

UM ESTUDO DE TÉCNICAS DE JUSTIÇA EM  
ALGORITMOS DE RANKING NO CONTEXTO  
DE BUSCA ACADÊMICA

BRENO DE SOUSA MATOS

UM ESTUDO DE TÉCNICAS DE JUSTIÇA EM  
ALGORITMOS DE RANKING NO CONTEXTO  
DE BUSCA ACADÊMICA

Monografia científica apresentada como requisito para a conclusão da Graduação em Ciência da Computação da Universidade Federal de Minas Gerais.

ORIENTADOR: RODRYGO LUIS TEODORO SANTOS

Belo Horizonte

Março de 2021

© 2021, Breno de Sousa Matos.  
Todos os direitos reservados.

de Sousa Matos, Breno

D1234p Um estudo de técnicas de justiça em algoritmos de ranking no contexto de busca acadêmica / Breno de Sousa Matos. — Belo Horizonte, 2021  
xii, 17 f. : il. ; 29cm

Monografia — Universidade Federal de Minas Gerais

Orientador: Rodrygo Luis Teodoro Santos

**1. Introdução 2. Referencial Teórico**  
**3. Avaliação 4. Experimentos 5. Resultados**  
**6. Conclusão**

CDU 519.6\*82.10

*Para Marcos e Rosângela.*

# Agradecimentos

A minha família pelo apoio incondicional. Aos amigos que melhoram até os piores dias.

*“E vou viver as coisas novas, que também são boas  
O amor, humor das praças cheias de pessoas  
Agora eu quero tudo, tudo outra vez ”*  
(Belchior)

# Resumo

**Palavras-chave:** Fairness, Information Retrieval, Diversification.

Algoritmos de *ranking* são amplamente utilizados na atualidade: para apresentar produtos em lojas *on-line*, exibir páginas da *web* como resultados de consultas em buscadores, ou até sugerir perfis de trabalho com os quais se conectar em redes sociais profissionais. Entretanto, se não acompanhados de forma cautelosa, os resultados dos *rankings* gerados podem conter vieses, como de etnia ou orientação sexual. Nesse contexto, mitigar possíveis preconceitos motiva a área de justiça em algoritmos de *ranking*. Justiça, porém, é distinta de diversidade no contexto de recuperação de informação. Neste trabalho, é apresentado um estudo de técnicas de justiça em algoritmos de *ranking* no contexto de busca acadêmica, utilizando abordagem de diversificação de resultados.

# Abstract

**Palavras-chave:** Fairness, Information Retrieval, Learning to Rank.

Ranking algorithms are widely used nowadays: to display web pages as search engines' results, recommend products on online stores, or even to suggest which profiles to connect to on professional network services. However, if not closely monitored, these algorithms may produce biased rankings, based on features such as ethnicity or sexual orientation. Mitigating possible prejudices motivates the study of fair ranking algorithms. Fairness, however, is different from diversity in information retrieval. In this work, we present a study of fair ranking algorithms for an academic search task, applying a diversification approach.



# Lista de Figuras

5.1	Resultados utilizando divisão de grupos por nível de desenvolvimento econômico . . . . .	13
5.2	Resultados utilizando divisão de grupos h-index . . . . .	14

# Lista de Tabelas

# Sumário

Agradecimentos	v
Resumo	vii
Abstract	viii
Lista de Figuras	ix
Lista de Tabelas	x
<b>1 Introdução</b>	<b>1</b>
<b>2 Referencial Teórico</b>	<b>3</b>
2.1 Rankings Justos . . . . .	4
2.2 Diversidade e Justiça . . . . .	4
2.3 Algoritmos de Ranking Justos. . . . .	4
2.4 Métricas de Avaliação. . . . .	5
<b>3 Avaliação</b>	<b>7</b>
3.1 Definição da Tarefa . . . . .	7
3.2 Métricas de Avaliação . . . . .	7
3.2.1 Justiça . . . . .	8
3.2.2 Relevância . . . . .	9
3.2.3 Parâmetros da competição . . . . .	9
3.3 Descrição dos Dados . . . . .	9
<b>4 Experimentos</b>	<b>11</b>
4.1 Abordagem Proposta . . . . .	11
4.1.1 Maximal Marginal Relevance . . . . .	11
4.1.2 Baselines . . . . .	12

4.2	Definições de Grupos . . . . .	12
<b>5</b>	<b>Resultados</b>	<b>13</b>
5.1	Divisão por nível de desenvolvimento econômico . . . . .	13
5.2	Divisão por nível h-index . . . . .	14
<b>6</b>	<b>Conclusão</b>	<b>15</b>
	<b>Referências Bibliográficas</b>	<b>16</b>

# Capítulo 1

## Introdução

Algoritmos de *ranking* possuem grande importância no mundo moderno: estão presentes nos mais variados tipos de *websites*, como serviços de *streaming*, redes sociais e sites de compras online. Além disso, estão presentes nos buscadores da Internet, como Google<sup>1</sup> e Bing,<sup>2</sup> que facilitam a navegação online, trazendo à tona os resultados mais relevantes às buscas dos usuários.

Tradicionalmente, esses algoritmos têm como objetivo fornecer resultados de utilidade máxima para os usuários, ordenando, de forma decrescente, os itens de acordo com sua relevância inferida. Dessa forma, itens mais relevantes devem ser, idealmente, posicionados no começo da lista.

Entretanto, quando esses algoritmos são utilizados para gerar *rankings* de pessoas e entidades, a ordem em que essas aparecem nas listas tem grande impacto em suas vidas, podendo propagar estereótipos do mundo real.

Os impactos desses algoritmos são cada vez mais presentes na sociedade. Por serem parte integrante e ubíqua do mundo digital moderno, esses algoritmos não estão isentos de falhas e erros: um problema existente é a presença de viés nos *rankings* gerados, que podem variar desde penalizar notícias de veículos pequenos, no caso de *ranking* de notícias, até discriminar candidatos em redes sociais de trabalho, como LinkedIn,<sup>3</sup> baseado em etnia e gênero. Com o surgimento desses problemas, iniciaram-se estudos para gerar resultados úteis para os usuários e justos para com as entidades que estão sendo buscadas [6].

Justiça, nesse contexto, pode ser definida como gerar listas de objetos de interesse do usuário (documentos, imagens, perfis de trabalho, entre outros) de forma indiferente

---

<sup>1</sup><https://google.com>

<sup>2</sup><https://www.bing.com/>

<sup>3</sup><https://about.linkedin.com/>

a variáveis sensíveis presentes nos dados (como informações sobre gênero, etnia e idade).

O estudo de algoritmos de *ranking* existe há décadas, entretanto, os avanços no desenvolvimento de soluções com foco em justiça ainda são recentes, apesar de crescerem nos últimos anos [6], abrindo espaço para criação de competições de algoritmos de *ranking* justos, como a TREC 2019 - Fair Ranking Track.<sup>4</sup> O objetivo da competição é avaliar exposição justa para diferentes grupos de autores [2] no contexto de busca acadêmica de artigos, uma tarefa em que, dada uma consulta  $q$ , são retornadas listas de artigos ordenadas por sua relevância de forma decrescente. Especificamente, a competição é focada em *reranking* de artigos científicos, ou seja, os participantes deveriam reordenar listas de artigos associados a consultas.

Neste trabalho, propomos um estudo de abordagens de *ranking* utilizando o arcabouço da competição supracitada. Comparamos três abordagens distintas: a primeira realiza um *reranking* aleatório. A segunda utiliza a função BM25. Por fim, propomos uma abordagem que combina BM25 com MMR (Maximal Marginal Relevance), com o objetivo de avaliar o impacto de diversificação sobre justiça e qualidade dos *rankings* obtidos.

---

<sup>4</sup><https://fair-trec.github.io/2019/index.html>

# Capítulo 2

## Referencial Teórico

Recuperação de informação é a área da computação que lida com a representação, armazenamento, organização e acesso a itens de informação [1], que podem ser textos, imagens, vídeos, páginas da *Web*, entre outros.

Gerar *rankings* é uma parte central de várias tarefas de recuperação de informação (como recuperação de documentos e sistemas de recomendação) e envolve diversas áreas da computação, como aprendizado de máquina e processamento de linguagem natural. Para gerar *rankings*, é possível utilizar diferentes abordagens, das mais clássicas, como TF-IDF e BM25 [1], até mais robustas como Learning to Rank [9].

*Learning to rank* é a tarefa de construir modelos de *ranking* utilizando aprendizado de máquina, tipicamente de forma supervisionada, semi-supervisionada ou por reforço. Como exemplos de abordagens, é possível citar *ranking aggregation* [8] (agregação de rankings diferentes para gerar uma lista final mais relevante), *single feature boosting* (promoção de documentos com base no valor de um de seus atributos, como número de citações) e *multi-feature learning to rank* [9] (treinamento de modelos de *ranking* utilizando diversos atributos distintos). Entretanto, por se tratarem de modelagens que utilizam *machine learning*, sofrem de problemas clássicos da área, como algoritmos enviesados e injustos.

Em aprendizado de máquina, um algoritmo é considerado justo se seus resultados são indiferentes a variáveis sensíveis, como gênero, etnia e orientação sexual. No contexto atual de algoritmos de *ranking* e recuperação de informação, esse é um tema cada vez mais presente na literatura [5], pois estudos nos últimos anos [6] mostram que os resultados gerados por esses algoritmos, quando mitigação de viés é ignorada, podem resultar em pouca diversidade, promover polarização e estereótipos.

## 2.1 Rankings Justos

Justiça em *rankings* é um conceito que pode ser definido de diversas formas, dependendo da abordagem utilizada. Porém, em [5], é proposto que um *ranking* justo deve ter, no mínimo, as seguintes características:

- Presença suficiente de itens de grupos diferentes, em especial, grupos considerados protegidos (socialmente desfavorecidos). A ideia por trás é evitar discriminação estatística de grupos específicos, garantindo que todos estejam presentes em quantidade suficientemente expressiva.
- Tratamento consistente para itens similares, tal que não haja discriminação individualizada.
- Representação adequada de itens, em especial, itens de grupos protegidos. O objetivo é mitigar danos representacionais, como presença de estereótipos em sugestões de consultas em máquinas de busca online.

## 2.2 Diversidade e Justiça

Apesar de ser um conceito bastante utilizado em recuperação de informação, diversidade não significa o mesmo que justiça, sendo necessário fazer a diferenciação dos termos.

Ao elaborar um modelo de *ranking*, diversidade de resultados é muito importante, pois torna o modelo mais robusto e útil. Entretanto, diversidade se preocupa com a utilidade do item para o usuário, enquanto justiça foca na perspectiva do item retornado pelo modelo de *ranking*. Por exemplo, no contexto de busca de perfis em redes sociais profissionais, diversidade se preocupa em garantir que haja múltiplos perfis úteis ao usuário que realiza a consulta, enquanto justiça busca garantir que membros de grupos protegidos (como minorias étnicas) estejam efetivamente presentes nos *rankings*.

Justiça pode ser vista como um conceito assimétrico em recuperação de informação, pois parte do pressuposto de que existem diferentes grupos, com diferentes graus de representação, mas que devem ser incluídos nos *rankings* sem discriminação.

## 2.3 Algoritmos de Ranking Justos.

Por ser uma área em crescimento, derivada da Recuperação de Informação tradicional, justiça em *ranking* possui suas próprias abordagens e métricas de avaliação. É



possível essas abordagens em três categorias: pré-processamento, pós-processamento e em-processamento.

As abordagens pré-processamento buscam reduzir o viés dos dados que serão utilizado para treino de algum modelo de *ranking*. Por exemplo, buscando garantir que grupos protegidos não sejam sub-representados no conjunto de treino.

Métodos do tipo em-processamento, geralmente, realizam a otimização de duas funções de perda, uma relativa à justiça e outra à relevância dos itens retornados para uma consulta qualquer. Um dos principais algoritmos desta classe é o DELTR [15], cuja função de perda otimiza, simultaneamente, duas funções:  $L$  e  $U$ .  $L$  captura o erro de treino, comparando, para uma consulta  $q$ , os *rankings* gerados pelo modelo e as saídas reais.  $U$ , por sua vez, captura o erro relacionado à exposição dos grupos protegidos nos *rankings* produzidos. Otimizando as duas funções, é possível obter *rankings* justos e com boa utilidade.

Por fim, existe o grupo de métodos pós-processamento, que buscam reordenar os itens de um *ranking* de acordo com restrições pré-definidas. Esse grupo de técnicas compreende a maioria das abordagens propostas na literatura. Uma das técnicas mais relevantes desse grupo é o algoritmo FA\*IR [14], que gera *rankings top-k* justos, priorizando a escolha de itens de grupos protegidos para as primeiras posições dos *rankings*. Também é possível citar a abordagem proposta em [6], em que o problema de obter *rankings* justos para múltiplos grupos protegidos é modelado como uma programação inteira.

## 2.4 Métricas de Avaliação.

Com as diferentes abordagens supracitadas, é necessário também definir formas de avaliação que levem justiça em consideração. Na literatura, existem dois grupos principais de métodos avaliativos: baseados em exposição e baseados em probabilidade.

Métodos baseados em exposição tentam mensurar a atenção que o usuário dá a diferentes itens, seja estimando uma distribuição de probabilidade ou utilizando algum tipo de *feedback*, como quantidade de cliques ou mapa de calor da exibição do *ranking*. Em [11, 12], é apresentada a noção de justiça de exposição, em que há dois grupos (protegido e em vantagem) e a justiça ocorre quando, para uma consulta qualquer  $q$ , a razão entre a exposição e a utilidade dos dois grupos, considerando os itens retornados para  $q$ , é a mesma. Desse grupo, também é possível citar as métricas rND (*Normalized discounted difference*), rDR (*Normalized discounted ratio*) e rKL (*Normalized*

*discounted KL-divergence*), propostos em [13].

Por outro lado, métodos do grupo baseado em probabilidade pressupõem um *ranking* gerado por um processo aleatório, para, em seguida, medir a diferença entre os atributos esperados para esse *ranking* e os observados como saída do modelo que está sendo avaliado. Em [13], é proposto um método baseado em probabilidade que utiliza experimentos de Bernoulli [10] para averiguar a justiça de um *ranking*.

# Capítulo 3

## Avaliação

Neste trabalho, utilizamos o arcabouço de avaliação da competição TREC 2019 - Fair Ranking Track [2], cujo objetivo é fornecer exposição justa a diferentes grupos de autores em uma tarefa de *reranking* de artigos acadêmicos. As definições de grupos utilizadas na competição não foram reveladas até o prazo de submissão dos participantes ser finalizado, com o objetivo de valorizar soluções robustas que valorizem justiça e utilidade. Em seguida, propomos uma abordagem que aplica diversificação de resultados a ser comparada com *baselines* tradicionais de Recuperação de Informação.

O arcabouço de avaliação utilizado neste estudo é apresentado na Seção 3.2. A descrição das abordagens avaliadas é apresentada na Seção 4.1.

### 3.1 Definição da Tarefa

Avaliações de justiça devem ser feitas considerando sequências de *rankings*, ao invés de *rankings* individuais [3, 12]. Portanto, a competição define uma sequência  $Q$  de consultas acompanhadas de listas (de tamanho variável) de documentos a serem reordenados.

Para cada consulta  $q \in Q$  e conjunto de documentos associados  $D_q$ , participantes devem produzir *rankings* considerando os documentos de  $D_q$ . Desta forma, a saída do sistema é composto de uma sequência de *rankings*, que devem ser otimizados para serem justos e relevantes.

### 3.2 Métricas de Avaliação

Descritas em [2], as métricas de avaliação da competição são descritas nas subseções a seguir.

### 3.2.1 Justiça

#### 3.2.1.1 Exposição de autores em um ranking qualquer

Para calcular exposição, é utilizada a métrica *Expected Reciprocal Rank* [7]. Para um *ranking*  $\pi$ , a exposição de um autor  $a$  é definida como:

$$e_a^\pi = \sum_{i=1}^n \left[ \gamma^{i-1} \prod_{j=1}^{i-1} (1 - p(s|\pi_j)) \right] I(\pi_i \in D_a) \quad (3.1)$$

Sendo:

- $n$ : quantidade de documentos no *ranking*  $\pi$
- $D_a$ : documentos que incluem  $a$  como autor
- $\pi_i$ : documento na posição  $i$
- $\gamma$ : probabilidade de continuação
- $p(s|d)$ : probabilidade de parar dado que o usuário examinou  $d$

Sendo  $\Pi$  a sequência de todos os *rankings* gerados pelo sistema sob avaliação, a exposição amortizada de  $a$  é descrita como:

$$e_a = \sum_{\pi \in \Pi} e_a^\pi \quad (3.2)$$

#### 3.2.1.2 Relevância de autores em um ranking qualquer

A relevância de um autor para um *ranking*  $\pi$  é definida como:

$$r_a^\pi = \sum_{d \in D_a} p(s|d) \quad (3.3)$$

#### 3.2.1.3 Justiça de Grupos

Suponha que cada autor pertence exatamente a um grupo  $g \in G$ , sendo  $G$  o conjunto de todos os grupos de autores. Seja  $A_g$  o conjunto de todos os autores no grupo  $g$ . Podemos calcular a exposição ( $\varepsilon$ ) e a relevância ( $R$ ) de um grupo como:

$$\varepsilon_g = \frac{\sum_{a \in A_g} e_a}{\sum_{g' \in G} \sum_{a \in A_{g'}} e_a} \quad (3.4)$$

$$R_g = \frac{\sum_{a \in A_g} r_a}{\sum_{g' \in G} \sum_{a \in A_{g'}} r_a} \quad (3.5)$$

Grupos devem receber exposição proporcional a sua relevância. Para quantificar desvios da exposição ideal, calculamos a diferença entre  $\varepsilon$  e  $R$ :

$$\Delta_g = \varepsilon_g - R_g \quad (3.6)$$

A Equação 3.6 descreve uma métrica de exposição normalizada por relevância. Por fim, é possível computar o grau de justiça da exposição dos grupos calculando a norma  $l_2$ :

$$\Delta = \sqrt{\sum_{g \in G} \Delta_g^2} \quad (3.7)$$

### 3.2.2 Relevância

A qualidade de um *ranking* é calculado como a utilidade esperada, de forma análoga à apresentada na Equação 3.1:

$$u^\pi = \sum_{i=1}^n \left[ \gamma^{i-1} \prod_{j=1}^{i-1} (i - p(s|\pi_j)) \right] p(s|\pi_i) \quad (3.8)$$

O cálculo final de relevância é obtido calculando a média de todas as relevâncias:

$$U = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} u^\pi \quad (3.9)$$

### 3.2.3 Parâmetros da competição

Na avaliação final das submissões dos competidores, foram utilizados os parâmetros:

- $\lambda = 0.5$  (presente nas Equações 3.1 e 3.8)
- $p(s|d) = f(r_d) = 0.7 \cdot r_d$  (presente nas Equações 3.1 e 3.3). O valor de  $r_d$  é binário e inferido por informação de cliques de usuários.

## 3.3 Descrição dos Dados

Além de um arcabouço de avaliação, a competição conta com um conjunto de dados disponível para os participantes, composto de um *corpus* de artigos, exemplo de definições de grupos (diferente da utilizada na avaliação final), além de consultas com uma

---

lista de identificadores dos artigos a serem reordenados. O *corpus* é o Semantic Scholar Open Corpus<sup>1</sup> e possui diversos dados e metadados de artigos científicos, como título, *abstract* e lista de autores. As definições de grupo fornecidas como exemplo são baseadas na quantidade de artigos com no mínimo 10 citações que um autor possui. Por fim, cada consulta possui um identificador único, texto da consulta e lista de identificadores dos artigos a serem reordenados.

---

<sup>1</sup><http://api.semanticscholar.org/corpus/>

# Capítulo 4

## Experimentos

Este capítulo apresenta a descrição das três abordagens avaliadas neste estudo, assim como as definições de grupos necessárias para utilização da metodologia de avaliação apresentada na Seção 3.2.

### 4.1 Abordagem Proposta

Para estudar o impacto de diversificação na justiça e qualidade dos *rankings*, propomos uma abordagem que combina BM25 [1] e MMR (Maximal Marginal Relevance)[4] para reordenar as listas de documentos associadas a cada consulta.

#### 4.1.1 Maximal Marginal Relevance

MMR, ou Maximal Marginal Relevance [4], é uma técnica utilizada para reduzir a redundância e aumentar a diversidade dos resultados, mantendo a relevância destes para a consulta.

A formulação matemática do MMR pode ser observada na Equação 4.1:

$$MMR = \arg \max_{D_i \in R/S} [\lambda Sim_1(D_i, Q) - (1 - \lambda) \max_{D_j \in S} Sim_2(D_i, D_j)] \quad (4.1)$$

Os parâmetros da equação são:

- $D_i$ : Documentos da coleção  $C$
- $Q$ : Consulta (*Query*)
- $R$ : Documentos relevantes em  $C$
- $S$ : Conjunto de resultados atual

- $R/S$ : Documentos de  $R$  ainda não selecionados para  $S$
- $\lambda$ : Parâmetro ajustável que varia entre 0 e 1. Valores de  $\lambda$  próximos de 1 valorizam acurácia. Valores de  $\lambda$  próximos de 0 valorizam diversidade.

Para alcançar a combinação ótima de diversidade e acurácia, utilizamos  $\lambda = 0.5$  nos experimentos realizados.

### 4.1.2 Baselines

Como *baseline*, utilizamos duas abordagens. A primeira realiza *reranking* aleatório dos documentos. O segundo *baseline* utiliza apenas BM25 para reordenar os documentos, com o objetivo de avaliarmos o impacto da diversificação proposta na Seção 4.1 sobre os resultados finais.

## 4.2 Definições de Grupos

A avaliação apresentada no capítulo anterior leva em consideração definições arbitrárias de grupos de autores. Neste trabalho, foram utilizadas as seguintes definições:

- Divisão por nível de desenvolvimento econômico do país de origem dos autores.
- Divisão por *h-index*, medida de produtividade científica que representa, para um autor, o número de artigos com citações maiores ou iguais a  $h$ .

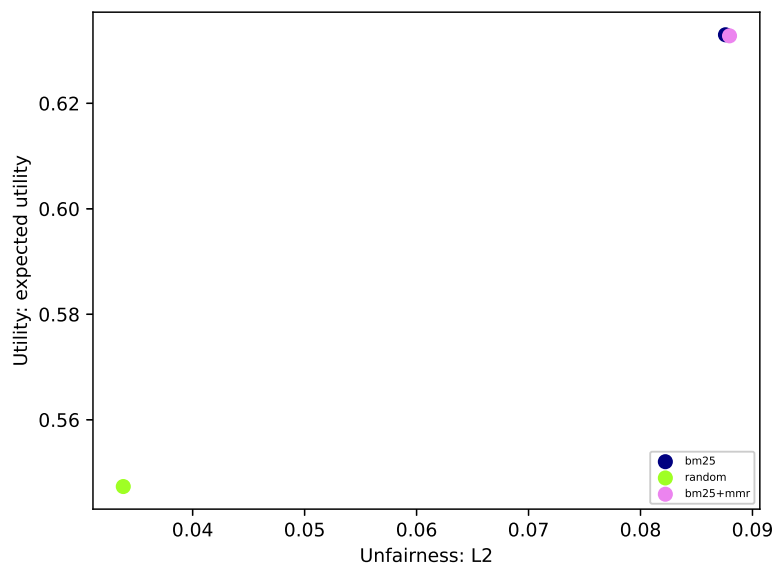


# Capítulo 5

## Resultados

Os resultados dos experimentos são apresentados nas seções abaixo, divididos nas duas definições de grupos de autores apresentadas na Seção 4.2.

### 5.1 Divisão por nível de desenvolvimento econômico



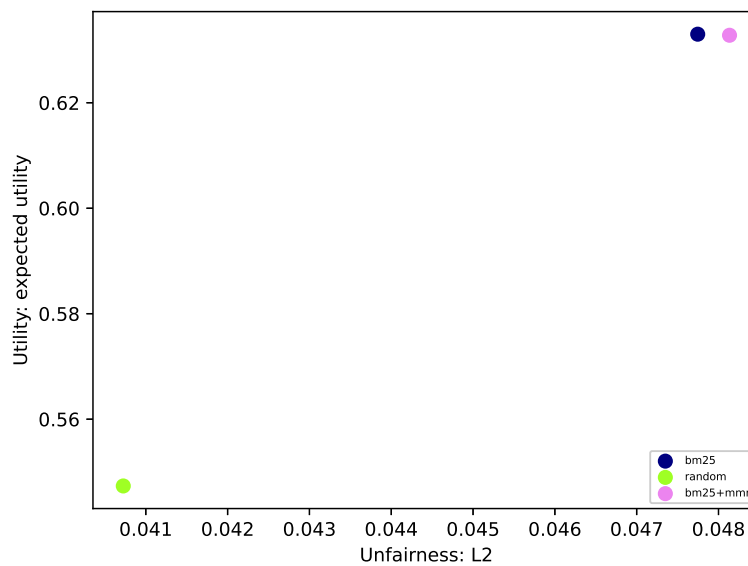
**Figura 5.1.** Resultados utilizando divisão de grupos por nível de desenvolvimento econômico

Na Figura 5.1, resultados finais de injustiça e utilidade para as três abordagens

utilizadas, considerando a definição de grupos que divide autores pelo nível de desenvolvimento econômico de seu país de origem.

A solução que combina BM25 e MMR obteve resultados similares a que usa apenas BM25, com diferença irrisória em injustiça e utilidade. Apesar de ser a menos útil, a abordagem aleatória gerou os *rankings* mais justos.

## 5.2 Divisão por nível h-index



**Figura 5.2.** Resultados utilizando divisão de grupos h-index

Na Figura 5.2, resultados considerando a definição de grupos que divide autores por h-index. A performance das três abordagens foi similar à apresentada na Seção 5.1, com a solução que utiliza apenas BM25 sendo a mais útil dentre as três. Em justiça, o *reranking* aleatório obteve a melhor performance.

# Capítulo 6

## Conclusão

A crescente utilização de algoritmos de *ranking* no dia-a-dia torna necessário reavaliar o papel que estas ferramentas desempenham na sociedade. Quando seu uso tradicional é atualizado e pessoas passam a ser um dos grandes focos desses algoritmos, é necessário tomar medidas para garantir que vieses sociais não sejam replicados por essas ferramentas. Para tentar solucionar esse problema, existem diversas abordagens de algoritmos de *ranking* justos propostas na literatura. Justiça, entretanto, é diferente de diversidade, cujo objetivo é ser capaz de satisfazer o maior número possível de usuários e, ao mesmo tempo, minimizar ou eliminar informações duplicadas e redundantes dos resultados de pesquisa.

Neste trabalho, foi apresentado um estudo de três abordagens distintas para *re-ranking* de artigos científicos. A primeira realiza um *reranking* aleatório. A segunda utiliza a função BM25. Por fim, foi proposta uma abordagem que combina BM25 com o método de diversificação MMR.

Cada abordagem foi avaliada de acordo com o arcabouço da competição *TREC 2019 - Fair Ranking Track*, que utiliza definições arbitrárias de grupos de autores para tentar avaliar a justiça e utilidade das submissões dos competidores.

A abordagem que combina BM25 e MMR não trouxe ganhos significativos para justiça ou utilidade, obtendo resultados similares à abordagem que utiliza apenas BM25 para todas as definições de grupo. Dentre as 3 abordagens, a mais justa foi a que implementa *reranking* aleatório, apesar de ser a menos útil dentre todas, considerando todas as definições de grupos.

Em trabalhos futuros, o objetivo é propor uma abordagem mais robusta, explorando mais técnicas de diversificação, como xQuAD, para complementar o estudo sobre o impacto de diversificação na justiça e qualidade de *rankings* gerados para busca acadêmica de artigos.

# Referências Bibliográficas

- [1] Baeza-Yates, R.; Ribeiro-Neto, B. et al. (1999). *Modern information retrieval*, volume 463.
- [2] Biega, A. J.; Diaz, F.; Ekstrand, M. D. & Kohlmeier, S. (2020). Overview of the trec 2019 fair ranking track.
- [3] Biega, A. J.; Gummadi, K. P. & Weikum, G. (2018). Equity of attention: Amortizing individual fairness in rankings. In *The 41st international acm sigir conference on research & development in information retrieval*, pp. 405--414.
- [4] Carbonell, J. & Goldstein, J. (1998). The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 335--336.
- [5] Castillo, C. (2019). Fairness and transparency in ranking. *SIGIR Forum*, 52(2):64--71. ISSN 0163-5840.
- [6] Celis, L. E.; Straszak, D. & Vishnoi, N. K. (2017). Ranking with fairness constraints.
- [7] Chapelle, O.; Metlzer, D.; Zhang, Y. & Grinspan, P. (2009). Expected reciprocal rank for graded relevance. In *Proceedings of the 18th ACM conference on Information and knowledge management*, pp. 621--630.
- [8] Kuhlman, C. & Rundensteiner, E. (2020). Rank aggregation algorithms for fair consensus. *Proceedings of the VLDB Endowment*, 13(12):2706--2719.
- [9] Liu, T.-Y. (2009). Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval*, 3(3):225--331. ISSN 1554-0669.
- [10] Papoulis, A. & Pillai, S. U. (2002). *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education.

- 
- [11] Singh, A. & Joachims, T. (2017). Equality of opportunity in rankings. In *Workshop on Prioritizing Online Content (WPOC) at NIPS*, p. 31.
- [12] Singh, A. & Joachims, T. (2018). Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2219--2228.
- [13] Yang, K. & Stoyanovich, J. (2017). Measuring fairness in ranked outputs. In *Proceedings of the 29th International Conference on Scientific and Statistical Database Management, SSDBM '17*, New York, NY, USA. Association for Computing Machinery.
- [14] Zehlike, M.; Bonchi, F.; Castillo, C.; Hajian, S.; Megahed, M. & Baeza-Yates, R. (2017). Fa\* ir: A fair top-k ranking algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pp. 1569--1578.
- [15] Zehlike, M. & Castillo, C. (2020). Reducing disparate exposure in ranking: A learning to rank approach. In *Proceedings of The Web Conference 2020*, pp. 2849--2855.