Fair Exposure of Documents in Information Retrieval: a Community Detection Approach

Adrian-Gabriel Chifu, Josiane Mothe, Md Zia Ullah

Fair Exposure of Documents in Information Retrieval: a Community Detection Approach

Adrian-Gabriel Chifu
Aix Marseille Univ, Université de Toulon, CNRS, LIS
Marseille, France
adrian.chifu@univ-amu.fr

Josiane Mothe
IRIT UMR5505 CNRS, INSPE, Univ. de Toulouse, Toulouse, France
Josiane.Mothe@irit.fr

Md Zia Ullah
IRIT UMR5505 CNRS, Toulouse, France
mdzia.ullah@irit.fr

ABSTRACT

While (mainly) designed to answer users' needs, search engines and recommendation systems do not necessarily guarantee the exposure of the data they store and index while it can be essential for information providers. A recent research direction so called "fair" exposure of documents tackles this problem in information retrieval. It has mainly been cast into a re-ranking problem with constraints and optimization functions. This paper presents the first steps toward a new framework for fair document exposure. This framework is based on document linking and document community detection; communities are used to rank the documents to be retrieved according to an information need. In addition to the first step of this new framework, we present its potential through both a toy example and a few illustrative examples from the 2019 TREC Fair Ranking Track data set.

KEYWORDS

Information systems, Information retrieval, Fair document exposure, Document network, Document communities, Document re-ranking

1 INTRODUCTION

While (mainly) designed to answer users' needs, search engines and recommendation systems (RS) do not necessarily guarantee the exposure of the data they store and index. For example, many algorithms are founded on both content and collaborative selection of items. They are likely to retrieve or recommend the items that other users have liked or seen on similar need topics. In the recommendation domain, this problem is known as the cold start problem where new items are unlikely to be recommended since no user clicked on them when just added [24, 26, 27]. In search engines, ranking algorithms also include information on past searches and users' preferences [15, 34] that favors the most popular documents to the detriment of others, new or less known, although potentially equally relevant. To solve this problem, algorithms both in Information Retrieval (IR) and RS have been developed and aim at diversifying the results. In IR, diversity is mainly considered in relation to ambiguity [9], where the problem is to provide the user documents that answer the various meanings of a word (e.g. Orange as a Telecom company, a French city, a color, or a fruit). In RS, many algorithms also diversify the results by adding items that would not be recommended otherwise [22, 26].

Another recent research direction is the so called "fair" exposure of documents. It has mainly been cast into a re-ranking problem with constraints and optimization functions [2, 29, 38]. In 2019, TREC also decided to tackle this problem in one of its track: the Fair Ranking Track. The rationale is that not only users are concerned by which information is retrieved, but also information producers who are certainly inclined to store their data on platforms that give them the best exposure to users. As defined by TREC organizers, the Fair Ranking Track considers scientific documents and the central goal is to provide fair exposure to different groups of authors. Fair exposure is certainly a much more complex problem than currently defined since when platforms disclose the algorithms they use, producers try to fit to be better exposed. However, this track as it is, corresponds to a first interesting step and is likely to result in important milestones related to IR transparency.

This paper presents the first steps toward a new framework for fair document exposure in information retrieval. This framework is based on information clustering and document linking that is then used to select the documents to be retrieved according to an information need. The rest of the paper is organized as follows: Section 2 includes some related work. Section 3 presents the first step of our method along with a toy example. Section 4 describes the preliminary results on some examples from the TREC Fair Ranking Track. Section 5 concludes this paper and presents our future work.

2 RELATED WORK

Fair algorithms. Designing fair algorithms has recently been emerged to tackle the underlying bias in the training data or algorithm itself. Chierichetti et al. [7] study the fair clustering algorithms (a protected class must have fair representation in different clusters) under the disparate impact [14]. The authors introduce fairlets that satisfy fair representation while maintaining the clustering objective. The authors then reduce the fair clustering problem into fairlet decomposition and use the classical clustering algorithm to obtain the balanced representation of protective class in different clusters. In further study, Chierichetti et al. [8] focus on the large class of fair algorithmic problems for optimization subject to matroid constraints such as cardinality (sets), connectivity (graph), or matching (subgraph). They found that matroid constraints are general enough to encode many different types of problems while relatively easy to optimize.

Document diversity. While not initially formulated as a fair exposure of data, diversity is nonetheless a way to provide document exposure.

https://fair-trec.github.io/

Diversification is the process of re-ranking the documents initially retrieved for the query or selected for recommendation, taking into account the coverage, importance, and novelty of the documents in an implicit or explicit approach to reduce the search bias [37, 40]. Result diversification is generally formulated as an optimization problem and diversification methods differ by how they implement the objective function.

Some methods consider the document content only such as Maximal Marginal Relevance (MMR) [5] which selects the best document incrementally by balancing the relevance with the query and topic novelty. Fusion diversification is another example of content-based methods where topic modeling is used to infer latent subtopics; the results of sub-topic rankers are then fused to obtain the final ranking [23].

Other methods use external cues which represent users' intention like, Xia et al. who proposed a diverse ranking approach based on continuous state Markov decision process to model the user perceived utility at different ranks [41]. Also considering users' perspective, Santos et al. proposed xQuAD, a greedy diversification algorithm that maximizes the coverage of explicitly mined query subtopics [36]. Wasilewski *et al.* proposed an intent-aware collaborative filtering based recommendation system where the system diversify the recommended items by integrating the relevance, user intentions, and item-aspect associations [40]. Abdollahpouri *et al.* introduced a method to diversify the recommended items by reducing the popularity bias and enhancing the novelty and coverage [1]. Küçüktunç [21] proposed an enhanced random-walk based diversification for citation recommendation using the citation graph to allow users to reach the old, fundamental and recent papers.

Fair document exposure. None of the above mentioned methods provides fair item exposure in their re-ranking step. Recently, different frameworks have been proposed for the formulation of fairness constraints on rankings in terms of exposure allocation [2, 38]. In [38], the authors aim to find rankings that would maximize the fairness with respect to various facets, by proposing a general framework that employs probabilistic rankings and linear programming.

In IR, the concept of fair exposure is seen as the average attention received by ranked items [13], as an optimization problem with fairness constraints [17], or as a diversity ranking problem [18]. Diaz *et al.* [13] argued that measuring and optimizing expected exposure metrics using randomization opens a new area for retrieval algorithms, while Gao and Shah [17] estimated the solution space and answered questions about the impact of fairness consideration in the system, or about the trade-off between various optimization strategies. Still, Gao and Shah [18] studied the top-k diversity fairness ranking in terms of statistical parity fairness (equal number of documents from different groups) and disparate impact fairness (unintentional discrimination to a group).

Some approaches consider the fair exposure as a re-ranking problem [25, 30]. Liu and Burke [25] generated the ranking list based on a linear combination trade-off between accuracy and provider coverage and argued that the algorithm can significantly promote fairness, however, with a slight loss in terms of accuracy. Natwar *et al.* [30] introduced the concept of representative diversity, which means the level of interest of the consumer in different categories. They noted that higher diversity and fairness would lead to increased user acceptance regarding the results.

From the evaluation point of view, the authors of [12] proposed an interpolation strategy to integrate relevance and fairness into a unique evaluation metric. They investigated how existing TREC corpora could be adapted for a fairness task. Castillo [6] reviewed the recent works on fairness and transparency in ranking.

In RS, Khenissi and Nasraoui modeled the user exposure in an iterative closed feedback loop and developed a de-biasing strategy to reduce the bias inherent in the recommendation system, such as [20]. Mehrotra *et al.* proposed a framework to evaluate the trade-off between the relevance of recommendation to the customers and fairness of representation of suppliers in a two-sided marketplace such as Airbnb or Spotify [29].

In TREC 2019 Fair Ranking Track, Mcdonald et al. [28] cast the fair ranking as a diversification problem and employed two diversification approaches based on xQuAD [36] by optimizing the authors influence in one run, and authors and venues influence in another run; the results show that both of the runs are effective in minimizing unfairness. Wang et al. [39] used the convolutional kernel-based neural ranking model (Conv-KNRM) to obtain the relevance oriented ranking and then applied the documents swapping for authors exposure. Bonart, another participant, used a basic learning-to-rank framework for solving the track, stressing that the main challenge was that multiple objectives (searchers' relevance and authors' group fairness) have to be optimized under the constraint of an unknown author-group-partition.

3 PROPOSED METHODOLOGY

While in this paper we are not providing a completed framework for fair information exposure, we are describing the first steps toward it. In this section, we present the main intuition that drove our proposal, the main principles as well as a toy example to illustrate the model.

Rationale of our model.

Rather than considering document ranking constraints, as in previous work [17, 38], in order to solve the issue of fair document exposure, our model relies on graph structure and community founding.

Fair document exposure can be considered on different views or considering different facets[38]. One of them is what we call fair *item* exposure. In that case, the items in the retrieved documents should expose as many items as possible within a small as possible document set. For example, if items are authors, then the objective is to expose as many authors as possible in the (relevant) retrieved documents. The second view is what we call fair *community* exposure. In that case, we want to provide a fair exposure of the different *communities* of documents. A document community is based on the links between documents as defined by their shared items. The main idea is that the top ranked documents should not belong to the same *document community* to ensure a fair exposure of the variety of documents in terms of their items (e.g., their authors, sources, affiliation, opinions).

Document communities are extracted from the *document network* which is composed of the initially retrieved documents as nodes. As such the ground of the model is similar to the popular page rank

algorithm [4]. However, rather than considering the in and out hyperlinks or explicit references between documents, document links are rather extracted from implicit links. The implicit links are based on meta-data that documents share (e.g., common authors). Such links can be weighted and eventually oriented, depending on what they represent. Indeed, there is a large variety of type of links that can be considered in between documents. For example, considering scientific publications, documents can be linked according to the authors they have in common: the more common authors in between two documents, the higher the weight between the two corresponding nodes in the network. Once the document network is built, the main document communities can easily be extracted using state of the art community detection algorithms such as Fast Greedy [10], Leading Eigenvector [32], Walktrap [33] or Infomap [35], for example.

The principle is then to define a document ranking that both merges the communities (top documents should belong to different communities) and ranks the documents within a given community while considering the topic-document relevance.

Using contextual networks & community detection. In general, the goal of community detection is to partition any network into communities to extract the subgroups of densely connected nodes [16, 19]. The elements from a given community are then considered as very close concerning the cue the edges represent.

The problem of fair exposure of documents is then cast to the problem of a fair exposure of each "community" of documents. Our model thus relies on a network where the nodes of the network correspond to documents, while the edges between nodes are a means to contextualize the type of exposure the model should give to documents. The edges and thus the document network are contextual and depend on the targeted type of document exposure. Indeed, documents can be linked in different ways, depending on the types of documents and metadata which are associated with them. The principle is also very close to the linked data principle. "The term *Linked Data* refers to a set of best practices for publishing and connecting structured data on the Web" [3]. As an example, linking documents according to the shared countries their authors belong to is a way to structure the document space and extract communities of documents that are authored by people from the same region of the world. As another example, linking the documents according to the venue where they have been published is another way to structure the document space. This document space representation has several advantages:

- First, it is very adaptive: the definition of the edges can depend on the type of documents or the type of exposure one wants to reveal;
- Second, different types of linking can be easily combined: it is possible to link documents both regarding the countries the authors belong to and at the same time, the venue where they have published;
- Third, well-known community detection algorithms can be easily applied; some are known to result in smaller but denser communities, while others yield larger communities [11, 31];
- Finally, the communities can be adaptive and can be defined on a document type-based, user-based or query-based manner to fit better the objectives or definition of the fairness of the exposure (e.g., changing the type(s) of links to consider between the retrieved documents or changing the community detection algorithms).

Problem formulation.

- Document communities:

D_k is the set of k documents $\{d_1, d_2, \cdots, d_k\}$. Each d_j becomes a node of the document network. An edge between two document nodes is typed and depends on the meta-data or items used. Let I_k^T be the set of items of type T extracted from D_k. The weight of the edge between d_i and d_j is defined as:

$$Weight\left(d_i, d_j\right) = \frac{\left|I_{d_i}^T \cap I_{d_j}^T\right|}{\left|I_{d_i}^T \cup I_{d_j}^T\right|} \tag{1}$$

- Fair exposure of items:

Let I be a set of items and d_j be a document. I_{d_j} is the set of items associated with a document d_j and $|I_{d_j}|$ is the number of those unique items. In the same way, $|I_{D_k}|$ is the number of unique items associated with D_k where D_k is the set of k documents $\{d_1, d_2, \cdots, d_k\}$.

$\mathcal{R}$ is the initial -assumed as non empty- list of N retrieved documents $\{d_1, d_2, \cdots, d_N\}$ ordered by document score, expressed as the probability of the document d_k at rank k to be relevant to the query q ($P(d_k|q)$).

$\mathcal{S}_{k-1}$ is the set of documents already selected with fair exposure to items at $(k-1)$-th iteration and initially empty ($\mathcal{S}_0 = \emptyset$ i.e., for k=1).

We want to optimize the document selection in an iterative way. d_k^* is the best fair document to be selected at the k-th iteration and is calculated as follows:

$$d_k^* = \underset{d_i \in \mathcal{R} \setminus \mathcal{S}_{k-1}}{\operatorname{argmax}} \; \gamma\, P(d_i|q) + (1-\gamma)\, E(d_i, \mathcal{S}_{k-1}) \tag{2}$$

where γ is a combining parameter and $\gamma \in [0, 1]$, $\mathcal{R} \setminus \mathcal{S}_{k-1}$ is the list of documents from $\mathcal{R}$ that has not been selected yet in $\mathcal{S}_{k-1}$ at the $(k-1)$-th iteration.

Then,

$$\mathcal{S}_k = \mathcal{S}_{k-1} \cup \left\{d_k^*\right\} \tag{3}$$

The function $E(d_i, S_{k-1})$ measures the exposure of items of the document d_i given the document set S_{k-1} and is defined as follows:

$$E(d_i, S_{k-1}) = \frac{\left|I_{d_i} \cup I_{S_{k-1}}\right|}{|I_{\mathcal{R}}|} \tag{4}$$

- Fair exposure of communities:

Let C be a set of communities detected from the graph of documents. Each community $c \in C$ contains a set of documents D_c. Like previously, I_{D_c} is the set of items in D_c and $|I_{D_c}|$ is the number of those items.

For fair exposure of communities, we want to optimize the document selection over the communities C. We define this selection in an iterative way. d_k^* is the document to be selected for fair exposure of communities at the k-th iteration and is calculated as a trade-off between document relevance and exposure as follows:

$$d_k^* = \underset{d_i \in \mathcal{R} \setminus \mathcal{S}_{k-1}}{\operatorname{argmax}} \; \gamma\, P(d_i|q) + (1-\gamma)\, E(d_i, \mathcal{S}_{k-1}|C) \tag{5}$$

Like in Equation 2, $P(d_i|q)$ is the probability of the document d_i to be relevant for q. The function $E(d_i, S_{k-1}|C)$ measures the exposure of items of the document d_i in the document set S_{k-1} considering the set of communities C and is defined as follows:

$$E(d_i, \mathcal{S}_{k-1}|C) = \sum_{c \in C} P(d_i, \mathcal{S}_{k-1}|c)\ P(c) \tag{6}$$

where $P(c)$ denotes the probability of the community c which is considered as uniform ($\frac{1}{|C|}$) of any community $c \in C$ and can be ignored for ranking objective.

In Equation 6, $P(d_i, \mathcal{S}_{k-1}|c)$ estimates the exposure of the items of document d_i given the exposure of the items of already selected documents $\mathcal{S}_{k-1}$ that belong to the community c and can be defined as follows:

$$P(d_i, \mathcal{S}_{k-1}|c) = P(d_i|c) \prod_{d_j \in \mathcal{S}_{k-1}} (1 - P(d_j|c)) \tag{7}$$

where $P(d_i|c) = \frac{|I_{d_i}|}{|I_{D_c}|} \times \mathbf{1}_{d_i}$ denotes the exposure of the items of document d ($|I_{d_i}|$) over the items of community c ($|I_{D_c}|$) if document d belongs to community c.

$\mathbf{1}_{d_i}$ makes the probability being 0 if the document does not belong to the community and is defined as follows:

$$\mathbf{1}_{d_i} = \begin{cases} 1 & \text{if } d_i \in c, \\ 0 & \text{otherwise.} \end{cases}$$

4 PRELIMINARY RESULTS

TREC Fair Ranking Track and data. In order to provide a preliminary evaluation and proof of concept, we used the Fair Ranking data as released by the track organizers.

In this shared task, participants are asked to re-rank query-based retrieved document sets in order to optimize both the relevance to the consumers and the fairness to the producers.

The data consists in:

- the Semantic Scholar (S2) Open Corpus from the Allen Institute for Artificial Intelligence which consists of 47 of 1GB data files. Most of the papers consist of the identifier, its DOI, title, abstract, authors along with their IDs, inbound, and outbound citations;
- the Query log provided by Semantic Scholar. For each query, the organizers provide the query ID, the query string, the query normalized frequency, as well as a list of documents to rerank. There are 652 training and 635 evaluation queries although not all contain a non-empty document list.

Preliminary results. In this first experiment, we do not provide exhaustive results on the entire collection but rather some illustrative examples.

To start with, we considered a few queries for which the set of associated documents contains at least 10 documents. We present two of them to illustrate the potential power of our model.

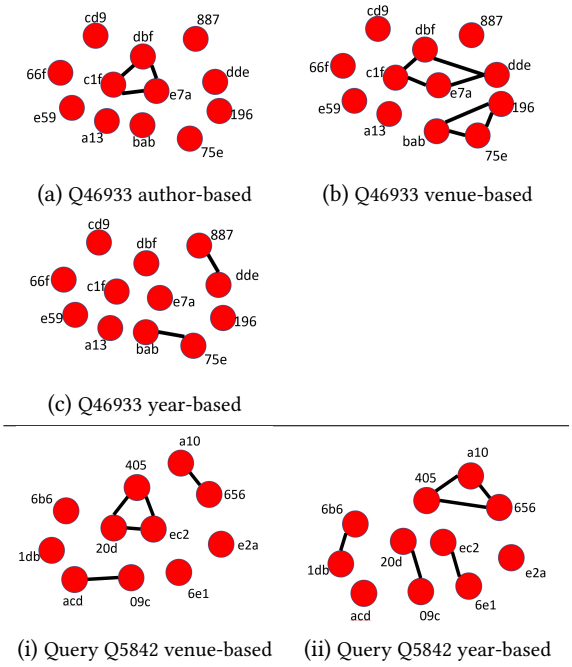
Query 38944 is one of them (detailed in the first row Table 1). For this query, the two first rows in Figure 1 display the document network and document communities based on (a) co-authorship

https://fair-trec.github.io/

http://api.semanticscholar.org/corpus/ (the 2019-01-31 version)

Query	Docs #	Authors #	Authors AVG	Entities #	Entities AVG	JN #
Q38944	12	64	5.58	122	14	7
Q5842	13	56	4.3	98	10.5	9

Table 1: Examples of Fair Ranking track queries along with some statistics. # stands for "number of different", AVG stands for "average number of ... per document". Docs corresponds to documents, while JN is for journals.



(a) Q46933 author-based (b) Q46933 venue-based

(c) Q46933 year-based

(i) Query Q5842 venue-based (ii) Query Q5842 year-based

Figure 1: Document communities extracted from the retrieved documents for two of the 2019 Fair Ranking Track queries. Document IDs have been shorten for readability (using the two first and the latest characters of the real IDs.

(b) venues where the papers were published (c) year of publication. In (a), two documents are linked if they have at least one author in common while in (b) (resp. (c)), two documents are linked if they have been published in the same venue (resp. on the same year). Other types of linking could have been made using the "In" or "Out" links, entities (key-words) for examples. For a fair exposure of documents associated to Query 38944, considering venues, the group "dbf", "c1f", "e7a" and "dde" for example should have a single of them being top ranked as the others are redundant regarding the venue. A fair exposure regarding the time line of the field could also be considered.

In a similar way, for Query Q5842 (second row in Table 1), we obtained the document communities as presented in Figure 1, when considering (i) the venues where the papers are published and (ii) the years of publication.

While this method does not include yet the final order of documents, it opens new ways of considering diversity or/and fair

exposure according to various criteria that can be either considered individually or combined. An interesting point of the model is that the criteria can evolve according to the needs.

5 CONCLUSION

In this paper, we paved the way for a new framework for fair document exposure. We presented the model.We also present some illustrative examples when considering the TREC Fair Ranking Track 2019 data.

In future work, first, we will develop the model in a formal way and complete the framework so that it includes a total document ordering. We will also evaluate it in a more general way by considering the entire 2019 TREC Fair Ranking Track the organizers provided as well as the track evaluation measures. As it has been mentioned by the organizers, in the 2019 collection a lot of queries have too few retrieved documents; a method such as ours make more sense when the list of documents is large enough. As a future work, we will also participate to the track in 2020 with our model.

REFERENCES

[1] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2019. Managing popularity bias in recommender systems with personalized re-ranking. In *The Thirty-Second International Flairs Conference*.

[2] Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. 2018. Equity of attention: Amortizing individual fairness in rankings. In *The 41st international acm sigir conference on research & development in information retrieval*. 405–414.

[3] Christian Bizer, Tom Heath, and Tim Berners-Lee. 2011. Linked data: The story so far. In *Semantic services, interoperability and web applications: emerging concepts*. 205–227.

[4] Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. (1998).

[5] Jaime Carbonell and Jade Goldstein. 1998. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proc. of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*. 335–336.

[6] Carlos Castillo. 2019. Fairness and transparency in ranking. In *ACM SIGIR Forum*, Vol. 52. 64–71.

[7] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. 2017. Fair clustering through fairlets. In *Advances in Neural Information Processing Systems*. 5029–5037.

[8] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvtiskii. 2019. Matroids, matchings, and fairness. In *The 22nd International Conference on Artificial Intelligence and Statistics*. 2212–2220.

[9] Charles LA Clarke, Maheedhar Kolla, Gordon V Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. 2008. Novelty and diversity in information retrieval evaluation. In *Proc. of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. 659–666.

[10] A. Clauset, M. Newman, and C. Moore. 2004. Finding community structure in very large networks. *Phys. Rev. E 70* 70, 066111 (December 2004). https://doi.org/10.1103/PhysRevE.70.066111

[11] Leon Danon, Albert Diaz-Guilera, Jordi Duch, and Alex Arenas. 2005. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment* 2005, 09 (2005), P09008.

[12] Anubrata Das and Matthew Lease. 2019. A Conceptual Framework for Evaluating Fairness in Search. *CoRR* abs/1907.09328 (2019). arXiv:1907.09328 http://arxiv.org/abs/1907.09328

[13] Fernando Diaz, Bhaskar Mitra, Michael D. Ekstrand, Asia J. Biega, and Ben Carterette. 2020. Evaluating Stochastic Rankings with Expected Exposure. arXiv:cs.IR/2004.13157

[14] Michael Feldman, Sorelle A Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. 2015. Certifying and removing disparate impact. In *proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 259–268.

[15] Larry Fitzpatrick and Mei Dent. 1997. Automatic Feedback Using Past Queries: Social Searching?. In *The 20th ACM SIGIR Conference*. ACM, 306–313.

[16] S. Fortunato. 2010. Community detection in graphs. *Physics Reports* 486 (2010), 75–174.

[17] Ruoyuan Gao and Chirag Shah. 2019. How Fair Can We Go: Detecting the Boundaries of Fairness Optimization in Information Retrieval. In *Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval* (Santa Clara, CA, USA) *(ICTIR '19)*. Association for Computing Machinery, New York, NY, USA, 229–236. https://doi.org/10.1145/3341981.3344215

[18] Ruoyuan Gao and Chirag Shah. 2020. Toward creating a fairer ranking in search engine results. *Information Processing and Management* 57, 1 (2020), 102138. https://doi.org/10.1016/j.ipm.2019.102138

[19] Mariam Haroutunian, Karen Mkhitaryan, and Josiane Mothe. 2018. f-Divergence measures for evaluation in community detection. In *W. on Collaborative Technologies and Data Science in Smart City Applications (CODASSCA)*. 137–144.

[20] Sami Khenissi and Olfa Nasraoui. 2020. Modeling and Counteracting Exposure Bias in Recommender Systems.

[21] Onur Küçüktunç, Erik Saule, Kamer Kaya, and Ümit V Çatalyürek. 2014. Diversifying citation recommendations. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 4 (2014), 1–21.

[22] Matevž Kunaver and Tomaž Požrl. 2017. Diversity in recommender systems–A survey. *Knowledge-Based Systems* 123 (2017), 154–162.

[23] Shangsong Liang, Zhaochun Ren, and Maarten de Rijke. 2014. Fusion helps diversification. In *Proc. of the 37th international ACM SIGIR conference on Research & development in information retrieval*. 303–312.

[24] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. 2014. Facing the cold start problem in recommender systems. *Expert Systems with Applications* 41, 4 (2014), 2065–2073.

[25] Weiwen Liu and Robin Burke. 2018. Personalizing Fairness-aware Re-ranking. *CoRR* abs/1809.02921 (2018). arXiv:1809.02921 http://arxiv.org/abs/1809.02921

[26] Jonathan Louëdec, Max Chevalier, Josiane Mothe, Aurélien Garivier, and Sébastien Gerchinovitz. 2015. A multiple-play bandit algorithm applied to recommender systems. In *The Twenty-Eighth International Flairs Conference*.

[27] Jie Lu, Dianshuang Wu, Mingsong Mao, Wei Wang, and Guangquan Zhang. 2015. Recommender system application developments: a survey. *Decision Support Systems* 74 (2015), 12–32.

[28] Graham McDonald, Thibaut Thonet, Iadh Ounis, Jean-Michel Renders, and Craig Macdonald. [n.d.]. University of Glasgow Terrier Team and Naver Labs Europe at TREC 2019 Fair Ranking Track. ([n. d.]).

[29] Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. 2018. Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In *Proc. of the 27th acm international conference on information and knowledge management*. 2243–2251.

[30] Natwar Modani, Deepali Jain, Ujjawal Soni, Gaurav Kumar Gupta, and Palak Agarwal. 2017. Fairness Aware Recommendations on Behance. In *Advances in Knowledge Discovery and Data Mining*, Jinho Kim, Kyuseok Shim, Longbing Cao, Jae-Gil Lee, Xuemin Lin, and Yang-Sae Moon (Eds.). Springer International Publishing, Cham, 144–155.

[31] Josiane Mothe, Karen Mkhitaryan, and Mariam Haroutunian. 2017. Community detection: Comparison of state of the art algorithms. In *2017 Computer Science and Information Technologies (CSIT)*. 125–129.

[32] M. Newman. 2006. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* 74, 036104 (September 2006). https://doi.org/10.1103/PhysRevE.74.036104

[33] P. Pons and M. Latapy. 2005. Computing Communities in Large Networks Using Random Walks. *Lecture Notes in Computer Science, Springer* 3733 (2005). https://doi.org/10.1007/11569596_31

[34] Antonio J Roa-Valverde and Miguel-Angel Sicilia. 2014. A survey of approaches for ranking on the web of data. *Information Retrieval* 17, 4 (2014), 295–325.

[35] M. Rosvall and C. Bergstrom. 2007. Maps of random walks on complex networks reveal community structure. *PNAS* 105, 4 (July 2007), 1118–1123. https://doi.org/10.1073/pnas.0706851105

[36] Rodrygo LT Santos, Craig Macdonald, and Iadh Ounis. 2010. Exploiting query reformulations for web search result diversification. In *Proc. of the 19th international conference on World wide web*. 881–890.

[37] Rodrygo LT Santos, Craig Macdonald, Iadh Ounis, et al. 2015. Search result diversification. *Foundations and Trends® in Information Retrieval* 9, 1 (2015), 1–90.

[38] Ashudeep Singh and Thorsten Joachims. 2018. Fairness of exposure in rankings. In *Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2219–2228.

[39] Meng Wang, Haopeng Zhang, Fuhuai Liang, Bin Feng, and Di Zhao. [n.d.]. ICT at TREC 2019: Fair Ranking Track. ([n. d.]).

[40] Jacek Wasilewski and Neil Hurley. 2018. Intent-Aware Item-Based Collaborative Filtering for Personalised Diversification. In *Proc. of the 26th Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, 81–89. https://doi.org/10.1145/3209219.3209234

[41] Long Xia, Jun Xu, Yanyan Lan, Jiafeng Guo, Wei Zeng, and Xueqi Cheng. 2017. Adapting Markov Decision Process for Search Result Diversification. In *Proc. of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery, 535–544. https://doi.org/10.1145/3077136.3080775