EQNet: A Post-Processing Approach to Manage Popularity Bias in Collaborative Filter Recommender Systems

Gabriel Balbio V. Machado[a], Wladmir C. Brandão[b] and Humberto T. Marques-Neto[c]
Department of Computer Science, Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, Brazil

Keywords: Recommender Systems, PageRank, Popularity Count, Collaborative Filtering, Popularity Bias.

Abstract: Recommendation systems play a pivotal role in digital platforms, facilitating novel user experiences by effectively sorting and presenting items that align with their preferences. However, these systems often suffer from popularity bias, a phenomenon characterized by the algorithm's inclination to favor a few popular items, resulting in the under-representation of the vast majority of items. Addressing this bias and enhancing the recommendation of long-tail items is of utmost importance. In this paper, we propose the EQNet, a re-ranking approach designed to mitigate popularity bias and improve the recommendation quality of an SVD-based recommendation system. EQNet leverages PageRank or Popularity Count outputs to re-rank items, and its effectiveness is evaluated using four metrics: average popularity, percentage of long-tailed items, coverage of long-tailed items, and recommendation quality. We incorporate the widely recognized bias mitigation algorithm FA*IR into our experimentation to establish a robust baseline. By comparing the performance of EQNet against this state-of-the-art approach, we show the efficiency of EQNet and highlight its potential to enhance existing methods for mitigating popularity bias.

1 INTRODUCTION

The ubiquity of digital content on the internet continues to expand, resulting in an overwhelming array of choices for users seeking to consume media, make purchases, or even engage in personal relationships. In this ever-expanding digital landscape, recommendation systems (RSs) play an indispensable role in guiding users through the vast sea of content (Castells et al., 2011; Taylor, 2023). However, an inherent challenge present in these systems is popularity bias, where certain items or content are recommended to users, perpetuating a cycle of limited diversity. This bias not only constrains the range of options presented to users but also poses profound implications for fairness and equity, as it amplifies the visibility of already popular items, often at the expense of less-known, high-quality ones (H. Abdollahpouri and Mobasher, 2017; Yao and Huang, 2017; Nematzadeh et al., 2018; Yalcin and Bilge, 2022).

Addressing popularity bias is an interesting challenge, as it deals with the delicate balance between popularity and relevance. Removing popularity-driven recommendations might risk introducing quality loss, as popular items often align with user preferences (Jannach, 2015; Kowald and Lacic, 2022). This dilemma is the focal point of our research, as we introduce the EQNet, an approach poised to improve the landscape of recommendation systems. We begin by delving into the challenges posed by popularity bias, outlining its effects on user experiences, content diversity, and the overall fairness of recommendation systems. The EQNet is founded upon a post-processing approach that leverages the power of network ranking algorithms to re-rank recommendation lists, ensuring that underrepresented and high-quality items receive the attention they deserve, without compromising recommendation quality.

To accomplish this goal, it is crucial to balance expanding the recommended list to include long-tail (LT) items and maintaining the recommendation system's (RS) accuracy. As Abdollahpouri's research suggests, each RS has a specific correlation between item popularity and the number of recommendations generated (H. Abdollahpouri and Mobasher, 2019). This study showed that the SVD algorithm has a more linear behavior toward popularity versus recommendation. Also, dot products can be a better default

[a] https://orcid.org/0000-0002-8080-488X
[b] https://orcid.org/0000-0002-1523-1616
[c] https://orcid.org/0000-0002-3164-8323

Machado, G., Brandão, W. and Marques-Neto, H.
EQNet: A Post-Processing Approach to Manage Popularity Bias in Collaborative Filter Recommender Systems.
DOI: 10.5220/0012612800003690
In *Proceedings of the 26th International Conference on Enterprise Information Systems (ICEIS 2024) - Volume 1*, pages 919-932
ISBN: 978-989-758-692-7; ISSN: 2184-4992

choice for a study since they are more relevant for the industry and have similar performance and application (J. Lu, 2017; Rendle et al., 2020; Anelli et al., 2022).

The EQNet can use the output from data processing with PageRank or Popularity Count algorithms as inputs for its operation. We used these algorithms to identify and attenuate the prominence of popular items while concurrently elevating the visibility of long-tail items within our recommendation lists since these two algorithms are well known for their propensity to prioritize popular items (Google, 2011; Sora, 2015; J. Lu, 2017; AWS, 2022). The re-ranked lists generated by the EQNet were compared with the ones generated by FA*IR ranking algorithm (Zehlike et al., 2017). In these experiments, we managed to evaluate the approach's efficiency and effectiveness.

By exploring the EQNet's capabilities, this paper tries to contribute to the ongoing discourse on fairness and quality in recommendation systems, offering a viable option for creating recommendation algorithms that cater to users' diverse and evolving preferences while ensuring fair content exposure.

We organized this paper into six sections. The first three sections discuss the scientific background and related work linked with collaborative filtering RSs, popularity bias behavior, and some ways to mitigate it. Then, we present the EQNet approach, how we developed it, and the experiment built to test it against a baseline in sections four and five. Finally, there is a conclusion and future work section to show possible paths to be explored and deepen the research.

2 BACKGROUND

With the immense amount of information available on the web, RSs capable of filtering, prioritizing, and delivering relevant information to users are indispensable in minimizing this overload (Taylor, 2023). RSs solve this problem by processing a large amount of data to list, rank and provide users with content and services quickly and practically (D. Jannach and Friedrich, 2010; F. Ricci and Kantor, 2011). Additionally, platforms can leverage these recommendation techniques to offer marketing products to other companies (Yin et al., 2012; Instagram, 2016; Sun and Xu, 2019; Didi et al., 2023). Three main types of algorithms used to build these filters (D. Jannach and Friedrich, 2010; F. Ricci and Kantor, 2011; Cano and Morisio, 2017).

- **Content-based filtering:** This method is based on the analysis of content features of items (e.g., text descriptions, keywords, metadata) to create user profiles based on their historical interactions or explicit preferences for certain content features. So, when a user interacts with the system, content-based algorithms recommend items that are like those the user has shown interest in. It relies on the idea that if a user has liked or interacted with certain content characteristics in the past, they will be interested in items with similar characteristics in the future.
- **Collaborative filtering:** The systems focus on finding similarities between users or items based on their historical interactions and preferences without considering item content. There are two main types of collaborative filtering: memory-based and model-based. Memory-based recommendation systems can be constructed based on user preferences or item attributes, adopting either a user-based or item-based approach. These systems identify analogous items within a portfolio and load them into memory, often incurring significant memory expenses. On the other hand, model-based recommendation systems exhibit higher processing costs but lower memory requirements. This is attributed to developing a recommendation system model using fundamental models such as clusterization, matrix decomposition, Bayesian networks, or neural networks. Leveraging these models enables the system to filter and rank extensive datasets without loading a substantial model into memory.
- **Hybrid filtering:** This method combines content-based and collaborative filtering methods to improve recommendation accuracy and overcome the limitations of each approach. Hybrid systems can operate in several ways. One common approach is to generate recommendations separately from content-based and collaborative filtering methods and combine them using weighted averages or other techniques. Alternatively, the system can use one method to enhance the results of the other method. The key is to leverage the strengths of each method to provide more personalized and accurate recommendations.

Since a recommendation algorithm works with ranking lists, they are supposed to have popular items delivered to the users as a sign the algorithm works properly (D. Jannach and Friedrich, 2010; F. Ricci and Kantor, 2011). However, when LT items are present alongside popular items in the recommendation, it not only enriches the user experience but also the diversity of user behavior data (Sun and Xu, 2019). Because of that, being able to control the popularity bias is a challenge for RSs. This bias needs to be accounted for and managed to optimize recommendation systems (H. Abdollahpouri and Mobasher, 2017; E. Mena-Maldonado, 2021; Yalcin and Bilge, 2022).

Also, incorporating collaborative filtering into our experimental design offers a well-grounded approach to investigating the management of popularity bias in recommendation systems. Collaborative filtering methods inherently exhibit popularity bias, favoring recommending popular items because of their reliance on user-item interaction data (Kowald and Lacic, 2022; Ahanger et al., 2022). This bias can hinder user satisfaction by limiting exposure to less popular but more relevant items, a critical concern in many practical recommendation scenarios (Nematzadeh et al., 2018; Yalcin and Bilge, 2022). Moreover, collaborative filtering serves as an ideal candidate for our study because of its widespread use and well-established understanding within the field of recommendation systems (D. Goldberg, 1992; J. Ben Schafer, 2007; Su and Khoshgoftaar, 2009; M. Ranjbar, 2015; Jalili et al., 2018; Rendle et al., 2020). Its inclusion ensures that our findings are not only applicable to real-world recommendation scenarios but also provide a benchmark for evaluating the effectiveness of bias reduction strategies, thus contributing to a more comprehensive understanding of popularity bias management.

2.1 Popularity Bias

Popularity bias exerts detrimental effects on multiple stakeholders within a recommender system environment, encompassing not only consumers but also providers and the overall system. The skewed preference towards popular items not only impacts consumer choices but also impedes the visibility and potential popularity growth of less popular items, thereby undermining recommendation fairness. Ramifications of this bias can be comprehensively explained by examining real-world instances. The prevalence of popularity bias leads to market homogenization, empowering a few item producers to dominate the market (H. Abdollahpouri and Mobasher, 2019). Consequently, this stifles opportunities for innovation and originality, curtailing diversity and limiting the scope for novel offerings.

This repetition of just a few items being recommended to the same user is very tiresome and also represents a significant experience issue (E. Mena-Maldonado, 2021). Psychological studies describe a trend in user behavior in which negative memories linked to a user experience are stronger and more lasting than good ones (D. Yin and Zhang, 2010), causing the negative experiences arising from Popularity Bias to be devastating to a platform in the long run. Hence, this type of bias can also worsen user experience and hinder the overall experience users might have in it.

The correlation between the popularity of an item and the number of times it is recommended varies depending on the recommendation technique applied. Previous studies showed that the SVD algorithm has a more linear behavior towards popularity versus recommendation (H. Abdollahpouri and Mobasher, 2019). Also, dot product can be a better default choice for combining embeddings than learned similarities using MLP or NeuMF, since they are more relevant for the industry (J. Lu, 2017; K. Li, 2019). Dot product similarity simplifies modeling and learning and provides a better alignment with other research areas, such as natural language processing or image models (Rendle et al., 2020). Therefore, the algorithm chosen to generate the matrix decomposition and create the recommendations was the SVD, as popularized by Simon Funk during the Netflix Prize (Koren, 2009; sup,).

Given the pervasive nature of popularity bias in recommendation systems, this study introduces the EQNet approach as a strategic response to manage this bias while preserving the quality of recommendations. The EQNet's core objective lies in enhancing recommendation diversity and engagement while preserving the quality of the recommendation. This fortifies the long-term sustainability and user satisfaction, making the EQNet a valuable contribution to advancing recommendation systems.

2.2 Popularity Parameters

In the experiment using the EQNet, we employed two established item classification algorithms to rerank the recommendations. The first algorithm selected was PageRank, which has been used in recommendation systems and has shown its effectiveness in addressing various challenges within the field (P. Nguyen, 2015; S. Park and Lee, 2019; Al_Sultany and Ghaidaa, 2022). Equation 1 provides a concise representation of the mathematical foundation of PageRank. Let us consider B_i as the set of adjacent items to i, j as a adjacent item in B_i, d as a damping factor (usually set to 0.85 to represent the probability that a user will continue navigating from the current item rather than jumping to a random one), and L_j as the number of outbound links in j.

$$PageRank_i = (1-d) + d \sum_{j \in B_i} \frac{PageRank_j}{L_j} \quad (1)$$

We selected the Popularity Count algorithm as our second algorithm of choice. The Popularity Count algorithm exhibits notable significance in our research, as it employs a ranking approach that incorporates user evaluations of items and the recency of those evaluations. By considering these factors, the Popularity Count algorithm offers valuable insights into the relative popularity of items within a given list, making it a compelling choice for our investigation (AWS, 2022).

Because of its characteristics, Popularity Count has gained usage in various online platforms, including Amazon. Borges and Stefanidis proposed a novel approach to address Popularity Bias, wherein they incorporated this popularity score to penalize item scores based on their historical popularity (Borges and Stefanidis, 2021). This methodology successfully mitigated the bias and fostered diversity, as observed in his paper. Equation 2 shows the evaluation of item (i) popularity using Popularity Rank, which takes into account the weight of the user interaction (w_u) and the recency of the interaction (t_u) across all users (u). Consequently, even if an item has a relatively lower number of interactions, it can receive a higher score if these interactions are recent and carry substantial significance within the specific application context under investigation.

$$PopCount_i = \sum_{u \in i} w_u * t_u \tag{2}$$

By incorporating the outputs of both ranking algorithms as a weighting component, we believed EQNet could effectively address popularity bias and significantly enhance fairness in recommendation systems. Since EQNet would leverage not only the hierarchical structure of items but also the implicit network information, such as user navigation patterns. Figure 1 illustrates how three user rating behaviors can be translated to a network for further processing via PageRank to extract each movie's relevance score. Therefore, unlike many existing re-ranking methods, the EQNet distinguishes itself by utilizing scalar values derived from this type of information rather than relying solely on explicit attributes, cluster parameters, or query-based re-ranking. This unique characteristic of EQNet provides an interesting approach to manage popularity bias and enhance RSs fairness (Adomavicius and Kwon, 2009; Ai et al., 2020).

3 RELATED WORK

In their scientific study, Adomavicus and Kwon conducted experiments with six re-ranking models applied to a recommendation system, aiming to identify the optimal approach for handling popularity (Adomavicius and Kwon, 2009). To assess the effectiveness of these models, they measured the impact on the presence of long-tail (LT) items in the recommendation lists, besides evaluating accuracy. It is widely recognized that accuracy alone does not encompass the entire user experience provided by a recommendation system. Hence, as long as the accuracy remains stable, post and pre-processing techniques can enhance the balance of LT items, thus improving the overall user experience (Knijnenburg et al., 2012; Raza and Ding, 2021; Karboua et al., 2022).

Besides direct popularity, alternative approaches for assessing the relevance of items and re-ranking them exist when considering algorithms applicable to complex networks. An example is the application of PageRank, a well-known algorithm for ranking web pages, which considers not only the inherent popularity of a page but also its interconnectedness with other popular pages. This approach offers a more comprehensive relevance assessment, incorporating intrinsic qualities and contextual relationships within the network structure (Bressan and Pretto, 2011).

Previous studies have evaluated the efficacy of utilizing PageRank algorithms in recommendation systems in the domains of movie databases and web pages, thereby providing a foundation for the present research (Al_Janabi and Kadiam, 2020; Sharma et al., 2022; Al_Sultany and Ghaidaa, 2022). These models use graph structures derived from user-item interactions to build hybrid recommender systems using personalized PageRank applications to rank the lists or build clusters of items and users. Another study has successfully mitigated the impact of popularity bias by employing clusterization techniques and centrality parameters to diminish the influence of nodes that are distant from the user's current navigation cluster (Ai et al., 2020). However, to the best of our knowledge, no previous research has explored the utilization of PageRank, a well-established algorithm renowned for its efficacy in identifying influential nodes within a network, as a re-ranking mechanism to prioritize recommendations.

Some studies propose algorithms to satisfy the fairness constraint as much as possible at each position (C. Geyik and Kenthapadi, 2019; Zehlike et al., 2017; Zehlike et al., 2022). Zehlike proposes a priority queue-based approach (Zehlike et al., 2017), called FA*IR, for item fairness scenarios where only two groups exist. FA*IR ensures that the number of protected candidates does not fall far below a required minimum percentage p at any point in the ranking by formulating this fairness as a statistical significance

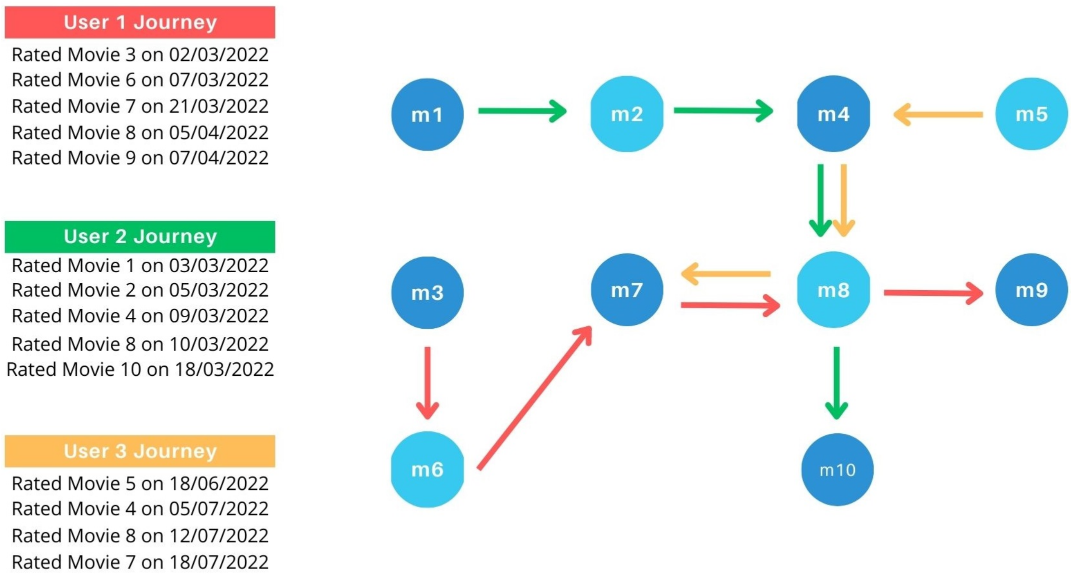

Figure 1: Certain nodes from the Complex Network illustrate the structure used for PageRank calculation. Specifically, in this example snapshot, users 1, 2, and 3 assessed a set of 10 hypothetical movies over the course of 2022.

test of whether a ranking was likely to have been produced by a Bernoulli process. If not, then select the item with the highest relevance in the protected queue; otherwise, compare the two queues and select the item with the highest relevance.

Furthermore, Zehlike improved the FA*IR algorithm to address scenarios involving multiple protected groups (Zehlike et al., 2022), necessitating the adoption of a statistical test based on a multinomial distribution rather than the binomial distribution employed in the original FA*IR framework. Notably, FA*IR keeps its capacity to preserve fairness when a protected group already enjoys helpful exposure, and the ranked group fairness condition is satisfied based on candidate ranking scores. In this context, FAIR ensures that a protected candidate may only experience a reduced exposure if another protected candidate from a different group ascends in the ranking, thus safeguarding against exposure loss because of non-protected candidates. This enhancement aligns FA*IR with the complexities inherent in multi-group fairness considerations, making it a valuable tool for addressing fairness concerns in RSs.

The FA*IR algorithm uses a fairness metric to achieve this goal. One commonly used fairness metric is the Demographic Disparity (DD), presented in Equation 3, where S_u represents the top-k ranking for user u, P_i represents the set of items associated with protected group i and U is the set of all users. The objective is to minimize the DD while optimizing the recommendation quality.

$$DD = \frac{S_u \cap P_i}{S_u} - \frac{P_i}{U} \tag{3}$$

Because of its recency, efficiency, and wide application, the FA*IR algorithm serves as an excellent state-of-the-art baseline for research on reducing unfairness by managing popularity bias in recommendation systems. In the presented research, which introduces the novel re-ranking algorithm called EQNet, FA*IR's prominence and application as a post-processing solution become an essential argument. Therefore, EQNet can also be used to build upon and enhance the existing methodologies, contributing to advancing fairness-oriented recommendation systems (Zehlike et al., 2022).

4 THE EQNet

EQNet is a re-ranking approach that leverages popularity values computed by ranking algorithms, such as PageRank, to reclassify the items in the portfolio. Including popularity values is crucial due to the presence of popularity bias, which arises from the variations in user interactions with different items on a platform, distinguishing between popular items and long-tail (LT) (H. Abdollahpouri and Mobasher, 2020; E. Mena-Maldonado, 2021). The concept of developing an approach that leverages intrinsic values associated with item popularity for their reclassification aligns with existing research in the field and enables us to adopt an approach for exploring additional possibilities. We selected Popularity Count and PageRank as the algorithms to evaluate item popularity because of their efficiency in identifying and highlighting relevant items within a collection and their importance in the area. By employing a new approach, we conducted experiments with these well-established algorithms to explore their effectiveness in mitigating popularity bias.

To calculate PageRank, it is necessary to construct a Complex Network that reflects the underlying data meaningfully. Inspired by Jiang's previous work (Al_Sultany and Ghaidaa, 2022), we designed a network structure where films served as nodes, and user evaluations, arranged chronologically, were used to establish directional edges connecting one film to another. Figure 1 visually represents the network structure in the study's context.

The EQNet approach can operate with both PageRank or Popularity Count output values as the main re-ranking variable (α_i), needing a prior normalization process to compress into values between 0 and 1. In Equation 4, we present the transformation process performed by EQNet, wherein the ranking score ($S_a i$) is converted into a rebalanced score ($S_b i$), suitable for re-ranking. Including a parameter (λ) enables fine-grained control over the intensity of the re-ranking procedure. This formulation allows us to adjust the degree to which recommendations are influenced by fairness considerations, ensuring that the resulting recommendations balance relevance and bias mitigation.

$$S_{bi} = \frac{S_{ai}}{\sqrt[\lambda]{\alpha_i}} \tag{4}$$

Through this approach, we successfully achieved a re-balancing of recommendation lists, specifically by diminishing the prominence of highly popular and central items while elevating the relevance scores of long-tail items. Figure 2 illustrates how the standard recommendation works after the SVD ranking, and Figure 3 illustrates the dynamic of the re-ranking process when the EQNet is applied. Essentially, the alterations induced by the reduction in the popularity of certain items and the increase in others lead to the substitution of previously dominant popular items with long-tail items that exhibit a strong user affinity, thus enhancing the diversity and consequently reducing the popularity bias.

5 VALIDATING EQNet

In this study, we comprehensively validated our proposed method using two widely recognized public datasets. The first dataset employed is the well-known MovieLens dataset, which encompasses a vast collection of user movie reviews (Harper and Konstan, 2016). Including this dataset enables us to evaluate the performance and effectiveness of our approach on a large-scale, real-world recommendation scenario. Additionally, we used a second dataset from the Netflix Prize, a benchmark used in various recommendation algorithm contests (Koren, 2009).

Under the data reduction method outlined, users with fewer than 20 ratings were excluded from the Netflix dataset (H. Abdollahpouri and Mobasher, 2017). This filtering process was conducted as we found that users with longer profiles were much more likely to have interacted with LT (long-tail) items. Users kept after this reduction were those that interacted more with the platform, thus being more likely to have interacted and rated LT items, enabling our training vs. testing scenarios to be executed.

To get a comprehensive understanding of each method employed in our experiments, we conducted the experiment using ten distinct values for the parameter λ, enabling a more comprehensive understanding. In the baseline's case, we evaluated the FA*IR algorithm's recommendation quality impact by testing it with various values for the proportion of protected candidates' parameter (p) and observed its influence on the popularity bias reduction. Additionally, we assessed the performance of EQNet as a post-processing factor to the FA*IR algorithm to investigate potential synergies between the two solutions.

5.1 Evaluation

We diligently computed and recorded metric values at each iteration step throughout the model testing process. To establish a meaningful baseline for comparison with the two EQNet variants, we adopted FA*IR, a top-k ranking algorithm known for its impressive performance on the selected databases, as supported by prior research (Wang et al., 2023; Krasanakis et al., 2021). To assess the Popularity Bias and overall efficacy of EQNet in our simulated recommendation system, we carefully selected four metrics for evaluation against the established baseline:

- **Average Recommendation Popularity (ARP):** Is used to analyze the average popularity of items in each recommendation list as a crucial metric. To quantify this, we define U_t as the total number of users, L_u as the total number of items in a recommendation list, and Φ as the total number of times we evaluated the item i in the training phase as presented in Equation 5 (Yin et al., 2012):

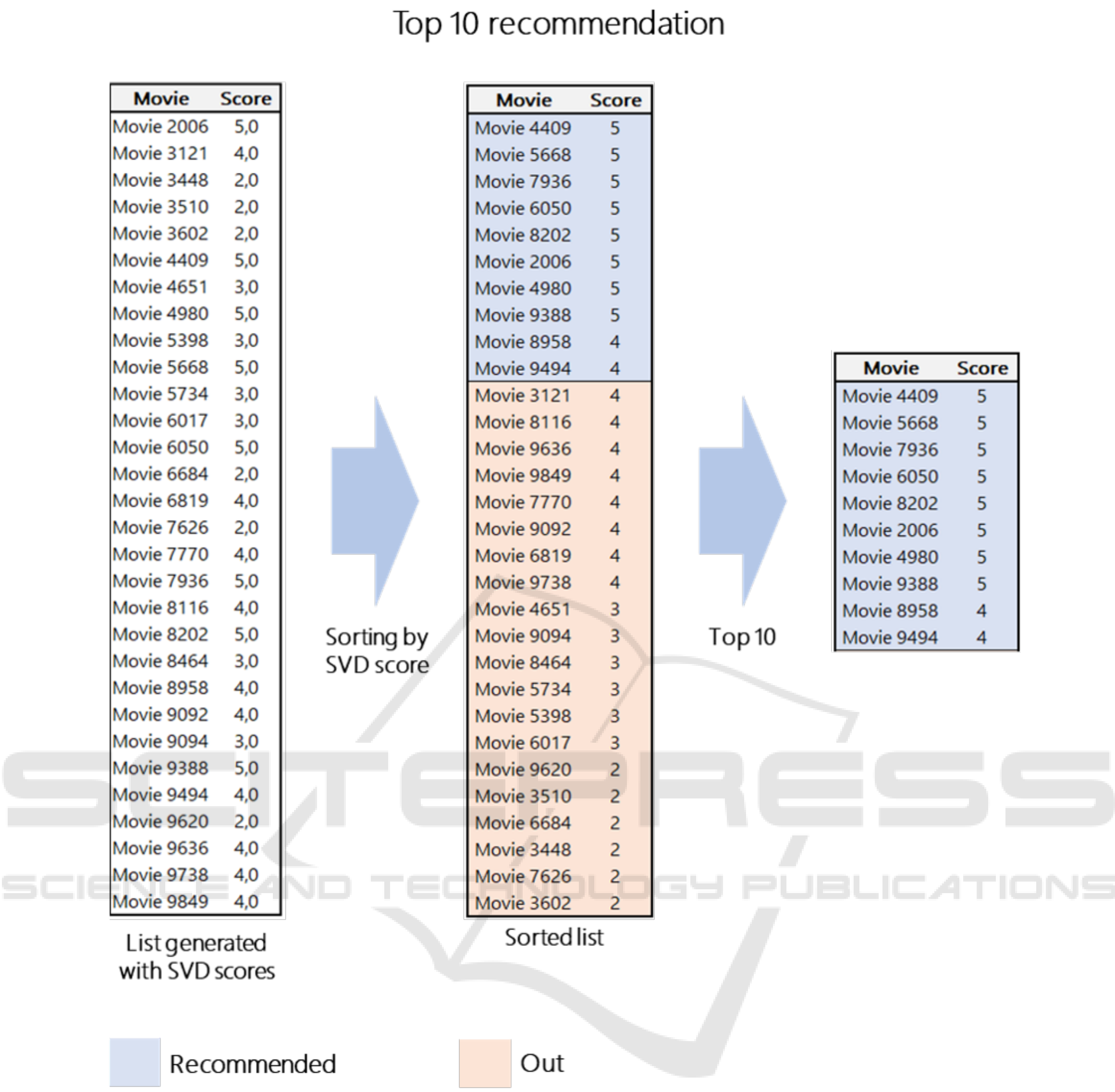

Figure 2: Representation of the standard SVD recommendation of a top-10 list from a universe containing 30 items.

$$ARP = \frac{1}{|U_t|} \sum_{u \in U_t} \frac{\sum_{i \in L_u} \Phi(i)}{|L_u|} \tag{5}$$

- **Average Percentage of Long-Tailed Items (APLT):** As previously proposed by Abdollahpouri et al. (H. Abdollahpouri and Mobasher, 2019), this metric is employed to assess the average percentage of long-tail (LT) items present in the recommendation list. As we can see in Equation 6, Γ denotes the group comprising all long-tail items.

$$APLT = \frac{1}{|U_t|} \sum_{u \in U_t} \frac{|i, i \in (L_u \cap \Gamma)|}{|L_u|} \tag{6}$$

- **Average Coverage of Long-Tailed Items (ACLT):** This metric complements the analysis of Average Proportion of Long-Tail (APLT) items and provides valuable insights into the diversity of recommendations (H. Abdollahpouri and Mobasher, 2019). By assessing whether the recommendation consistently lists the same set of LT items, the metric, as shown in Equation 7, has $1(i \in \Gamma)$ where item i belongs to set Γ, serves as an indicator of recommendation fairness and diversity. A value of 1 is assigned when item i is present in set Γ, allowing us to shed some light on the degree to which LT items are exposed in the recommendation system.

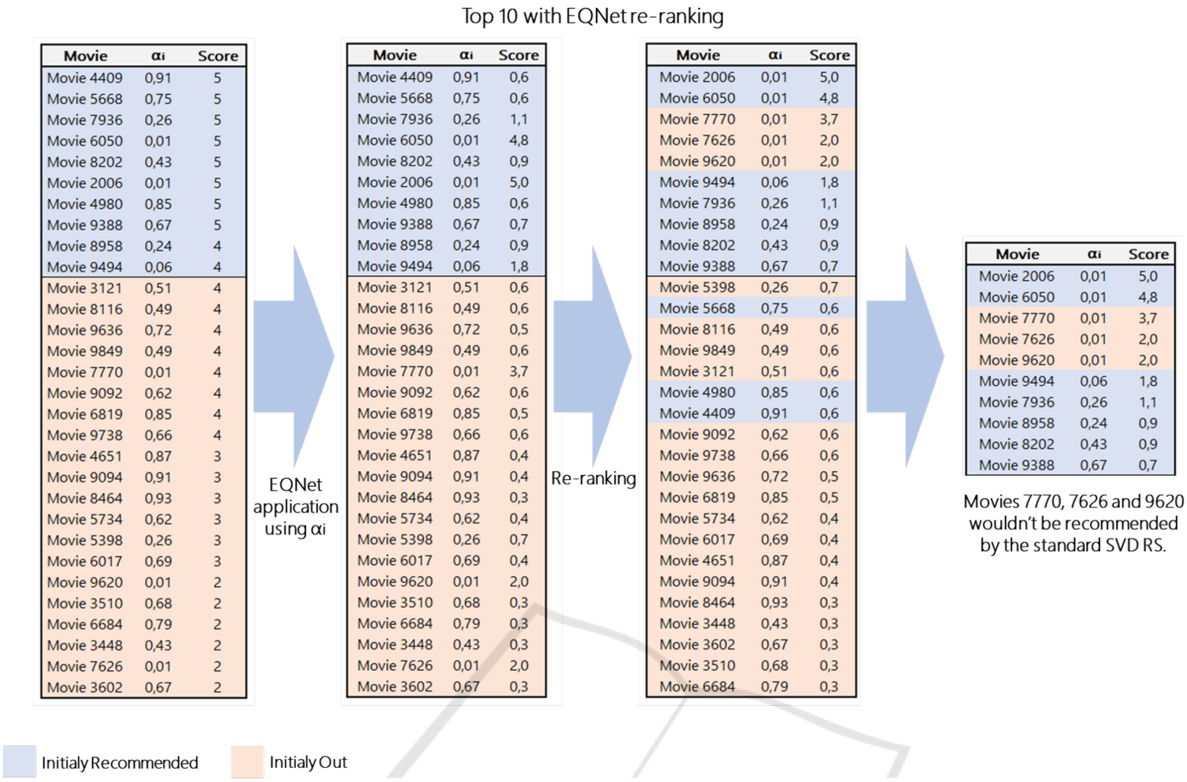

Figure 3: Representation of the EQNet approach to a top-10 recommendation list from a universe with 30 items already ranked by SVD. This involves the re-ranking process using the popularity score in column (α_i) to systematically adjust the item scores inversely to their popularity levels.

$$ACLT = \frac{1}{|U_t|} \sum_{u \in U_t} \sum_{i \in L_u} 1(i \in \Gamma) \tag{7}$$

- **Normalized Discount Cumulative Gain (NDCG):** The accuracy of recommendations, commonly employed as a metric to assess the system's performance, holds significant value in evaluating the effectiveness of the application. While acknowledging that accuracy might not directly reflect user experience, measuring this metric remains crucial in providing insights into the recommendation system's overall efficiency (Wang et al., 2013).

With the EQNet parameters configured, we conducted the experimental study. Initially, we performed the recommendation using Singular Value Decomposition (SVD) on the databases without any re-ranking approach. Subsequently, we applied three distinct techniques, namely the FA*IR, the EQNet with Popularity Count, and the EQNet with PageRank, as well as both EQNet applications after FA*IR. Collecting results from each experimental batch and comparing the evaluation metrics of the recommendation system before and after the re-ranking executions, we systematically assessed the overall evolution and efficacy of the EQNet approach in improving recommendation fairness.

The experimental framework depicted in Figure 4 outlines the essential steps undertaken to gather data and evaluate the results of our study. Initially (1), we performed a dataset reduction by filtering out users with fewer than 20 ratings, thereby focusing on the behavior of highly active users with a greater likelihood of interaction with the recommended items (H. Abdollahpouri and Mobasher, 2017). Subsequently (2), we conducted recommendation runs based on SVD and recorded the results of ARP, APLT, ACLT, and NDCG. For the third step (3), we iteratively applied a re-ranking method, varying parameters across ten iterations to amass a more extensive dataset for our analysis of recommendation quality and bias reduction. When EQNet was used with PageRank, an additional step was necessary to construct a complex network that enabled the calculation of PageRank values for each item (4). The outcomes of the three re-ranking methods were studied, organized, and presented in the most appropriate

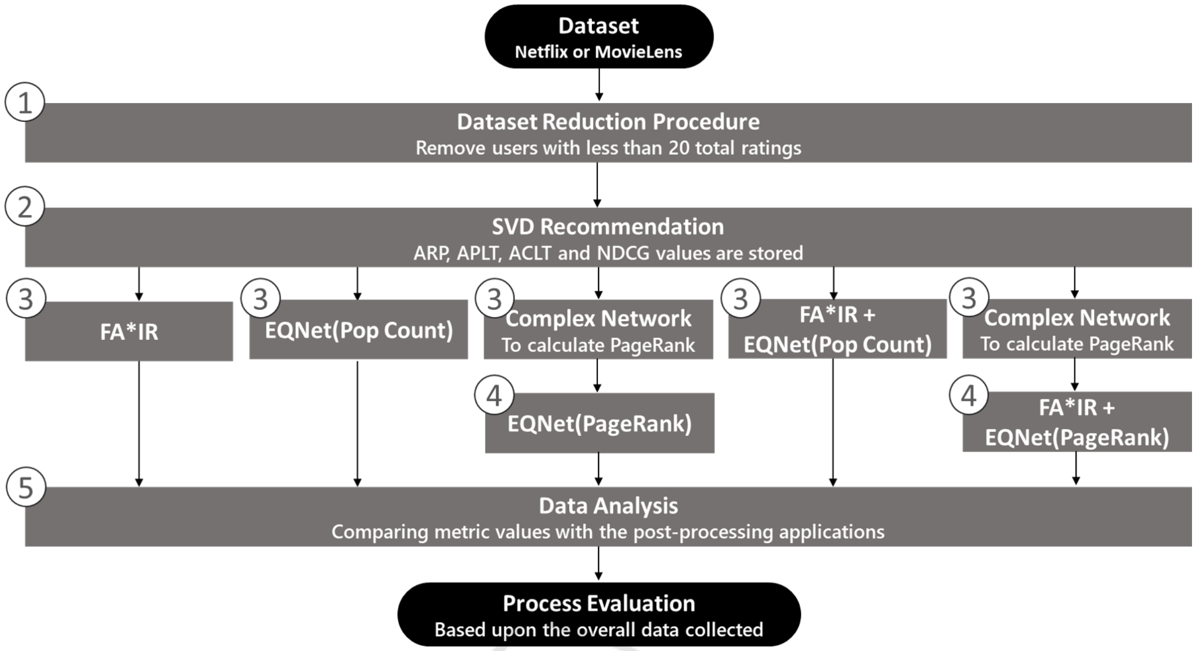

Figure 4: Summarized flowchart with the main steps of the experimental process.

manner for a comprehensive analysis (5).

5.2 Result Analysis

For each method, we registered the values of nine iterations versus the initial value. At the end of each iteration, we compared the values of ARP, APLT, ACLT, and NDCG with the ones from the SVDS recommendation with no optimization and used this relative value to generate our analysis. For example, the APLT variation from an iteration t would be $\Delta APLT_t = \frac{APLT_t}{APLT_0}$, and we would calculate it for each iteration to register the total variation versus the initial value. Also, for the EQNet approaches, we used five different values of λ to capture more behavior nuances.

In our experiment, we recorded the values of all four key metrics (ARP, APLT, ACLT, and NDCG), across nine iterations relative to their initial values for each of the five methods: FA*IR, EQNet with PageRank, EQNet with Popularity Count, FA*IR followed by EQNet with PageRank and FA*IR followed by EQNet with Popularity Count. Through this iterative analysis, we gained valuable insights into the behavior of the tradeoff between recommendation quality and bias reduction. Moreover, by comparing the performance of the FA*IR followed by EQNet approaches with their isolated counterparts, we could discern and quantify the specific impact of the combined method in addressing bias and enhancing recommendation quality.

The values for the basic Singular Value Decomposition (SVD) recommendation model, without a post-processing re-rank, are presented in the following list:

– NDCG = 0.225971
– ARP = 0.155042
– APLT = 0.505874
– ACLT = 5.058735

The Normalized Discounted Cumulative Gain (NDCG) values exhibit consistent alignment with those reported in related studies within the field when employing Singular Value Decomposition (SVD) as a recommendation technique and comparing the train and test groups sourced from offline data. Notably, the evaluation includes a subset of long-tailed items, ensuring a comprehensive assessment of the algorithm's performance across diverse item distributions (D. M. Ferrari and Cremonesi, 2022; Valcarce et al., 2018). These values provided a crucial reference for evaluating the outputs with the post-processing approaches.

To gain a comprehensive insight into EQNet's impact on the RS outcomes, we conducted an analysis of the variations between the original recommendation metrics and the post-processed results. In the context of ARP, negative variations are favorable, as they signify a reduction of the overall popular items within the recommendation lists. Additionally, we delved into metrics such as APLT and ACLT which measure the involvement and coverage of long-tail items,

Table 1: Table depicting the metrics percent variation for each method using the MovieLens Database.

Methods	ARP var. (%)	APLT var. (%)	ACLT var. (%)	NDCG var. (%)
FA*IR	-32.73	18.23	17.78	-1.39
EQNet (PageRank)	-79.39	44.93	44.93	-2.01
EQNet (PopCount)	-57.95	65.77	65.77	-2.53
FA*IR + EQNet (PageRank)	-54.18	44.62	44.62	-1.66
FA*IR + EQNet (PopCount)	-13.34	85.70	85.70	-1.91

respectively. Here, a higher positive variation indicates improved performance. With the NDCG, our goal was to minimize loss while maximizing gains in other variations. These examinations provide valuable insights into the method's impact on popularity bias management and recommendation quality.

Table 1 and Table 2 showcase the comparative performance of each approach in terms of percent variation, with a focus on Average Percentage and Average Coverage of Long-Tailed Items (APLT and ACLT) as well as Average Recommendation Popularity (ARP) while employing EQNet and FA*IR algorithms. The results indicate the EQNet outperforms FA*IR in reducing ARP and simultaneously improving APLT and ACLT with only a slight reduction in Normalized Discount Cumulative Gain (NDCG). Moreover, when combining EQNet with FA*IR, the experiments exhibit even more promising outcomes, achieving further enhancements in APLT and ACLT, albeit at a modest cost to NDCG. These findings underscore the effectiveness of EQNet as a powerful tool for mitigating bias and enhancing fairness in recommendation systems, with the potential for complementary utilization alongside FA*IR to achieve superior performance in optimizing multiple fairness metrics.

Also, when comparing both tables, it's possible to see that each EQNet performed better at a given database. Considering that after the database reductions, MovieLens had more balance between the number of users and the number of movies, while the Netflix database had significantly more users than movies. This indicates that both algorithms worked well with EQNet, and each has its own application niche.

Upon comparing the results in both tables, it becomes clear that each variant of the EQNet exhibited superior performance on specific databases. After reducing the database, the MovieLens database resulted in a more balanced distribution between the number of users and movies, while the Netflix database exhibited a considerable imbalance, with significantly more users than movies. This observation suggests that both EQNet variations effectively handled popularity bias concerns, and their respective strengths and niche applications were apparent.

Furthermore, the analysis of scatter-plot charts in Figure 5 and Figure 6 provides valuable insights into the entire spectrum of NDCG variation for each of the three popularity metrics. These visualizations offer a comprehensive view of the behavior of post-processing methods regarding changes in their parameters and the corresponding sensitivity of each metric to incremental adjustments. Figure 6 reveals a compelling observation where the combined method of FAIR followed by EQNet with PageRank shows superior retention of recommendation quality while effectively mitigating Popularity Bias and enhancing Fairness, compared to the standalone FAIR approach. This observation underscores the efficacy and potential synergistic benefits of using the proposed EQNet re-ranking algorithm alongside FA*IR for addressing popularity bias and fairness concerns in recommendation systems.

In our experiment, the EQNet effectively reduced popularity bias while only marginally affecting RS accuracy. The evaluation of EQNet using two distinct ranking algorithms yielded interesting results, revealing high variations in ARP, APLT, and ACLT with marginal NDCG loss in both databases. Moreover, in comparative assessments against the state-of-the-art FA*IR algorithm, EQNet exhibited substantial potential in handling balanced databases independently and, in tandem with other ranking algorithms, displayed superior performance, contingent on the specific ranking algorithm employed in the ensemble.

6 CONCLUSION AND FUTURE WORK

Collaborative filtering is a compelling approach for enhancing user experience by recommending relevant items. In this scientific paper, we introduced the EQNet, an innovative approach designed to mitigate popularity bias in experimental data, leading to potential improvements in recommendation accuracy. Using two distinct databases, each characterized by different user profiles, and employing two algorithms to

Table 2: Table depicting the metrics percent variation for each method using the Netflix Database.

Methods	ARP var. (%)	APLT var. (%)	ACLT var. (%)	NDCG var. (%)
FA*IR	-31.54	17.23	16.88	-1.37
EQNet (Pagerank)	-33.20	7.71	7.71	-0.06
EQNet (PopCount)	-18.51	12.96	12.96	-1.73
FA*IR + EQNet (PageRank)	-52.90	15.42	15.42	-0.20
FA*IR + EQNet (PopCount)	-51.09	14.19	14.19	-1.45

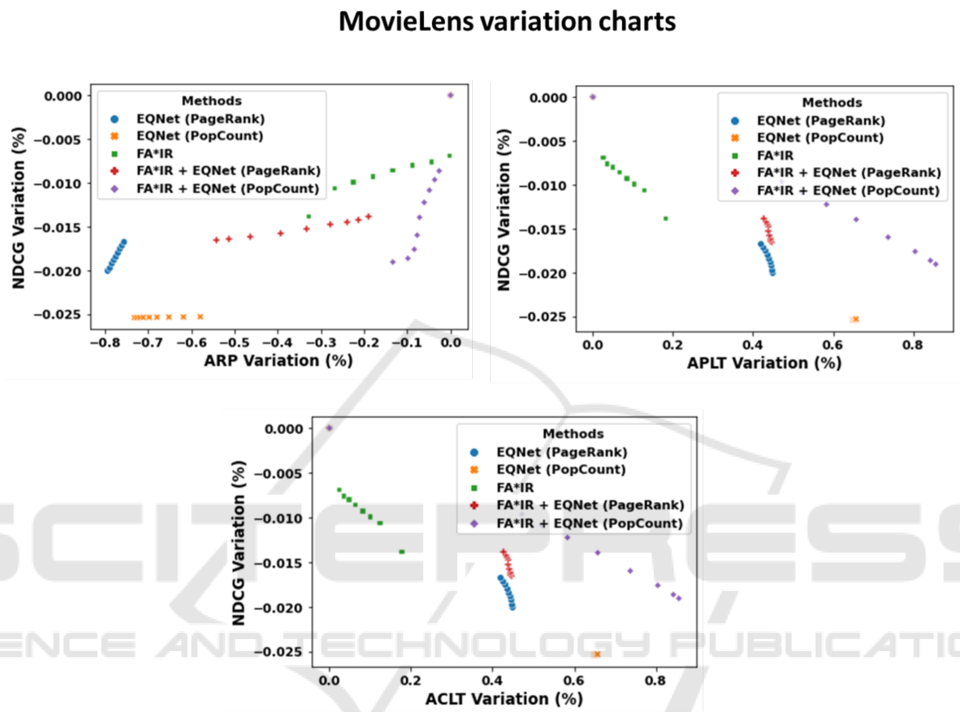

Figure 5: Graphs comparing the variations of NDCG with ARP, APLT, and ACLT using the five post-processing methods with the MovieLens database.

rank the items by popularity, we gained valuable insights into EQNet's behavior, uncovering underlying intrinsic factors that contribute to its performance.

We present a new technique for mitigating popularity bias in Recommender Systems (RSs) using EQNet with PageRank and Popularity Count outputs to reevaluate nodes. Through comprehensive evaluations of the evolution of NDCG (Normalized Discounted Cumulative Gain), ARP (Average Recommendation Popularity), APLT (Average Popularity of the Last T recommendations), and ACLT (Average Clicks on the Last T recommendations), we compare our proposed EQNet algorithm with the renowned FA*IR algorithm, examining their performance in tandem. Our experimental results show that EQNet effectively addresses popularity bias in both databases with only marginal recommendation quality loss. Additionally, EQNet exhibits the potential to enhance the overall performance of the FA*IR algorithm.

Due to its efficiency, simplicity, and low computational complexity, EQNet presents as a viable module for controlling popularity bias behavior and enhancing fairness within recommendation systems. While the workflow employed in this study may require adaptation to other recommendation methods, EQNet, and its parameter calibration principles are expected to remain unchanged. The re-ranking approach, incorporating popularity and network-related parameters and metrics, offers a compelling avenue to strike a balance in the recommendation list, enabling a combination of multiple factors to address more intricate scenarios. These findings validate the efficacy of the EQNet in managing popularity bias and advancing state-of-the-art fairness-oriented recommendation systems.

We expect several potential avenues for future research and practical applications of the EQNet, including:

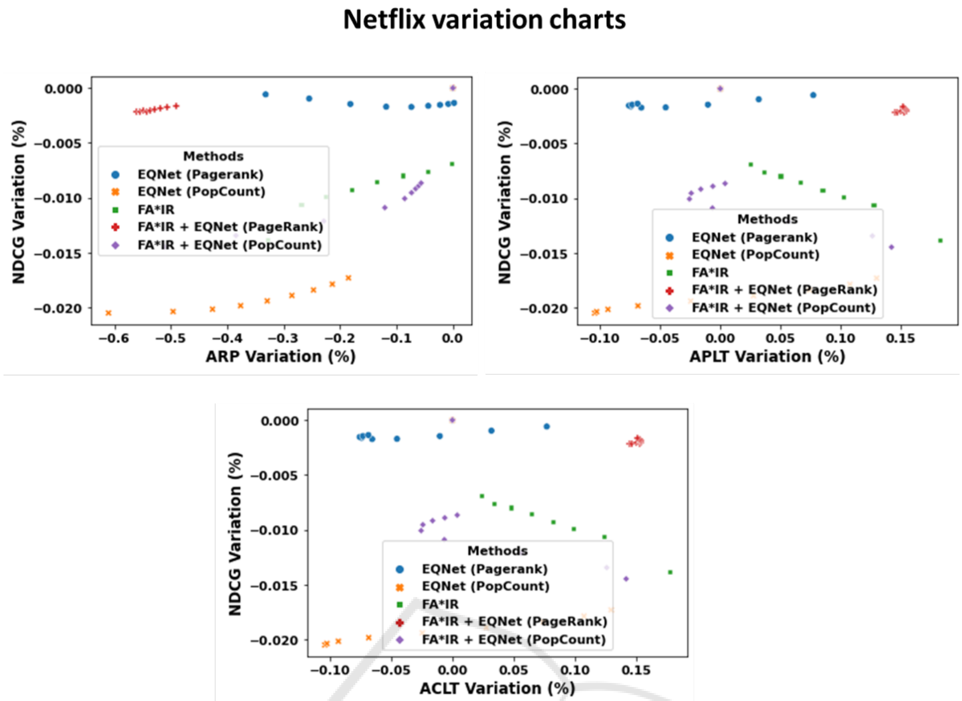

Figure 6: Graphs comparing the variations of NDCG with ARP, APLT and ACLT using the five post-processing methods with the Netflix database.

- **Does the EQNet Work with Other Types RSs?** It is crucial to explore the behavior of various recommendation systems (RSs) beyond just SVD-based approaches when subjected to the EQNet re-ranking algorithm. Regardless of the system's complexity, a comprehensive evaluation should involve metrics assessing the alignment between the recommendations and user preferences, such as RMSE and NMAE, besides popularity-based metrics like ARP and ACLT. Integrating these metrics will ensure a fair and robust assessment of the EQNet's impact on recommendation fairness across a diverse range of RSs.
- **How Other Complex Network Graph Structures Can Change the PageRank Influence in EQNet?** Understanding the efficacy of the EQNet with diverse PageRank network structures is essential to find out its versatility and potential for achieving fairness in recommendations across a wide range of real-world use cases. By exploring and evaluating these configurations, we can gain valuable insights into the adaptability of the EQNet.
- **What Other Ranking Algorithms Can Be Used to Produce Satisfactory Results with EQNet?** Understanding the behavior of other ranking algorithms with EQNet and their performance in various recommendation scenarios is crucial for a comprehensive assessment and applicability in real-world settings.
- **How Does More Than One Instance of EQNet Work in Tandem?** Since this paper initially sought to comprehend how the EQNet behaves and conduct comparative assessments with other contemporary state-of-the-art methodologies, it is pertinent to investigate the collaborative behavior of distinct EQNet instances employing diverse ranking algorithms. For instance, a comparative analysis of an EQNet configured with PageRank against another configured with Popularity Count warrants examination. Such an investigation promises valuable insights into the synergistic effects and differential performance exhibited by these EQNet variants when deployed in tandem.

ACKNOWLEDGEMENTS

The authors thank the Pontifícia Universidade Católica de Minas Gerais – PUC Minas and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – (Grant PROAP 88887.842889/2023-00 – PUC/MG, Grant

PDPG 88887.708960/2022-00 – PUC/MG - INFORMÁTICA and Finance Code 001).

REFERENCES

surprise documentation. https://surprise.readthedocs.io/en/stable/matrix_factorization.html, note = Accessed: 2023-07-31.

Adomavicius, G. and Kwon, Y. (2009). Toward more diverse recommendations: Item re-ranking methods for recommender systems. *Workshop on Information Technologies and Systems*.

Ahanger, A., Aalam, S., Bhat, M., and Assad, A. (2022). Popularity bias in recommender systems - a review. pages 431–444.

Ai, J., Su, Z., Wang, K., Wu, C., and Peng, D. (2020). Decentralized collaborative filtering algorithms based on complex network modeling and degree centrality. *IEEE Access*, 8:151242 – 151249.

Al_Janabi, S. and Kadiam, N. (2020). Recommendation system of big data based on pagerank clustering algorithm. pages 149–171.

Al_Sultany, G. and Ghaidaa, A. (2022). Enhancing recommendation system using adapted personalized pagerank algorithm. In *2022 5th International Conference on Engineering Technology and its Applications (IICETA)*, pages 1–5.

Anelli, V. W., Bellogín, A., Noia, T. D., Jannach, D., and Pomo, C. (2022). Top-n recommendation algorithms: A quest for the state-of-the-art.

AWS (2022). Aws developer guide. https://docs.aws.amazon.com/personalize/latest/dg/native-recipe-popularity.html.

Borges, R. and Stefanidis, K. (2021). On mitigating popularity bias in recommendations via variational autoencoders. *36th Annual ACM Symposium on Applied Computing*.

Bressan, M. and Pretto, L. (2011). Local computation of pagerank. *20th ACM International Conference on Information and Knowledge Management*.

C. Geyik, S. A. and Kenthapadi, K. (2019). Fairness-aware ranking in search & recommendation systems with application to linkedin talent search. *25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.

Cano, E. and Morisio, M. (2017). Hybrid recommender systems: A systematic literature review. *Intelligent Data Analysis*, 21(6):1487–1524.

Castells, P., Vargas, S., and Wang, J. (2011). Novelty and diversity metrics for recommender systems: Choice, discovery and relevance. *Proceedings of International Workshop on Diversity in Document Retrieval (DDR)*.

D. Goldberg, e. a. (1992). Using collaborative filtering to weave an information tapestry. *Communication of the ACM*, 35(12):61–70.

D. Jannach, M. Zanker, A. F. and Friedrich, G. (2010). Recommender systems, an introduction.

D. M. Ferrari, N. F. and Cremonesi, P. (2022). Offline evaluation of recommender systems in a user interface with multiple carousels. *Mathematical Problems in Engineering*, pages 1–13.

D. Yin, S. B. and Zhang, H. (2010). Are bad reviews always stronger than good? asymmetric negativity bias in the formation of online consumer trust. *31st International Conference on Information Systems (ICIS'10)*, pages 1 – 18.

Didi, T., Guy, I., Livne, A., Dagan, A., Rokach, L., and Shapira, B. (2023). Promoting tail item recommendations in e-commerce. page 194–203.

E. Mena-Maldonado, e. a. (2021). Popularity bias in false-positive metrics for recommender systems evaluation. *ACM Transactions on Information Systems*, 39(3).

F. Ricci, L. Rokach, B. S. and Kantor, P. (2011). Recommender systems handbook. pages 73 – 140.

Google (2011). Facts about google. *www.google.com/competition/howgooglesearchworks*.

H. Abdollahpouri, R. B. and Mobasher, B. (2017). Controlling popularity bias in learning-to-rank recommendation. *RecSys'17*.

H. Abdollahpouri, R. B. and Mobasher, B. (2019). Managing popularity bias in recommender systems with personalized re-ranking. pages 242 – 251.

H. Abdollahpouri, M. Mansoury, R. B. and Mobasher, B. (2020). The connection between popularity bias, calibration, and fairness in recommendation. *RecSys 2020*.

Harper, F. and Konstan, J. (2016). The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems*, 5.

Instagram, B. T. (2016). Increasing website conversions with instagram. *https://business.instagram.com/blog/increasing-website-conversions-with-instagram*.

J. Ben Schafer, Dan Frankowski, J. H. . S. S. (2007). Collaborative filtering recommender system. *The Adaptive Web*, 17(6):291–324.

J. Lu, e. a. (2017). Trust-enhanced matrix factorization using pagerank for recommender system.

Jalili, M., Ahmadian, S., Izadi, M., Moradi, P., and Salehi, M. (2018). Evaluating collaborative filtering recommender algorithm: A survey. *IEEE Access*, 6:74003–74024.

Jannach, D. (2015). What recommenders recommend: An analysis of recommendation biases and possible countermeasures. *User-Modeling and User-Adapted Interaction*, 25:427 – 491.

K. Li, e. a. (2019). Deep probabilistic matrix factorization framework for online collaborative filtering.

Karboua, S., Harrag, F., Meziane, F., and Boutadjine, A. (2022). Mitigation of popularity bias in recommendation systems: A selective review.

Knijnenburg, B., Willemsen, M., Gantner, Z., Soncu, H., and Newell, C. (2012). Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction*, page 441–504.

Koren, Y. (2009). The bellkor solution to the netflix grand prize.

Kowald, D. and Lacic, E. (2022). Popularity bias in collaborative filtering-based multimedia recommender systems. pages 1–11.

Krasanakis, E., Papadopoulos, S., and Kompatsiaris, I. (2021). Applying fairness constraints on graph node ranks under personalization bias. In Benito, R. M., Cherifi, C., Cherifi, H., Moro, E., Rocha, L. M., and Sales-Pardo, M., editors, *Complex Networks & Their Applications IX*, pages 610–622, Cham. Springer International Publishing.

M. Ranjbar, e. a. (2015). An imputation-based matrix factorization method for improving accuracy of collaborative filtering systems. *Engineering Applications of Artificial Intelligence*, 46:58–66.

Nematzadeh, A., Ciampaglia, G., Menczer, F., and Flammini, A. (2018). How algorithmic popularity bias hinders or promotes quality. *Scientific Reports*, 8.

P. Nguyen, e. a. (2015). An evaluation of simrank and personalized pagerank to build a recommender system for the web of data.

Raza, S. and Ding, C. (2021). News recommender system: A review of recent progress, challenges, and opportunities.

Rendle, S., Krichene, W., Zhang, L., and Anderson, J. (2020). Neural collaborative filtering vs. matrix factorization revisited. *RecSys '20: Proceedings of the 14th ACM Conference on Recommender Systems*.

S. Park, W. Lee, B. C. and Lee, S. (2019). A survey on personalized pagerank computation algorithms. *IEEE Access*, PP:1–1.

Sharma, U., Sajeev, G. P., and Rani, S. S. (2022). Personalized fashion recommendation using nearest neighbor pagerank algorithm. pages 1–6.

Sora, I. (2015). A pagerank based recommender system for identifying key classes in software systems. *International Symposium on Applied Computational Intelligence and Informatics*.

Su, X. and Khoshgoftaar, T. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 2009.

Sun, C. and Xu, Y. (2019). Topic model-based recommender system for long-tailed products against popularity bias. *4th Int. Conf. on Data Science in Cyberspace (DSC)*.

Taylor, P. (2023). Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025. www.statista.com/statistics/871513/worldwide-data-created/. Accessed: 2023-09-03.

Valcarce, D., Bellogín, A., Parapar, J., and Castells, P. (2018). On the robustness and discriminative power of information retrieval metrics for top-n recommendation. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18, page 260–268, New York, NY, USA. Association for Computing Machinery.

Wang, Y., Ma, W., Zhang, M., Liu, Y., and Ma, S. (2023). A survey on the fairness of recommender systems. *ACM Trans. Inf. Syst.*, 41(3).

Wang, Y., Wang, L., Li, Y., He, D., Liu, T., and Chen, W. (2013). A theoretical analysis of ndcg type ranking measures.

Yalcin, E. and Bilge, A. (2022). Evaluating unfairness of popularity bias in recommender systems: A comprehensive user-centric analysis. *Information Processing & Management*, 59(6):103100.

Yao, S. and Huang, B. (2017). Beyond parity: Fairness objectives for collaborative filtering. page 2925–2934.

Yin, H., Cui, B., Li, J., Yao, J., and Chen, C. (2012). Challenging the long tail recommendation. *Proceedings of the VLDB Endowment*, 5(9):896 – 907.

Zehlike, M., Bonchi, F., Castillo, C., Hajian, S., Megahed, M., and Baeza-Yates, R. (2017). Fa*ir: A fair top-k ranking algorithm. page 1569–1578.

Zehlike, M., Sühr, T., Baeza-Yates, R., Bonchi, F., Castillo, C., and Hajian, S. (2022). Fair top-k ranking with multiple protected groups. *Information Processing & Management*, 59(1):102707.