Maximizing the Engagement: Exploring New Signals of Implicit Feedback in Music Recommendations

Andres Ferraro
Music Technology Group - Universitat Pompeu Fabra
Barcelona, Spain
andres.ferraro@upf.edu

Sergio Oramas
Pandora
Oakland, United States

Massimo Quadrana
Pandora
Oakland, United States

Xavier Serra
Music Technology Group - Universitat Pompeu Fabra
Barcelona, Spain

ABSTRACT

Music recommendation engines play a pivotal role in connecting artists with their listeners. Optimizing myopically only for user satisfaction may lead systems to recommend just a small fraction of all the available artists, or to recommend artists to users who might engage with them only in the short-term. In this work, we investigate such effects by exploring different signals of implicit feedback provided by users when using a music service (i.e., counting the number of tracks, days or times a user listens to an artist) and propose novel combined signals. Our approaches are evaluated over four different datasets, combining traditional user-centered evaluation metrics with artist-based ones, which allows us to measure the quality of the recommendations and the potential engagement with the recommended artists. Our experiments reveal that the selection of the implicit feedback signal has a significant impact on the quality of the recommendations. In addition, we show that the proposed signals increase the chances of a higher engagement between users and the artists they get recommended.

CCS CONCEPTS

• **Information systems → Recommender systems**.

KEYWORDS

Recommender Systems, Multi-stakeholder Recommendation, Implicit Feedback

1 INTRODUCTION

Streaming platforms play a fundamental role nowadays in music consumption. Recommender systems are a fundamental part of these platforms, allowing users to explore their music collections, which otherwise would be unmanageable given their size. Following recent trends in the field of Recommender Systems[1, 6, 7] it is clear the importance of taking into account the interests of all the stakeholders involved when making recommendations (e.g., users, artists, record labels or the service itself). In this work, we explore how user implicit feedback can be leveraged by a music recommender system to provide more value to both users and artists.

Proceedings of the ImpactRS Workshop at ACM RecSys '20, September 25, 2020, Virtual Event, Brazil.

We start from the assumption that users bring different value to the artists the system recommends to them. Artists likely prefer to be recommended to those users who will actively engage more with their production, such as listening to their music, buying their latest releases, merchandising and concert tickets. Unfortunately, such strong engagement metrics can hardly be tracked by the existing music streaming platforms, which instead rely on implicit interaction signals such as play counts or session lengths to gauge the engagement and satisfaction of users with the recommended content [13]. Specifically to artist recommendation, most music recommender systems consider the number of times a user plays a track or an artist (the *playcount*) as the main engagement signal [10]). However, playcounts alone can hardly gauge all the different ways in which listeners "consume" an artist's production. E.g., a listener who frequently listens to only the same few tracks will unlikely be attracted by new releases by that artist or attend to their concerts. On the same line, listeners who played a few albums by the artist only for a few days in the past are likely less engaged with the artist than listeners who constantly listen to the artist's production over long periods. For these reasons, we believe it is crucial to use implicit interaction signals beyond the simple frequency of interaction with the artist. Therefore, we introduce novel signals that capture both the *breadth* of the listener's engagement with the artist's production, computed as the number of distinct artist's tracks played by the listener (*trackcounts*), and the *temporal extent* over which the listener engaged with the artist, computed as the number of days a user listens to an artist (*daycounts*).

We study the behavior of state-of-the-art Implicit Matrix Factorization (ALS) [8] over these new engagement signals, both from the listener's and artist's perspective. We evaluate both the case in which these relevance functions are used as implicit relevance values to train ALS, and when they are used as evaluation metrics over test data in combination with other traditional offline evaluation metrics, such as MAP and NDCG. Our experiments over four different datasets show that the selection of the implicit relevance metric used to train ALS has huge implications on the extent to which listeners will engage with the recommended artists and that there is no single optimal engagement metric for all datasets.

2 RELATED WORK

If we want to generate recommendations that will lead to more fans for an artist, the first difficulty that we have to solve is how to measure that. The most common signal used for measuring how much a user likes an artist is playcounts [17]. However, this signal doesn't capture relevant information about the interaction of users with artists (e.g., during how much time, how many times a day, how many songs of the artist).

In the music domain, previous work uses different signals as a way to describe the behavior of the users from what they listen to. Farrahi et al. [5] compares the listening activities of the users in terms of playcounts, diversity and *mainstreamness*. Vigliensoni and Fujinaga [21] goes beyond and from the listeners' activity extract the *exploratoryness*, *mainstreamness* and *genderness* which defines how the users interact with the content. Oliveira et al. [14] propose a multiobjective optimization approach to find a balance for diversity in the recommendation. As far as we know there is no prior work that tries to capture using the implicit feedback signals how much a user is engaged with a music artist.

Recently, Dacrema et al. [3] shows that not much improvement has been achieved for top-n recommendations with the introduction of deep learning approaches. Based on this idea, instead of proposing a better algorithm, we hypothesize that improving the pre-processing of the input signals will lead to an improvement in the quality of the recommendations.

For the evaluation of the recommendations, there are strong limitations in the offline evaluation of recommender systems that we also face. One of the strongest limitations is that recommending something outside of the ground truth items does not mean that the user would not like it. In addition, offline evaluation usually has a popularity bias, favoring the algorithm that recommends more popular items [2, 7, 16, 19], which can also vary depending on demographic aspects of the users [4], and can have a disparate bias for different user groups [11].

3 IMPLICIT ENGAGEMENT SIGNALS

In this section, we describe the implicit engagement signals that we use to train and evaluate artist recommendations generated by Implicit Matrix Factorization (ALS). We consider both *raw signals* that are extracted directly from the user listening logs, and *composite signals* which are combinations of the raw ones.

3.1 Raw signals

Given a listener u and artist a, we extract the following raw signals from listening logs:

- $playcounts(u, a)$ is the number of tracks of a played by u;
- $binary(u, a)$ is the binarized playcount, i.e. $binary(u, a) = \mathbb{1}\{placounts(u, a) \geq 1\}$;
- $trackcounts(u, a)$ is the number of distinct tracks of a played by u;
- $daycounts(u, a)$ is the number of distinct days in which u listened to at least one track by a.

3.2 Composite signals

To capture multiple aspects of the listener's behavior with the artist in a single implicit signal, we combine the raw signals into two novel composite signals named *engagement* and $fidelity$.

- $engagement(u, a)$ is a discounted weighted combination of the playcounts accumulated by the listener u over the days they have listened to a. Specifically, weight plays on the first days of listening less than plays happening later down by using the following formula:

$$engagement(u, a) = \sum_{d=0}^{D} playcounts(u, a, d) * log(d)$$

Where D is the number of days a user listens to an artist and $playcounts(u, a, d)$ is the number of tracks of a played by u on day d.

- $fidelity(u, a)$ combines *engagement* with *trackcounts* into a single metric in the following way:

$$fidelity(u, a) = \alpha * trackcounts(u, a) + (1 - \alpha) * engagement(u, a)$$

The motivation for the given definition of Engagement is that we want to weight play interactions differently, according to the day they were played. We assume that plays on the first day are less valuable for the artist than plays on the subsequent days. We apply logarithm to the number of days to soften the impact of large day numbers, making this factor more determinant in the first days of listening. For example, the difference between the first and the third day of listening is larger than between the tenth and the twentieth. *Fidelity* combines the three raw signals by a linear combination of *trackcounts* and *engagement*, which is already combining *playcounts* and *daycounts*.

4 EVALUATION METRICS

To understand the quality of the recommendations both from the listeners' and from the artists' perspectives, we compute both *listener-centric* and *artist-centric* metrics. Listener-centric metrics are the following traditional offline accuracy metrics: Mean Average Precision (MAP) and normalized Discounted Cumulative Gain (*nDCG*) [17].

The artist-centric metrics are instead the average values of playcounts (PLAYS@K), trackcounts (TRACKS@K) and daycounts (DAYS@K) over all the artists that were recommended. We also compute the coverage (C@10) of the recommended artists as in [15]. Let A and U be the sets of artists and users in the dataset respectively, $A = \{a_1, a_2, ..., a_n\}$ and $U = \{u_1, u_2, ..., u_n\}$. We define the above artist-centric metrics as follows:

$$PLAYS@K = \frac{1}{|A|} \sum_{a \in A} \frac{\sum_{u \in U} hit@K(u, a) \cdot playcounts(u, a)}{\sum_{u \in U} hit@K(u, a)}$$

$$TRACKS@K = \frac{1}{|A|} \sum_{a \in A} \frac{\sum_{u \in U} hit@K(u, a) \cdot trackcounts(u, a)}{\sum_{u \in U} hit@K(u, a)}$$

$$DAYS@K = \frac{1}{|A|} \sum_{a \in A} \frac{\sum_{u \in U} hit@K(u, a) \cdot daycounts(u, a)}{\sum_{u \in U} hit@K(u, a)}$$

where $hit(a, u)$ returns 1 iff. a was recommended to u *and* belongs to the test set of that user. Finally, given L_u as the top-k artists recommended to user u, we define the catalog coverage of recommended artists as $C@K = \frac{\bigcup_u L_u}{|A|}$

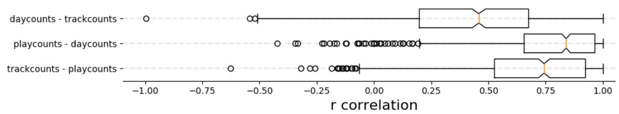


Figure 1: Correlation for *playcounts*, *daycounts* and *trackcounts* on 1000 artists.

5 DATASETS

In this work, we use four datasets of user-artist interactions with timestamps (see Table 1). Original datasets are filtered according to the following constraints. First, we discard all artists having less than 3 interactions and all users that interacted with less than 10 artists. Having users with less of these interactions would make the recommendations harder to evaluate. Then, we split each dataset on a temporal-basis by first sorting the interactions by timestamp, and then assign 80% of the events to the training set and the remaining 20% to the test set. Finally, since our goal is to study the impact of recommendations of artists that were not previously listened by the user, and symmetrically to shed a light on how artists can reach new audiences through recommendations, for each user we removed from the test set all the artists occurring in their training set. The resulting number of artists and users used on each dataset is detailed in Table 2[1].

If we analyze the distribution of playcounts for user-artist interactions, we observe that some datasets have a higher distribution of values bigger than 10, such as the lfm-1b. It is important to highlight this since, we can expect from lfm-1b dataset richer information regarding the interaction between the users and the artists compared to the nowplaying dataset for instance, where there are more values around one.

6 QUALITATIVE ANALYSIS OF THE RAW SIGNALS

We provide here a qualitative analysis of the raw signals introduced in Section 3. We hypothesize that these signals capture different and complementary aspects of how listeners engage with artists.

We measured the correlation between the described raw signals for 1000 random artists of the lfm-1b dataset. In Figure 1 we see that the highest average correlation is between *playcounts* and *daycounts*. However, for some artists, these values are not very correlated, which means that for those artists there could be a higher benefit of using other signals than only *playcounts*. Also note that for *daycounts* and *trackcounts* the correlation is lower for most of the artists.

We further illustrate this with two artists taken from this set, which have similar popularity (i.e., number of users) but different music styles. (a) Roland Pontinen is a pianist and composer of chamber music from Sweden and (b) The Honeycombs was a British band from the '60s influenced by The Beatles. In Figure 2, we plot the correlation between the raw signals for these two artists. Those graphs highlight some interesting differences in the way users engage with both artists: *trackcounts* and *daycounts*, and *trackcounts* and *playcounts* are more correlated for (b) (r=0.32) than for (a) (r=0.06), whereas *playcounts* and *daycounts* are more correlated for (a) (r=0.99) than for (b) (r=0.43). From these two analyses, we observe that the correlation of the raw signals can be different between artists. Therefore, they provide complementary information that could be useful for generating recommendations.

7 RECOMMENDATIONS USING ENGAGEMENT SIGNALS

At this stage of the work, we are mainly interested into knowing how the usual "listener-centric" training of recommenders impacts the artists who are recommended and, at the same time, we would like to understand whether the alternative formulation of implicit engagement signals that we proposed in Section 3 can be favorable to artists while keeping acceptable levels of (offline) recommendation quality to listeners. For these reasons, we decided to study the behavior of the Implicit Matrix Factorization with Alternating Least Squares (ALS) [8] in this scenario. ALS is known to be one of the most used collaborative filtering algorithms and a *de-facto* industrial standard. While we cannot know what algorithms are used by the various online music services available nowadays, the choice of ALS surely extends the applicability of our experimental results to many real-world music recommendation scenarios. We trained ALS[2] on all training datasets with each of the implicit engagement signals defined in Section 3 as relevance functions. For the case of *fidelity*, we decided to give the same weight to *engagement* and *trackcounts* (α=0.5) to simplify the experiments, but further optimization of these weights may lead to improved results. To measure the performance of the recommendations, we generate a list of 10 artists for each user (K=10) and use the metrics defined in Section 4. We tuned only the number of latent dimensions for each (dataset, relevance function) combination. The final number of dimensions used are 200, 200, 50 and 30 for lfm-1b, Streaming-service, 30music and nowplaying respectively.

In Table 3, we show the performance according to the listener-centric metrics (MAP@10 and nDCG@10) and the artist-centric metrics (PLAYS@10, TRACKS@10, DAYS@10 and C@10). Due to space restrictions, we do not report Precision@10. However, its values are in line with the reported user-centric metrics. It is worth noting that, by design, ALS optimizes for the ranking of the relevant items in the user's recommendation list.

The results show that there is no single relevance function that performs the best on all the datasets in terms of listener-centric metrics. Function *daycounts* performs the best for the 30music dataset, while *trackcounts* is the best relevance function for the nowplaying dataset. The *engagement* function, which is a discounted weighted combination of an artist's *trackcounts* over the days it was played by the listener, performs the best on the Streaming-service dataset. Interestingly, *binary* input obtains the worst performance for all the previous three datasets and the highest for lfm-1b.

From these results, we cannot say that an implicit engagement signal will give always the best accuracy from the listener's perspective. We hypothesize that it is related to the nature of each dataset, as they all present a different distribution of values.

However, we see more consistent results in all datasets according to artist-centric metrics. The *engagement* relevance function

[1]For reproducibility purposes code is provided: https://github.com/andrebola/artist-engagement

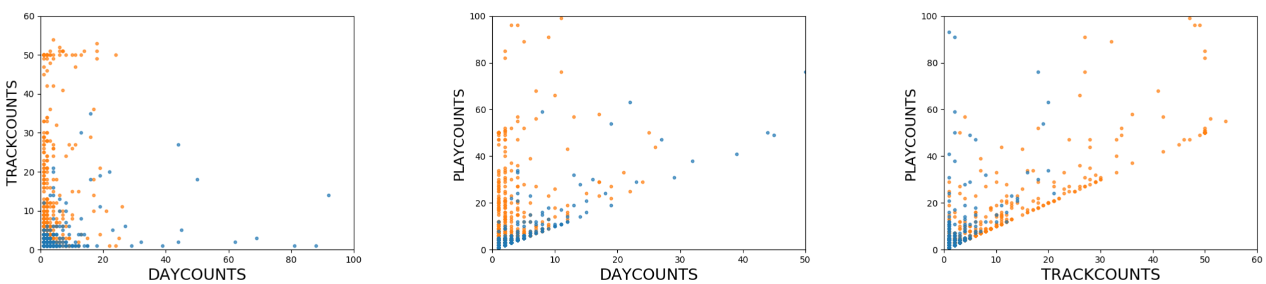
[2]We used the implementation available at https://implicit.readthedocs.io/en/latest/#

Table 1: Datasets used in the comparison.

lfm-1b	Large dataset with over a billion listening events containing playcounts and timestamp extracted from last.fm [18]
Streaming-service	Dataset obtained from [name-omitted] music streaming service for 6 months in 2019
Nowplaying	This dataset contains listening logs collected from Twitter [23]. We use a subset of the original dataset published by Ludewing et al. [12]
30music	Collected from last.fm [20] with the main purpose of session recommendations

Table 2: Information about the datasets.

Dataset	Users	Artists		Density	# Days		User-Artist Interaction	
		train	test	train	train	test	train	test
lfm-1b	119,692	693,436	111,086	0.0007	3073	122	61,443,465	517,903
Streaming-service	25,981	22,667	17,189	0.0028	147	36	1,655,600	235,653
Nowplaying	7,198	13,213	7,921	0.0033	428	107	318,250	25,124
30music	33,462	112,354	97,274	0.0010	292	73	3,888,882	1,307,575



(a) Users' *trackcounts* and *daycounts* values. **(b) Users' *playcounts* and *daycounts* values.** **(c) Users' *trackcounts* and *playcounts* values.**

Figure 2: Distribution of implicit raw signals for 'The Honeycombs' (blue) and 'Roland Pontinen' (orange) in the LFM-1b dataset.

outperforms all other relevance functions in terms of C@10 on all datasets. This suggests that using *engagement* as relevance function increases the fraction of artists that are effectively recommended in the top-10 wrt. all the other relevance functions. Function *engagement* performs particularly well also in terms of DAYS@10, for which it is the best function in all but the 30music, where it is the second-best. The *engagement* also performs particularly well in terms of PLAYS@10, for which it is the best function in all but the lfm-1b, where it is the second-best. This suggests that *engagement* tends to recommend artists to users who will likely engage with them for *longer* time and *more frequently*. While we do not observe consistently the same behavior for TRACKS@10 on all datasets, *trackcounts* seems to give the highest performance and *engagement* has a notable performance on this metric as well. Interestingly, *fidelity* does not perform incredibly well on any of the datasets and metrics. This suggests that the simple linear combination of *engagement* and *trackcounts* is not sufficient, and will be investigated in future works.

8 CONCLUSIONS

In this work, we proposed new signals for listener engagement in music recommender systems. We used these signals both as relevance functions to train Implicit Matrix Factorization, and as evaluation metrics to gauge how traditional "listener-centric" recommender systems impact listeners and artists differently. Our results suggest that listener-centric quality is highly dependent on the choice of the relevance function and of the dataset they are tested on. It is therefore an important parameter to optimize when designing a music recommender system. Looking at the results from the artists' perspective, the proposed *engagement* relevance function, which combines *playcounts* and *daycounts*, performs better in most datasets, providing in general a higher average consumption of the artists' music in terms of the number of plays and number of days. It also notably increases the fraction of recommended artists overall. However, regarding distinct tracks played per artist, *trackcounts* still performs better in some datasets, suggesting that it is an important implicit signal to capture when optimizing for a wider consumption of the artists' catalog. More investigation is needed

Table 3: Evaluation of the recommendations in all the datasets

		listener-centric		artist-centric			
dataset	Rel. fun.	MAP@10	nDCG@10	PLAYS@10	TRACKS@10	DAYS@10	C@10
lfm-1b	*binary*	**0.0290**	**0.0640**	6.0294	3.6433	1.7735	0.0128
	playcounts	0.0256	0.0580	**8.5717**	**4.3151**	1.8706	0.0291
	daycounts	0.0287	0.0632	7.1185	3.8260	1.8619	0.0235
	trackcounts	0.0279	0.0623	7.6089	4.2643	1.8420	0.0210
	engagement	0.0240	0.0545	8.5211	4.2640	**1.8887**	**0.0324**
	fidelity	0.0253	0.0574	8.3988	4.1912	1.8483	0.0292
Streaming-service	*binary*	0.0519	0.1024	2.7665	1.7372	1.6663	0.1099
	playcounts	0.0610	0.1177	3.4088	1.9597	1.7535	0.1697
	daycounts	0.0585	0.1136	3.1650	1.8468	1.7200	0.1560
	trackcounts	0.0573	0.1122	3.3188	**2.0064**	1.6983	0.1372
	engagement	**0.0619**	**0.1193**	**3.4272**	1.9669	**1.7620**	**0.1826**
	fidelity	0.0615	0.1185	3.2919	1.9666	1.7209	0.1695
nowplaying	*binary*	0.0527	0.1019	2.5673	1.9380	1.6365	0.0673
	playcounts	0.0553	0.1071	3.2953	2.0027	1.7635	0.0892
	daycounts	0.0540	0.1044	2.6502	1.8291	1.7155	0.0901
	trackcounts	**0.0563**	**0.1088**	3.2008	**2.1385**	1.6348	0.0745
	engagement	0.0522	0.1019	**3.9596**	2.0120	**1.8485**	**0.0951**
	fidelity	0.0553	0.1063	3.9416	2.0553	1.7928	0.0892
30music	*binary*	0.0659	0.1360	4.0627	3.9759	1.4631	0.0123
	playcounts	0.0679	0.1403	5.5051	5.2998	1.4678	0.0214
	daycounts	**0.0703**	**0.1432**	4.7363	4.5380	**1.5012**	0.0177
	trackcounts	0.0680	0.1406	5.5297	5.3305	1.4702	0.0213
	engagement	0.0669	0.1384	**5.7237**	**5.4944**	1.4921	**0.0228**
	fidelity	0.0687	0.1411	4.9575	4.7234	1.4903	0.0134

to properly combine the three individual implicit signals in a single one.

As future work, we would like to get a trade-off between the interests of both users and artists. This can be achieved by optimizing at the same time for the users' and the artists' metrics. However, to properly assess that, a more comprehensive online evaluation with real users for a long period would be required. Given the challenge of performing this type of evaluation, simulation-based techniques have shown to be effective to study the impact that recommender systems can have on users' behaviour [9, 22, 24]. In future work, we plan to use these simulation techniques to evaluate the impact that recommendations may have on both users and artists.

ACKNOWLEDGMENTS

This research has been partially supported by Kakao Corp.

REFERENCES

[1] Himan Abdollahpouri, Gediminas Adomavicius, Robin Burke, Ido Guy, Dietmar Jannach, Toshihiro Kamishima, Jan Krasnodebski, and Luiz Pizzato. 2020. Multistakeholder recommendation: Survey and research directions. *User Modeling and User-Adapted Interaction* (2020).

[2] Alejandro Bellogin, Pablo Castells, and Ivan Cantador. 2011. Precision-oriented evaluation of recommender systems: an algorithmic comparison. In *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. ACM, New York, NY, USA, 333–336.

[3] Maurizio Ferrari Dacrema, Paolo Cremonesi, and Dietmar Jannach. 2019. Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems (RecSys '19)*. 101–109.

[4] Michael D Ekstrand, Mucun Tian, Ion Madrazo Azpiazu, Jennifer D Ekstrand, Oghenemaro Anuyah, David McNeill, and Maria Soledad Pera. 2018. All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In *Proceedings of the Conference on Fairness, Accountability and Transparency*. 172–186.

[5] Katayoun Farrahi, Markus Schedl, Andreu Vall, David Hauger, and Marko Tkalcic. 2014. Impact of Listening Behavior on Music Recommendation. In *Proceedings of the 15th International Society for Music Information Retrieval Conference*. 483–488.

[6] Andres Ferraro. 2019. Music cold-start and long-tail recommendation: bias in deep representations. In *Proceedings of the 13th ACM Conference on Recommender Systems* (Copenhagen, Denmark) *(RecSys '19)*. ACM, New York, NY, USA, 586–590.

[7] Andres Ferraro, Dmitry Bogdanov, Xavier Serra, and Jsson Yoon. 2019. Artist and style exposure bias in collaborative filtering based music recommendations. In *Proceedings of the 1st Workshop on Designing Human-Centric Music Information Research Systems (wsHCMIR '19)*. 8–10.

[8] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE International Conference on Data Mining*.

[9] Dietmar Jannach, Lukas Lerche, Iman Kamehkhosh, and Michael Jugovac. 2015. What recommenders recommend: an analysis of recommendation biases and possible countermeasures. *User Modeling and User-Adapted Interaction* 25, 5 (2015), 427–491.

[10] Dietmar Jannach, Lukas Lerche, and Markus Zanker. 2018. Recommending based on implicit feedback. In *Social Information Access*. Springer, 510–569.

[11] Kun Lin, Nasim Sonboli, Bamshad Mobasher, and Robin Burke. 2019. Crank up the volume: preference bias amplification in collaborative recommendation. *arXiv preprint arXiv:1909.06362* (2019).

[12] Malte Ludewig, Noemi Mauro, Sara Latifi, and Dietmar Jannach. 2019. Performance comparison of neural and non-neural approaches to session-based recommendation. In *Proceedings of the 13th ACM Conference on Recommender Systems (RecSys '19)*. 462–466.

[13] Rishabh Mehrotra, Mounia Lalmas, Doug Kenney, Thomas Lim-Meng, and Golli Hashemian. 2019. Jointly Leveraging Intent and Interaction Signals to Predict User Satisfaction with Slate Recommendations. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. ACM, New York, NY, USA, 1256–1267. https://doi.org/10.1145/3308558.3313613

[14] Ricardo S Oliveira, Caio Nóbrega, Leandro Balby Marinho, and Nazareno Andrade. 2017. A Multiobjective Music Recommendation Approach for Aspect-Based Diversification. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR '17)*. 414–420.

[15] Sergio Oramas, Vito Claudio Ostuni, Tommaso Di Noia, Xavier Serra, and Eugenio Di Sciascio. 2016. Sound and music recommendation with knowledge graphs. *ACM Transactions on Intelligent Systems and Technology (TIST)* 8, 2 (2016), 1–21.

[16] Yoon-Joo Park and Alexander Tuzhilin. 2008. The long tail of recommender systems and how to leverage it. In *Proceedings of the 2nd ACM Conference on Recommender systems (RecSys '08)*. ACM, New York, NY, USA, 11–18.

[17] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2011. Introduction to recommender systems handbook. In *Recommender systems handbook*. Springer, 1–35.

[18] Markus Schedl. 2016. The LFM-1b dataset for music retrieval and recommendation. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*. 103–110.

[19] Harald Steck. 2011. Item popularity and recommendation accuracy. In *Proceedings of the 5th ACM Conference on Recommender Systems (RecSys '11)*. ACM, New York, NY, USA, 125–132.

[20] Roberto Turrin, Massimo Quadrana, Andrea Condorelli, Roberto Pagano, and Paolo Cremonesi. 2015. 30Music Listening and Playlists Dataset.. In *RecSys Posters*.

[21] Gabriel Vigliensoni and Ichiro Fujinaga. 2016. Automatic Music Recommendation Systems: Do Demographic, Profiling, and Contextual Features Improve Their Performance?.. In *Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR '16)*. 94–100.

[22] Friederike Wall. 2016. Agent-based modeling in managerial science: an illustrative survey and study. *Review of Managerial Science* 10, 1 (Jan. 2016), 135–193.

[23] Eva Zangerle, Martin Pichl, Wolfgang Gassler, and Günther Specht. 2014. # nowplaying music dataset: Extracting listening behavior from twitter. In *Proceedings of the First International Workshop on Internet-Scale Multimedia Management*. 21–26.

[24] Jingjing Zhang, Gediminas Adomavicius, Alok Gupta, and Wolfgang Ketter. 2020. Consumption and Performance: Understanding Longitudinal Dynamics of Recommender Systems via an Agent-Based Simulation Framework. *Information Systems Research* forthcoming (2020).