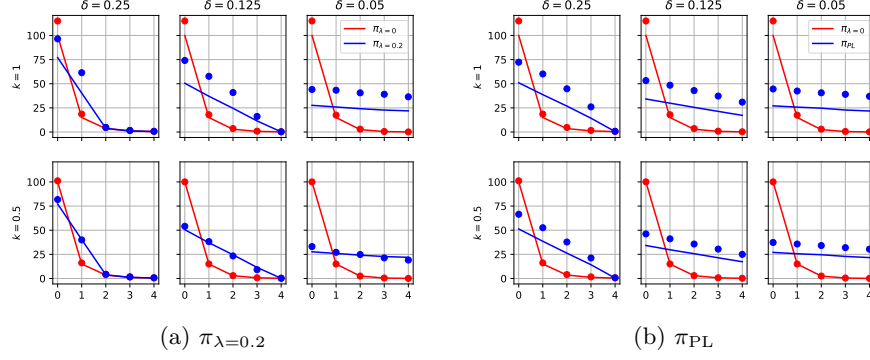
Fig. 4: Same analysis as Figure 3 with different ranking policies.

tain if duplicated. Different rows in Figure 3 correspond to a different relevance multiplier for duplicates, namely $k = 1$ (top) and $k = 0.5$ (bottom).

For every combination of parameters k and δ considered and for each item u_i, duplicates are always rewarded more under a fairness-aware policy $\pi_{\lambda=0.5}$ than under a policy solely focused on relevance $\pi_{\lambda=0}$. This finding suggests that fairness in rankings may be gamed by providers who duplicate their items. Moreover, in the presence of duplicates or near-duplicates, fairness of rankings may be at odds with diversity. Duplicated items, especially top-scoring ones, end up obtaining a significant amount of extra-attention. In turn, this may incentivize item providers to duplicate their listings. If redundancy in candidate items increases, it becomes harder for a ranking system to achieve diverse rankings, with potential repercussions on user satisfaction [11] and perception [16]. As expected, however, the benefits of duplication become smaller as its cost increases (bottom panels).

Figure 4a summarizes the same analysis for $\lambda = 0.2$, which corresponds to a more balanced ranking policy. In general, policy $\pi_{\lambda=0.2}$ is more similar to $\pi_{\lambda=0}$, i.e. it is more focused on relevance and rewards duplicates less than $\pi_{\lambda=0.5}$. The most relevant items still obtain a sizeable benefit from duplication, especially when the copying process does not affect item relevance (top panels). Finally, we evaluate the extent to which a policy based on PL sampling rewards duplicates. Figure 4b reports the extra-attention obtained by duplicated items under π_{PL}. These results are similar to those obtained under policy $\pi_{\lambda=0.5}$ in Figure 3, showing that duplicates are likely to receive a sizeable extra-exposure also under the stateless PL-based policy. This finding is not surprising given that, in order for $\pi_{\lambda=0.5}$ and π_{PL} to achieve similarly low unfairness for frequent queries (Figure 2b), they must distribute item exposure in a similar fashion.

4   Conclusions

In this work we have shown that duplicates are a potential blind spot in the normative reasoning underlying common fair ranking criteria. On one hand, fairness-aware ranking policies, both stateful and stateless, may be at odds with diversity due to their potential to incentivize duplicates more than policies solely focused on relevance. This can be an issue for system owners, as diversity of search results is often associated with user satisfaction [11]. On the other hand, allowing providers who duplicate their items to benefit from extra-exposure seems unfair for the remaining providers. Finally, system users (item consumers) may end up being exposed to redundant items in low-diversity search results; this would be especially critical in situations where items convey opinion.

While technical solutions for near-duplicate detection and removal are certainly available [1,7], they may not always be viable, as nearly identical listings can be posted in accordance with system regulation, e.g. to stress slight differences in products. Control over near-duplicates is even weaker in web page collections indexed by search engines. Therefore, it is important to consider the entities and subjects who benefit from exposure of an item and factor them into the normative reasoning underlying a fair ranking objective. While in marketplaces beneficiaries of exposure are more easily identifiable, for document collections the situation is surely nuanced, including for instance the writer, publisher and subject of a document.

Future work should comprise a more detailed study, including cross-query measures, considering different user browsing models and richer models for duplication and its cost. Moreover, it will be interesting to systematically assess the relationship between provider-side fairness and diversity of search results in the presence of duplicates, and the extent to which these desirable objectives are in conflict with one another.

References

1. Amazon: Potential duplicates. `https://sellercentral.amazon.com/gp/help/external/G202105450`, accessed: 2020-12-06
2. Amazon: Split different products into different pages. `https://sellercentral.amazon.com/gp/help/external/help.html?itemID=201950610`, accessed: 2020-12-06
3. Barocas, S., Hardt, M., Narayanan, A.: Fairness and Machine Learning. fairmlbook.org (2019), `http://www.fairmlbook.org`
4. Bernstein, Y., Zobel, J.: Redundant documents and search effectiveness. In: Proc. of the 14th ACM International Conference on Information and Knowledge Management. p. 736–743. CIKM '05, ACM, New York, NY, USA (2005)
5. Biega, A.J., Diaz, F., Ekstrand, M.D., Kohlmeier, S.: Overview of the trec 2019 fair ranking track. In: The Twenty-Eighth Text REtrieval Conference (TREC 2019) Proceedings (2019)

6. Biega, A.J., Gummadi, K.P., Weikum, G.: Equity of attention: Amortizing individual fairness in rankings. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. p. 405–414. SIGIR '18, ACM, New York, NY, USA (2018)

7. Broder, A.Z.: Identifying and filtering near-duplicate documents. In: Annual Symposium on Combinatorial Pattern Matching. pp. 1–10. Springer (2000)

8. Chapelle, O., Metlzer, D., Zhang, Y., Grinspan, P.: Expected reciprocal rank for graded relevance. In: Proc. of the 18th ACM Conference on Information and Knowledge Management. p. 621–630. CIKM '09, ACM, New York, NY, USA (2009)

9. Clarke, C.L., Kolla, M., Cormack, G.V., Vechtomova, O., Ashkan, A., Büttcher, S., MacKinnon, I.: Novelty and diversity in information retrieval evaluation. In: Proc. of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 659–666. SIGIR '08, ACM, New York, NY, USA (2008)

10. Diaz, F., Mitra, B., Ekstrand, M.D., Biega, A.J., Carterette, B.: Evaluating stochastic rankings with expected exposure. In: Proc. of the 29th ACM International Conference on Information & Knowledge Management. p. 275–284. CIKM '20, ACM, New York, NY, USA (2020)

11. Ekstrand, M.D., Harper, F.M., Willemsen, M.C., Konstan, J.A.: User perception of differences in recommender algorithms. In: Proc. of the 8th ACM Conference on Recommender Systems. p. 161–168. RecSys '14, ACM, New York, NY, USA (2014)

12. Fröbe, M., Bevendorff, J., Reimer, J.H., Potthast, M., Hagen, M.: Sampling bias due to near-duplicates in learning to rank. In: Proc. of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 1997–2000. SIGIR '20, ACM, New York, NY, USA (2020)

13. Fröbe, M., Bittner, J.P., Potthast, M., Hagen, M.: The effect of content-equivalent near-duplicates on the evaluation of search engines. In: Jose, J.M., Yilmaz, E., Magalhães, J., Castells, P., Ferro, N., Silva, M.J., Martins, F. (eds.) Advances in Information Retrieval. pp. 12–19. Springer International Publishing, Cham (2020)

14. Geyik, S.C., Ambler, S., Kenthapadi, K.: Fairness-aware ranking in search & recommendation systems with application to linkedin talent search. In: Proc. of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. p. 2221–2231. KDD '19, ACM, New York, NY, USA (2019)

15. Joachims, T., Radlinski, F.: Search engines that learn from implicit feedback. Computer **40**(8), 34–40 (Aug 2007)

16. Kay, M., Matuszek, C., Munson, S.A.: Unequal representation and gender stereotypes in image search results for occupations. In: Proc. of the 33rd Annual ACM Conference on Human Factors in Computing Systems. p. 3819–3828. CHI '15, Association for Computing Machinery, New York, NY, USA (2015)

17. Luce, R.D.: Individual Choice Behavior. Wiley (1959)

18. Morik, M., Singh, A., Hong, J., Joachims, T.: Controlling fairness and bias in dynamic learning-to-rank. In: Proc. of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 429–438. SIGIR '20, ACM, New York, NY, USA (2020)

19. Nishimura, N., Tanahashi, K., Suganuma, K., Miyama, M.J., Ohzeki, M.: Item listing optimization for e-commerce websites based on diversity. Frontiers Comput. Sci. **1**, 2 (2019)

20. Plackett, R.L.: The analysis of permutations. Journal of the Royal Statistical Society: Series C (Applied Statistics) **24**(2), 193–202 (1975)

21. Singh, A., Joachims, T.: Fairness of exposure in rankings. In: Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. p. 2219–2228. KDD '18, ACM, New York, NY, USA (2018)
22. Singh, A., Joachims, T.: Policy learning for fairness in ranking. In: Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., Garnett, R. (eds.) Advances in Neural Information Processing Systems. vol. 32, pp. 5426–5436. Curran Associates, Inc. (2019)
23. Thonet, T., Renders, J.M.: Multi-grouping robust fair ranking. In: Proc. of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 2077–2080. SIGIR '20, ACM, New York, NY, USA (2020)