Quantification of the Impact of Popularity Bias in Multi-Stakeholder and time-aware environment

Francisco Guíñez
Pontificia Universidad Católica
Santiago, Chile
fhguinez@uc.cl

María Ignacia Sánchez
Pontificia Universidad Católica
Santiago, Chile
mcsanchez@uc.cl

Javier Ruiz
Pontificia Universidad Católica
Santiago, Chile
jiruiz2@uc.cl

ABSTRACT

In this paper we make an analysis of the impact of the popularity bias over time, for different stakeholders, considering different algorithms. The dataset Last.FM-1B was used, which has musical interactions of different users. The algorithms to be compared were ALS and BPR while Most Popular and random recommendations where used as baseline. More specifically, the main contributions of this paper consisted in providing a new way of measuring popularity in a variable way over time and analyzing which algorithms maintain a popularity bias over time that negatively impacts stakeholders. The analysis shows that the popularity bias is not static over time and that the ALS algorithm obtains more stable results with less unfairness than BPR among different groups of stakeholders.

KEYWORDS

Popularity bias over time, stakeholders, multi-stakeholders, time-aware

1 INTRODUCTION

Popularity bias is one of the main biases present in recommendation algorithms. This consists in the fact that the most popular items are over-recommended by the algorithms (in contrast to the recommendations that would really ensure their popularity), while items with less interactions are invisible. This generates a *snowball effect*, in which recommendations increase the popularity of already popular items and do not give the less popular ones a chance to emerge. In general, efforts to decrease this bias have been placed on those who consume these recommendations, i.e. the users. However, in platforms where there are multiple stakeholder groups, it is important to consider what impact each one is having, otherwise some of these groups may have incentives to stop using the platform. For example, if Netflix made only popular recommendations due to the bias of its algorithm, only the directors of the most popular movies would win. In turn, those who would be most interested in the platform would be those users who have more popular tastes. This would hurt those directors and users who create/consume less popular movies, which would eventually give them incentives to leave the platform.

Within this context, it is also important to consider the time in which an item is popular as a variable. That is, items that are popular today may not be popular tomorrow. Thus, users who at one time had popular tastes may eventually become users with less common tastes; and similarly, artists may go from being unpopular to becoming popular. While metrics have been constructed to measure the impact of popularity bias, they have not considered the time dimension of interactions.

Recommender Systems, December 2020,

Thus, the main contributions of this paper are:

- Propose a way to measure the popularity of the items but considering the dynamics of time. This will also allow the cataloguing of stakeholders in dynamic groups of popularity over time.
- Perform a popularity bias analysis over time, considering for this the state of the art metrics on the subject, which measure the impact of the popularity bias on users and providers.

For this, Last.fm's interaction dataset was used, where in this case the stakeholders turn out to be the users who consume albums and the artists who provide them.

2 RELATED WORK

If we examine the related work, it is possible to identify three main areas of research. Research with a focus on multiple stakeholders, analysis of time-aware recommendations, and the presence of bias and injustice in multiple contexts.

2.1 Multistakeholders

As described above, a stakeholder is defined as a person or group of people interested in a business who are directly affected by the decisions made in it. Thus, within the context of recommendation, considering multiple stakeholders implies making recommendations taking into account the differentiated interests between the different groups and considering the needs of each one [1]. The research in this subject arises after the academy has historically focused on the interests of the end user when proposing recommending models, but the complexity of these made it necessary to recognize that they serve multiple objectives [2], various models have been proposed in this direction in the last decade [3–6].

Specifically, an interesting multiple approach is presented in [7], where, based on a music dataset, the quality of recommendations for both end users and composing artists are considered, in addition to subdividing these groups according to their popularity in high, medium and low.

This field of research is related to the search for new and more diverse evaluation metrics for recommending systems, because of the need to evaluate groups as well as individuals. [8]

2.2 Time-aware recommendations

Over the years, this topic has been approached from two main perspectives, which are independent and complementary [9].

On the one hand, we have the notion that the popularity of an item may change over time, this occurs intrinsically in every context but is also influenced by external events. Examples of this perspective are the prediction of values on a timeline through before and after interactions [10] and the incorporation of time-dependent weights that decrease the importance of old items [11, 12].

On the other hand, there is the temporal evolution of the users, the way they interact with their environment is not constant and therefore their tastes or opinions vary depending on the temporal context [13, 14]. In particular, this has motivated research that proposes models with a time dimension. Among them we have, for example: models aware of the day of the week [15], models aware of the time of the day [16] and models where the importance of the interactions is penalized according to how many newer interactions with the same item the user has [17].

2.3 Bias and unfairness

Finally, it is important to mention that both the presence of multiple stakeholders and the time dimension are sources of bias and injustice when making recommendations [18].

When we recognize stakeholder groups differentiated by popularity, it is common to keep in mind a popularity bias that is unfair to groups describing unpopular segments [19]. There have been efforts to measure this injustice by means of knowledge graphs [20] and considering the efficiency of pareto [21], while other works have analyzed the implications that unfair models have on people [22, 23].

Likewise, not considering the notions of temporal evolution described in the Section 2.2 almost always implies the presence of a time bias that is impossible to measure [13]. Therefore, in the literature we see methods such as the one proposed in [9], which counteracts this bias by means of penalizers proportional to the measured bias.

3 DATASET

In order to be able to carry out the desired study, the dataset of interactions of Last.FM called 'LFM-1B' [24] was used, which has the reproductions made by users to different songs. This dataset was pre-processed in order to be able to adapt the amount of available data to the computational capacity that the team had by following these steps:

(1) First, we selected a random sample of 40,000,000 data from the initial dataset.
(2) The data belonging to the interactions carried out in 2011 were selected.
(3) Three groups were defined depending on the number of interactions in each album. The first group owned the albums with the top 20% of the accumulated interactions. The second group had the albums with the lowest 20% of the accumulated interactions. Finally, the third group consisted of the rest of the albums.
(4) One sixteenth of each group was selected at random..
(5) Only the interactions belonging to these new groups were maintained.
(6) Then, one half of the dataset was used for training of the recommending systems and the other half for testing and parameter analysis.
(7) Finally, the resulting dataset was divided into 6 parts, each of which had interactions up to a certain point in time. The first part had interactions occurring until the end of the sixth month of 2011, the second until the end of the seventh month of 2011, the third until the end of the eighth month of 2011, and so on.

After the pre-processing, a dataset with 177,854 interactions between a user and a specific album was obtained. The matrix of interactions associated with the dataset can be seen in the Table 1.

Number of users	32,353	**Average of albums per artist**	1,52
Number of artists (suppliers)	26,860	**Average of albums per user**	1,26
Number of albums (items)	40,856	**Average of users per album**	0.79
Sparsity artists-albums	99.996%	**Average of users per artist**	1.2
Sparsity users-albums	99.99%	**Average of artists per user**	0.83
Sparsity users-artists	99.986%	**Average of artists per album**	0.66

Table 1: Summary of the matrix of interactions

4 PROPOSED METHOD

Intuitively, when we talk about the popularity of a song for example, we consider that it becomes popular when suddenly many users start listening to it. Thus, most works in the area consider the number of interactions of an item as a key factor to define its popularity. In other words, the popularity of an item is measured by the amount of attention it receives.

On the other hand, from a mathematical perspective, when we are facing a continuous change in time it is natural to model the problem as a derivative, understood as the ratio of instantaneous change of a variable [25]. Common examples of the application of the derivative concept are the change of motion in time, which translates into speed, and the change of speed in time, which translates into acceleration.

Having said this and considering the state of the art, given that the objective of this work is to measure the popularity bias considering the existence of a time variable, a new way of measuring the popularity of an item is proposed, which interprets popularity as the ratio of instantaneous change related to the number of interactions that an item achieves in time. Having this popularity value for each item, we will proceed to classify the set of items into three subgroups of items: populars ($\mathcal{H}^t$), moderately popular ($\mathcal{M}^t$) and of low popularity ($\mathcal{T}^t$). A similar procedure will also be carried out to group users and artists as stakeholders.

4.1 Time popularity metrics

We will define N_i^t as the number of interactions achieved by the album i until a period of time t. For example, for an album i, $N_i^{2018} = 1000$ will imply that until 2018 the album has achieved a number of 1000 reproductions. This variable is discrete, and easily obtainable through the dataset. In our case and as described in the section 3 ,$t \in \{7, \ldots, 12\}$ are monthly periods from July to December in 2011.

For this, there are many methods of approximation of functions with different degrees of accuracy. However, since the main focus of this project is an analysis at the bias level, a polynomial adjustment was chosen, which proves to be effective and fast without sacrificing too much accuracy. The importance of maintaining a low computational cost in this work is emphasized, since in certain cases it is necessary to adjust more than a million items. For this purpose, the library *statsmodels* was used [26], which contains what is needed to create a non-linear model based on dataframes in a simple way, and allows to perform a polynomial regression using least squares. Another benefit of this library is that it allows to adjust the degree d of the polynomial with which the regression is sought, which will allow to obtain coefficients to be able to model different variants of the function in case it is required. In this case, it was considered that $d = 3$ gave an error small enough for this context.

Once the coefficients of this regression are obtained, the function $N_i(t)$ is obtained that adjusts to the variable N_i^t with an error of ε_i, which will be used to build the form item popularity function:

$$\textsc{Pop}_i(t) = \frac{\partial}{\partial t} N_i(t) \tag{1}$$

This popularity function will allow you to calculate the popularity of an item at any instant of time. It should be noted that, depending on the dataset, other measures of popularity could be proposed. For example, instead of considering the number of interactions accumulated by an item N_i^t, you could consider the number of user interactions achieved by this item B_i^t. That is, for each user, a 1 is added to the variable B_i^t if an interaction was made with the item i until the period t. Other approaches could be to divide these variables by the number of total users, so as to capture the growth of the item within a specific context. All these variants were implemented and tested, but their results were similar to (1), so the latter was finally chosen given its low computational time compared to the others.

Sequential recommendations will be made, in particular, recommendations will be generated in each period; in order to collect the set points, the trainset associated to each time interval was used, where $t \in \{7, \ldots, T\}$, with T the month until which the interactions of the original dataset are collected.

4.2 Dynamic grouping of items and stakeholders

Once these popularity functions $\textsc{Pop}_i(t)$ for all i items have been obtained, the dynamic grouping of items and *stakeholders* (users and artists) is carried out. To make this grouping in the items (albums), as in [7], the Pareto Principle will be used [27]. Given a period of time t, the items are ordered from most to least popular according to $\textsc{Pop}_i(t)$. The first 20% of these items will represent the group $\mathcal{H}^t$. From 20% to 80% will belong to the group $\mathcal{M}^t$. The last 20% of the items will belong to the $\mathcal{T}^t$ group.

With regard to the grouping of users and artists, we will proceed to calculate the simple average of popularity of the albums that have been listened to or created, respectively. Then a similar procedure will be carried out for grouping, using the same percentages of cuts as for item grouping. Let $\mathrm{W}_u(t)$ be the average level of popularity heard by the user u over time t, and $\mathrm{P}_a(t)$ be the popularity of the artist k over time t, then:

$$\mathrm{W}_u(t) = \frac{\sum_{i \in E_u^t} \textsc{Pop}_i(t)}{|E_u^t|} \tag{2}$$

$$\mathrm{P}_a(t) = \frac{\sum_{i \in C_a^t} \textsc{Pop}_i(t)}{|C_a^t|} \tag{3}$$

Where E_u^t are the albums heard by the user u until time t, and C_a^t are the albums created by the artist a until time t. We will call the groups derived from this procedure as $\mathcal{U}_1^t$, $\mathcal{U}_2^t$, $\mathcal{U}_3^t$ for the users and $\mathcal{A}_1^t$, $\mathcal{A}_2^t$, $\mathcal{A}_3^t$ for the artists, where the highest rate represents the most popular group.

5 MEASURING THE UNFAIRNESS OF STAKEHOLDER RECOMMENDATIONS

In order to measure the impact of the recommendations on the different stakeholders, the Item Popularity Deviation (IPD), Supplier Popularity Deviation (SPD) and User Popularity Deviation (UPD) measures will be taken [1] proposed by [7], but adapted to a version that considers temporality. On the other hand, in order to measure the coverage of items that do not belong to the popular group, the well-known metrics of *Aggregate Diversity* (Agg-Div) and *Long Tail Coverage* (LC), in addition to the metric *Average Percentage of Long-tail Items* (APL) proposed by [7]. That said, the following variables are defined:

- ℓ_u^t: List of album recommendations delivered by an algorithm to a user u in time t.
- L^t: Set of recommended albums in the time period t.
- $V(i)$: Function that return the artist of the item i
- $\mathcal{U}^t = \{\mathcal{U}_1^t, \mathcal{U}_2^t, \mathcal{U}_3^t\}$: Set with the popularity groups of users, defined in a time t.
- U^t: Group of all users on the platform, up to the time t.
- $\mathcal{A}^t = \{\mathcal{A}_1^t, \mathcal{A}_2^t, \mathcal{A}_3^t\}$: Sets with the popularity groups of artists, defined in a time t.
- $\mathcal{I}^t = \{\mathcal{H}^t, \mathcal{M}^t, \mathcal{T}^t\}$: Set with the popularity groups of the items, defined in a time t.
- ρ_u^t: List of albums that user u listened before time t.
- n: Number of recommendations.

$$APL = \frac{1}{|U^t|} \sum_{u \in U^t} \frac{|\{i, i \in (\ell_u^t \cap (\mathcal{M}^t \cup \mathcal{T}^t))\}|}{|\ell_u^t|} \tag{4}$$

$$SPD = \frac{\sum_{s \in \mathcal{A}^t} |SPD(s)|}{|\mathcal{A}^t|} \tag{5}$$

[1] With respect to UPD, a small modification in the way it is calculated will be considered, but it follows the same idea proposed by [7].

$$UPD = \frac{\sum_{g \in \mathcal{U}^t} |UPD(g)|}{|\mathcal{U}^t|} \tag{6}$$

$$IPD = \frac{\sum_{c \in \mathcal{I}^t} |IPD(c)|}{|\mathcal{I}^t|} \tag{7}$$

Here, $SPD(s)$ and $IPD(c)$ are calculated as proposed in [7], with the difference that the popularity groups, instead of being defined statically by the number of interactions, were defined according to the proposed new popularity metric, which considers a time-varying subdivision. In addition, the variables associated with the recommendations given to a user and the interactions of a user are also considered in a variable way in time. On the other hand, for $UPD(g)$ a slight variation of the formula was considered with respect to what was proposed by [7], but it was decided to maintain the same idea proposed to calculate SPD and IPD and to average over the groups the subtraction between the proportion of recommendations achieved by a group of users and the proportion of interactions achieved by that same group, i.e:

$$UPD(g) = qu(g) - pu(g) \tag{8}$$

$$qu(g) = \frac{\sum_{u \in U^t} \sum_{j \in \ell_u^t} \mathbb{1}(V(j) \in s)}{n \times |U^t|} \tag{9}$$

$$qu(g) = \frac{\sum_{u \in U^t} \sum_{j \in \rho_u^t} \mathbb{1}(V(j) \in s)}{|\rho_u^t|} \tag{10}$$

For these last three metrics, lower values mean that there is a smaller average difference between the proportion of the recommended and the proportion of the actual interactions achieved per popularity group, so the algorithm would be fair to the different popularity groups.

6 PARAMETERS ANALYSIS

In order to find an ideal configuration for the recommendation algorithms and to enable them to do their job better, different combinations of values for their hyperparameters were tested.

The hyperparameters studied, both for ALS and BPR, were the following:

- Number of Factors: Defines the number of latent factors to be used when performing matrix factoring. It was varied between 50, 100 and 200.
- Regularization parameter: Defines the weighting associated with the regularization effect. It was varied between 0.01 and 0.1

In addition, the following hyperparameter was studied for BPR:

- Learning Rate: Defines the learning rate at which Stochastic Gradient Descent updates should be performed during training. It was varied between 0.001, 0.01 and 0.1.

The decision was made to make 10 recommendations per user, since a smaller number does not allow us to adequately analyze the capacity of the algorithms to recommend less popular items, since the most popular options monopolize all the recommendations. On the other hand, a higher number of recommendations would not be representative of the context to be studied, since very few users have listened to more than 10 different albums, adding noise to the metrics.

Then, for each specific hyper parameter configuration, MAP@10 and nDCG@10 were calculated for each monthly dataset. Finally, the average of the metrics for each set of parameters was obtained in order to select the one that delivers a higher value.

The results obtained for the metrics by each recommendation algorithm and the configuration chosen to make the final recommendations will be described in the following sections.

6.1 ALS

The results obtained for the average of the monthly datasets by configuration of hyperparameters for the ALS algorithm can be reviewed in Figure 1. The values for the chosen parameters were 50 latent factors and 0.1 as a regularization parameter, because with this configuration higher values were obtained in both MAP@10 and nDCG@10.

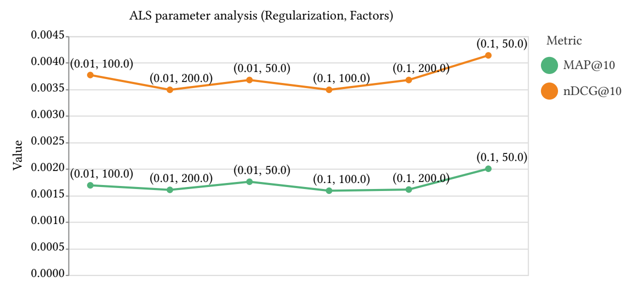

Figure 1: Results of MAP@10 and nDCG@10 on ALS

6.2 BPR

The results obtained for the average of the monthly datasets by configuration of hyperparameters for the BPR algorithm can be reviewed in Figure 2. The values for the chosen parameters were 200 latent factors, 0.01 as a regularization parameter and 0.01 as a learning rate parameter because with this configuration higher values were obtained in both MAP@10 and nDCG@10.

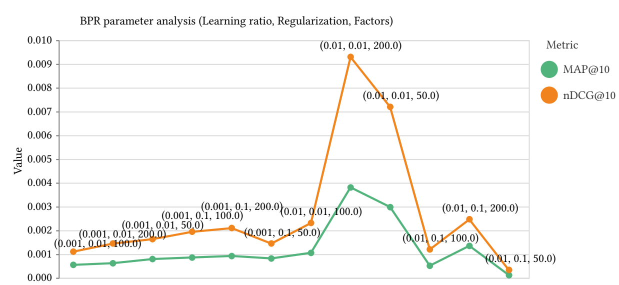

Figure 2: Results of MAP@10 and nDCG@10 on BPR

7 RESULTS

Once the best hyper-parameter configurations were defined, recommendations were made for each user in each monthly dataset, resulting in six groups of recommendations, one per month. Then, for these recommendations the unfairness was measured using the metrics described in Section 5. In addition, recommendations were made with the Most Popular and Random algorithms to have a baseline of values for the studied metrics, since the first algorithm should deliver the highest values of popularity bias, while the second illustrates a moderate unfairness. It is important to note that Random should not have completely fair values, since it negatively affects the recommendations of the most popular items. The results obtained can be reviewed in Figure 3.

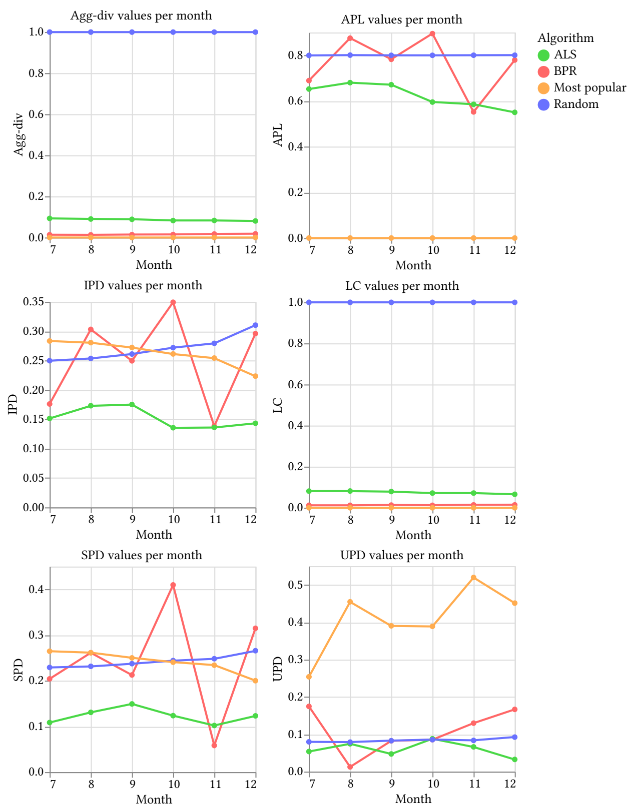

Figure 3: Unfairness metric results

As can be seen in Figure 3, except for Agg-div and LC which tend to be more linear, all the metrics have uneven variations as they move forward in time, which shows that the bias is not static, that is, it can both increase and decrease over time.

In addition, the ALS recommendation algorithm manages to overcome BPR in Agg-div, LC, APL and IPD metrics in all months. In addition, for the SPD metric, ALS manages to exceed BPR in the same way, except in month 11, where the latter algorithm achieves a decrease in the value of this metric. This means that ALS manages to give a higher priority to recommendations of less popular items compared to its counterpart, making it a less unfair algorithm for any group of stakeholders.

On the other hand, for BPR, IPD and SPD values above 0.2 were obtained during most months, while for ALS the values tend to remain between 0.05 and 0.2. These values demonstrate the existence of a popularity bias in both algorithms, and even for BPR, it tends to obtain results similar to Most Popular, the latter being an algorithm that has a clear popularity bias due to the way it operates to make recommendations.

It is interesting to note that the algorithm that manages to maintain better metrics over time with respect to Agg-Div, LC, IPD, APL and SPD is ALS. This means that, in general, ALS manages to give higher priority in recommendations to less popular items compared to BPR from a time perspective.

With respect to the inequality of recommendations among popularity groups, the UPD, SPD and IPD metrics show that this inequality is maintained more or less over time in the case of ALS, which means that the injustice of recommending is constantly present. On the other hand, these metrics have non-linear variations from one month to another for BPR, which shows that the injustice varies over time.

8 FUTURE WORK

As the objective of this work was to explore and make a first approach to the analysis of popularity bias in algorithms considering a temporal dimension, it would be interesting to address the following tasks in the future:

- Due to the computational demands of the number of albums, it was decided to abstract the problem to monthly time points, which translates into an adjustment error that could not be less. Thus, in the future the adjustment of each function $N_i(t)$ could be better and more accurate. For example, a greater number of points could be considered to obtain a better adjustment, or a adjustment error of $\varepsilon_i = \varepsilon \;\; \forall i$ could be set and the number of degrees of the polynomial d_i could be increased until achieving a value less than or equal to this error.
- Given that the focus was on temporal analysis of popularity bias, temporary recommendations were made using a basic algorithm. In the future, more sophisticated and less costly methods could be used in making them. Furthermore, it would be interesting to analyze how other algorithms besides ALS and BPR behave with respect to their temporal popularity bias.
- Analyze in detail the possibilities provided by the item popularity metric based on the derivative. By pre-computing the popularity of the items over time, it is possible to know at any time which texts belong to the groups $\mathcal{H}^t$, $\mathcal{M}^t$ and $\mathcal{T}^t$. With this it is possible to build an algorithm that allows rewarding recommendations from less popular groups. Naturally, in order to control this benefit, an adjustment parameter α would be required, which would allow weighting the importance given to this aspect. It would be interesting to analyze how such a metric behaves in contrast to conventional metrics to measure coverage and variability of recommendations.
- Currently, recommendability is defined as the number of interactions. This approach can change and interpret it as a reason for instantaneous change in time, generating the possibility of creating new metrics to measure popularity bias over time.

9 CONCLUSIONS

This paper demonstrated that the popularity bias is not static in time, this highlights the need to build time-conscious recommendations, since timeless analyses do not give a complete picture of the problem to be addressed.

A dataset of musical interactions was used as object of study, with which recommendations were made considering a sequence of instants in time. With the results obtained, it is concluded that ALS is less unfair than BPR in its recommendations.

The main difficulties arose from the high computational cost of estimating the popularity functions for each item, which was overcome by subsampling the information. It is concluded that this decision did not greatly affect the results of this research and that a similar analysis can be carried out in the future with better hardware.

REFERENCES

[1] M. Chelliah, Y. Zheng, and S. Sarkar, "Recommendation for multi-stakeholders and through neural review mining," in *Proceedings of the 28th ACM international conference on information and knowledge management*, pp. 2979–2981, 2019.

[2] H. Abdollahpouri, G. Adomavicius, R. Burke, I. Guy, D. Jannach, T. Kamishima, J. Krasnodebski, and L. Pizzato, "Beyond personalization: Research directions in multistakeholder recommendation," *arXiv preprint arXiv:1905.01986*, 2019.

[3] M. Rodriguez, C. Posse, and E. Zhang, "Multiple objective optimization in recommender systems," in *Proceedings of the sixth ACM conference on Recommender systems*, pp. 11–18, 2012.

[4] S. Wang, M. Gong, H. Li, and J. Yang, "Multi-objective optimization for long tail recommendation," *Knowledge-Based Systems*, vol. 104, pp. 145–155, 2016.

[5] P. Nguyen, J. Dines, and J. Krasnodebski, "A multi-objective learning to re-rank approach to optimize online marketplaces for multiple stakeholders," *arXiv preprint arXiv:1708.00651*, 2017.

[6] Y. Zheng and A. Pu, "Utility-based multi-stakeholder recommendations by multi-objective optimization," in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pp. 128–135, IEEE, 2018.

[7] H. Abdollahpouri, *Popularity Bias in Recommendation: A Multi-stakeholder Perspective.* PhD thesis, 07 2020.

[8] M. Morik, A. Singh, J. Hong, and T. Joachims, "Controlling fairness and bias in dynamic learning-to-rank," *arXiv preprint arXiv:2005.14713*, 2020.

[9] Y. Koren, "Collaborative filtering with temporal dynamics," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 447–456, 2009.

[10] Y. Zhang, Z. Zheng, and M. R. Lyu, "Wspred: A time-aware personalized qos prediction framework for web services," in *2011 IEEE 22nd International Symposium on Software Reliability Engineering*, pp. 210–219, 2011.

[11] Y. Ding and X. Li, "Time weight collaborative filtering," in *Proceedings of the 14th ACM international conference on Information and knowledge management*, pp. 485–492, 2005.

[12] D. Garg, P. Gupta, P. Malhotra, L. Vig, and G. Shroff, "Sequence and time aware neighborhood for session-based recommendations: Stan," in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1069–1072, 2019.

[13] L. Xiang and Q. Yang, "Time-dependent models in collaborative filtering based recommender system," in *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, vol. 1, pp. 450–457, IEEE, 2009.

[14] P. G. Campos, F. Díez, and I. Cantador, "Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols," *User Modeling and User-Adapted Interaction*, vol. 24, no. 1-2, pp. 67–119, 2014.

[15] G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," in *Recommender systems handbook*, pp. 217–253, Springer, 2011.

[16] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "Time-aware point-of-interest recommendation," in *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pp. 363–372, 2013.

[17] M. Pavlovski, J. Gligorijevic, I. Stojkovic, S. Agrawal, S. Komirishetty, D. Gligorijevic, N. Bhamidipati, and Z. Obradovic, "Time-aware user embeddings as a service," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3194–3202, 2020.

[18] R. Baeza-Yates, "Bias in search and recommender systems," in *Fourteenth ACM Conference on Recommender Systems*, pp. 2–2, 2020.

[19] H. Abdollahpouri, M. Mansoury, R. Burke, and B. Mobasher, "The unfairness of popularity bias in recommendation," 08 2019.

[20] Z. Fu, Y. Xian, R. Gao, J. Zhao, Q. Huang, Y. Ge, S. Xu, S. Geng, C. Shah, Y. Zhang, *et al.*, "Fairness-aware explainable recommendation over knowledge graphs," *arXiv preprint arXiv:2006.02046*, 2020.

[21] L. Xiao, Z. Min, Z. Yongfeng, G. Zhaoquan, L. Yiqun, and M. Shaoping, "Fairness-aware group recommendation with pareto-efficiency," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pp. 107–115, 2017.

[22] H. Abdollahpouri, M. Mansoury, R. Burke, and B. Mobasher, "The impact of popularity bias on fairness and calibration in recommendation," *arXiv preprint arXiv:1910.05755*, 2019.

[23] W. Liu, J. Guo, N. Sonboli, R. Burke, and S. Zhang, "Personalized fairness-aware re-ranking for microlending," in *Proceedings of the 13th ACM Conference on Recommender Systems*, pp. 467–471, 2019.

[24] M. Schedl, "The lfm-1b dataset for music retrieval and recommendation," *ICMR*, 2016.

[25] J. Stewart, *Cálculo de una variable. Trascendentes tempranas.* Cengage Learning, 2012.

[26] S. Seabold and J. Perktold, "statsmodels: Econometric and statistical modeling with python," in *9th Python in Science Conference*, 2010.

[27] R. Sanders, "The pareto principle: Its use and abuse," *, Journal of Services Marketing*, vol. 1, pp. 37–40, Feb. 1987.