Kowald et al. *EPJ Data Science* (2021) 10:14
https://doi.org/10.1140/epjds/s13688-021-00268-9

EPJ Data Science
a SpringerOpen Journal

EPJ.org

REGULAR ARTICLE Open Access

Support the underground: characteristics of beyond-mainstream music listeners

Dominik Kowald[1*], Peter Muellner[1], Eva Zangerle[2], Christine Bauer[3], Markus Schedl[4,5] and Elisabeth Lex[6*]

*Correspondence:
dkowald@know-center.at;
elisabeth.lex@tugraz.at
[1]Know-Center GmbH, Graz, Austria
[6]Graz University of Technology, Graz, Austria
Full list of author information is available at the end of the article

Abstract

Music recommender systems have become an integral part of music streaming services such as Spotify and Last.fm to assist users navigating the extensive music collections offered by them. However, while music listeners interested in mainstream music are traditionally served well by music recommender systems, users interested in music beyond the mainstream (i.e., non-popular music) rarely receive relevant recommendations. In this paper, we study the characteristics of beyond-mainstream music and music listeners and analyze to what extent these characteristics impact the quality of music recommendations provided. Therefore, we create a novel dataset consisting of Last.fm listening histories of several thousand beyond-mainstream music listeners, which we enrich with additional metadata describing music tracks and music listeners. Our analysis of this dataset shows four subgroups within the group of beyond-mainstream music listeners that differ not only with respect to their preferred music but also with their demographic characteristics. Furthermore, we evaluate the quality of music recommendations that these subgroups are provided with four different recommendation algorithms where we find significant differences between the groups. Specifically, our results show a positive correlation between a subgroup's openness towards music listened to by members of other subgroups and recommendation accuracy. We believe that our findings provide valuable insights for developing improved user models and recommendation approaches to better serve beyond-mainstream music listeners.

Keywords: Music recommender systems; Acoustic features; Last.fm; Clustering; User modeling; Fairness; Popularity bias; Beyond-mainstream users

1 Introduction

In the digital era, users have access to continually increasing amounts of music via music streaming services such as Spotify and Last.fm. Music recommender systems have become an essential means to help users deal with content and choice overload as they assist users in searching, sorting, and filtering these extensive music collections. Simultaneously, both music listeners and artists benefit from the employed segmentation and personalization approaches that are typically leveraged in music recommendation approaches [1]. As a result, users with different preferences and needs can be targeted in various ways with the

Springer

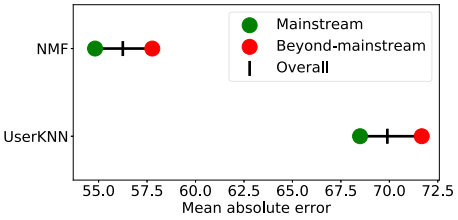

Figure 1 Recommendation accuracy measured by the mean absolute error (MAE) of a non-negative matrix factorization-based approach (i.e., NMF [10]) and a neighborhood-based approach (i.e., UserKNN [11]) for mainstream and beyond-mainstream user groups in Last.fm. We see that beyond-mainstream users receive a substantially lower recommendation quality (i.e., higher MAE) compared to mainstream music listeners. Thus, for recommender systems, it is harder to provide high-quality recommendations to beyond-mainstream than to mainstream listeners

goal that all users are presented the information and content that they need or prefer. This also means that current recommendation techniques should serve all users equally well, independent of their inclination to popular content.

Present work In the paper at hand, we focus on music consumers who listen to music beyond the mainstream (i.e., users who listen to non-popular music) in the music streaming platform Last.fm.[1] As highlighted in Fig. 1, current recommender systems do not work well for consumers of beyond-mainstream music (see Sect. 3.5 for details on this analysis). In contrast, music consumers who listen to popular music seem to get better recommendations. This finding is not essentially new. In fact, it is a widely-known problem that recommender systems (and those based on collaborative filtering, in particular) are prone to popularity bias, which leads to the behavior that long-tail items (i.e., items with few user interactions) have little chance being recommended. This phenomenon is also present across different application domains such as movies [2] or music [3].

Our previous work [4] has shown that users interested in beyond-mainstream music tend to have larger user profile sizes (i.e., individual users show a high(er) number of distinct artists they have listened to) compared to users interested in mainstream music. The observation that beyond-mainstream music listeners produce a substantial amount of digital footprints motivates the need to improve the recommendation quality for this group. However, although related research has already studied the long-tail recommendation problem (e.g., [5–8]; see Sect. 2 for a more detailed discussion of related work), it is still a fundamental challenge to understand and identify the characteristics of beyond-mainstream music and beyond-mainstream music listeners. Additionally, related work [9] has shown that the group-specific concepts of openness and diversity influence recommendation quality, where openness is defined as across-group diversity (i.e., do users of one group listen to the music of other groups?) and diversity is defined as within-group variability (i.e., how dissimilar is the music listened to by users within groups?). Thus, we are also interested in the correlation between the characteristics of beyond-mainstream music and music listeners with openness and diversity patterns as well as with recommendation quality. Concretely, our work is guided by the following research question:

RQ: What are the characteristics of beyond-mainstream music tracks and music listeners, and how do these characteristics correlate with openness and diversity patterns as well as with recommendation quality?

[1] https://www.last.fm/

To address this research question, we create, provide, and analyze a novel dataset called *LFM-BeyMS*, which contains complete listening histories of more than 2000 beyond-mainstream music listeners mined from the Last.fm music streaming platform. Besides, our dataset is enriched with acoustic features and genres of music tracks. Using this enriched dataset, we identify different types of beyond-mainstream music via unsupervised clustering applied to the acoustic features of music tracks. We then characterize the resulting music clusters using music genres. Then, we assign beyond-mainstream users to the clusters to further divide the beyond-mainstream users into subgroups. We study how the characteristics of these beyond-mainstream subgroups correlate with openness and diversity patterns as well as with recommendation quality measured through prediction accuracy.

Findings and contributions We identify four clusters of beyond-mainstream music in our dataset: (i) C_{folk}, music with high acousticness such as "folk", (ii) C_{hard}, high energy music such as "hardrock", (iii) C_{ambi}, music with high acousticness and high instrumentalness such as "ambient", and (iv) C_{elec}, music with high energy and high instrumentalness such as "electronica". By assigning users to these clusters, we get four distinct subgroups of beyond-mainstream music listeners: (i) U_{folk}, (ii) U_{hard}, (iii) U_{ambi}, and (iv) U_{elec}. We also find that these groups differ considerably with respect to the accuracy of recommendations they receive, where group U_{ambi} gets significantly better recommendations than U_{hard}. When relating our results to openness and diversity patterns of the subgroups, we find that U_{ambi} is the most open but least diverse group, while U_{hard} is the least open but most diverse group. This is in line with related research [9], which has shown that openness is stronger correlated with accurate recommendations than diversity. This means that users are more likely to accept recommendations from different groups (i.e., openness) rather than varied within a group (i.e., diversity).

Summed up, our contributions are:

- We identify more than 2000 beyond-mainstream music listeners on the Last.fm platform and enrich their listening profiles with acoustic features and genres of music tracks listened to (Sects. 3.1–3.4).
- We validate related research by showing that beyond-mainstream music listeners receive a significantly lower recommendation accuracy than mainstream music listeners (Sect. 3.5).
- We identify four clusters of beyond-mainstream music using unsupervised clustering and characterize them with respect to acoustic features and music genres (Sect. 4.1).
- We define four subgroups of beyond-mainstream music listeners by assigning users to the music clusters and discuss the relationship between openness, diversity, and recommendation quality of these groups (Sect. 4.2).
- To foster reproducibility of our research, we make available our novel *LFM-BeyMS* dataset via Zenodo[2] and the entire Python-based implementation of our analyses via Github.[3]

We believe that our findings provide useful insights for creating user models and recommendation algorithms that better serve beyond-mainstream music listeners.

[2] https://doi.org/10.5281/zenodo.3784764

[3] https://github.com/pmuellner/supporttheunderground

2 Related work

We identify three strands of research that are relevant to our work: (i) modeling of music preferences, (ii) long-tail recommendations, and (iii) popularity bias in music recommender systems.

Modeling of music preferences A multitude of factors [12] influences musical tastes and musical preferences of users. Characteristics of music listeners and music preferences have been studied in various research domains [13], ranging from music sociology [14] and psychology [15] to music information retrieval and music recommender systems [1]. Studies on music listening behavior showed that personal traits and long-term music preferences are correlated as people tend to prefer music styles that align with their personalities [16, 17]. Furthermore, related work found a relationship between music and motivation [18], music and emotion [19–22] or both personality and emotion [23]. Openness, a personality trait from the Five Factor Model [24], has also been shown to positively influence a user's preference for music recommendations [9]. Specifically, the authors of [9] found that people tend to prefer recommendations from different kinds of music (i.e., openness) rather than varied within a specific kind of music (i.e., diversity). Others showed that familiarity has a positive influence on music preferences [25, 26] and that music preferences may change over time [27]. Another strand of research on modeling users' music preferences leverages content features, e.g., acoustic features. It has been shown that the distribution of acoustic features of a user's preferred genre substantially influences the user's choice of music within other genres [28]. Also, acoustic features have been utilized to model users' preferences under different contextual conditions, in order to refine recommendation quality [29]. Based on tracks' acoustic features, the authors of [30] identified several types of music, and subsequently modeled each user by linearly combining the acoustic features of the music types. In contrast to these works, we focus on using acoustic features of music tracks for modeling and clustering beyond-mainstream music. Additionally, we link these beyond-mainstream music clusters to music genres and users in our Last.fm data sample.

Long-tail recommendations Related research [6, 7] has found that individual music consumption is biased towards popular music and that usage data for less popular music is scarce. Due to the scarcity problem, items with no or few ratings (i.e., long-tail items) have little chance of being recommended [5]. As a consequence, users that particularly favor items with few ratings or interactions are less likely to be recommended those items that they like [3]. That is problematic because many users, from time to time, prefer niche music [8]. Therefore, such users are not well served as a result of their preference for less popular items. That has been attributed to *popularity bias*, which corresponds to over-representation of popular items in the recommendation lists [31–33]. Abdollahpouri et al. [2] studied popularity bias in a dataset of movies (i.e., the MovieLens 1M dataset [34]) from the user perspective. Their study showed that commonly used recommendation techniques tend to deliver worse recommendations to users who prefer less popular movies. In our work [4], we found evidence for popularity bias in a Last.fm dataset and showed that traditional personalized recommendation algorithms such as collaborative filtering deliver worse recommendations for consumers of niche music. In the present work, we aim to gain a deeper understanding of the behavior and preferences of this

beyond-mainstreaminess user group. Thus, in contrast to existing works in long-tail recommendations, we focus on the user rather than the item perspective.

Popularity bias in music recommender systems Music recommender systems [1] are crucial tools in online streaming services such as Last.fm, Pandora, or Spotify. They help users find music that is tailored to their preferences. The basis of music recommender systems are user models derived from users' listening behavior, user properties such as personality (e.g., [35]), content features of music, or hybrid combinations of both, e.g., [36–39]. As discussed earlier, due to insufficient amounts of usage data for less popular items, many music recommendation algorithms do not provide useful recommendations for consumers of less popular and niche items. As a remedy, in [40], an approach is suggested that divides music consumers into experts and novices according to their long tail distribution in their playlists. These experts are then converted to nodes with bidirectional links connecting all the experts. These links are created to perform link analysis on the graph and to assign fine-grained weights to songs. The presented approach helps add music from the long-tail into the recommendation list. In our previous research [41, 42], we have used a framework [43] that employs insights from human memory theory to design a music recommendation algorithm that provides more accurate recommendations than collaborative filtering-based approaches for three groups of users, i.e., low-mainstream, medium-mainstream and high-mainstream users. While the awareness of popularity bias in music recommender systems increases (e.g., [44]), the characteristics of music consumers whose preferences lie beyond popular, mainstream music are still not well understood. In the present work, we shed light on the characteristics of such beyond-mainstream music consumers and relate them to openness and diversity patterns as well as recommendation quality. With this, we aim to provide useful insights for creating novel music recommendation models that mitigate popularity bias.

3 Preliminaries

We investigate the characteristics of beyond-mainstream music listeners in a dataset mined from Last.fm, a popular music streaming platform. We characterize the tracks in our dataset with acoustic features. Besides, we compare the recommendation accuracy of beyond-mainstream music listeners with the one of mainstream music listeners to motivate our subsequent analysis of the characteristics of beyond-mainstream music listeners.

3.1 Acoustic music features

For our analyses, we characterize music tracks using acoustic features that describe the content of a given track. Following the lines of, e.g., [30, 45–47], we rely on acoustic features provided by the Spotify API as a compact characterization of tracks.[4] The following eight features are extracted from the audio signal of a track:

Danceability captures how suitable a track is for dancing and is computed based "on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity".

Energy describes the perceived intensity and activity of a track and is based on the dynamic range, perceived loudness, timbre, onset rate, and general entropy of a track.

[4] https://developer.spotify.com/web-api/get-several-audio-features/

Speechiness captures the presence of spoken words in a track. High speechiness values indicate a high degree of spoken words (e.g., an audiobook), whereas medium values indicate tracks with both music and speech (e.g., rap music). Low values represent typical music tracks.

Acousticness measures the probability that the given track only contains acoustic instruments.

Instrumentalness quantifies the probability that a track contains no vocals, i.e., the track is instrumental.

Tempo measures the rate of the track's beat in beats per minute.

Valence describes the "emotional positiveness" conveyed by a track (i.e., cheerful and euphoric tracks reach high valence values).

Liveness measures the probability that a track was performed live, i.e., whether an audience is present in the recording.

3.2 Enriched dataset of music listening events

To study characteristics of beyond-mainstream users and their listening preferences, we create a novel dataset called *LFM-BeyMS* that contains the required information for such analyses. We base our work on a dataset gathered from the Last.fm music platform, which we considerably enrich with the music tracks' acoustic features (see Sect. 3.1) [48]. Additionally, we combine this data with mainstreaminess information of Last.fm users (see Sect. 3.3) as well as music genre information to identify beyond-mainstream listeners and music (see Sect. 3.4).

An overview of our new *LFM-BeyMS* dataset and its data sources is depicted in Fig. 2. As shown, the starting point for our new dataset is the publicly available *LFM-1b* dataset[5] of music listening information shared by users of the online music platform Last.fm [49]. *LFM-1b* contains listening histories of 120,322 users; their listening records (or "listening events") have been created between January 2005 and August 2014. They sum up to over 1.1 billion listening events (LEs), where each LE is described by an (anonymous) user identifier, the artist name, the album name, the track name, and the timestamp of the listening event. Also, the *LFM-1b* dataset includes demographics of some users (i.e., country, age, and gender).

To enrich the *LFM-1b* dataset to suit our task, we utilize our previously created *CultMRS* music recommendation dataset [50]. This dataset contains 55,191 users, who have listened to a total of 26,022,625 distinct tracks, captured by a total of 807,890,921 LEs [48].

To further enrich the dataset with music acoustic features, we gather the acoustic features described in Sect. 3.1 for the tracks remaining in the dataset after the filtering described above. To this end, we rely on the Spotify API to gather content-based acoustic features for each track. Particularly, we search tracks using the ⟨track, artist, album⟩ triples extracted from the *LFM-1b* dataset using the Spotify search API[6] to gather the Spotify track URI of each track by using all three parts of the triple in a conjunctive query. In total, this allowed gathering 4,326,809 Spotify URIs. For the remainder of the tracks, we were not able to retrieve a URI. We attribute this to two factors: either the searched track is not provided by Spotify or the track, artist, and album information cannot be matched to

[5] http://www.cp.jku.at/datasets/LFM-1b/

[6] https://developer.spotify.com/web-api/search-item/

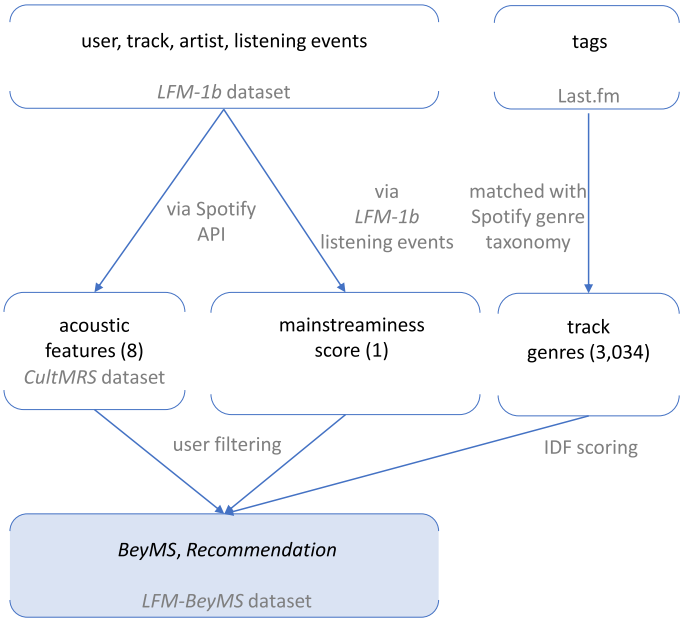

Figure 2 Overview of our new *LFM-BeyMS* dataset and its data sources. We depict the different features, their origin, and relation, and show the feature groups with the number of contained features in brackets. *LFM-BeyMS* contains *BeyMS*, i.e., data to study the beyond-mainstream user group, and *Recommendation*, i.e., data to conduct recommendation experiments of beyond-mainstream and mainstream music listeners

a Spotify track unambiguously. Subsequently, we use the obtained track URI to query the acoustic features API,[7] which returns the acoustic features of a given track (cf. Sect. 3.1). In a subsequent cleaning step, we remove all tracks for which the Spotify API did not provide the full set of acoustic features.

That procedure provides us with a set of 3,478,399 unique tracks and their acoustic features. Within the LFM-1b dataset, this amounts to 13.36% of the distinct tracks. Overall, these account for as much as 48.89% of all listening events (i.e., the tracks listened to by users) of the LFM-1b dataset. The resulting dataset, now enriched by acoustic music descriptors, comprises a total of approximately 394 million listening events of 55,149 users. In Table 1 (column "*CultMRS*"), we provide further descriptive statistics of the *CultMRS* dataset. We refine this dataset to create our new *LFM-BeyMS* dataset (column "*BeyMS* in Table 1), which consists of *BeyMS*, i.e., data to study the characteristics of beyond-mainstream music listeners, and *Recommendation*, i.e., data to conduct recommendation experiments of beyond-mainstream and mainstream music listeners.

3.3 Identifying beyond-mainstream music listeners

To identify beyond-mainstream music listeners, for each user, we compute a mainstreaminess score, which is generally defined as the overlap between a user's individual listening history and the aggregated listening history of all Last.fm users in the dataset. In this vein, the mainstreaminess score reflects a user's inclination to music listened to by the Last.fm mainstream listeners (i.e., the "average" Last.fm listener in the dataset). In [51], several

[7] https://developer.spotify.com/web-api/get-several-audio-features/

Table 1 Descriptive statistics of the *CultMRS* dataset and our novel *LFM-BeyMS* dataset. *CultMRS* comprises acoustic features of tracks. *LFM-BeyMS* is based on *CultMRS* and consists of *BeyMS* and *Recommendation*. Our analyses of beyond-mainstream music listeners utilize *BeyMS* and our recommendation experiments utilize *Recommendation*, which includes listening events of both users with beyond-mainstream and mainstream music taste

Item	*CultMRS* [50]	*LFM-BeyMS* (our novel dataset)	
		BeyMS	*Recommendation*
Users	55,149	2074	4148
Tracks	3,478,399	157,444	1,084,922
Artists	337,840	14,922	110,898
Listening Events (LEs)	394,944,868	4,916,174	16,687,363
Min. LEs per user	1	3	9
Q_1 LEs per user	1442	1254	2604
Median LEs per user	5667	2048	3766
Q_3 LEs per user	9738	3239	5252
Max. LEs per user	399,210	10,536	11,177
Avg. LEs per user	7161.41 ($\pm$ 10,326.91)	2371.526 ($\pm$ 1520.629)	4,022.990 ($\pm$ 1898.060)

measures of user mainstreaminess are defined. Out of these, we choose the *M-global-R-APC* definition since it yielded good results in context-based music recommendation experiments for the *LFM-1b* dataset, as evidenced in [51]. The *M-global-R-APC* measure approximates a user's mainstreaminess score by computing Kendall's τ [52] rank correlation between the user's vector of artist play counts and the global vector of artist play counts (aggregated over all users in the dataset). This definition also explains the name of the measure, where "M" refers to mainstreaminess, "global" indicates the global perspective, "R" stands for rank correlation, and "APC" refers to artist play counts.

Next, we describe how we identify our beyond-mainstream users via filtering the users by the number of listening events (see Fig. 3 and Sect. 3.3.1) and by mainstreaminess scores (see Fig. 4 and Sect. 3.3.2).

3.3.1 Filtering users by the number of listening events

For our study, we select the users so that listeners of different levels of "listening activity" are equally represented. We conduct a Gaussian kernel density estimation (KDE) [53] on the distribution of listening events over users to estimate the continuous probability density function (PDF) [54]. However, KDE estimates the PDF via discrete bins and hence, it is necessary to approximate the gradient via the principle of finite differences. The gradient of the PDF helps us identifying regions of increasing or decreasing probability.

Figure 3 shows that two large subsets of users exist that exhibit either very few or an abundance of listening events. For our analysis, we consider only users who are not in one of the subsets as mentioned earlier. On the one hand, we exclude users with too little data available for studying their listening behavior; and on the other hand, we exclude so-called power listeners who might bias our analyses. Furthermore, such users with a very high number of listening events are often radio stations, which do not contribute reliable data to our investigations.

Hence, we define lower and upper bounds regarding the number of users' listening events to include in our study, such that the rate of change in terms of the number of listening events is minimal and stable within these boundaries. That requires the gradient of the region within the lower and upper bound to be near zero (i.e., $\pm 10^{-6}$). By computing the second-order accurate central differences [55], we obtain an approximation of the

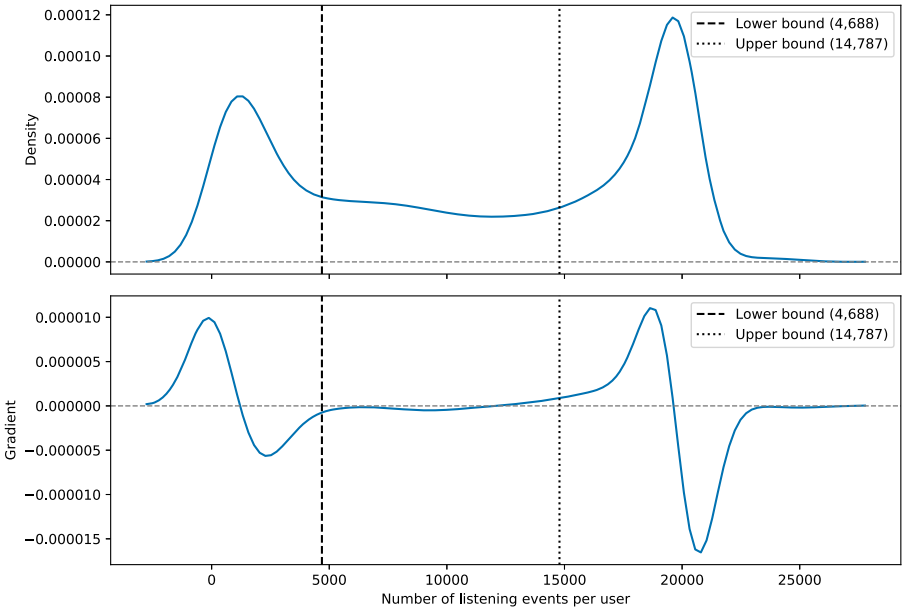

Figure 3 Distribution of listening events in our set of Last.fm users. We set the lower and upper bound marked as dashed and dotted lines, respectively based on the gradient, which results in 12,814 users with a similar number of listening events

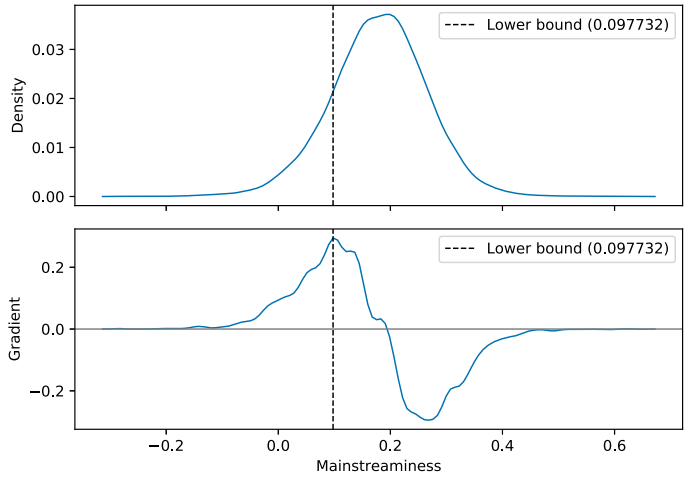

Figure 4 Mainstreaminess distribution of the 12,814 users illustrated in Fig. 3. Based on the maximum gradient, we select an upper bound of 0.097732 to identify the 2074 beyond-mainstream users of the *BeyMS* user group

gradient and find the longest cohesive region fulfilling the requirements between a lower bound of 4688 and an upper bound of 14,787 listening events per user, which leads to 12,814 users.

3.3.2 Filtering users by mainstreaminess scores

Figure 4 illustrates the mainstreaminess distribution of the 12,814 users that we have extracted based on the number of listening events. Here, mainstreaminess is defined accord-

ing to the *M-global-R-APC* definition taken from [51] (explained in Sect. 3.3). By setting an appropriate upper bound, we aim to exclude mainstream music listeners. In other words, we aim to set the upper bound to the beginning of the distribution's bulk, which is motivated as follows: Firstly, the first inflection point (i.e., maximal gradient) of a Gaussian distribution is found at $\mathbb{E}[X] - \text{std}(X)$, where $\mathbb{E}[X]$ is the expectation, and $\text{std}(X)$ is the standard deviation of the Gaussian random variable X. Secondly, the first inflection point of a Gaussian distribution is equivalent to the 15.9-percentile. By setting the mainstreaminess threshold to this point, we intend to omit the majority of users and hence, only consider the 15.9% of users with the lowest mainstreaminess scores. Utilizing this upper bound on the mainstreaminess score, we obtain a set of 2074 beyond-mainstream users. Furthermore, the Gaussian assumption can be strengthened by the observation that the 2074 beyond-mainstream users represent 16.19% of users. In the remainder of this paper, we refer to this set of beyond-mainstream music listeners as *BeyMS*.

3.4 Identifying beyond-mainstream music

We aim to study beyond-mainstream listeners in terms of their music taste. We characterize music via its acoustic features, as described in Sect. 3.1, and also investigate genres as an alternative way to describe a music track via conventional categories. As the *LFM-1b* dataset does not contain genre annotations of tracks and the Spotify API only provides genres on artist level,[8] we leverage the tags assigned to each track by Last.fm users to identify genre annotations. To obtain these tags, we use the respective Last.fm API endpoint.[9] After having fetched the tags for each track, we de-capitalize them and remove all non-alpha-numeric characters. Since not all tags used by Last.fm users correspond to actual music genres (e.g., the "seenlive" tag is used to indicate that a user has seen an artist performing this track live), we use a fine-grained music genre taxonomy consisting of 3034 genres that are also utilized by Spotify, which we gather from the EveryNoise service (2019-07-24).[10] Specifically, for each track listened to by any of our *BeyMS* users, we remove all tags that are not part of the EveryNoise genre taxonomy, using a case-insensitive matching approach.

We note that Last.fm users tend to assign very general genre tags to a large number of tracks, such as "pop" or "rock". To remove these coarse-grained genres and to identify fine-grained beyond-mainstream music genres, we calculate the inverse document frequency (IDF) [56] metric of our genre-track distribution by treating genres as terms and tracks as documents, i.e., $\text{IDF}(g) = \log_{10} \frac{|T|}{|\{t \in T \text{ with } g \in G_t\}|}$. More precisely, the IDF-score of genre g is determined by relating the number of all tracks $|T|$ to the number of tracks annotated with genre g where $|G_t|$ is the set of genres assigned to track t. This way, a coarse-grained genre receives a small IDF-score, while a fine-grained genre receives a high IDF-score. Figure 5 shows the IDF-score distribution of the top-100 genres in ascending order (i.e., from coarse-grained to fine-grained). Here, we identify two groups of genres, where the first group consists of 6 genres with small IDF-scores, and the second group consists of 94 genres with high IDF-scores. The visual inspection of Fig. 5 indicates that the lower bound of 0.90 serves as a discriminant between these two groups of coarse-grained and

[8] https://developer.spotify.com/documentation/web-api/reference-beta/#endpoint-get-an-artist

[9] https://www.last.fm/api/show/track.getTopTags

[10] http://everynoise.com/

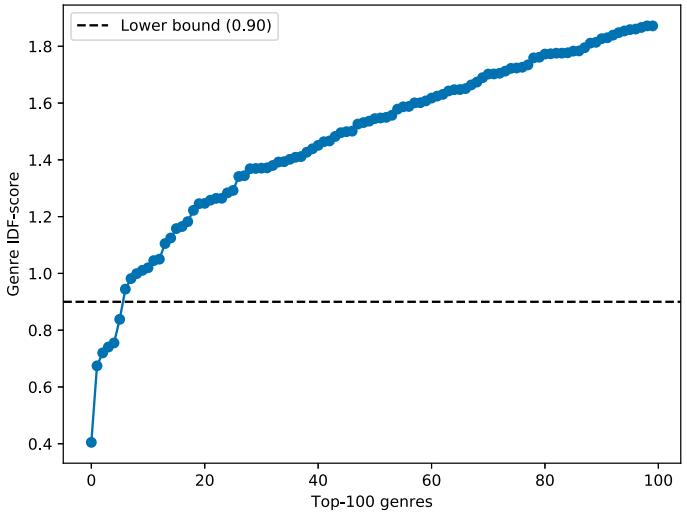

Figure 5 IDF-score distribution of the top-100 genres in ascending order (i.e., from coarse-grained to fine-grained). The 6 coarse-grained genres below the lower bound of 0.90 are removed from the genre assignments, i.e., "rock", "pop", "electronic", "metal", "alternativerock", "indierock"

fine-grained genres. Consequently, we remove the 6 coarse-grained genres (i.e., "rock", "pop", "electronic", "metal", "alternativerock", "indierock") from the genre assignments of our tracks, which leads to 157,444 out of 799,659 tracks listened to by *BeyMS* users with at least one remaining genre. In total, these tracks are annotated with 1418 unique genre identifiers.

We are aware of the fact to our track filtering procedure leads to incomplete listening profiles of users. Since we rely on genres to describe beyond-mainstream music, these filtering steps are necessary for our study. To ensure that the *BeyMS* users' reduced listening profiles are still representative of their music preferences, we further investigate the consequences of the filtering procedure. Here, we find that a user's listening history (i.e., the entirety of a user's listening events) is reduced to 61% on average. However, we also find that there are only 62 of the 2074 *BeyMS* users, for whom the listening history is reduced to less than 20%. For these users most affected by the filtering, we compare the acoustic feature distributions of their listened tracks before and after the filtering steps, and find that filtering only marginally affects the acoustic feature distributions (i.e., average change in mean = 0.0098 ± 0.0148). This means that the acoustic feature distribution contained in the user's profile is highly robust against the filtering. The statistics of *BeyMS* are summarized in column "*BeyMS*" in Table 1.

3.5 Recommendations for beyond-mainstream music listeners

In order to compare the recommendation accuracy of recommendations received by the users of our *BeyMS* group and by mainstream users, we construct a dataset consisting of *BeyMS*'s listening events and the listening events of an equally-sized group of mainstream users. Therefore, we define the *MS* user group as 2074 (i.e., the size of our *BeyMS* group) randomly-chosen users with a mainstreaminess score that is higher than the upper bound for low mainstreaminess, identified in Fig. 4. Furthermore, the *MS* users are also in between the lower and upper bounds for listening events identified in Fig. 3. As shown in

Table 1 (column "*Recommendation*"), the dataset used for the evaluation of recommendations contains data of 4148 distinct *BeyMS* and *MS* users, 1,084,922 distinct tracks, and 16,687,363 listening events.

We use the Python-based open-source recommendation library Surprise[11] to compute and evaluate recommendations. One advantage of using Surprise is that it provides built-in recommendation algorithms as well as a standardized evaluation pipeline, which enhances the reproducibility of our research. Since Surprise is focused on rating prediction, we formulate our music recommendation scenario also as a rating prediction problem, in which we predict the preference of a target user u for a target track t. As done in [57], we model the preference of t for u by scaling the play count (i.e., number of listening events) of t by u to a range of $[1; 1000]$ using min-max normalization. We perform this normalization on the individual user level to ensure that all users share the same preference value ranges. Thus, with this method, we ensure that each user's most listened track has a preference value of 1000, while their least listened track has a preference value of 1. To ensure that this min-max normalization procedure does not disrupt the play count distribution of our users, we compare the original play count distribution with the normalized distribution and find that both distributions are strongly right-skewed. Specifically, we find very similar distributions for large amounts of our play count data.

We utilize a selection of Suprise's built-in recommendation methods consisting of one baseline approach (i.e., UserItemAvg), two neighborhood-based approaches (i.e., UserKNN and UserKNNAvg), and one matrix factorization-based approach (i.e., NMF). Specifically, UserItemAvg predicts the average play count in the dataset by also accounting for deviations of u and t, for example, if a user u tends to have more listening events than the average Last.fm user [58]. UserKNN [11] is a user-based collaborative filtering approach and is calculated using $k = 40$ nearest neighbors and the cosine similarity metric, which are the default settings of Surprise. UserKNNAvg is an extension of UserKNN [11] that also takes the average rating of target user u into account. Finally, NMF, i.e., non-negative matrix factorization [10], is calculated using 15 latent factors, which is the default parameter in the Surprise library. As shown in our previous work [4], NMF is also capable of recommending non-popular items from the long tail and should therefore especially be of interest for our beyond-mainstream recommendation setting.

We use Surprise's default parameters and refrain from performing any hyperparameter tuning since we are only interested in assessing (relative) performance differences between the two user groups *BeyMS* and *MS*, and not in outperforming any state-of-the-art algorithm. This is also the reason why we focus on traditional algorithms instead of investigating the most recent deep learning architectures, which would also require a much higher computational effort.

The resulting mean absolute error (MAE) results can be observed in Table 2 (and correspond to the ones already shown in Fig. 1). We favor MAE over the commonly used root mean squared error (RMSE) due to several pitfalls, especially regarding the comparison of groups with different numbers of observations [59]. Here, we perform 5-fold cross-validation leading to 5 different 80/20 train-test splits and average the MAE over the 5 folds. NMF clearly outperforms UserItemAvg as well as the two neighborhood-based methods (i.e., UserKNN and UserKNNAvg) both for the two user groups (see

[11] http://surpriselib.com/

Table 2 Mean absolute error (MAE) results for the two user groups *MS* and *BeyMS* of different mainstreaminess and a selection of standard recommendation algorithms. A one-tailed Mann–Whitney-U test ($\alpha = 0.0001$) provides significant evidence, indicated by ***, that all algorithms perform worse on *BeyMS* than on *MS* in terms of MAE. Furthermore, NMF (as shown in bold) outperforms the other three approaches UserItemAvg, UserKNN and UserKNNAvg

User group	UserItemAvg	UserKNN	UserKNNAvg	NMF
BeyMS	63.4608***	71.6694***	67.5770***	**57.7703***
MS	61.2562	68.4894	63.3985	**54.8182**
Overall	62.2315	69.8962	65.2469	**56.2492**

rows "*BeyMS*" and "*MS*") separately and overall without distinguishing between the user groups (see row "Overall"). Additionally, we conduct a one-tailed Mann–Whitney-U test ($\alpha = 0.0001$), where we define the null-hypothesis as the MAE for *MS* being larger than or equal to the MAE for *BeyMS*. Results marked with *** indicate that the null-hypothesis was rejected for every fold. Thus, all algorithms (including NMF) provide a significantly larger error for *BeyMS* than for *MS*. In other words, recommendation quality is significantly better for users with mainstream taste than for users who prefer beyond-mainstream music across all recommendation approaches.

These initial results underpin the need to study the characteristics of the *BeyMS* user group that receives worse recommendations. The corresponding experiments are presented in the next section.

4 Characteristics of beyond-mainstream music and listeners

We identify the types of beyond-mainstream music using unsupervised clustering and characterize these types with respect to acoustic features and music genres. Besides, we detect subgroups of beyond-mainstream music listeners by assigning users to these clusters and evaluate the recommendation quality obtained for these subgroups. Finally, we discuss the recommendation quality with respect to openness and diversity. For this, we relate to the definitions given by [9]:

Openness is the across-groups diversity (or categorical diversity) and describes if users of one group also listen to the music of other groups.

Diversity is the within-groups diversity (or thematic diversity) and describes the dissimilarity of music listened to by users within groups.

Based on the findings of [9], we would expect that subgroups with high openness should receive more accurate recommendations than subgroups with high diversity.

4.1 Clustering and characterizing beyond-mainstream music

To study the different types of music listened to by the users in our *BeyMS* group, we conduct a cluster analysis. Specifically, we cluster the 157,444 tracks listened to by *BeyMS* users, where each track is described by the eight acoustic features danceability, energy, speechiness, acousticness, instrumentalness, tempo, valence, and liveness (see Sect. 3.1). We scale the value ranges of these features to $[0, 1]$ using min-max normalization. The use of latent representations of musical elements such as tracks was shown to be efficient in the area of music information retrieval [30, 60, 61]. Furthermore, for visually analyzing the obtained music clusters and decreasing computation time, we favor a reduction of dimensionality to two dimensions.

We conduct experiments with a broad body of dimensionality reduction methods, i.e., linear and nonlinear principal component analysis (PCA) [62], locally linear embedding

[63], multidimensional scaling [64], Isomap [65], spectral embedding [66], t-distributed stochastic neighbor embedding (t-SNE) [67] and uniform manifold approximation and projection (UMAP) [68]. We visually inspected the 2-dimensional feature spaces created by these methods with regards to the clustering quality, and we obtained the visually most homogeneous results with UMAP. Moreover, UMAP has already been successfully used in the music domain [30] and thus, we use it for the remainder of our experiments. Specifically, we utilize the open-source implementation of UMAP [69], which requires four parameters: (i) the distance metric M in the input space, (ii) the number of latent dimensions D, (iii) the minimum distance of points in the latent space d_{min}, and (iv) the number of neighbors of a point N. Based on experimentation and related literature (e.g., [69]), we set the distance metric M to the Euclidean distance, the number of latent dimensions D to 2, the distance d_{min} to 0.1 and the number of neighbors N to 15.

In a next step, we perform clustering on the dimensionality-reduced acoustic features of tracks. Again, we conduct experiments with various clustering methods, i.e., DBSCAN [70], K-Means [71], Gaussian mixture models [72], affinity propagation [73], spectral clustering [74], hierarchical agglomerative clustering [75], OPTICS [76] and HDBSCAN* [77]. Here, we obtain the best results with respect to cluster cohesion and separation using HDBSCAN*. Furthermore, HDBSCAN* was also already used by related work to cluster music items [78]. We employ the open-source implementation of HDBSCAN* [79] that requires four parameters: (i) the minimum cluster size s_{min} that defines the minimum size of a group of points to consider a cluster, (ii) the minimum number of samples in the neighborhood of a core point N_{min}, which quantifies how conservative the clustering is, (iii) ε, which enables the recovery of DBSCAN clusters if the s_{min} value is not reached, and (iv) the scaling of the distance α, which is another measure of the clustering's conservativeness. In detail, α scales the distance between two points, which determines whether these points are merged into a cluster. This scaling is used in the construction of HDBSCAN*'s hierarchy of clusterings. Again, we find the best-suited parameters based on experimentation and related literature (e.g., [77]). Specifically, we require each cluster to comprise a sufficiently large number of tracks to increase the level of significance of our subsequent experiments. We expect the existence of very small music clusters and thus, search for the optimal value of the minimal cluster size s_{min} in the search space of $\{1000; 1025; \ldots; 1475; 1500\}$, where we obtain the best results with respect to the within-cluster variance for $s_{\text{min}} = 1375$. Furthermore, tightly packed clusters without any contribution of noise should be favored. In other words, all points within a cluster should be within the neighborhood of at least one core point. Thus, we set the minimal number of samples in the neighborhood $N_{\text{min}} = s_{\text{min}} = 1375$. The remaining two parameters are set to their default values, i.e., $\varepsilon = 0$ and $\alpha = 1$.

Figure 6 shows the results of the clustering process using HDBSCAN* and UMAP for the 2-dimensional mapping. This process leads to four music clusters. Here, the green cluster (hatch: +) is the largest one with 92,798 tracks, followed by the pink cluster (hatch: x) with 30,379 tracks and the blue cluster (hatch: /) with 12,148 tracks. The smallest cluster is the orange one (hatch: o) as it contains 7629 tracks. The remaining 14,490 of our 157,444 *BeyMS* tracks have not been assigned to a cluster and thus, will not be included in further analyses and interpretations. Next, we describe how we name these clusters based on their music genre distributions.

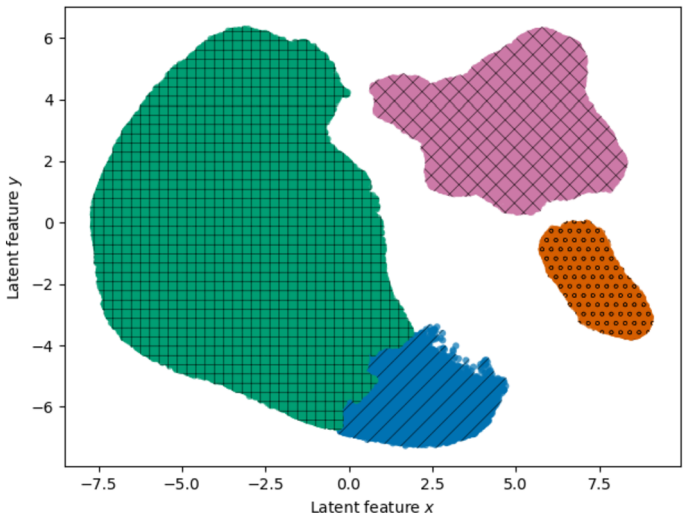

Figure 6 Music clustering results obtained with HDBSCAN* and UMAP for the 2-dimensional mapping. The outputs are four clusters with the following cluster sizes: 12,148 (blue, hatch: /), 92,798 (green, hatch: +), 7629 (orange, hatch: o) and 30,379 (pink, hatch: x) tracks. 14,490 of our 157,444 *BeyMS* tracks have not been assigned to a cluster

4.1.1 Genre distributions

In Fig. 7, we illustrate the top-10 genres of the four music clusters. For this, we refer to the genre IDF-scores presented in Sect. 3.4 and weight each genre assigned to a track in a cluster with its corresponding IDF-score. For example, if a genre with an IDF-score of 1.4 is assigned to 1000 tracks in a cluster, it is visualized as an aggregated genre IDF-score of 1400 in the corresponding plot of Fig. 7. Based on the genre distributions, we label each cluster according to its top genre.

With respect to the blue cluster (hatch: /) in Plot (a), we find top genres such as "folk" and "singersongwriter", which typically reflect music with high acousticness. In the remainder of this paper, therefore, we refer to this cluster as C_{folk}. The top genres of the green cluster (hatch: +) in Plot (b) are typical high energy music genres such as "hardrock", "punk", "poprock", and "hiphop". Based on this, we name this cluster C_{hard}.

For the orange cluster (hatch: o) in Plot (c), we find genres that reflect music with high acousticness and high instrumentalness such as "ambient", "experimental", "newage", and "postrock". As "ambient" clearly dominates the genre distribution for this cluster, we name this cluster C_{ambi}. Similarly to C_{folk}, this cluster contains music with high acousticness; yet, while C_{folk} is characterized by low instrumentalness music, C_{ambi} is characterized by a high level of instrumentalness. Finally, Plot (d) shows the genre distribution of the pink cluster (hatch: x) with "electronica" as the top genre, which leads to the name C_{elec} for this cluster.

Thus, both, C_{elec} and C_{hard}, consist of high energy music but in contrast to C_{hard}, C_{elec} also comprise high instrumentalness values. This also makes sense when looking at other top genres of C_{elec} such as "deathmetal" and "blackmetal" where guttural vocal techniques are often mistakenly classified as another type of instrument [80].

To compare the genre distributions among the four music clusters, we illustrate the relative genre frequency distribution of the clusters in Fig. 8. The relative frequency of a genre g depicts the fraction of listening events of tracks within a cluster c that are annotated with g. Here, we only show genres with a minimum relative genre frequency of 0.1. We

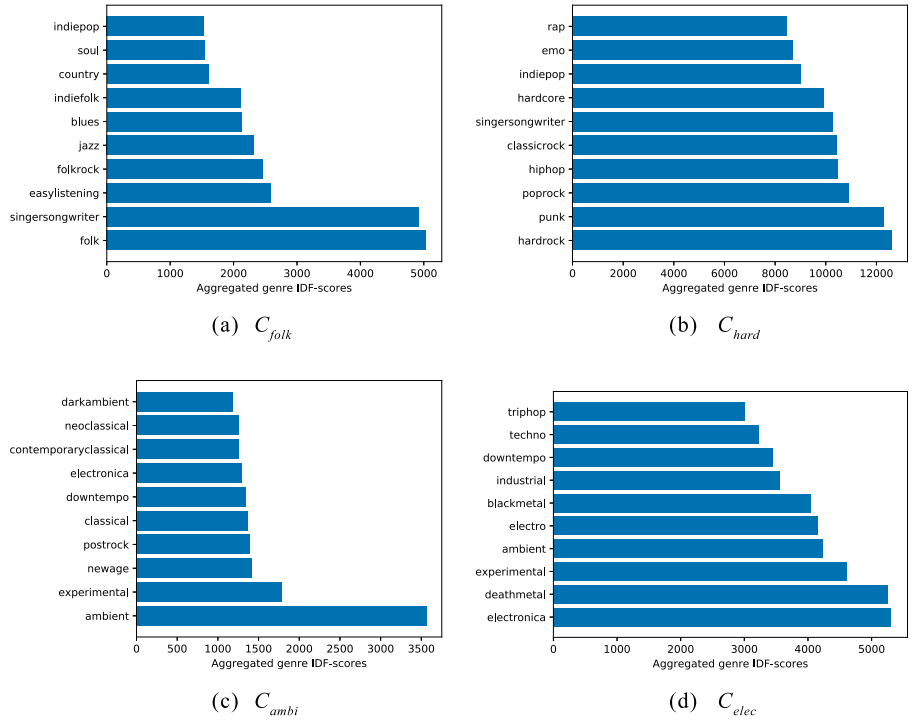

Figure 7 Top-10 genres of the four music clusters C_1–C_4 according to the aggregated genre IDF-scores. We name the clusters according to the top genre, i.e., (**a**) blue (hatch: /) → C_{folk} ("folk"), (**b**) green (hatch: +) → C_{hard} ("hardrock"), (**c**) orange (hatch: o) → C_{ambi} ("ambient"), and (**d**) pink (hatch: x) → C_{elec} ("electronica")

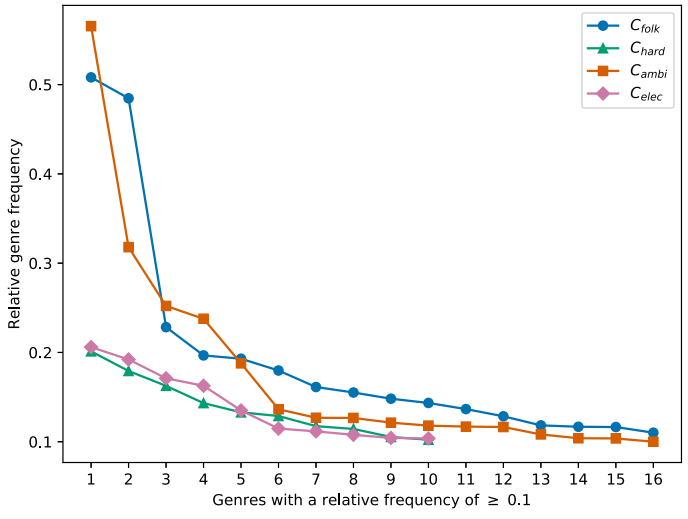

Figure 8 Relative genre frequency distribution of the four music clusters. While there are dominating genres in C_{folk} and C_{ambi}, the genre distribution is more diverse in C_{hard} and C_{elec}

see that there are clearly dominating genres in C_{folk} and C_{ambi}, whereas the genre distributions in C_{hard} and C_{elec} are more evenly distributed. When relating this finding to the findings of Fig. 7, we clearly see that the results correspond to each other: C_{hard} and C_{elec} contain a more diverse genre spectrum (e.g., "hardrock" and "hiphop" are both part of

C_{hard}'s top genres) than C_{folk} and C_{ambi} (e.g., in C_{ambi}'s top genres, we find "ambient" and "darkambient").

4.1.2 Acoustic feature distributions

To understand the musical content of these four music clusters, we analyze the acoustic feature distributions of the four music clusters using boxplots in Fig. 9. This visualization does not show any obvious differences with respect to danceability and tempo among the four clusters. For the acoustic features energy, speechiness, acousticness, valence, and liveness, there are similar values for the cluster pairs C_{folk} and C_{ambi}, and C_{hard} and C_{elec}. We observe differences between these two cluster pairs with respect to energy and acousticness. While C_{hard} and C_{elec} provide high energy values and small acousticness values, C_{folk} and C_{ambi} feature small energy values and high acousticness values.

In contrast, for instrumentalness, we see similar values for the cluster pairs C_{folk} and C_{hard} as well as for C_{ambi} and C_{elec}. We observe very high values for C_{ambi} and C_{elec}, and very small values for C_{folk} and C_{hard}. This difference is also visible in Fig. 6 in the form of the gap between C_{folk} and C_{hard} on the left, and C_{ambi} and C_{elec} on the right.

Summing up, in C_{folk}, we find music with low energy, high acousticness, and low instrumentalness; C_{hard} contains music with high energy, low acousticness, and low instrumentalness; in C_{ambi}, we observe music with low energy, high acousticness, and high instru-

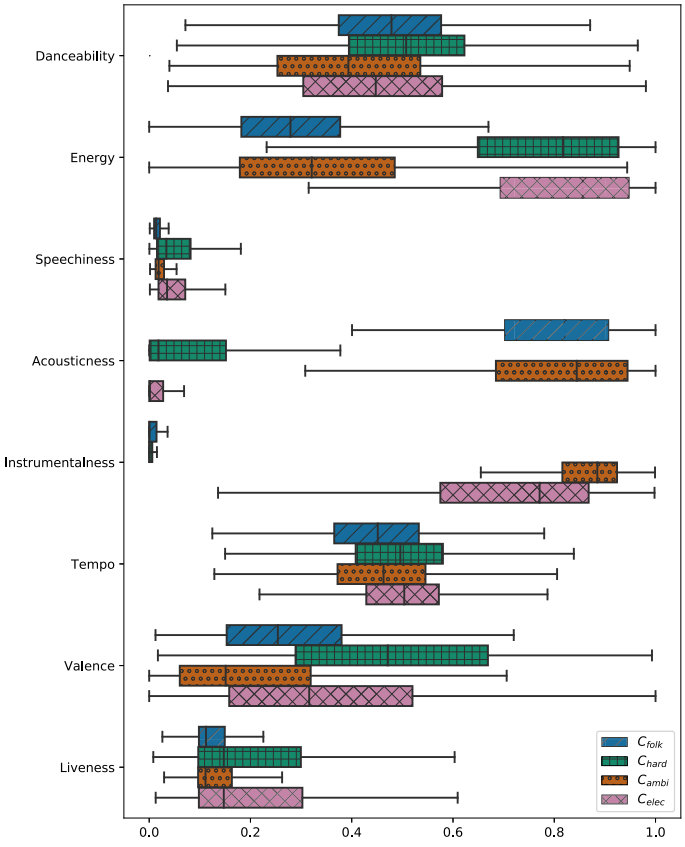

Figure 9 Distribution of the eight acoustic features for the four music clusters. While the clusters do not show obvious differences with respect to danceability and tempo, we find large differences with respect to energy, acousticness and instrumentalness

mentalness; and in C_{elec}, we find high energy, low acousticness, and high instrumentalness. Thus, these findings are in line with the genre distributions presented in Fig. 7.

4.2 Assigning and studying beyond-mainstream music listeners

In the next step, we assign the 2074 *BeyMS* users to the four music clusters to categorize them into four distinct beyond-mainstream subgroups for further analyses.

For each user u, we count the number of listening events $LE_{u,c}$ that u has contributed to the tracks in each cluster c, where $c \in C = \{C_{\text{folk}}, C_{\text{hard}}, C_{\text{ambi}}, C_{\text{elec}}\}$. Then, we assign u to the cluster c for which the number of contributed listening events $LE_{u,c}$ is the highest. However, because we have varying cluster sizes, the probability of u listening to a track t of the two larger clusters C_{hard} and C_{elec} is much higher than for the two smaller clusters C_{folk} and C_{ambi}, although C_{folk} and C_{ambi} could be more representative choices for u. Thus, similar to the IDF distribution of genres (see Fig. 5), we take advantage of the IDF scoring to reduce the influence of the larger clusters and to assign higher weights to the smaller clusters. Specifically, these cluster IDF-scores are given by $\text{IDF}(c) = \log_{10} \frac{|T|}{|\{t \in T \text{ with } c_t\}|}$, i.e., by relating the number of all tracks $|T|$ to the number of tracks in cluster c where c_t is the music cluster assigned to track t. That lets us define the user–cluster weight $w_{u,c}$ for user u and cluster c as $w_{u,c} = \text{IDF}(c) \cdot LE_{u,c}$.

Consequently, users are assigned to the highest weighted music cluster and thus, a subgroup U_c for cluster c is given by $U_c = \{u \in U : \arg\max_{c \in C}(w_{u,c})\}$.

Out of the 2074 *BeyMS* users, we can assign 2073 users to these subgroups. Thus, only 1 user listened to tracks not contained in any cluster in Fig. 6. Similar to the naming scheme of music clusters, we label the subgroups according to the name of their assigned music cluster. Hence, we obtain four subgroups U_{folk}, U_{hard}, U_{ambi}, and U_{elec}.

Table 3 provides basic descriptive statistics of these four resulting subgroups. Here, U_{hard} is the largest subgroup with $|U| = 919$ users, followed by U_{elec} with $|U| = 642$ users, U_{folk} with $|U| = 369$ users, and U_{ambi} with $|U| = 143$ users. The differences with respect to the number of users also correspond to the differences regarding the number of artists $|A|$, the number of tracks $|T|$, and the number of listening events $|LE|$ contained in the clusters. In the case of the number of genres $|G|$, this differs slightly because the users in the smaller U_{ambi} cluster listen to more genres (i.e., 918) than the bigger U_{folk} cluster (i.e., 811). This indicates that the users in U_{ambi} listen to a broader set of music than the users in U_{folk}.

Considering the average number of listening events per user (i.e., $\overline{|LE_u|}$) and the average number of tracks per user (i.e., $\overline{|T_u|}$), we see that, while there is little difference between U_{hard} and U_{elec} with respect to $\overline{|LE_u|}$, $\overline{|T_u|}$ is much higher for U_{elec} (i.e., 670.402) than for U_{hard} (i.e., 557.470). This indicates that, although the number of listening events is nearly the same, users of U_{elec} tend to listen to a wider set of tracks than users of U_{hard}. With

Table 3 Descriptive statistics of the four subgroups. Here, $|U|$ is the number of users, $|A|$ is the number of artists, $|T|$ is the number of tracks, $|LE|$ is the number of listening events, $|G|$ is the number of genres, $\overline{|LE_u|}$ is the average number of listening events per user, $\overline{|T_u|}$ is the average number of tracks per user and $\overline{Age}$ is the average age (along with the standard deviation) of users in the group

Subgroup	$\lvert U\rvert$	$\lvert A\rvert$	$\lvert T\rvert$	$\lvert LE\rvert$	$\lvert G\rvert$	$\overline{\lvert LE_u\rvert}$	$\overline{\lvert T_u\rvert}$	$\overline{Age}$ (std.)
U_{folk}	369	9559	72,663	702,635	811	1904.160	549.650	27.599 (± 10.369)
U_{hard}	919	11,966	107,952	2,150,246	1274	2339.767	557.470	23.867 (± 8.912)
U_{ambi}	143	6869	39,649	224,327	918	1568.720	473.308	29.571 (± 14.138)
U_{elec}	642	11,814	105,907	1,416,354	1005	2206.159	670.402	24.639 (± 7.886)

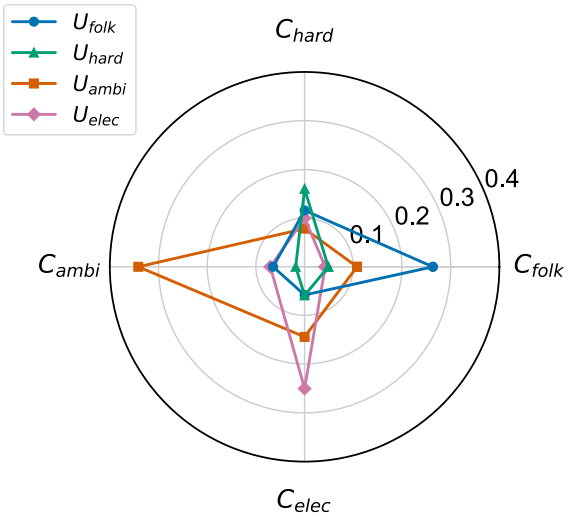

Figure 10 Radar plot illustrating the contribution of each music cluster to a subgroup. While the weight distribution of U_{hard} and U_{elec} is rather narrow, it is more broad in case of U_{folk} and U_{ambi} suggesting that these groups are more open to music outside the own music cluster

respect to the average age of the users $\overline{Age}$, we see that the users of U_{folk} and U_{ambi} are the oldest ones, and users of U_{hard} and U_{elec} are the youngest ones. However, it is worth noting that the group with the highest average age (i.e., U_{ambi}) also shows by far the highest standard deviation of age (i.e., 14.138 years).

In Fig. 10, we show the contribution of each music cluster to each subgroup in the form of a radar plot. For this, we use the user-cluster weights $w_{u,c}$ introduced before and calculate the average weight over all users in cluster c. One consequence of the IDF scoring applied to $w_{u,c}$ is that the weight contributions of a user group to the four clusters does not sum up to 1, which eventually influences the interpretation of the values shown in Fig. 10. However, in return, these values account for the varying cluster sizes and can also be interpreted as preference weights for a user group towards a specific music cluster.

We observe that the weight distribution of the two larger subgroups U_{hard} and U_{elec} is rather narrow, which indicates that these users do not listen to many tracks of other clusters. Contrary to that, the weights of the two smaller subgroups U_{folk} and U_{ambi} are more broadly distributed over the four music clusters. This suggests that users of U_{folk} and U_{ambi} are more open to music outside of their own music cluster than users of U_{hard} and U_{elec}.

4.2.1 Correlation of music clusters and beyond-mainstream subgroups

To better understand the correlations and connections between the music clusters and subgroups, we plot the Pearson correlation matrix of the four music clusters as a heatmap in Fig. 11. Here, we represent each music cluster c by a 2073-dimensional vector (i.e., one entry for each user) consisting of the user–cluster weights $w_{u,c}$, introduced before. Each element in the matrix is then calculated using the Pearson correlation measure based on these cluster vectors. For example, if there is a positive correlation between two clusters, we assume that a user who enjoys music from the one cluster likely also enjoys music from the other cluster. This can give us also an indication of the openness of a subgroup

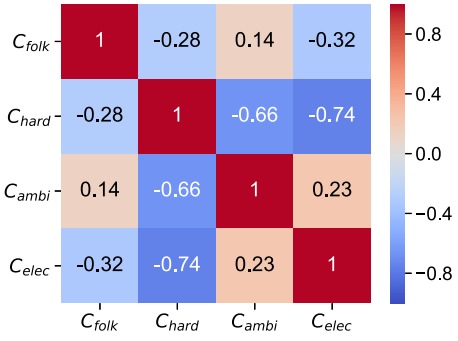

Figure 11 Pearson correlation matrix of the four music clusters. While C_{hard} has solely negative correlations with all other clusters, and thus, listeners of C_{hard} seem to be the most closed subgroup, C_{ambi} has positive correlations with C_{folk} and C_{elec}, and thus, listeners of C_{ambi} seem to be the most open subgroup

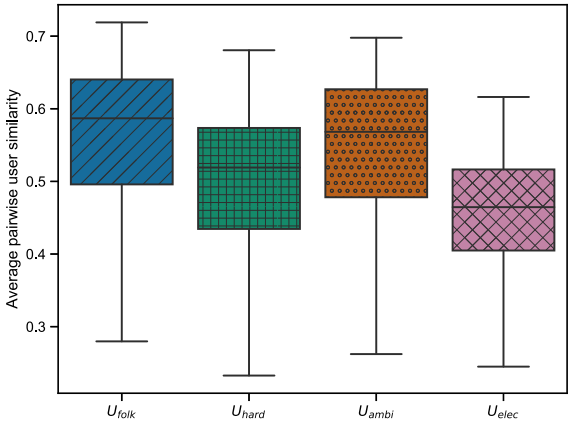

Figure 12 Boxplots showing the average pairwise user similarity of the four subgroups using the cosine similarity calculated on the users' genre distributions. While the users in U_{hard} and U_{elec} exhibit a more diverse listening behavior, users in U_{folk} and U_{ambi} tend to listen to more similar, i.e., less diverse, music genres

for music mainly listened to by other subgroups. Specifically, for C_{folk}, we see a positive correlation between C_{folk} and C_{ambi}, and a negative correlation between C_{folk} and both, C_{hard} as well as C_{elec}. Users listening to the music of C_{hard} seem to represent the most closed subgroup as C_{hard} because it solely has negative correlations with all other clusters, especially with C_{ambi} and C_{elec}. In contrast, users listening to the music of C_{ambi} seem to represent the most open subgroup as C_{ambi} has positive correlations with two other clusters, i.e., C_{folk} and C_{elec}. The fourth cluster, C_{elec}, is negatively correlated with C_{folk} and especially with C_{hard}, and positively correlated with C_{ambi}. These results are also in line with the ones shown in Fig. 10, in which we identify the users of U_{ambi} as more open music listeners than the ones of U_{hard}.

In order to relate the openness of the subgroups to the diversity of the users within the subgroups, we calculate the average pairwise user similarity using the cosine similarity metric computed on the users' genre distributions, i.e., number of listening events per genre. Figure 12 shows the resulting boxplots for the four identified subgroups (i.e., C_{folk}, C_{hard}, C_{ambi}, and C_{elec}). Figure 12 shows that users in U_{hard} and U_{elec} have a rather small average pairwise user similarity and, thus, exhibit a more diverse listening behavior, whereas users in U_{folk} and U_{ambi} tend to listen to more similar music genres and, thus, have a narrow listening behavior within the group. Summed up, we find pronounced differences with respect to openness and diversity across the subgroups. Although U_{ambi} is the most open

subgroup (i.e., also listens to music of other subgroups), it is also the least diverse subgroup (i.e., the users within the group listen to very similar music). That observation is in line with what is shown in Figs. 7, and Fig. 8. Here, we see that C_{ambi}, i.e., the most tightly connected music cluster to U_{ambi}, contains the dominating genre "ambient" as well as genres that are strongly associated with this dominating genre (e.g., "darkambient"). For U_{hard}, we observe the opposite. While it is the least open subgroup, it is also the most diverse one (e.g., it contains "hardrock" as well as "hiphop" listeners).

4.2.2 Recommendations for beyond-mainstream user subgroups

In Sect. 3.5, we have shown that the recommendation accuracy of four personalized recommendation algorithms is significantly worse for *BeyMS* users than for *MS* users. Now, we extend this analysis and evaluate the recommendation accuracy of these algorithms for the four subgroups (i.e., U_{folk}, U_{hard}, U_{ambi}, and U_{elec}).

Table 4 shows our results with respect to the mean absolute error (MAE). Additionally, we analyze these results with respect to statistically significant differences in Table 5 by performing ANOVA ($\alpha = 0.01$) and a subsequent Tukey-HSD test ($\alpha = 0.05$). Here, we report pairwise differences as significant (marked with **), if both ANOVA and Tukey-HSD were significant across all five folds (see Sect. 3.5 for details on the experimental setup).

We see that among all algorithms, the significantly worst accuracy results (i.e., the highest MAE scores) are achieved for the U_{hard} subgroup. Next, U_{folk}, U_{ambi} and U_{elec} reach significantly better (i.e., lower MAE scores) than U_{hard} for all algorithms. However, there is no statistically significant difference between the recommendation accuracy of U_{folk} and U_{elec}. The overall best accuracy results (i.e., lowest MAE scores) are reached for the U_{ambi} subgroup. These results are also statistically significant when compared with the other subgroups for the NMF algorithm. NMF also gives the overall best accuracy results for

Table 4 Mean absolute error (MAE) measurements for the four subgroups and four personalized recommendation algorithms. NMF (in bold) outperforms all other algorithms for all subgroups. Among the subgroups, the best accuracy results (i.e., lowest MAE scores) are reached by U_{ambi}, while the worst accuracy results (i.e., highest MAE scores) are reached by U_{hard}. To facilitate comparison, we also show the MAE measurements for the *BeyMS* and *MS* user groups

Subgroup	UserItemAvg	UserKNN	UserKNNAvg	NMF
U_{folk}	63.2143	70.3049	67.4406	**57.2278**
U_{hard}	65.1464	73.1949	69.2855	**59.6887**
U_{ambi}	60.5558	69.8315	65.5708	**54.2073**
U_{elec}	62.2894	71.0387	66.1499	**56.6209**
BeyMS	63.4608	71.6694	67.5856	**57.7703**
MS	61.2562	68.4894	63.3985	**54.8182**

Table 5 Statistically significant differences between pairs of subgroups, as determined by ANOVA ($\alpha = 0.01$) and a subsequent Tukey-HSD test ($\alpha = 0.05$)

Subgroup	UserItemAvg				UserKNN				UserKNNAvg				NMF			
	U_{folk}	U_{hard}	U_{ambi}	U_{elec}	U_{folk}	U_{hard}	U_{ambi}	U_{elec}	U_{folk}	U_{hard}	U_{ambi}	U_{elec}	U_{folk}	U_{hard}	U_{ambi}	U_{elec}
U_{folk}		**	**			**				**				**	**	
U_{hard}	**		**	**	**		**	**	**		**	**	**		**	**
U_{ambi}	**	**				**				**			**	**		**
U_{elec}		**				**				**				**	**	

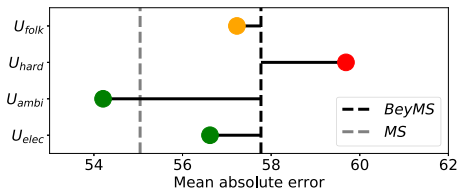

Figure 13 Comparison of the mean absolute error (MAE) scores reached by NMF for the four subgroups with the ones reached by NMF for *BeyMS* (black dashed line) and *MS* (gray dashed line). While specific subgroups (i.e., U_{hard}) are treated in an unfair way by recommendation algorithms, others (i.e., U_{ambi}) are not

all subgroups, which is in line with our results presented in Sect. 3.5 and in our previous work [4].

Furthermore, we find a relationship between openness, diversity, and recommendation quality. Here, U_{hard} is the least open but most diverse subgroup and gets the worst recommendations, while U_{ambi} is the most open but least diverse subgroup and gets the best recommendations. This is in line with the findings of [9], who have shown that users are more likely to accept recommendations from different groups (i.e., openness) rather than varied within a group (i.e., diversity). Thus, we find a relationship between the quality of recommendations provided to beyond-mainstream music listeners and openness as well as diversity patterns of these users.

Finally, in Fig. 13, we visually compare the MAE scores reached by the best performing approach NMF for the four subgroups. Additionally, we depict the MAE score for *BeyMS* as a black dashed line and the MAE score for *MS* as a gray dashed line. We see that U_{hard} reaches worse results than *BeyMS* while U_{folk} and U_{elec} reach slightly better results than *BeyMS*. Interestingly, U_{ambi} not only reaches better results than *BeyMS* but also better results than *MS*. Although this improvement over *MS* is not statistically significant (according to a one-tailed Mann–Whitney-U test with $\alpha = 0.0001$), it shows that there is a large variety among *BeyMS* users, where specific subgroups (i.e., U_{hard}) are disadvantaged in terms of recommendation accuracy by recommendation algorithms while others (i.e., U_{ambi}) are not.

5 Conclusions and future work

In this paper, we shed light on the characteristics of beyond-mainstream music and music listeners. As our first contribution, we identified 2074 beyond-mainstream music listeners (i.e., *BeyMS*) in the Last.fm platform, and subsequently created a novel dataset called *LFM-BeyMS* based on the listening histories of these users. We further enriched this dataset with (i) acoustic features of music tracks gathered from Spotify, and (ii) genre information of tracks derived from Last.fm tags and matched with the Spotify microgenre taxonomy. Additionally, for reasons of comparability, *LFM-BeyMS* contains data of 2074 Last.fm users listening to mainstream music. Using this dataset, as our second contribution, we validated related research by showing that beyond-mainstream music listeners receive a significantly lower recommendation accuracy than mainstream music listeners by four standard recommendation algorithms (i.e., UserItemAvg, UserKNN, UserKNNAvg and NMF).

As our third contribution, we applied the clustering algorithm HDBSCAN* on the acoustic features of tracks listened by *BeyMS* and identified four clusters of beyond-mainstream music: (i) C_{folk}, music with high acousticness such as "folk", (ii) C_{hard}, high energy music such as "hardrock", (iii) C_{ambi}, music with high acousticness and instrumen-

talness such as "ambient", and (iv) C_{elec}, music with high energy and instrumentalness such as "electronica".

As our fourth contribution, we mapped these clusters to our *BeyMS* users, which led to four beyond-mainstream subgroups: (i) U_{folk}, (ii) U_{hard}, (iii) U_{ambi}, and (iv) U_{elec}. We analyzed these subgroups with respect to their openness (i.e., across-groups diversity—do users of one group listen to music of other groups?) and diversity (i.e., within-groups diversity—how dissimilar is the music listened to by users within groups?). Here, we found large differences between U_{hard} and U_{ambi}. Although U_{hard} is the most closed subgroup (i.e., users do not listen to music of other subgroups), it is also the most diverse subgroup (i.e., users listen to a diverse set of genres such as "hardrock" and "hiphop"). For U_{ambi}, we get opposite results: while it is the most open subgroup (i.e., users listen to music of other subgroups as well), it is also the least diverse one (i.e., the users within the group listen to very similar music such as "ambient" and "darkambient"). We related these characteristics of the subgroups to the recommendation quality of the four recommendation algorithms UserItemAvg, UserKNN, UserKNNAvg and NMF. Here, we found that U_{hard} got music recommendations with lowest accuracy, while U_{ambi} got music recommendations with highest accuracy. This is in line with related research [9], which has shown that openness is stronger correlated with accurate recommendations than diversity. U_{ambi} even received better recommendations than the group of mainstream music listeners. This result highlights that there are large differences between the subgroups of beyond-music listeners. Finally, to foster reproducibility of our research, we provide our novel *LFM-BeyMS* dataset via Zenodo as well as our source code via Github.

We believe that our findings provide useful insights for creating user models and recommendation algorithms that better serve beyond-mainstream music listeners. As it was shown in [4], beyond-mainstream music listeners tend to have larger user profile sizes than users interested in mainstream music, which means that they provide a substantial amount of listening interaction data for services such as Last.fm and Spotify. We assume that improving the recommendation quality for this active user group also leads to another effect, namely a more prominent exposure of (long-tail) music artists due to a better-connected recommendation network [81]. We leave such investigations to future work.

Limitations Despite the merits of this work, we are aware of its limitations. The first limitation we recognize is that our analyses are based on a sample of the Last.fm community. The extent to which their listening behavior is representative of the Last.fm community at large, or similar music streaming communities such as Spotify, needs further investigation.

Next, since we conducted a comparative study of the accuracy of recommender systems algorithms—and were therefore not interested to beat state-of-the-art algorithms—we focused on traditional algorithms (e.g., KNN-based collaborative filtering) instead of investigating the most current deep learning architectures, which would also require a much higher computational effort. Furthermore, an award-winning-paper by Dacrema et al. [82] has recently shown that traditional algorithms are able to outperform almost all deep learning architectures.

Future work While our work serves as a first milestone towards better characterizing beyond-mainstream music and listeners of such music, future work should focus on user modeling techniques to individually target the different subgroups, for example by integrating knowledge about openness and diversity. With respect to analyzing openness and

diversity of users and user groups, we would also like to work on a more formal definition of these dimensions, which would not only allow us to measure them more precisely but also to integrate them into the recommendation calculation process.

Additionally, since previous research has shown that the listener's cultural background impacts the quality of music recommendations [48], we plan to compare the cultural and socioeconomic aspects of beyond-mainstream and mainstream music listeners. We plan to employ these aspects by means of Hofstede's cultural dimensions [83] and the World Happiness Report [84].

Finally, another avenue for future work is the research in the area of fair music recommender systems. Here, we plan to build user models that are capable of accounting for the complex characteristics of beyond-mainstream music listeners presented in this paper. While we believe that more specialized user models could help to provide better recommendations for users who currently receive worse recommendations (e.g., the U_{hard} subgroup identified in this paper), we also aim to highlight that such user models still need to be generalizable to avoid any unfair treatment of other users. Hence, future research should work on achieving a specialization-generalization trade-off in music recommender systems. We hope that our open *LFM-BeyMS* dataset as well as our source code will be of use to the scientific community for subsequent analyses.

Acknowledgements
This work is supported by the Know-Center GmbH within the COMET—Competence Centers for Excellent Technologies Programme, funded by the Austrian Federal Ministry for Transport, Innovation and Technology (bmvit), the Austrian Federal Ministry for Digital and Economic Affairs (bmdw), the Austrian Research Promotion Agency (FFG), the province of Styria (SFG) and partners from industry and academia. The COMET Programme is managed by FFG. program.

Funding
This work is funded by the TU Graz Open Access Publishing Fund and the Austrian Science Fund (FWF): V579.

Availability of data and materials
The *CultMRS* dataset utilized during the current study is made available on Zenodo, accessible via the following DOI: https://doi.org/10.5281/zenodo.3477842 and further described in [50]. Additionally, we provide the novel *LFM-BeyMS* dataset on Zenodo: https://doi.org/10.5281/zenodo.3784764. The entire Python-based implementation of the experiments conducted in this study is publicly available at https://github.com/pmuellner/supporttheunderground.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
All authors contributed to manuscript revision, read, and approved the submitted version.

Author details
[1]Know-Center GmbH, Graz, Austria. [2]University of Innsbruck, Innsbruck, Austria. [3]Utrecht University, Utrecht, The Netherlands. [4]Johannes Kepler University Linz, Linz, Austria. [5]Linz Institute of Technology AI Lab, Linz, Austria. [6]Graz University of Technology, Graz, Austria.

Publisher's Note

Received: 7 August 2020 Accepted: 23 February 2021 Published online: 30 March 2021

References

1. Schedl M, Knees P, McFee B, Bogdanov D, Kaminskas M (2015) Music recommender systems. In: Recommender systems handbook, pp 453–492
2. Abdollahpouri H, Mansoury M, Burke R, Mobasher B (2019) The unfairness of popularity bias in recommendation. In: RMSE workshop held in conjunction with the 13th ACM conferenceon recommender systems (RecSys)
3. Celma O (2009) Music recommendation and discovery in the long tail. PhD thesis, Universitat Pompeu Fabra
4. Kowald D, Schedl M, Lex E (2020) The unfairness of popularity bias in music recommendation: a reproducibility study. In: European conference on information retrieval. Springer, Berlin, pp 35–42
5. Celma O, Cano P (2008) From hits to niches?: or how popular artists can bias music recommendation and discovery. In: Proceedings of KDD '2018 (Netflix price workshop)

6. Celma O (2010) Music recommendation and discovery—the long tail, long fail, and long play in the digital music space. Springer
7. van den Oord A, Dieleman S, Schrauwen B (2013) Deep content-based music recommendation. In: Proceedings of NIPS '2013. Curran Associates, Red Hook, pp 2643–2651
8. Goel S, Broder A, Gabrilovich E, Pang B (2010) Anatomy of the long tail: ordinary people with extraordinary tastes. In: Proceedings of the third ACM international conference on web search and data mining, pp 201–210
9. Tintarev N, Dennis M, Masthoff J (2013) Adapting recommendation diversity to openness to experience: a study of human behaviour. In: Carberry S, Weibelzahl S, Micarelli A, Semeraro G (eds) User modeling, adaptation, and personalization. Springer, Berlin, pp 190–202
10. Luo X, Zhou M, Xia Y, Zhu Q (2014) An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. IEEE Trans Ind Inform 10(2):1273–1284
11. Herlocker JL, Konstan JA, Terveen LG, Riedl JT (2004) Evaluating collaborative filtering recommender systems. ACM Trans Inf Syst 22(1):5–53
12. Schedl M, Zamani H, Chen C-W, Deldjoo Y, Elahi M (2018) Current challenges and visions in music recommender systems research. Int J Multimed Inf Retr 7(2):95–116
13. Haas R, Brandes V (2010) Music that works: contributions of biology, neurophysiology, psychology, sociology, medicine and musicology. Springer Science & Business Media
14. Adorno TW (1988) Introduction to the sociology of music. Burns & Oates
15. Deutsch D (2013) Psychology of music. Elsevier
16. Laplante A (2014) Improving music recommender systems: what can we learn from research on music tastes? In: Proceedings of the International Society for Music Information Retrieval conference (ISMIR)
17. Rentfrow PJ, Gosling SD (2007) The content and validity of music-genre stereotypes among college students. Psychol Music 35(2):306–326
18. Kim Y, Aiello LM, Quercia D (2020) Pepmusic: motivational qualities of songs for daily activities. EPJ Data Sci 9(1):13
19. Juslin PN, Sloboda JA (2001) Music and emotion: theory and research. Oxford University Press
20. Zentner M, Grandjean D, Scherer KR (2008) Emotions evoked by the sound of music: characterization, classification, and measurement. Emotion 8(4):494
21. Juslin PN, Laukka P (2004) Expression, perception, and induction of musical emotions: a review and a questionnaire study of everyday listening. J New Music Res 33(3):217–238
22. Yang Y-H, Chen HH (2011) Music emotion recognition. CRC Press
23. Ferwerda B, Schedl M, Tkalcic M (2015) Personality & emotional states: understanding users' music listening needs. In: Late-breaking results of 23rd international conference on user modeling, adaptation and personalization (UMAP)
24. Goldberg LR (1993) The structure of phenotypic personality traits. Am Psychol 48(1):26
25. Schubert E (2007) The influence of emotion, locus of emotion and familiarity upon preference in music. Psychol Music 35(3):499–515
26. Pereira CS, Teixeira J, Figueiredo P, Xavier J, Castro SL, Brattico E (2011) Music and emotions in the brain: familiarity matters. PLoS ONE 6(11):e27241
27. Moore JL, Chen S, Turnbull D, Joachims T (2013) Taste over time: the temporal dynamics of user preferences. In: Proceedings of the International Society for Music Information Retrieval conference (ISMIR), pp 401–406
28. Barone MD, Bansal J, Woolhouse MH (2017) Acoustic features influence musical choices across multiple genres. Front Psychol 8:931
29. Gong B, Kaya M, Tintarev N (2020) Contextual personalized re-ranking of music recommendations through audio features. Master's thesis, TU Delft
30. Zangerle E, Pichl M (2018) Content-based user models: modeling the many faces of musical preference. In: Proceedings of the 19th International Society for Music Information Retrieval conference 2018 (ISMIR 2018), pp 709–716
31. Ekstrand MD, Tian M, Azpiazu IM, Ekstrand JD, Anuyah O, McNeill D, Pera MS (2018) 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
32. Brynjolfsson E, Hu YJ, Smith MD (2006) From niches to riches: anatomy of the long tail. Sloan Manag Rev 47(4):67–71
33. Jannach D, Lerche L, Kamehkhosh I, Jugovac M (2015) What recommenders recommend: an analysis of recommendation biases and possible countermeasures. User Model User-Adapt Interact 25(5):427–491
34. Harper FM, Konstan JA (2015) The movielens datasets: history and context. ACM Trans Interact Intell Syst 5(4):1–19
35. Cheng R, Tang B (2016) A music recommendation system based on acoustic features and user personalities. In: Pacific-Asia conference on knowledge discovery and data mining. Springer, Berlin, pp 203–213
36. Kaminskas M, Ricci F, Schedl M (2013) Location-aware music recommendation using auto-tagging and hybrid matching. In: Proceedings of RecSys '2013. ACM, Hong Kong, pp 17–24
37. Donaldson J (2007) A hybrid social-acoustic recommendation system for popular music. In: Proceedings of RecSys '2007. ACM, New York, pp 187–190
38. Aggarwal CC (2016) Ensemble-based and hybrid recommender systems. In: Recommender systems, pp 199–224
39. Zangerle E, Pichl M (2018) Content-based user models: modeling the many faces of musical preference. In: 19th International Society for Music Information Retrieval conference (ISMIR)
40. Lee K, Lee K (2011) My head is your tail: applying link analysis on long-tailed music listening behavior for music recommendation. In: Proceedings of the 5th ACM conference on recommender systems, pp 213–220
41. Lex E, Kowald D, Schedl M (2020) Modeling popularity and temporal drift of music genre preferences. Trans Int Soc Music Inf Retr 3(1):17–30
42. Kowald D, Lex E, Schedl M (2019) Modeling artist preferences of users with different music consumption patterns for fair music recommendations. In: Late-breaking-results of the 20th annual conference of the International Society for Music Information Retrieval (ISMIR)
43. Kowald D, Kopeinik S, Lex E (2017) The tagrec framework as a toolkit for the development of tag-based recommender systems. In: Adjunct publication of the 25th conference on user modeling, adaptation and personalization, pp 23–28
44. Bauer C (2019) Allowing for equal opportunities for artists in music recommendation. In: 1st workshop on Designing Human-Centric Music Information Research systems in conjunction with ISMIR confernce

45. Pichl M, Zangerle E, Specht G (2016) Understanding playlist creation on music streaming platforms. In: IEEE international symposium on multimedia, ISM 2016, pp 475–480
46. Andersen JS (2014) Using the echo nest's automatically extracted music features for a musicological purpose. In: 4th international workshop on cognitive information processing (CIP), pp 1–6
47. McVicar M, Freeman T, De Bie T (2011) Mining the correlation between lyrical and audio features and the emergence of mood. In: Proceedings of the 11th International Society for Music Information Retrieval conference (ISMIR), pp 783–788
48. Zangerle E, Pichl M, Schedl M (2020) User models for culture-aware music recommendation: fusing acoustic and cultural cues. Trans Int Soc Music Inf Retr 3(1):1–16. https://doi.org/10.5334/tismir.37
49. Schedl M (2016) The lfm-1b dataset for music retrieval and recommendation. In: Proceedings of the 2016 ACM on international conference on multimedia retrieval. ACM, New York, pp 103–110
50. Zangerle E Culture-aware music recommendation dataset. https://doi.org/10.5281/zenodo.3477842
51. Bauer C, Schedl M (2019) Global and country-specific mainstreaminess measures: definitions, analysis, and usage for improving personalized music recommendation systems. PLoS ONE 14(6):e0217389. https://doi.org/10.1371/journal.pone.0217389
52. Kendall MG (1938) A new measure of rank correlation. Biometrika 30(1/2):81–93
53. Sheather SJ (2004) Density estimation. Stat Sci 19(4):588–597
54. Davis RA, Lii K-S, Politis DN (2011) Remarks on some nonparametric estimates of a density function. In: Selected works of Murray Rosenblatt. Springer, New York, pp 95–100
55. Quarteroni A, Sacco R, Saleri F (2007) Numerical mathematics. Springer, New York
56. Jones KS (1972) A statistical interpretation of term specificity and its application in retrieval. J Doc 28(1):11–21
57. Schedl M, Bauer C (2017) Distance-and rank-based music mainstreaminess measurement. In: Adjunct publication of the 25th conference on user modeling, adaptation and personalization. ACM, New York, pp 364–367
58. Koren Y (2010) Factor in the neighbors: scalable and accurate collaborative filtering. ACM Trans Knowl Discov Data 4(1):1
59. Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (mae) over the root mean square error (RMSE) in assessing average model performance. Clim Res 30(1):79–82
60. Moore JL, Chen S, Joachims T, Turnbull D (2012) Learning to embed songs and tags for playlist prediction. In: Proceedings of the 12th International Society for Music Information Retrieval conference (ISMIR), vol 12, pp 349–354
61. Levy M, Sandler M (2008) Learning latent semantic models for music from social tags. J New Music Res 37(2):137–150
62. Tipping ME, Bishop CM (1999) Mixtures of probabilistic principal component analyzers. Neural Comput 11(2):443–482
63. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500):2323–2326
64. Kruskal JB (1964) Nonmetric multidimensional scaling: a numerical method. Psychometrika 29(2):115–129
65. Tenenbaum JB, De Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290(5500):2319–2323
66. Ng AY, Jordan MI, Weiss Y (2002) On spectral clustering: analysis and an algorithm. In: Advances in neural information processing systems, pp 849–856
67. van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 9:2579–2605
68. McInnes L, Healy J, Melville J (2018) UMAP: uniform manifold approximation and projection for dimension reduction. J Open Sour Softw
69. McInnes L, Healy J, Saul N, Grossberger L (2018) UMAP: uniform manifold approximation and projection. J Open Sour Softw 3(29):861
70. Ester M, Kriegel H-P, Sander J, Xu X et al (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD, vol 96, pp 226–231
71. Bishop CM (2006) Pattern recognition and machine learning Springer, New York, pp 424–429
72. Reynolds D (2015) Gaussian mixture models. In: Encyclopedia of biometrics, pp 827–832
73. Frey BJ, Dueck D (2007) Clustering by passing messages between data points. Science 315(5814):972–976
74. Shi J, Malik J (2000) Normalized cuts and image segmentation. Departmental Papers (CIS), 107
75. Murtagh F, Legendre P (2014) Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? J Classif 31(3):274–295
76. Ankerst M, Breunig MM, Kriegel H-P, Sander J (1999) Optics: ordering points to identify the clustering structure. In: ACM sigmod record, vol 28. ACM, New York, pp 49–60
77. McInnes L, Healy J (2017) Accelerated hierarchical density based clustering. In: Data mining workshops (ICDMW), 2017 IEEE international conference on. IEEE, New York, pp 33–42
78. Yoo S, Lee K (2017) A data-driven approach to identifying music listener groups based on users' playrate distributions of listening events. In: Adjunct publication of the 25th conference on user modeling, adaptation and personalization, pp 77–81
79. McInnes L, Healy J, Astels S (2017) hdbscan: hierarchical density based clustering. J Open Sour Softw 2(11):205
80. York W (2004) Voices from hell—the dark, not-so-dulcet cookie monster vocals of extreme metal. The San Francisco Bay Guardian, 14–20
81. Lamprecht D, Strohmaier M, Helic D (2017) A method for evaluating discoverability and navigability of recommendation algorithms. Comput Soc Netw 4(1):9
82. Dacrema MF, Cremonesi P, Jannach D (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 2019, Copenhagen, Denmark, September 16–20, 2019, pp 101–109
83. Hofstede G, Hofstede GJ, Minkov M (2010) Cultures and organizations: software of the mind, 3rd edn. McGraw-Hill, New York
84. Helliwell JF, Layard R, Sachs J (2016) World happiness report 2016 update. Sustainable Development Solutions Network, New York