Popularity Bias as Ethical and Technical Issue in Recommendation: A Survey

Anastasiia Klimashevskaia[1]

University of Bergen, Norway anastasiia.klimashevskaia@uib.no

Abstract. Recommender Systems have become omnipresent in our everyday life, helping us making decisions and navigating in the digital world full of information. However, only recently researchers have started discovering undesired and harmful effects of automated recommendation and began questioning how fair and ethical these systems are, while influencing our day-to-day decision making, shaping our online behaviour and tastes. In the latest research works, various biases and phenomena like filter bubbles and echo chambers have been uncovered among the resulting effects of recommender systems and rigorous work has started on solving these issues. In this narrative survey, we investigate the emergence and progression of research on one of the potential types of biases in recommender systems, i.e. Popularity Bias. Many recommender algorithms have been shown to favor already popular items, hence giving them even more exposure, which can harm fairness and diversity on the platforms using such systems. Such a problem becomes even more complicated if the object of recommendation is not just products and content, but people, their work and services. This survey describes the progress in this field of study, highlighting the advancements and identifying the gaps in the research, where additional effort and attention is necessary to minimize the harmful effect and make sure that such systems are build in a fair and ethical way.

Keywords: Recommender Systems · Fairness · Popularity Bias · Ethics

1 Introduction

Within the past decade, it has been discovered that Recommender Systems (RS) can have not only a positive impact on our everyday life, but might also influence us with certain undesired effects [21,34,46]. Previous research has been mostly concentrated on the improvements of RS accuracy and efficiency in an attempt to make the recommendations as precise as possible. However, with the growing interest in the ethics and fairness issues in Artificial Intelligence (AI), the scope of research on RSs has broadened as well. Questions about fairness and privacy in the context of RSs have been getting more attention recently, leading researchers to uncovering various types of biases [17] and other related phenomena such as Filter Bubbles [7,21], Echo Chambers, persuasion and manipulation. Perhaps one of the most researched types of bias within RS is Popularity Bias [2,22], which could cause the so-called "Matthew effect", i.e., "the rich are getting

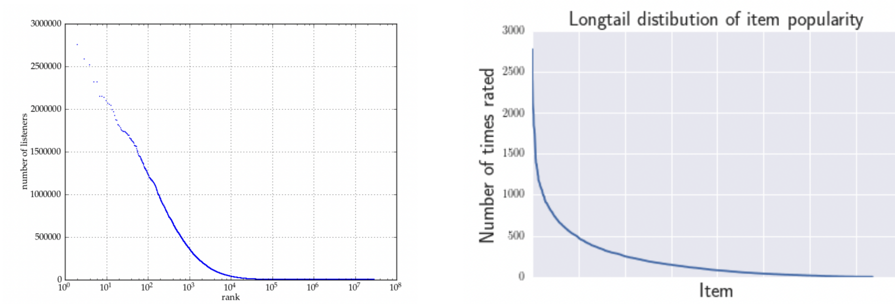

Fig. 1. The "Long Tail" of recommendation - user interactions with each item are plotted to demonstrate that a small share of items being extremely popular and interacted with by an immense number of users, while the rest majority remains unexposed. This effect is observed in different domains as well - i.e. music [31] or movie recommendation [2].

richer", when only already popular recommendations receive most of the exposure.
This work is aiming at understanding the current state of the popularity bias issue in RS by reviewing some of the recent scientific works on the topic. We examine the definitions and terminologies used in this field, drawing comparisons in different domains, as well as investigating the available techniques for mitigation of popularity bias and quantifying metrics. This survey is expected to aid in recognizing the achievements and the missing parts of current research, which can help in defining future work and directions.

2 Preliminaries

Generally recommendation approaches can be categorized into three groups: *Collaborative Filtering* (CF) methods [42,19] that utilize historical data of user

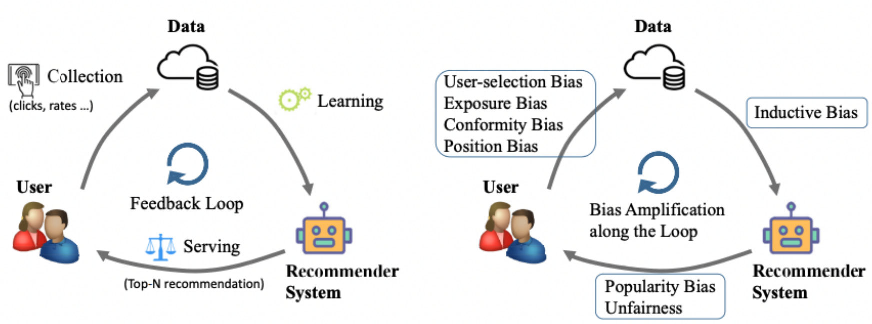

Fig. 2. Feedback loop in RS [17]. The graphic demonstrates the stages of recommendation process, where bias can enter the system or become reinforced.

content consumption to make recommendation based on the implicit or explicit interactions between users and items; *Content-Based Filtering* (CBF) methods [38] working with item features and contents to define similarity between them, as well as modeling user profiles with traits; and finally, a *hybrid approach*, utilizing the working principles of both of the above mentioned techniques. CF methods seem to be used most frequently in the literature and most of the research concentrates on these algorithms [28]. The core of a RS is most commonly an algorithmic model trained on pre-existing data. Such a model serves the user with recommendations, and receives explicit (ratings) or implicit (interactions) feedback from them. This new data is used afterwards to update and re-train the model, and the cycle goes on. This is the so-called "*feedback loop*" (Fig. 2 left side). It has been previously shown that bias can enter the system at any stage of this loop [17], and even get amplified and propagated by the recommender algorithm [32] (Fig. 2 right side).
CF algorithms are well known for amplifying the popularity of the already well-known items (i.e.*"short head"*), over-representing them in recommendation lists, while the majority of the less popular (i.e. *"long tail"*) items barely receive any exposure through recommendation and remain undiscovered and unknown to the users [28,2,17,22]. The popularity distribution among the items is approaching closely the *power law*, when the number of interactions or ratings for the most popular items is growing exponentially compared to the less popular ones.
The general aim of RSs is to recommend items from a catalogue that the user might like. Fairness in the scope of RSs has been described in [20] as consistent accuracy across different groups of users. These groups can be identified by any characteristics the user might have - gender, age, nationality, as well as preferences, including their interest towards mainstream or niche items. Several groups can be usually identified among the users of a RS based on the popularity of the content they consume [2] - usually they are split into "blockbuster-focused" users who are mainly interested in most popular items, "niche" users who prefer less-known long tail items, and the rest of the users with diverse tastes including the items of varying popularity. Since many CF algorithms tend to recommend more popular items, it leads to a decrease in recommendation quality for the users with more niche tastes. These users end up being treated less fairly by a RS than the mainstream consumers. Furthermore, the harm of such an effect can increase greatly if the RS is recommending businesses [12] or even people - job seekers [1][2], artisans [3][4], and freelancers [5]. For example, without any special treatment the newcomers to the platform with an RS would never receive the equal exposure they deserve. Moreover, RSs that rely strongly on item popularity may be abused by malicious users [5], in an attempt to use the algorithm to their advantage and "outplay" the system, by boosting their popularity artificially and

[1] `https://www.linkedin.com/`
[2] `https://www.wcc-group.com/employment/`
[3] `https://www.amazon.com/Handmade/`
[4] `https://www.etsy.com/`
[5] `https://www.fiverr.com/`

gaining better chances of being recommended due to that [29].
The way how omnipresent RSs are in our everyday life, it is without doubt that they shape our tastes and behaviour to some extent [34]. One of the main goals of a RS is to encourage exploration and discovery. At the same time, that requires the system to influence and persuade the user to pick certain items. Unfortunately, there is no clear distinction when benign nudging turns into malicious coercion and manipulation, "addicting" the users of a RS to certain types of popular recommendation items. RSs are very frequently used by corporations and industry for profit, and then the boundary between user and corporate interests become blurred and the recommendation tuning might be too often driven more towards the corporate goals, infringing on the autonomy of the user [12,47] and their right to receive fair recommendations based on their preference and not dictated by a higher agenda or the opinion of the majority [48].
Last but not least, targeting only accuracy in hopes for a better user retention and engagement without any account for fairness and responsibility can lead to further issues in domains like news. For example, the YouTube engine has been well known for recommending “fake news” and dubious and biased content just because it is effective at engaging and retaining the attention of the user, causing controversy and heated discussion [16]. While it might lead to good performance metrics of a RS, it is obviously a questionable tactic in terms of ethics and responsibility. Additionally, popularity-based recommendation in news topics like politics can substantially hurt the state of pluralism in society [26], polarizing it and creating filter bubbles and echo chambers.
All the aforementioned issues are concerning the receiving side within the RS. Additionally to that, it is important to realize that many RSs can be perceived as a multi-stakeholder environment, which makes it an even more complicated system, where the fairness towards every participant has to be accounted for in an ideal case [10].
While accuracy and performance of RSs have been studied thoroughly for a while, the field of research on the ethical issues of RSs is very new and there is plenty of ambiguity in definitions, metrics and ways to evaluate [41]. A solid common ground is needed within the community to bring this research on the new level, establishing boundaries and instruments to distinguish ethical from unethical.
This work is organized as follows: In Chapter 3 we discuss other related literature surveys in the field of RSs in general. Chapter 4 is split into several parts: in Subchapter 4.1 the approach to literature collection and reviewing is presented; in 4.2 a general overview on the domains is given, as well as categories and types of works investigated, expanding on various stages of research and its progression in Subchapters 4.3, 4.4, 4.5. Lastly, the survey is completed with a final discussion and conclusions in Chapter 5.

3 Connected Literature Surveys

Popularity bias has been previously often brought up as a subtopic in different works related to biases [17], undesired effects [21] or fairness issues in RSs in general [1]. A thorough investigation of various biases and debiasing techniques is conducted in [17], offering a classification of various effects and bias types in RS. Furthermore, the authors present diverse debiasing approaches, giving an overview on the state of the art in this field of study. The authors in [21] raise concerns regarding the issues and undesired effects of RSs within the media domain. The authors in [6] have collected a technical overview on popularity bias mitigation techniques and some metrics. However, to our knowledge, there has not been any attempts of conducting a literature review on fairness and ethics issues of popularity bias as a stand-alone topic. Thus, it appears to be a gap in the research in this field and we believe that such a review can help understand the current state of the field better, find common definitions and possible misconceptions, provide a better overview on how this topic has been evolving over time.

4 Literature review

4.1 Research Methodology

For this survey we have utilized search engines in digital libraries like ACM[6], Springer[7], ScienceDirect[8] and Web Of Science[9], as well as Google Scholar[10] to search first for the most prominent publications on the topic of popularity bias in RSs. We have performed the search by key phrases such as "*popularity bias*" and "*recommender systems*" plus such additional keywords as "*fairness*", "*ethic*", "*responsible*". After browsing through the most cited and recent publications and hand-picking notable and thought-provoking ones, the range of the literature has been expanded through citations in these works and by investigating connected publications through tools like Connected Papers [11]. Such an approach allows to find connection points between works and seeing how opinions on certain topics have been changing over time. Thus, this survey contains not only some of the works that are retrieved as "most relevant" and have received the most citations, but also less well-known publications that still bring up important discussion points and draw attention to thought-provoking details. The author would like to point out that this work is not a systematic review, but a one of a more narrative nature, thus not every existing work on the topic has been included, but the ones that assist in unraveling the narrative and help in the argumentation.

---

[6] `https://dl.acm.org/search/advanced`
[7] `https://link.springer.com/advanced-search`
[8] `https://www.sciencedirect.com/search`
[9] `https://www.webofscience.com/wos/woscc/basic-search`
[10] `https://scholar.google.com/`
[11] `https://www.connectedpapers.com/`

4.2 Overview

After the search described above 46 works total have been retrieved, with 3 marked as not entirely related, but worthy of discussing and inclusion, and 7 entirely excluded. The papers are stemming from various fields of computer science judging by the conferences and journals they were published in - RSs, information retrieval, data mining, expert systems. Furthermore, researchers from economics and management seem to also have interest in long tail exposure, willing to investigate the long term effects on customers, users, society. Last but not least, the newest papers concerning popularity bias and fairness issues appear more frequently in journals like *"AI and Ethics"*, *"AI and Society"*, or conferences focusing on fairness, accountability and transparency of AI technologies. This gives a clear understanding how the focus has extended from only accuracy and performance to beyond-accuracy and fairness issues as well. Bias within RSs is a topic that has gained increasingly more interest and visibility in the past years, and appears to continue doing so [17].

There is no unified attitude towards popularity bias among the reviewed literature - some works clearly portray it as a harmful phenomenon, while the others simply describe it as an existing factor that can be quantified and measured. A small number of papers, however, claims that the presence of popularity bias to a degree can be beneficial and improve both accuracy and fairness/diversity [51,39]. In terms of fairness, the majority of the earlier papers have been considering only user fairness first. However, with the emergence of the "*multi-stakeholder*" concept of RSs [11], item, group or producer fairness have also started getting attention in research. Latest works are attempting at combining multiple viewpoints as well. The majority of authors apply their techniques in movie or music domain, as well as e-commerce, likely due to easily accessible public datasets [25,13,52,35].

Lastly, the papers can be also divided into multiple categories by the type of research. There are authors attempting to quantify the phenomenon of popularity bias and measure the effects. They either perform data-based investigation, trade-off analysis or run simulations measuring long-term effects and changes in the system. New metrics can be proposed as a result of such work, adding new facets and dimensions to the field of research. Other works describe novel methods and techniques for mitigation of the popularity bias, comparing them to baselines to assess effectiveness of the proposed algorithms and pipelines. Such methods can be generally split into two categories[12]: *model-based* and *post-processing* mitigation techniques. These techniques can be divided into groups based on the personalization level as well - either platform-level bias mitigation, when every user receives the same treatment, or personalized, "calibrated" approaches, when each user's recommendation adjustment is based on their particular personal features and properties.

[12] Additionally, a third category is mentioned in the literature as well - pre-procesing. It is, however, mostly not applicable to tackling the popularity bias issue [3].

4.3 Accuracy-Oriented Popularity Bias Research

The early research focused on long tail recommendation was interested in leveraging tail items in recommendations solely to improve the quality of the recommendation before the term “popularity bias” was coined. The concerns regarding user or item fairness has not emerged yet, and hence the research works primarily aimed at recommendation relevance. A notable example can be the method proposed in [37] focused on clustering of tail items and predicting the ratings for each item based on the overall ratings in each cluster. Such a technique is claimed to leverage cold-start items, i.e. new items that have very few to no ratings. Instead of claiming that RSs favor popular items over unpopular, the authors just state that RSs discard the unpopular tail part as “cold-start” items that are unknown to the system and cannot be included properly in the recommendation process due to the difficulty of the rating prediction. An important point is brought up in this work - how to define the popularity threshold in order to identify the head and tail part? Selecting the right cutting point, right number of clusters, carefully choosing any parameters for a method - all this can be crucial for the performance assessment. This leads us to one of the main pitfalls in the current RS community research - very often there is no common ground on definitions and constants, thus no direct comparison can be made between methods. This can hinder reproducibility and slow down the progress, forcing the researchers to recreate instead of reusing.

Some of the early works have already started delving into beyond-accuracy research on RSs - the authors of [4], for example, investigated methods to improve aggregate diversity of the recommendation. Exposure or aggregate diversity is an important factor for RS, many researchers have argued about the power of diversity, claiming that “diverse exposure can be valued simply because it extends individual choice and affords individuals more opportunities to realize their interests” [27]. The authors of this work are suggesting that a recommendation covering more varying (in terms of popularity) items gives a better opportunity for the user to receive a more personalized and idiosyncratic recommendation. At the same time, it is being shown that such diversity can be attained by recommending more of the tail items. A notable point - the users are generally capable of finding the "bestsellers" themselves, while the recommendation mechanism should be aimed at providing a wider range of items instead and not only focus on the very relevant, safe and highly-popular options (which many RSs generally tend to gravitate towards). As a result of the work, various ranking criteria are proposed to escape from popularity-based recommendation. Additionally, the evaluation has shown that a significant improvement in diversity can be already achieved at a fairly small loss in accuracy, which can possibly address the concerns about the accuracy-fairness/diversity trade-off.

It has been shown that the novelty of recommendations can be also connected directly to its popularity, as it is demonstrated in [36]. The work suggests to consider user tendencies in terms of the popularity and *“mainstreamness”* of the content. The authors argue that using simple popularity-based recommendations makes the RS “ineffective” and “obvious”, while making a bigger emphasis

on novelty disregards the tastes of the users who prefer popular items. Thus, the user preference, or *Personal Popularity Tendency*, can be the key to adjusting the novelty and popularity of recommendations and still retain high accuracy. This work not only suggests a metric to gauge this user preference, but also a technique to measure the difference between user history and recommendation.
An interesting observation has been made in [50]: the authors claim that the popularity of the item is directly connected to the difficulty of finding it within the system - the more popular the item is, the easier it is for the user to discover it. At the same time, the more difficult it is to find it, the more valuable and significant is the fact of the user interaction with it. A concept of *"opinion"* is brought up for a finer weighting approach. When comparing two user profiles, a popular item co-rated by both users should be weighted more and given higher significance than a popular item rated by only one of them. The same applies to popular items with varying opinions about them - such items should be enhanced in comparison to the commonly well-accepted popular blockbusters. [13]
The differentiation between more and less "important" popularity demonstrates already at the beginning of popularity bias research the complexity of popularity as a phenomenon and the conclusion that some of it should be penalized while the rest could be utilized for the better. The authors claim that equal debiasing treatment of the whole catalogue is impossible and a certain priority must be given to some group of the items for the sake of eventual equality and right of exposure.

4.4 Wait A Minute... But Is It Fair?

Over time, it has been shown that optimizing only for accuracy without any consideration for other important objectives like fairness can actually decrease the utility and usefulness of recommendation [33]. Further on, more concerns started emerging considering the fairness of popularity-based recommendation. As shown in [2], many users with more diverse or niche taste can potentially be treated unfairly by RSs, receiving recommendations over-concentrated on popular items. The term *"popularity bias"* starts appearing in the literature [23,14,49]. The scope of research works expands greatly, varying from quantification approaches, data-based and trade-off analysis to popularity bias mitigation techniques, simulations, user studies and other experiments. New metrics emerge for better quantification and capturing the influence and effects of the bias.
Understanding causality and development of undesired effects within RSs is crucial to mitigate them. It is necessary to be aware of how such phenomena change under the influence of the user input, how the algorithm adapts (if it does at all) to certain biases or debiasing processes, whether the system really takes all the aspects of the user preference into consideration and to what extent. The work in [15] is one of the many which attempted to assess how much the RSs propagate the observable user preference for popularity. As a result, the authors have

[13] The idea of weight amplification has been previously introduced in [8] for transforming the predicted weights for penalizing or leveraging.

come to a conclusion that at least the collaborative filtering approaches that have been tested in this work have very little consideration for these particular user preferences and lack personalization in this sense. However, this does not necessarily mean it has only adverse effects - such a feature of a RS can make it resilient to other failure conditions and protect against over-personalization and filter bubble effects.

The authors in [20] consider popularity bias from a standpoint of RS evaluation, in a combination with fairness towards user groups of different demographics, characteristics and sizes. The work brings up important points, such as the fact that different biases can be correlated, but it is not necessarily always true, and correcting the RS for one undesired effect can actually lead to exacerbation of another one. The authors argue that popularity bias could be also influenced by the way we evaluate and measure. The recommendation algorithms tend to favour more popular items attempting to optimize for accuracy, but, as how it's stated in this work, "user satisfaction in a recommender system depends on more than accuracy".

In [40] the authors analyze two datasets in the point-of-interest domain, demonstrating the negative effects of popularity bias on fairness and propagation of it in various recommendation algorithms. In contrast to the previous works, which have investigated either consumer or producer fairness separately, this work attempts to study the interplay of both of the factors with recommendation accuracy. The experiments demonstrate and confirm that the investigated methods suffer from underlying popularity bias, and that in most of the cases there is a trade-off between user and item fairness. Most of the selected techniques fail to balance between these two factors and retain good accuracy of recommendation at the same time.

The research in [9] is not targeted at studying popularity bias directly, however, gives insights into the influence of popularity and social ties on the market and the users of a RS. The authors create a mathematical model in order to understand how strongly consumers can be influenced by recommendation or by social aspects and the "web of kinship" between the users. A model is built to simulate the behaviour of the users taking into account various factors: social ties, recency, popularity, and super influential "celebrity" users of the RS. It clearly demonstrates, how the popularity of the recommendation can affect the market or system where it is implemented in a significant way. From the simulations and experiments the researchers draw conclusions that social networks between users can even overcome the significance of "influencers" and "celebrities" on the platform, drawing the recommendation potentially away from high popularity. Nowadays many implementations of RSs lack consideration for the social aspect between the users (also outside of the platform), and their output is typically generated and presented in a non-transparent way, without demonstrating which of the other users' histories influence it. On the one hand, that leaves the user only to the mercy of a popularity-based recommendation, also lowering the transparency and explainability of the recommendation process. On the other hand, exposing other users' preferences through explanation can become a threat to

privacy and an ethical issue. Nevertheless, it is important to remember that a RS within a platform is not the only driver of the content consumption - an actual real-life social network always exists outside of these platforms, with friend or family word-of-mouth recommendations, and that is still an influential factor.
The experiments conducted in [30] have shown that currently used state-of-the-art popularity bias metrics might not capture the phenomenon in full extent. Deeper analysis with more detailed metrics can reveal some patterns and tendencies within the system, giving more understanding of what might be causing popularity bias. Moreover, the authors have investigated the effect of popularity bias on different user gender groups. Like many other works involving user features, this experiment has shown as well that different user groups are affected in a different way depending on the group size and representation.

4.5 Popularity: Harmful and Useful

With the research on popularity of the recommended items and bias expanding, it has also become apparent that there is no uniform one-size-fits-all solution to the emerging issues. Moreover, in some cases item popularity can have different meaning and origin, and even be used to an advantage actually improving fairness or diversity, depending on the settings and the application domain.
The work of [51] offers to leverage temporal information to disentangle true quality popularity of items and conformity effects, when the popularity is short lived. The hypothesis of this work says that item popularity can be influenced by two factors - item quality, which can be observed more or less continuously over an extended period of time; and conformity effect, which forces users to some extent to behave according to the group norms. The authors believe that blindly removing popularity bias can actually hurt the quality of recommendation [14]. Moreover, they claim that the information about the item quality can be extracted from the observation of its popularity over time. They introduce a certain *temperature* parameter, that would help utilize the benign item quality effect and control the harmful conformity effect. It is shown later in several experiments that creating such a time-aware system is an effective technique for high quality recommendations. Since the authors are attempting to partially leverage popularity bias, it is not being mitigated or quantified as well, as it is not considered inherently bad.
The research in [39] revolves around the idea that recommending popular news is effective in case of cold-start users and helps avoiding too similar recommendations. The authors assert that there is enough topical diversity among the most popular news articles and it can help to get the users covered with not only the topics they are solely interested in, but also broaden their horizons. The proposed technique is to incorporate popularity measures into recommendation,

[14] Such an opinion has been already previously brought up in RS research - for example, *"a pretty healthy dose of (unpersonalized) popularity"* has been mentioned in [24] as a necessary part of recommendation. The authors in [45] state that including some very popular/well-known items helps with establishing user's trust towards a RS.

adjusting the influence of it according to the user profile size and embedding learned from the user reading history. Such ideas help to understand that popularity is not inherently bad and can be utilized in the recommendation process to further improve quality.

5 Discussion and Conclusions

There is still a lot of research required to be enable us to understand and predict how and to what extent the RSs affect the user choices and decision making process. There seem to be no consensus in the community - the research in [9] demonstrates that social ties and word-of-mouth can overcome the influence of RSs, while the authors in [44] claim that online recommendations can be more influential than human-made ones. Prior works such as [18] have investigated factors like familiarity, context and additional information and the way they effect decisions made by the customer using a RS. Such human-computer interaction appears to be very complex and may have a lot of hardly or non-quantifiable aspects to it like user personality, temperament, credulousness. Nevertheless, a huge responsibility task for software architects and computer scientists is to ensure that the RS they create does not abuse any of these factors and truly encourages exploration without exploitation. Automated recommendation can nudge the user in a certain direction, but it should not entrap them in a certain belief, habit or preference, especially a potentially harmful one. In food and recipe RSs bad eating habits can become reinforced unless the recommendation is diversified enough with less popular but healthier options. News readers might get "stuck" in polarized beliefs and ideals if the political news articles they receive in a recommendation are just catering to their previous reading history.
Promoting the content of varying popularity is important, it is crucial not only for the profit goals of the industry side of the recommendation, but for the social benefit and fairness. Keeping various types of diversity in mind while developing an RS will ensure that it is being created ethically and responsibly. The following list includes a number of possible conclusion points:

- Long tail item exposure is not only the issue of accuracy, but has become a point of interest for the research area focused on fairness as well. Multi-objective approaches are the current state-of-the-art, attempting to bridge the gaps in research and establish a connection between different aspects of recommendation.
- Debiasing techniques should not necessarily be applied uniformly and in the same way to every user - that leads to disrespect and disregard of user's interests and preference. Furthermore, item popularity is not inherently bad - it is a complex concept, some parts of it can be utilized for actually achieve higher quality and diversity of recommendation.
- There is undoubtedly a correlation and connection between item popularity, novelty and diversity. Studying and researching these facets of recommendation qualities collectively could help understand them better and come up with more sophisticated bias mitigation and fairness-oriented techniques.

– Common baselines, concepts and evaluation protocols are required to truly bring the research in this field to a new level. Hyperparameter tuning is crucial as well and should be standardized.
– The research appears to be lacking work and data in the fields where ethical and fairness issues might be the most harmful - medical sphere, social services recommendation, banking, etc. Data is required to research and investigate the effects of recommendation in such cases, however, it is very often too protected or unavailable due to privacy concerns. Appropriate scientific solutions are long overdue and required in this field of recommendation.
– RSs appear to have been created to simulate real life word-of-mouth recommendation, when a person with similar tastes or needs can "recommend" you something they liked. It is a societal phenomenon [43], and the society in itself lacks fairness, meaning mimicking it would inherently lead to a creation of an unfair and possibly unethical system. Just like with biased AI, extra work, awareness, attention and effort is required to bring such systems to a new level, elevating them above the societal standards and improving the fairness within.

6 Acknowledgements

This research was supported by industry partners and the Research Council of Norway with funding to MediaFutures: Research Centre for Responsible Media Technology and Innovation, through the Centres for Research-based Innovation scheme, project number 309339.

References

1. Abdollahpouri, H., Adomavicius, G., Burke, R., Guy, I., Jannach, D., Kamishima, T., Krasnodebski, J., Pizzato, L.: Multistakeholder recommendation: Survey and research directions. User Modeling and User-Adapted Interaction **30**(1) (2020)
2. Abdollahpouri, H., Mansoury, M., Burke, R., Mobasher, B.: The unfairness of popularity bias in recommendation. arXiv preprint arXiv:1907.13286 (2019)
3. Abdollahpouri, H., Mansoury, M., Burke, R., Mobasher, B.: Addressing the multistakeholder impact of popularity bias in recommendation through calibration. arXiv preprint arXiv:2007.12230 (2020)
4. Adomavicius, G., Kwon, Y.: Improving aggregate recommendation diversity using ranking-based techniques. IEEE Transactions on Knowledge and Data Engineering **24**(5), 896–911 (2011)
5. Aggarwal, A.: Detecting and mitigating the effect of manipulated reputation on online social networks. In: Proceedings of the 25th International Conference Companion on World Wide Web. pp. 293–297 (2016)
6. Ahanger, A.B., Aalam, S.W., Bhat, M.R., Assad, A.: Popularity bias in recommender systems-a review. In: International Conference on Emerging Technologies in Computer Engineering. pp. 431–444. Springer (2022)
7. Aridor, G., Goncalves, D., Sikdar, S.: Deconstructing the filter bubble: User decision-making and recommender systems. In: Fourteenth ACM Conference on Recommender Systems. pp. 82–91 (2020)

8. Breese, J.S., Heckerman, D., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. arXiv preprint arXiv:1301.7363 (2013)
9. Bressan, M., Leucci, S., Panconesi, A., Raghavan, P., Terolli, E.: The limits of popularity-based recommendations, and the role of social ties. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 745–754 (2016)
10. Burke, R.: Multisided fairness for recommendation. arXiv preprint arXiv:1707.00093 (2017)
11. Burke, R.D., Abdollahpouri, H., Mobasher, B., Gupta, T.: Towards multi-stakeholder utility evaluation of recommender systems. UMAP (Extended Proceedings) **750** (2016)
12. Calvo, R.A., Peters, D., Vold, K., Ryan, R.M.: Supporting human autonomy in ai systems: A framework for ethical enquiry. In: Ethics of Digital Well-Being, pp. 31–54. Springer (2020)
13. Celma, O.: Music Recommendation and Discovery in the Long Tail. Springer (2010)
14. Celma, O.: Foafing the music: Bridging the semantic gap in music recommendation. In: International Semantic Web Conference. pp. 927–934. Springer (2006)
15. Channamsetty, S., Ekstrand, M.D.: Recommender response to diversity and popularity bias in user profiles. In: The 13th International FLAIRS Conference (2017)
16. Chaslot, G.: How algorithms can learn to discredit "the media" (Jul 2020), `https://guillaumechaslot.medium.com/how-algorithms-can-learn-to-discredit-the-media-d1360157c4fa`
17. Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., He, X.: Bias and debias in recommender system: A survey and future directions. arXiv preprint arXiv:2010.03240 (2020)
18. Cooke, A.D., Sujan, H., Sujan, M., Weitz, B.A.: Marketing the unfamiliar: the role of context and item-specific information in electronic agent recommendations. Journal Of Marketing Research **39**(4), 488–497 (2002)
19. Ekstrand, M.D., Riedl, J.T., Konstan, J.A.: Collaborative filtering recommender systems. Now Publishers Inc (2011)
20. Ekstrand, M.D., Tian, M., Azpiazu, I.M., Ekstrand, J.D., Anuyah, O., McNeill, D., Pera, M.S.: All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In: Conference on Fairness, Accountability and Transparency. pp. 172–186. PMLR (2018)
21. Elahi, M., Jannach, D., Skjærven, L., Knudsen, E., Sjøvaag, H., Tolonen, K., Holmstad, Ø., Pipkin, I., Throndsen, E., Stenbom, A., et al.: Towards responsible media recommendation. AI and Ethics pp. 1–12 (2021)
22. Elahi, M., Kholgh, D.K., Kiarostami, M.S., Saghari, S., Rad, S.P., Tkalčič, M.: Investigating the impact of recommender systems on user-based and item-based popularity bias. Information Processing & Management **58**(5), 102655 (2021)
23. Fleder, D.M., Hosanagar, K.: Recommender systems and their impact on sales diversity. In: Proceedings of the 8th ACM conference on Electronic commerce. pp. 192–199 (2007)
24. Gomez-Uribe, C.A., Hunt, N.: The netflix recommender system: Algorithms, business value, and innovation. ACM Transactions on Management Information Systems (TMIS) **6**(4), 1–19 (2015)
25. Harper, F.M., Konstan, J.A.: The movielens datasets: History and context. ACM Transactions on Interactive Intelligent Systems (TIIS) **5**(4), 1–19 (2015)
26. Heitz, L., Lischka, J.A., Birrer, A., Paudel, B., Tolmeijer, S., Laugwitz, L., Bernstein, A.: Benefits of diverse news recommendations for democracy: A user study. Digital Journalism pp. 1–21 (2022)

27. Helberger, N., Karppinen, K., D'acunto, L.: Exposure diversity as a design principle for recommender systems. Information, Communication & Society **21**(2)
28. Jannach, D., Zanker, M., Ge, M., Gröning, M.: Recommender systems in computer science and information systems–a landscape of research. In: International conference on electronic commerce and web technologies. pp. 76–87. Springer (2012)
29. Lam, S.K., Riedl, J.: Shilling recommender systems for fun and profit. In: Proceedings of the 13th WWW conference. pp. 393–402 (2004)
30. Lesota, O., Melchiorre, A., Rekabsaz, N., Brandl, S., Kowald, D., Lex, E., Schedl, M.: Analyzing item popularity bias of music recommender systems: Are different genders equally affected? In: 15th ACM RecSys. pp. 601–606 (2021)
31. Levy, M., Bosteels, K.: Music recommendation and the long tail. In: 1st WOMRAD, ACM RecSys, 2010, Barcelona, Spain. Citeseer (2010)
32. Mansoury, M., Abdollahpouri, H., Pechenizkiy, M., Mobasher, B., Burke, R.: Feedback loop and bias amplification in recommender systems. In: Proceedings of the 29th ACM CIKM. pp. 2145–2148 (2020)
33. McNee, S.M., Riedl, J., Konstan, J.A.: Being accurate is not enough: how accuracy metrics have hurt recommender systems. In: CHI'06 extended abstracts on Human factors in computing systems. pp. 1097–1101 (2006)
34. Milano, S., Taddeo, M., Floridi, L.: Recommender systems and their ethical challenges. Ai & Society **35**(4), 957–967 (2020)
35. Ni, J., Li, J., McAuley, J.: Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 188–197 (2019)
36. Oh, J., Park, S., Yu, H., Song, M., Park, S.T.: Novel recommendation based on personal popularity tendency. In: 2011 IEEE 11th International Conference on Data Mining. pp. 507–516. IEEE (2011)
37. Park, Y.J., Tuzhilin, A.: The long tail of recommender systems and how to leverage it. In: Proceedings of the 2008 ACM RecSys. pp. 11–18 (2008)
38. Pazzani, M.J., Billsus, D.: Content-based recommendation systems. In: The adaptive web, pp. 325–341. Springer (2007)
39. Qi, T., Wu, F., Wu, C., Huang, Y.: Pp-rec: News recommendation with personalized user interest and time-aware news popularity. arXiv:2106.01300 (2021)
40. Rahmani, H.A., Deldjoo, Y., Tourani, A., Naghiaei, M.: The unfairness of active users and popularity bias in point-of-interest recommendation. arXiv preprint arXiv:2202.13307 (2022)
41. Said, A., Bellogín, A.: Comparative recommender system evaluation: benchmarking recommendation frameworks. In: The 8th ACM RecSys. pp. 129–136 (2014)
42. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: The adaptive web, pp. 291–324. Springer (2007)
43. Schwartz, R.D.: Artificial intelligence as a sociological phenomenon. Canadian Journal of Sociology/Cahiers canadiens de sociologie pp. 179–202 (1989)
44. Senecal, S., Nantel, J.: The influence of online product recommendations on consumers' online choices. Journal of Retailing **80**(2), 159–169 (2004)
45. Swearingen, K., Sinha, R.: Beyond algorithms: An hci perspective on recommender systems. In: ACM SIGIR 2001 RecSys workshop. vol. 13, pp. 1–11 (2001)
46. Tang, T.Y., Winoto, P.: I should not recommend it to you even if you will like it: the ethics of recommender systems. New Review of Hypermedia and Multimedia **22**(1-2), 111–138 (2016)
47. Varshney, L.R.: Respect for human autonomy in recommender systems. arXiv preprint arXiv:2009.02603 (2020)

48. Yao, S., Huang, B.: Beyond parity: Fairness objectives for collaborative filtering. Advances in Neural Information Processing Systems **30** (2017)
49. Zhang, M., Hurley, N.: Niche product retrieval in top-n recommendation. In: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. vol. 1, pp. 74–81. IEEE (2010)
50. Zhao, X., Niu, Z., Chen, W.: Opinion-based collaborative filtering to solve popularity bias in recommender systems. In: International Conference on Database and Expert Systems Applications. pp. 426–433. Springer (2013)
51. Zhao, Z., Chen, J., Zhou, S., He, X., Cao, X., Zhang, F., Wu, W.: Popularity bias is not always evil: Disentangling benign and harmful bias for recommendation. arXiv preprint arXiv:2109.07946 (2021)
52. Ziegler, C.N., McNee, S.M., Konstan, J.A., Lausen, G.: Improving recommendation lists through topic diversification. In: Proceedings of the 14th International Conference on World Wide Web. pp. 22–32 (2005)