

Bias and Feedback Loops in Music Recommendation: Studies on Record Label Impact*

PETER KNEES, TU Wien, Faculty of Informatics, Austria and Georgia Institute of Technology, School of Music, USA

ANDRES FERRARO, Mila – Quebec Artificial Intelligence Institute, Canada and McGill University, Canada

MORITZ HÜBLER, TU Wien, Faculty of Informatics, Austria

We investigate the dimension of record labels in music recommendation datasets and study their impact on recommender systems. While music recommender systems research traditionally focuses on dimensions and metadata such as artist or genre, other dimensions such as popularity and gender have recently drawn increased interest. We argue that also the role of record labels deserves consideration in this process. To study their effect, we present a multi-stage web crawling approach that retrieves record label information for individual albums as well as an assignment to a major record company (Universal, Sony, Warner, or Independent). Using this information, we augment existing datasets to enable further analyses. We present analyses of record label diversity on two datasets, namely the Spotify Million Playlist Dataset and the LFM-2b dataset using Last.fm listening profiles. Based on the additional information, we can show different characteristics and identify particular biases. Additionally, we present the results of first experiments with regard to feedback loop simulation and the stability of record label distribution in the recommendation process.

CCS Concepts: • **Information systems** → **Recommender systems**; **Music retrieval**.

Additional Key Words and Phrases: music recommender systems, bias, feedback loops, music record labels

1 INTRODUCTION AND RELATED WORK

With the sheer amount of music available on commercial online music streaming services,¹ music recommendation has not only become a commodity in music listening and discovery but a necessity. In their most common implementation and use case, music recommender systems have the goal to deliver the most suited tracks or artists to users in the right context [27]. For research, the problem under investigation is therefore often reduced to take into consideration the information (i.e. metadata) of artist name, track title, album title, user id, and context, e.g. timestamps or location of listening events, or playlist that contains a track and its given label (see [26] for a recent overview of existing research datasets). While this simplified representation fits well to the general domain-agnostic algorithms and models in recommender systems research, it neglects the complexity of the process of music distribution and the broad spectrum and goals of involved stakeholders [2, 14].

Existing work therefore investigates the possibilities of making recommendations *fair* with regard to different actors [8], such as *item providers*, specifically music artists in case of [20], especially with a focus on attributes such as gender [10, 28]. This also involves identifying feedback loop patterns that give rise to or amplify bias in the data over several cycles of recommendations by incorporating recommended items into the user profile [10, 19]. One long-term consequence of feedback loops can be the reduction of diversity in the recommendations made. Such a lack of diversity can manifest in real-world consequences such as decreased exposition of items and their providers (artists, copyright owners, and labels) to users, therefore resulting in less revenue for providers, as well as impact on the shaping of musical tastes, cf. [5, 9, 23].

These implications are but a few of the multiple aspects of music recommendation, leading to multiple objectives to consider or be aware of in this process. For instance, another very important—nonetheless often neglected—aspect

* Copyright 2022 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Presented at the MORS workshop held in conjunction with the 16th ACM Conference on Recommender Systems (RecSys), 2022, in Seattle, USA.

¹80–90 million tracks as of 2022, cf. <https://newsroom.spotify.com/company-info/>, <https://www.apple.com/apple-music/>

and stakeholder with respect to licensing and availability of music on streaming platforms (therefore a different *item provider*, who requests “fair” treatment) is the distributing record label. Record labels can be independent or owned by (or associated with) one of the three worldwide operating major record labels Sony Music Entertainment, Universal Music Group, and Warner Bros. Records. In comparison, major labels have a dominant position over independent labels and also exert their power to influence and expand their shares and presence on music streaming platforms [4, 11, 12, 24], ultimately shaping listening preferences and consumption data. Regardless of its importance, however, except for an exploratory study [15], the dimension of record labels has not yet received much attention in recommender systems research.

In this work, we take first steps in this direction. In particular, similar to, but going beyond [15], we first describe a strategy for obtaining the major record label for individual albums. While fine-grained sub-label information can often be obtained as metadata, e.g. via the Spotify API, deriving to which major record company (if any) this label belongs to is far more complicated. Some reasons for this is are the complex dependencies between (local) branches of globally acting record companies, changing distribution agreements, and strategies for “portfolio management” of assets within record companies, cf. [21]. To yield an assignment of a track/artist to a label and from there to a major record label or identify as an independent, we developed a multi-stage web crawling approach involving different sources. More precisely, from the retrieved record label information, we derive an assignment to one of the three worldwide major record companies Universal, Sony, and Warner or as independent distributor by incorporating information from Spotify, Discogs, and Wikipedia (Section 2). Using this assignment, we augment two publicly available datasets to analyze their distribution with regard to record label diversity and identify different characteristics and particular biases, as described in Section 3. Additionally, we are interested in the development and stability of the record label distribution throughout the recommendation process. To this end, in Section 4, we present the results of first simulation experiments to identify possible feedback loops wrt. record labels. In Section 5, we discuss our findings and point out possible implications that deserve further investigation in future work.

2 AUGMENTING MUSIC DATASETS WITH RECORD LABEL INFORMATION

In this first analysis, we focus on two datasets, namely the Spotify Million Playlist Dataset [6] and the LFM-2b dataset [25] built upon publicly available Last.fm listening profiles.

The **Spotify Million Playlist Dataset**² contains 1 million playlists created by US-based users on the Spotify. It includes playlist titles and track metadata, including Spotify track URIs. Overall, it comprises 2 million unique tracks by almost 300,000 artists. The typical use case for this dataset is playlist continuation, i.e. recommending tracks to fit a given playlist.

The **LFM-2b dataset**³ is a large-scale dataset consisting of listening histories of 120,000 Last.fm⁴ users, totalling over 2 billion listening events. Additional information provided comprises tags, lyrics features, and basic demographics of user. For record label assignment, we resort to the provided matching with Spotify track URIs, effectively reducing the dataset from comprising 50 million unique tracks to 2.4 million. Listening events without known Spotify track URIs are therefore removed from analyses. We deliberately further remove all listening profiles with less than 30 unique tracks listened to, to exclude less interesting cases wrt. diversity analyses (748 users). As this dataset contains comprehensive listening histories of users, the typical use case for this dataset is user taste profiling and listening continuation.

²<https://www.aicrowd.com/challenges/spotify-million-playlist-dataset-challenge>

³<http://www.cp.jku.at/datasets/LFM-2b/>

⁴<https://www.last.fm>

Going beyond the work of Knees and Hübler [15] in terms of coverage and incorporated sources, we propose to augment existing datasets with record label information and derive information on their relation with a major record label. Our approach consists of multiple steps that build on top of each other and include information from additional sources, as sketched in the following:

- (1) **Preprocessing:** Initially, we obtain low-level record label information and/or copyright information per track or album, e.g. by using the Spotify API for a given Spotify track URI. On this level, as starting point, we identify 170,000 and 110,000 unique record labels, to appear in MPD and LFM-2b, resp.
- (2) **Mapping trivial cases:** In a first pass, already trivial cases are mapped based on matching tokens, e.g. the low-level record label *Universal Group* belongs to the major label *Universal Music Group*;
- (3) **Discogs label crawler:** Discogs is a public, user-generated music information platform and marketplace with detailed metadata.⁵ We harvest information to link and classify low-level record labels using the provided API.
- (4) **Wikipedia label crawler:** Similar to the previous step, we harvest label information from Wikipedia.⁶ In comparison to Discogs, Wikipedia provides information in a less structured way. Therefore, we resort to infoboxes of pages on artists and record labels, in particular the items *parent company*, *distributors*, and *labels*.
- (5) **Interim label mapping:** Evaluation and incorporation of the additionally collected information from the previous crawler steps. Beyond mere similarity matching as performed in the previous steps, this involves traversing the label hierarchies extracted to identify top-level companies or previously classified labels.
- (6) **Copyright classification:** To recheck assignments made, we further analyze copyright information obtained in the first step to create an alias dictionary of frequent and decisive copyright tokens. The idea is that this information is usually more descriptive, hence by identifying frequent terms for known major assignments, additional links can be uncovered. This is used for both, classification of still unassigned labels and correction of previous assignments.
- (7) **Final label mapping:** For all still unknown and unclassified low-level record labels, we assume no connection to a major label and hence classify them as *Independent*. In this final step, also a manual check-up and possible corrections can be applied, if resources and domain knowledge are available.

The progress of assignment and distribution of major labels throughout the steps of incorporating information from different sources can be seen in Fig. 1. For the MPD, we can see how each individual step and additional source adds more information to the data, with the distribution among majors remaining consistent. Note that the assignment for LFM-2b builds upon the gained information from MPD, i.e. derived major labels for overlapping and thus already known Spotify track URIs, in the first (trivial) assignment step. Therefore coverage is already high. Nonetheless it can be seen that the distribution of majors vs. Independent differs substantially, i.e., the fraction of independent tracks is much higher in LFM-2b than in the MPD. Whether this is a bias introduced by the filtering down to provided tracks URIs in the LFM-2b dataset needs further investigation and can not be answered at this point.

Among the major labels, we see a similar distribution with Universal taking the largest share of tracks in the sets, followed by Sony and Warner, broadly reflecting worldwide market shares. The relative occurrence of major labels after the final assignment (rightmost columns in Fig. 1) is as follows, ordered as Universal – Sony – Warner – Independent: 41.11% – 25.87% – 18.97% – 14.05% for MPD and 24.70% – 18.53% – 14.43% – 42.34% for LFM-2b.

⁵<https://www.discogs.com>

⁶<https://www.wikipedia.org>

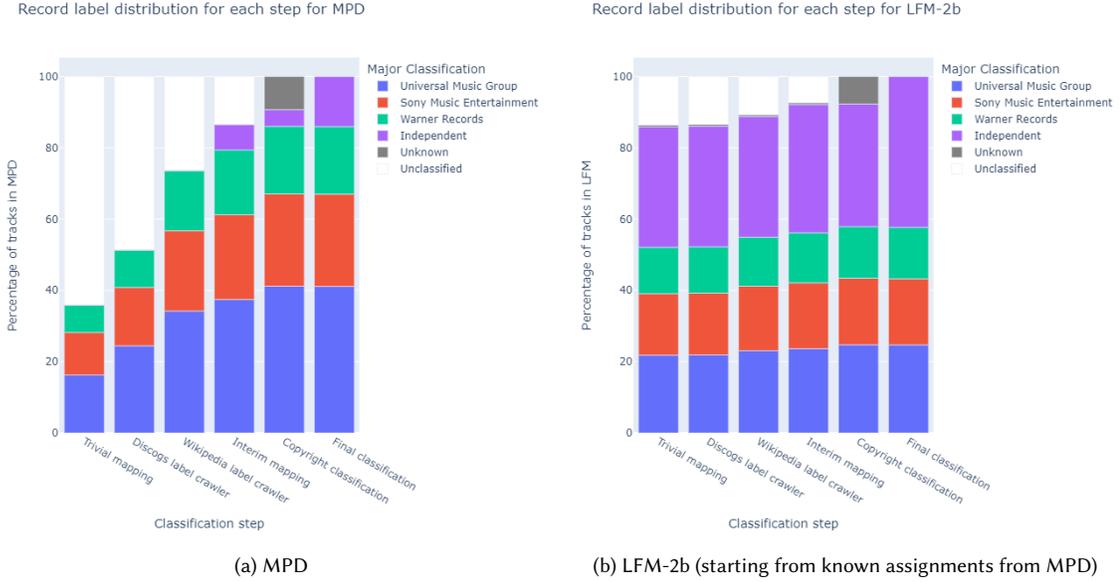


Fig. 1. Development of major label assignment over across individual steps of the crawler.

The source code as well as the obtained label classification (original label and copyright as well as derived major company) is made available as resource for the community.⁷

3 MEASURING RECORD LABEL DIVERSITY

Following the analysis conducted by Knees and Hübler [15], in the next step we are interested in how diverse individual playlist in MPD and user listening profiles in LFM-2b are according to major record labels. As such we also calculate the Simpson index per playlist (MPD) or user profile (LFM-2b), which measures the probability that two tracks within a playlist/profile belong to the same major label: $\lambda = \sum_{i=1}^R p_i^2$, where R describes the richness of the classes, in our case this is the number of major labels including *Independent*, i.e. $R = 4$, p_i is the probability for a class i that a randomly drawn track belongs to this major label. A low λ value therefore stands for high diversity, a high λ for low diversity [29]. In Fig. 2 we can see the distribution of Simpson indexes in both datasets, portraying the diversity of major labels per playlist/user profile. For both datasets we see the modal value in the lower range of λ , indicating that the majority of playlists and listening profiles have high diversity regarding record labels. Confirming the overall trend uncovered in [15], the MPD shows an exponential decay regarding diversity with a curious outlier peak at the high end of the scale. For LFM-2b, we observe an almost linear decay with very few (almost) completely homogeneous listening histories.

To further analyze the extreme cases of low diversity playlists and listening histories, we focus on the distribution of majors in playlists with a high Simpson index. Fig. 3 shows the distribution of tracks of major/independent labels as we increase a threshold for the Simpson index from 0.7 to 1, analysing how many tracks belong to a major label in the selected subgroup of playlists. Specifically, for each label, we show their fraction in the increasingly homogeneous playlists and histories as the threshold increases.

⁷https://github.com/nostromo7/MT_label_crawler

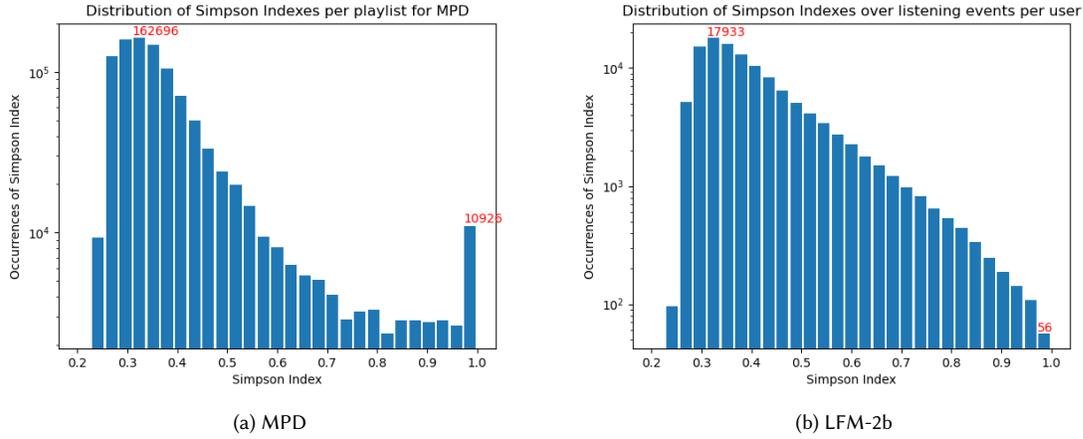


Fig. 2. Distribution of Simpson index of major label diversity per playlist (MPD) and user profiles (LFM-2b)

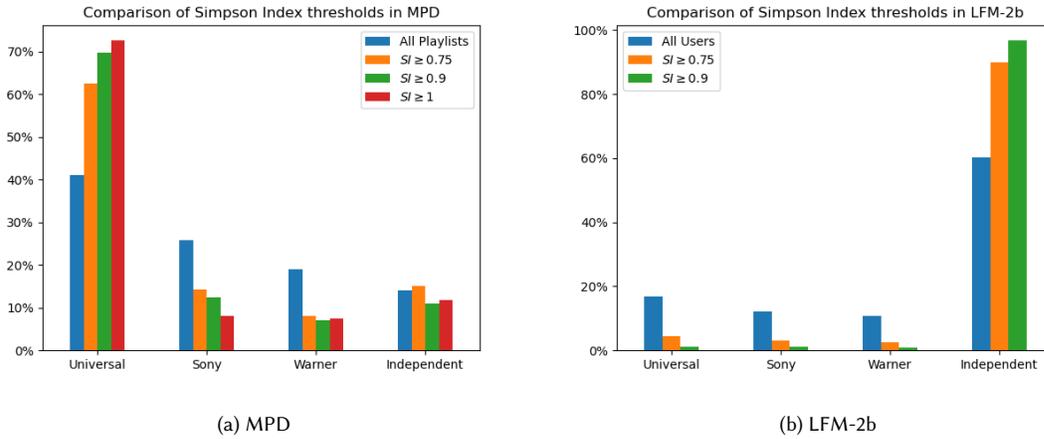


Fig. 3. Comparison of major distribution for different Simpson index (SI) thresholds

For MPD (Fig. 3a), we can observe a dominance of tracks by Universal that further increases as record label diversity gets lower. For the rightmost bars, i.e. playlists consisting exclusively of tracks by one major record label, the fraction of Universal-exclusive playlists is 72.6%. Thus, we can observe a trend that the overall most present label becomes exceedingly more present as diversity decreases. We will discuss this in more detail in Section 5. The same trend can be observed on LFM-2b (Fig. 3b), although the absolute numbers of low diversity profiles are much lower in comparison to the outlier peak in MPD. Here we can observe a trend towards Independent, which is also the most present label in the overall distribution. With low diversity, users seem to have a preference for independently distributed music, or possibly for music that can not be assigned a major label. However, as the absolute number of these cases is very low, a deeper investigation will have lower priority in the future.

Major Label	Iteration 0 (1st cycle)		Iteration 1 (2nd cycle)	
	<i>Recommended</i>	<i>First</i>	<i>Recommended</i>	<i>First</i>
Universal	45.70	1.78	45.66	1.73
Sony	27.32	5.22	27.39	5.26
Warner	20.44	9.05	20.37	8.89
Independent	9.11	31.37	9.11	30.65

Table 1. Result of feedback loop simulation on full MPD dataset

4 INVESTIGATING FEEDBACK LOOPS

With the additional information of record labels available, we also want to investigate the effects that recommender systems have on their distribution. To this end, we conducted first simulation experiments over multiple iterations of a recommender system in line with the experimental setup by Ferraro et al. [10] used to investigate feedback loops and their impact on gender distribution in recommendations. Here, we present first results on the effects of matrix factorization based collaborative filtering recommendation using the Alternating Least Squares algorithm (ALS) [31] for implicit feedback data [13] as implemented in Spark.⁸

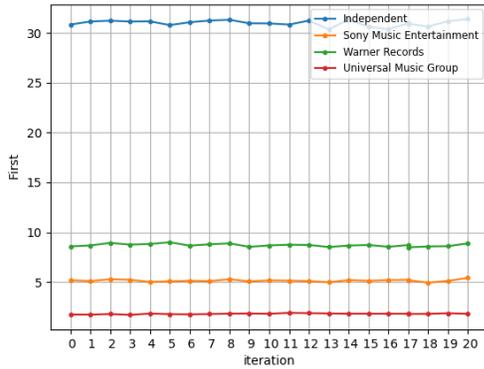
For lack of a truly user behavior driven simulation strategy, we adopt the following scenario to model the impact of recommendations on user behavior and subsequently its effect on recommendations. Starting from the user profiles initially given in the dataset, we generate a list of the top 100 recommendations for each user by means of ALS. To mimic user behavior, we assume that each user accepts the top 10 of the 100 recommended tracks, increasing the interactions in the user-item matrix for these 10 tracks for each user. The model is retrained after each iteration using the ALS algorithm and the full process is repeated up to n iterations. The metrics used to investigate recommendation behavior measuring the probability and representation of a record label in the recommendations are: *First*, indicating the first position of a specific label in the 100 recommendations, averaged over all users; and *Recommended*, referring to the percentage of how many of the 100 recommendations are represented by a specific label, again averaged over all users.

For the MPD, each playlist is interpreted as a user and the tracks from the dataset as items for the user-item matrix. The first experiment, where the pool of tracks to recommend from consists of all tracks in the dataset with the possibility for tracks to appear repeatedly in a playlist, was stopped due to limited resources. To cut complexity, the experiment on the MPD was repeated in a reduced format, only recommending songs to randomly selected 1% of the playlists (Table 1 shows the detailed results for the first two iterations). For LFM-2b, we use the given set of users with the further restriction of removing tracks that appear less than 15 times in the full dataset. In contrast to MPD, no tracks are re-recommended.

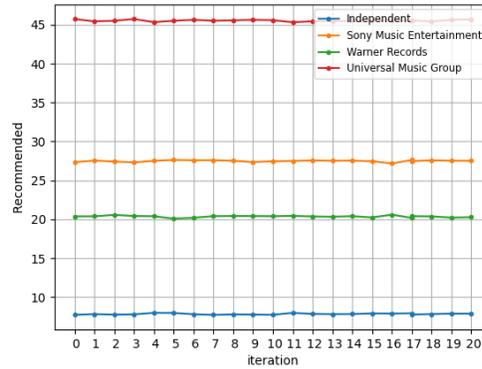
Results over iterations regarding first position (*First*) and representation (*Recommended*) can be found in Figs. 4 and 5 for MPD and LFM-2b, resp. For the MPD results remain stable and similar to the overall distribution over iterations and no amplifications or feedback loops could be discovered when running the full dataset (for only few iterations) or the reduced setup (only recommending songs to 1% of playlists). Given these limitations in the conducted experiments, we refrain from concluding that feedback loops do not exist in this scenario.

For LFM-2b, the *Recommended* metric (Fig. 5b) shows a very different representation than the overall distribution in the dataset. While in the first iteration (0), Independent is represented strongest with Universal as close second (31.5% vs. 30%; compared to 42.34% vs. 24.70% overall), within three iterations, Universal takes first place, largely on par with Independent just below 31%. For Sony and Warner, we also see an over-representation in relation to the overall

⁸see <https://spark.apache.org/docs/3.3.0/ml-collaborative-filtering.html>

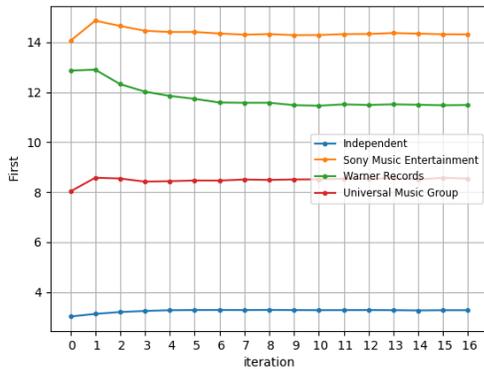


(a) First position of major label

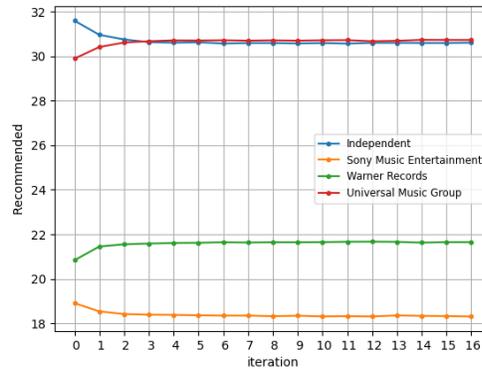


(b) Representation of major label in recommendations

Fig. 4. Results of reduced feedback simulations with the MPD, recommending songs only for 1% of the playlists. *First* and *Recommended* are averaged over all users.



(a) First position of major label



(b) Representation of major label in recommendations

Fig. 5. Results of the feedback simulations with the LFM-2b. *First* and *Recommended* are averaged over all users.

distribution, however with positions switched (Warner above Sony). Over multiple iterations, the distance between Warner and Sony even increases, giving further representation to Warner through recommendations.

5 DISCUSSION, CONCLUSIONS, AND FUTURE WORK

We have presented but the first results and efforts into the direction of investigating the impact and role of major record labels in music recommendation. Starting with two datasets of different origin, namely MPD as a playlist dataset and LFM-2b as a listening dataset, we could show very different characteristics and biases wrt. record label distribution. Of particular interest in that regard are the non-diverse outliers identified in MPD. While a much deeper analysis is needed in future work, upon first inspection, we could identify some of these playlists to contain collections of movie

soundtracks, thus exhibiting diverse artists, while being published under the same major label. This might partly reflect the different uses of playlists on platforms, such as structuring personal collections, which differ inherently from a log of listening events as found on Last.fm, cf. [17, 22].

With regard to the effects of recommender systems on record label distribution (i.e., one type of item providers), we could identify first feedback loop effects. Despite the dominance of independent labels in the LFM-2b set, major labels are over-represented in the recommendation process, with Universal's and Warner's over-representation even being further amplified over iterations. Further analyses need to be also linked to effects of popularity. While we can observe much more diversity in the overall distribution of LFM-2b, for recommendation, the popularity bias of the most successful tracks is most likely driving the process, cf. [3, 16]. While various strategies exist to control popularity bias and debias feedback loops, e.g. [1, 30], the role of record labels remains a complex one, and presents but one objective within the multi-objective task(s) of music recommendation. In this light, uncovering an overall definition of "recommendation fairness" (cf. [7, 18]) in this context is a longer-term objective, and will potentially merely guide a process of reflection on the status quo in music distribution. It is clear that facing a situation where different market participants hold different market shares is not per se an "unfair" scenario. Nonetheless, assessing and modeling the different stakeholders in a recommendation scenario such as music recommendation is essential and the steps of analytics, recommendation impact and feedback loop analysis on existing resources informs these reflections and subsequent discussions—in particular as the prevalent means of music distribution can lead to non-transparent strategies in terms of opportunity and remuneration [4].

Future work will therefore pragmatically first expand the scope of datasets augmented and analyzed, as well as continue to investigate the effects that recommender systems have on diversity and representation in recommender results. This comprises investigating alternative measures of diversity, and revisiting and questioning the assumptions and methods that have been used in the experimental setup, and the simulation of the recommendation feedback loops, including the user behavior modeling and its assumption, and the choice of the recommendation algorithm and strategy. Beyond that, we are also interested in investigating whether, and if, to which extent, record labels and bias brought into the data can impact and steer the recommendation process itself and its implications for users and their experience, as well as on other stakeholders such as the artists. In the bigger context, this extends to the overall objective in terms of multi-stakeholder and multi-objective recommender systems research, in which we are interested how existing power mechanisms play into the process of recommendation, and whether different recommendation techniques are even instrumental in this process.

ACKNOWLEDGMENTS

This research was funded in whole, or in part, by the Austrian Science Fund (FWF) [P33526]. For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

REFERENCES

- [1] Himan Abdollahpouri, Robin Burke, and Bamshad Mobasher. 2017. Controlling Popularity Bias in Learning-to-Rank Recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems* (Como, Italy) (*RecSys '17*). Association for Computing Machinery, New York, NY, USA, 42–46. <https://doi.org/10.1145/3109859.3109912>
- [2] Himan Abdollahpouri and Steve Essinger. 2017. Multiple Stakeholders in Music Recommender Systems. arXiv:1708.00120 [cs.CY]
- [3] Himan Abdollahpouri and Masoud Mansoury. 2020. Multi-sided Exposure Bias in Recommendation. arXiv:2006.15772 [cs.IR]

- [4] Daniel Antal, Amelia Fletcher, and Peter Ormosi. 2021. Music Streaming: Is It a Level Playing Field? *CPI Antitrust Chronicle* (23 2 2021). <https://www.competitionpolicyinternational.com/music-streaming-is-it-a-level-playing-field/>
- [5] Georgina Born. 2020. Diversifying MIR: Knowledge and Real-World Challenges, and New Interdisciplinary Futures. *Transactions of the International Society for Music Information Retrieval* 3, 1 (2020), 193–204. <https://doi.org/10.5334/tismir.58>
- [6] Ching-Wei Chen, Paul Lamere, Markus Schedl, and Hamed Zamani. 2018. Recsys Challenge 2018: Automatic Music Playlist Continuation. In *Proceedings of the 12th ACM Conference on Recommender Systems* (Vancouver, British Columbia, Canada) (*RecSys '18*). ACM, New York, NY, USA, 527–528. <https://doi.org/10.1145/3240323.3240342>
- [7] Yashar Deldjoo, Dietmar Jannach, Alejandro Bellogin, Alessandro Difonzo, and Dario Zanzonelli. 2022. Fairness in Recommender Systems: Research Landscape and Future Directions. ResearchGate. *User Modeling and User-Adapted Interaction* (under review) (2022).
- [8] Karlijn Dinmissen and Christine Bauer. 2022. Fairness in music recommender systems: a stakeholder-centered mini review. *Frontiers in Big Data* (2022). <https://doi.org/10.3389/fdata.2022.913608>
- [9] Andres Ferraro, Dmitry Bogdanov, Xavier Serra, and Jason Yoon. 2019. Artist and style exposure bias in collaborative filtering based music recommendations. arXiv:1911.04827 [cs.IR]
- [10] Andres Ferraro, Xavier Serra, and Christine Bauer. 2021. Break the Loop: Gender Imbalance in Music Recommenders. In *Proceedings of the 2021 Conference on Human Information Interaction and Retrieval* (Canberra ACT, Australia) (*CHIIR '21*). Association for Computing Machinery, New York, NY, USA, 249–254. <https://doi.org/10.1145/3406522.3446033>
- [11] David Hesmondhalgh and Leslie Meier. 2015. Popular music, independence and the concept of the alternative in contemporary capitalism. In *Media Independence: Working with Freedom or Working for Free?*, J. Bennett (Ed.). Routledge. <https://eprints.whiterose.ac.uk/82310/>
- [12] David Hesmondhalgh, Richard Osborne, Hyojung Sun, and Kenny Barr. 2021. Music Creators' Earnings in the Digital Era. *Intellectual Property Office Research Paper* (2021). <https://doi.org/10.2139/ssrn.4089749>
- [13] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. In *2008 Eighth IEEE International Conference on Data Mining*. 263–272. <https://doi.org/10.1109/ICDM.2008.22>
- [14] Dietmar Jannach and Gediminas Adomavicius. 2016. Recommendations with a Purpose. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) (*RecSys '16*). ACM, New York, NY, USA, 7–10. <https://doi.org/10.1145/2959100.2959186>
- [15] Peter Knees and Moritz Hübler. 2019. Towards Uncovering Dataset Biases: Investigating Record Label Diversity in Music Playlists. In *Proceedings of the 1st Workshop on Designing Human-Centric MIR Systems*. Delft, The Netherlands.
- [16] Dominik Kowald, Markus Schedl, and Elisabeth Lex. 2020. The Unfairness of Popularity Bias in Music Recommendation: A Reproducibility Study. In *Advances in Information Retrieval*, Joemon M. Jose, Emine Yilmaz, João Magalhães, Pablo Castells, Nicola Ferro, Mário J. Silva, and Flávio Martins (Eds.). Springer International Publishing, Cham, 35–42. https://doi.org/10.1007/978-3-030-45442-5_5
- [17] Jin Ha Lee, Hyerim Cho, and Yea-Seul Kim. 2016. Users' Music Information Needs and Behaviors: Design Implications for Music Information Retrieval Systems. *Journal of the Association for Information Science and Technology* 67, 6 (jun 2016), 1301–1330. <https://doi.org/10.1002/asi.23471>
- [18] Yunqi Li, Hanxiong Chen, Shuyuan Xu, Yingqiang Ge, Juntao Tan, Shuchang Liu, and Yongfeng Zhang. 2022. Fairness in Recommendation: A Survey. arXiv:2205.13619 [cs.IR]
- [19] Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher, and Robin Burke. 2020. Feedback Loop and Bias Amplification in Recommender Systems. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (Virtual Event, Ireland) (*CIKM '20*). Association for Computing Machinery, New York, NY, USA, 2145–2148. <https://doi.org/10.1145/3340531.3412152>
- [20] Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. 2018. Towards a Fair Marketplace: Counterfactual Evaluation of the Trade-off between Relevance, Fairness & Satisfaction in Recommendation Systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (Torino, Italy) (*CIKM '18*). Association for Computing Machinery, New York, NY, USA, 2243–2251. <https://doi.org/10.1145/3269206.3272027>
- [21] Keith Negus. 2001. The corporate strategies of the major record labels and the international imperative. In *Global Repertoires*, Andreas Gebesmar and Alfred Smudits (Eds.). Routledge, 21–31.
- [22] Martin Pichl, Eva Zangerle, and Günther Specht. 2015. Towards a Context-Aware Music Recommendation Approach: What is Hidden in the Playlist Name?. In *Proceedings of the 2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. 1360–1365. <https://doi.org/10.1109/ICDMW.2015.145>
- [23] Lorenzo Porcaro, Carlos Castillo, and Emilia Gómez. 2021. Diversity by Design in Music Recommender Systems. *Transactions of the International Society for Music Information Retrieval* 4, 1 (2021), 114–126. <https://doi.org/10.5334/tismir.106>
- [24] Robert Prey, Marc Esteve Del Valle, and Leslie Zwerwer. 2022. Platform pop: disentangling Spotify's intermediary role in the music industry. *Information, Communication & Society* 25, 1 (2022), 74–92. <https://doi.org/10.1080/1369118X.2020.1761859>
- [25] Markus Schedl, Stefan Brandl, Oleg Lesota, Emilia Parada-Cabaleiro, David Penz, and Navid Rekasaz. 2022. LFM-2b: A Dataset of Enriched Music Listening Events for Recommender Systems Research and Fairness Analysis. In *ACM SIGIR Conference on Human Information Interaction and Retrieval* (Regensburg, Germany) (*CHIIR '22*). Association for Computing Machinery, New York, NY, USA, 337–341. <https://doi.org/10.1145/3498366.3505791>
- [26] Markus Schedl, Peter Knees, Brian McFee, and Dmitry Bogdanov. 2022. Music Recommendation Systems: Techniques, Use Cases, and Challenges. In *Recommender Systems Handbook* (3rd ed.), Francesco Ricci, Lior Rokach, and Bracha Shapira (Eds.). Springer US, New York, NY, 927–971. https://doi.org/10.1007/978-1-0716-2197-4_24

- [27] Markus Schedl, Peter Knees, Brian McFee, Dmitry Bogdanov, and Marius Kaminskas. 2015. Music Recommender Systems. In *Recommender Systems Handbook* (2nd ed.), F. Ricci, L. Rokach, and B. Shapira (Eds.). Springer US, Boston, MA, 453–492. https://doi.org/10.1007/978-1-4899-7637-6_13
- [28] Dougal Shakespeare, Lorenzo Porcaro, Emila Gomez, and Carlos Castillo. 2020. Exploring Artist Gender Bias in Music Recommendation. In *Proceedings of the ImpactRS Workshop at ACM RecSys '20*, Vol. 2697. CEUR-WS.
- [29] Edward H. Simpson. 1949. Measurement of Diversity. *Nature* 163, 4148 (1949), 688–688. <https://doi.org/10.1038/163688a0>
- [30] Wenlong Sun, Sami Khenissi, Olfa Nasraoui, and Patrick Shafto. 2019. Debiasing the Human-Recommender System Feedback Loop in Collaborative Filtering. In *Companion Proceedings of The 2019 World Wide Web Conference* (San Francisco, USA) (*WWW '19*). Association for Computing Machinery, New York, NY, USA, 645–651. <https://doi.org/10.1145/3308560.3317303>
- [31] Yunhong Zhou, Dennis Wilkinson, Robert Schreiber, and Rong Pan. 2008. Large-Scale Parallel Collaborative Filtering for the Netflix Prize. In *Algorithmic Aspects in Information and Management*, Rudolf Fleischer and Jinhui Xu (Eds.). Springer, Berlin, Heidelberg, 337–348.