Introducing Global and Regional Mainstreaminess for Improving Personalized Music Recommendation

Markus Schedl
Johannes Kepler University
Linz, Austria
markus.schedl@jku.at

Christine Bauer
Johannes Kepler University
Linz, Austria
christine.bauer@jku.at

ABSTRACT

The *music mainstreaminess of a user* reflects how strong a user's listening preferences correspond to those of the larger population. Considering that music mainstream may be defined from different perspectives and on various levels, e.g., geographical (charts of a country), genre ("Indie charts"), or distribution channel (radio charts vs. download charts), we study how the user's music mainstreaminess influences the quality of music recommendations.

The paper's contribution is three-fold. First, we propose 11 novel mainstreaminess measures characterizing music listeners, considering both a global and a country-specific basis. To this end, we model *preference profiles* (as a vector over artists) for users, countries, and globally, incorporating artist frequency, listener frequency, and a newly proposed TF-IDF-inspired weighting function, which we call artist frequency–inverse listener frequency (AF-ILF). The resulting preference profile for each user u is then related to the respective country-specific and global preference profile using fraction-based approaches, symmetrized Kullback-Leibler divergence, and Kendall's τ rank correlation, in order to quantify u's mainstreaminess. Second, we demonstrate country-specific peculiarities of these mainstreaminess definitions. Third, we show that incorporating the proposed global and country-specific mainstreaminess measures into the music recommendation process can notably improve accuracy of rating prediction.

CCS CONCEPTS

• **Information systems → Multimedia information systems**; **Personalization**; **Recommender systems**; *Personalization*; *Rank aggregation*; *Clustering and classification*; Collaborative filtering; Social recommendation; • **Social and professional topics → Cultural characteristics**; • **Applied computing** → *Sound and music computing*;

KEYWORDS

Music mainstreaminess, music recommender systems, artist frequency-inverse listener frequency, popularity

ACM Reference format:
Markus Schedl and Christine Bauer. 2017. Introducing Global and Regional Mainstreaminess for Improving Personalized Music Recommendation. In *Proceedings of The 15th International Conference on Advances in Mobile Computing & Multimedia, Salzburg, Austria, December 4–6, 2017 (MoMM '17),* 8 pages.
https://doi.org/10.1145/3151848.3151849

MoMM '17, December 4–6, 2017, Salzburg, Austria

ACM ISBN 978-1-4503-5300-7/17/12.
https://doi.org/10.1145/3151848.3151849

1 INTRODUCTION

In the era of digitalization, music has become easier to access than ever: a tremendous number of musical recordings are readily available to consume on online music platforms such as YouTube, Spotify, or iTunes. This opportunity to access a large number of musical works, though, results in information overload and users require novel mechanisms and strategies to choose from the "deep blue sea of musical works" [30]. Music recommender systems (MRS) have, thus, become a significant topic both in research as well as in industry, over the past few years [5, 8, 33].

In general, recommender systems are meant to assist users in searching, sorting, and filtering the vast amount of information available [22]. MRS are specifically built to assist users in navigating through the myriad of available musical recordings and provide them with suggestions (and/or automatically generated playlists) that would fit the respective user's interest [33]. Thereby, "[t]he success of a music recommender system (RS) depends on its ability to propose the right music, to the right user, at the right moment" [19].

Various automatic approaches to music recommendation have been proposed [35]. Thereby, most MRS rely mainly on collaborative filtering [20] or on information about music items (i.e., content-based filtering [4]) [35]. For instance, content-based MRS may consider acoustic similarity information on the song level [38], or genre or artist similarity [21]. MRS employing collaborative filtering do not require exogenous information about neither users nor music items. Instead, a user is suggested music listened to by users with similar preferences and/or listening patterns [25].

An approach that is particularly applicable in hit-driven domains such as the music industry is popularity-based recommendation. This approach assumes that a random user is more likely to like a very popular music item than one of the far less popular items [8, 34]. Popularity-based MRS approaches are widely adopted to complement other approaches in cold start situations, when there is limited information about new users and/or items available in the system [10, 39]. One approach for considering popularity in the music domain is to describe music listeners "in terms of the degree to which they prefer music items that are currently popular or rather ignore such trends" [29]. Harnessing music mainstreaminess in combination with collaborative filtering techniques tends to deliver better results with respect to music recommendation accuracy and rating prediction error than pure collaborative filtering approaches alone [13, 32, 34, 37].

However, a limitation of existing work on quantifying a user's music mainstreaminess is that music mainstream is viewed from a global perspective. There exist regional peculiarities to mainstream, though, as music consumption behavior is affected by culturally influenced music preferences, market regulations, local radio airplay, etc. (e.g., [7, 16, 26, 36]). With respect to the music recommendation research domain, the definition of specific measures that can capture a user's mainstreaminess (i) on both, a global and a country-specific level, and (ii) in ways that can easily be operationalized in music recommendation is a new target of research.

Calling on this, the main contributions of this paper are threefold: (i) the definition of several novel measures for user mainstreaminess, considering both a global and a regional, country-specific basis, (ii) illustrating country-specific peculiarities of these mainstreaminess definitions, and (iii) analyzing the performance of the proposed mainstreaminess measures for personalized music recommendation.

The remainder of the paper is organized as follows. In Section 2, we provide a brief overview over existing work on mainstreaminess and popularity in music recommendation. We then detail our proposed mainstreaminess measures in Section 3 and provide examples that show their value to distill the regional mainstream, in addition to a global one. Section 4 shows how to exploit the proposed mainstreaminess measures in collaborative filtering recommendation and highlights the additional values of doing so. Eventually, we round off the paper in Section 5 with a conclusion and directions for future research.

2 CONCEPTUAL FOUNDATIONS AND RELATED WORK

2.1 Music Popularity and Mainstreaminess

In the context of recommender systems, popularity-based approaches are widely adopted in numerous domains, including music [10, 18, 39], news [40], or product recommendation in e-commerce in general [1]. Popularity is thereby typically constructed as a general consensus of a group's attitude about entities [18].

While various ways exist to define and measure popularity (for instance, in terms of sales figures, media coverage, etc.), in the field of MRS, music popularity is frequently characterized by using the total playcounts of a music item, cf. [8].

With respect to music popularity by using playcounts, the *long tail concept* as described in [2] is specifically applicable to the (online) music industry [9]; on online music platforms there is a concentration of playcounts on the most popular music items (the head) and then there is a long tail of less popular items [6, 8].

A more general concept to popularity concentration is referred to as *mainstream*. Although literature in the field of popular music studies and popular music cultures references to *mainstream* frequently, the term itself remains rather poorly defined, cf. e.g., [3]. According to the Oxford Dictionaries, *mainstream* is defined as "The ideas, attitudes, or activities that are shared by most people and regarded as normal or conventional". Due to the strong connection of the concepts, the terms *mainstream* and *long tail* are often interchangeably used. The mainstream is thereby frequently also referred to with other terms and phrases (e.g., *hits* [8], *the head* [12]) to circumscribe the phenomenon; the overall concept is also called, for instance, the hit-driven paradigm [8], the long-tail concept [2, 8], etc.

In MRS research, the user feature *music mainstreaminess of a user* [13, 34] essentially describes whether and how strong a user's music listening preferences correspond to those of the overall population. While other listening-centric features, for instance, serendipity [41] or novelty [11], are frequently exploited when modeling a user's music consumption behavior and providing music recommendations, music mainstreaminess is a rather new target of research [13, 34, 37]. Thereby, the mainstreaminess feature is used to analyze a user's ranking of music items and compare it with the overall ranking of artists, albums, or tracks [37].

2.2 Related Work on the Quantification of Music Mainstreaminess

Formal definitions to measure the level of music mainstreaminess of a user are scarce in literature (e.g., [32, 34, 37]). Most existing approaches quantify music mainstreaminess as fractions of the target user's playcounts among the playcounts of the overall population. A limitation of this approach is that it disproportionately privileges the absolute top hits [32], which is problematic in long-tail economies [2] such as the music economy, where there is a high concentration of demands on the most popular items and a long tail of less popular items. Privileging the top hits leads to lower performance when considering fraction-based user models of mainstreaminess in collaborative filtering approaches [32].

To overcome this limitation, Schedl and Bauer [32] proposed measurement approaches based on *rank-order correlation* and *Kullback-Leibler (KL) divergence*. However, also their work shares with existing fraction-based approaches to quantify mainstreaminess that music mainstream is viewed from a global perspective and does not take regional peculiarities of music mainstream into account.

2.3 Music Mainstreaminess and Cultural Aspects

As human preferences and behavior are rooted and embodied in culture [17], also music preferences and music consumption behavior are affected by cultural aspects [14, 16, 36]. Not only cultural aspects, but also other regional (e.g., country-specific) mechanisms that affect music consumption (e.g., market regulations, access to music items, radio airplay), shape the regional mainstream.

Against this background, we focus on country-specific differences in the paper at hand.

Closest to our work is the study presented in [37], which analyzes the recommendation performance of mainstreaminess (spelled "mainstreamness") and a user's country, among other features. Our work significantly differs from [37] in various regards: First, we use an open dataset to allow for replication. Second, [37] propose only one global mainstreaminess measure that compares a user's preferences to the overall dataset (global population), while we define mainstreaminess in various ways (based on fractional, divergence, and rank correlation functions) and at various levels (global and country-specific). Third, we also propose a novel weighting approach based on "inverse listening frequency" that highlights artists popular in a specific country, thus, contributing to its mainstream, but not necessarily on a global level.

Artist	AF	Artist	LF
The Beatles	2,985,509	Radiohead	24,829
Radiohead	2,579,453	Nirvana	24,249
Pink Floyd	2,351,436	Coldplay	23,714
Metallica	1,970,569	Daft Punk	23,661
Muse	1,896,941	Red Hot Chili Peppers	22,609
Arctic Monkeys	1,803,975	Muse	22,429
Daft Punk	1,787,739	Queen	21,778
Coldplay	1,755,333	The Beatles	21,738
Linkin Park	1,691,122	Pink Floyd	21,129
Red Hot Chili Peppers	1,627,851	David Bowie	20,602

Table 1: Global top artists in the LFM-1b dataset, according to artist frequency (AF) and listener frequency (LF), considering the 53,258 users with country information.

Artist	AF
Stam1na	105,633
In Flames	97,645
CMX	90,032
Kotiteollisuus	82,309
Turmion Kätilöt	78,722
Amorphis	78,159
Nightwish	75,742
Mokoma	73,453
Muse	69,507
Metallica	69,499
Artist	**LF**
Metallica	703
Nightwish	695
Muse	693
Daft Punk	675
Queen	671
System of a Down	663
Coldplay	634
Nirvana	614
Pendulum	613
Iron Maiden	609
Artist	**AF-ILF**
St. Hood	70.526
The Sun Sawed in 1/2	67.490
tiko-μ	66.546
Worth the Pain	66.058
Cutdown	65.247
Katariina Hänninen	64.955
Game Music Finland	64.835
Daisuke Ishiwatari	63.565
Altis	63.235
Redrum-187	62.428

(a) Finland (1,407 users)

Artist	AF
Radiohead	68,160
The Beatles	65,498
Pink Floyd	60,558
Fabrizio De André	53,928
Muse	48,168
Depeche Mode	42,586
Afterhours	42,473
Verdena	42,338
Sigur Rós	41,748
Arctic Monkeys	39,755
Artist	**LF**
Radiohead	556
Pink Floyd	539
The Beatles	505
David Bowie	500
Muse	500
Nirvana	497
Coldplay	475
The Cure	466
Depeche Mode	459
Daft Punk	457
Artist	**AF-ILF**
CaneSecco	68.451
DSA Commando	66.049
Veronica Marchi	65.864
Train To Roots	65.459
Alessandro Raina	64.228
Machete Empire	63.915
Danti	62.958
Dargen D'Amico	62.453
宝塚歌劇団・宙組	62.228
Aquefrigide	61.663

(b) Italy (972 users)

Artist	AF
Pink Floyd	68,887
Metallica	42,784
Daft Punk	42,020
Iron Maiden	34,174
Radiohead	31,390
Massive Attack	30,669
The Beatles	27,951
Opeth	25,744
Depeche Mode	25,075
Dream Theater	24,286
Artist	**LF**
Pink Floyd	292
Radiohead	289
Metallica	268
Coldplay	261
Nirvana	251
Massive Attack	249
The Beatles	240
Red Hot Chili Peppers	240
Queen	238
Led Zeppelin	236
Artist	**AF-ILF**
Cüneyt Ergün	64.473
Floyd Red Crow Westerman	61.955
Fırat Tanış	58.666
Acil Servis	58.439
Taste (Rory Gallager)	58.366
Mezarkabul	57.799
Rachmaninoff Sergey	57.733
Mabel Matiz	57.619
Grup Yorum	56.855
Yüzyüzeyken Konuşuruz	56.748

(c) Turkey (479 users)

Table 2: Top artists for selected countries, according to artist frequency (AF), listener frequency (LF), and artist frequency–inverse listener frequency (AF-ILF).

3 FORMALIZING MAINSTREAMINESS

When describing how well a user's listening preferences reflect those of an overall population, e.g., globally or within a country, what is considered *mainstream* depends on the selection of a population; this is a phenomenon which we will also show in our analysis. Consequently, we propose several quantitative measures for user mainstreaminess, both on a global and on a country-specific level, depending on the selection of the population against which the user is compared. Our approach is inspired by the well-established monotonicity assumptions in text processing and information retrieval [28] – the TF-IDF (term frequency–inverse document frequency) weighting. Based on this assumption, our proposed mainstreaminess measures rely on the concepts of *artist frequency (AF)*, *listener frequency (LF)*, and *artist frequency–inverse listener frequency (AF-ILF)*.

We define $AF_{a,U}$ as the sum of the number of tracks by artist a listened to by a set of users U. Note that U may be a single user u, all users in a country c, or the entirety of users in the collection (i.e., the global population g). Accordingly, we define $LF_{a,U}$ as the number of

Abbr.	Formula
$F_{g:AF,u:AF}$	$1 - \frac{1}{\|A\|} \cdot \sum_{a \in A} \frac{\|\widehat{AF_{a,u}} - \widehat{AF_{a,g}}\|}{\max\left(\widehat{AF_{a,u}}, \widehat{AF_{a,g}}\right)}$
$F_{g:AF,u:AF \cdot ILF}$	$1 - \frac{1}{\|A\|} \cdot \sum_{a \in A} \frac{\|\widehat{AF \cdot ILF_{a,u,g}} - \widehat{AF}_{a,g}\|}{\max\left(\widehat{AF \cdot ILF_{a,u,g}}, \widehat{AF}_{a,g}\right)}$
$F_{g:AF \cdot ILF,u:AF \cdot ILF}$	$1 - \frac{1}{\|A\|} \cdot \sum_{a \in A} \frac{\|\widehat{AF \cdot ILF_{a,u,g}} - \widehat{AF \cdot ILF_{a,g,g}}\|}{\max\left(\widehat{AF \cdot ILF_{a,u,g}}, \widehat{AF \cdot ILF_{a,g,g}}\right)}$
$F_{c:AF,u:AF}$	$1 - \frac{1}{\|A\|} \cdot \sum_{a \in A} \frac{\|\widehat{AF_{a,u}} - \widehat{AF_{a,c}}\|}{\max\left(\widehat{AF_{a,u}}, \widehat{AF_{a,c}}\right)}$
$F_{c:AF \cdot ILF,u:AF \cdot ILF}$	$1 - \frac{1}{\|A\|} \cdot \sum_{a \in A} \frac{\|\widehat{AF \cdot ILF_{a,u,c}} - \widehat{AF \cdot ILF_{a,c,g}}\|}{\max\left(\widehat{AF \cdot ILF_{a,u,c}}, \widehat{AF \cdot ILF_{a,c,g}}\right)}$
$D_{g:AF,u:AF}$	$\frac{1}{2} \cdot \left(\sum_{a \in A} \widehat{AF_{a,u}} \cdot \log \frac{\widehat{AF_{a,u}}}{\widehat{AF_{a,g}}} + \sum_{a \in A} \widehat{AF_{a,g}} \cdot \log \frac{\widehat{AF_{a,g}}}{\widehat{AF_{a,u}}} \right)^{-1}$
$D_{c:AF,u:AF}$	$\frac{1}{2} \cdot \left(\sum_{a \in A} \widehat{AF_{a,u}} \cdot \log \frac{\widehat{AF_{a,u}}}{\widehat{AF_{a,c}}} + \sum_{a \in A} \widehat{AF_{a,c}} \cdot \log \frac{\widehat{AF_{a,c}}}{\widehat{AF_{a,u}}} \right)^{-1}$
$D_{c:AF \cdot ILF,u:AF \cdot ILF}$	$\frac{1}{2} \cdot \left(\sum_{a \in A} \widehat{AF \cdot ILF_{a,u,g}} \cdot \log \frac{\widehat{AF \cdot ILF_{a,u,g}}}{\widehat{AF \cdot ILF_{a,c,g}}} + \sum_{a \in A} \widehat{AF \cdot ILF_{a,c,g}} \cdot \log \frac{\widehat{AF \cdot ILF_{a,c,g}}}{\widehat{AF \cdot ILF_{a,u,g}}} \right)^{-1}$
$C_{g:AF,u:AF}$	$\tau\left(ranks\left(PP_g^{AF}\right), ranks\left(PP_u^{AF}\right)\right)$
$C_{c:AF,u:AF}$	$\tau\left(ranks\left(PP_c^{AF}\right), ranks\left(PP_u^{AF}\right)\right)$
$C_{c:AF \cdot ILF,u:AF \cdot ILF}$	$\tau\left(ranks\left(PP_{u,c}^{AF \cdot ILF}\right), ranks\left(PP_{c,g}^{AF \cdot ILF}\right)\right)$

Table 3: Proposed music mainstreaminess measures on the user level. Terms denote the following: F stands for the fraction-based approach, D refers to the symmetrized Kullback-Leibler divergence approach, and C is used as abbreviation for the approaches based on rank-order correlation according to Kendall's τ. A is a list of all artists; $\widehat{AF}$ denotes the sum-to-unity normalized AF value; $ranks(PP_u^W)$ represents the real-valued preference profile converted to ranks, i.e. the vector containing all normalized item frequencies of user u, with respect to the frequency weighting approach W (AF or LF); in case of $AF \cdot ILF$, $ranks\left(PP_u^W\right)$ is extended to $ranks\left(PP_{u,c}^{AF \cdot ILF}\right)$, i.e. AF computed for user u, ILF on country c, or $ranks\left(PP_{c,g}^{AF \cdot ILF}\right)$, i.e. AF computed on country c, ILF globally. Note that we invert the values of some measures (F and D) in order to ensure that higher values always indicate closer to the mainstream.

listeners of artist a within a user population U. And we eventually define $AF \cdot ILF_{a,U_1,U_2}$ as in Equation 1. We set $AF \cdot ILF_{a,U_1,U_2} = 0$ iff $LF_{a,U_2} = 0$.

$$AF \cdot ILF_{a,U_1,U_2} = \log\left(1 + AF_{a,U_1}\right) \cdot \log\left(1 + \frac{|U_2|}{LF_{a,U_2}}\right) \quad (1)$$

Note that U_1 and U_2 may represent a single user, all users in the same country, or all users in the dataset. Therefore, this definition allows us to easily formalize both the global and the regional definitions of mainstreaminess, by varying U_1 and U_2. The ILF weighting term can be integrated when computing the *preference profile* for a user or for a country, e.g., $AF \cdot ILF_{a,u,c}$, where U_1 contains only the user u and U_2 all users in country c (to which u belongs), or $AF \cdot ILF_{a,c,g}$, where U_1 is composed of all users in country c (to which u belongs) and U_2 of all users in the dataset. Using ILF is motivated by the fact that, when determined by $AF_{a,c}$ or $LF_{a,c}$, the top artists in each country c are often identical or very similar to the global top artists (cf. Tables 1 and 2). In order to uncover the respective country-specific mainstream, we therefore use $ILF_{a,g}$ to penalize globally popular artists.

Table 2 illustrates the effect of this weighting. It shows the top artists for Finland, Italy, and Turkey, in terms of $AF_{a,c}$, $LF_{a,c}$, and

$AF \cdot ILF_{a,c,g}$, i.e., AF computed on the country level, ILF on the global level. Please note that artist IDs (on the x-axis) are sorted with respect to their *global* popularity in regards to the respective measure (AF, LF, or AF-ILF). As can be seen, the AF and even more the LF measures are not suited well to distill the essential mainstream of a country, except maybe for countries such as Finland that show a very specific music taste far away from the global taste [31]. In contrast, AF-ILF is capable of identifying those artists that are popular in a specific country, but not worldwide.

Based on the above definitions, we compute preference profiles globally (PP_g), for a country (PP_c), and for a user (PP_u). Given the LFM-1b dataset [30], these profiles are 585,095-dimensional vectors containing the AF, LF, or AF-ILF scores over all artists in the dataset. Figure 1 provides an example by visualizing the preference profiles including the top 50,000 artists for Finland, a country that does particularly not correspond to the global music mainstream. The black lines in the plots indicate the *global* AF, LF, or AF-ILF scores. As can be seen, the distributions of the AF- and LF-based preference profiles largely follow a trend similar to the global one. However, a second curve is indicative of a country-specific mainstream, in this case for Finland. In contrast, the AL-ILF weighting considerably increases the importance of globally less popular, but country-wise more popular artists (also see Table 2).

Exploiting the profiles, we propose three categories of mainstreaminess measures on the user level: fraction-based (F), symmetrized Kullback-Leibler divergence (D), and rank-order correlation according to Kendall's τ (C). The adoption of fraction-based measures is motivated by their easy interpretability (due to the share of overlap between a user's and the global or a country's preference profiles). Kullback-Leibler divergence is a well-established method to compare distributions (discrete preference profiles in our case); we employ rank correlation because conversion of feature values to ranks has already been proven successful for music similarity tasks [24].

We provide formulas for the specific measures in Table 3, where $\widehat{X}$ denotes the sum-to-unity normalized vector X and $ranks(PP_U^W)$ represents the real-valued preference profile converted to ranks, i.e. the vector containing all normalized item frequencies of user u, with respect to the frequency weighting approach W (AF or LF). When using $AF{\cdot}ILF$, $ranks\left(PP_u^W\right)$ is extended to $ranks\left(PP_{u,c}^{AF{\cdot}ILF}\right)$, i.e. AF computed for user u, ILF on country c, or $ranks\left(PP_{c,g}^{AF{\cdot}ILF}\right)$, i.e. AF computed on country c, ILF globally. Note that we invert the results of the fraction-based formulations and the symmetrized KL-divergences in order to be consistent in that higher values always indicate closer to the mainstream, while lower ones indicate farther away from the mainstream.

4 MUSIC RECOMMENDATION TAILORED TO USER MAINSTREAMINESS

To evaluate the proposed mainstreaminess measures (cf. Section 3) with respect to their ability to improve performance in music recommendation, we conduct rating prediction experiments, which is a common approach to recommender systems evaluation. For this evaluation, we use the LFM-1b dataset of user-generated listening events from Last.fm [30], as detailed in the following.

4.1 Data Preparation

The LFM-1b dataset [30] covers 1,088,161,692 listening events of 120,322 unique users, who listened to 32,291,134 unique tracks by 3,190,371 unique artists. The core component of the dataset is the cleaned user-artist-playcount matrix (UAM) containing the number of listening events of 120,175 users to 585,095 unique artists. The distribution of listening events of the Last.fm data corresponds to a typical long-tail distribution [8].

As 65,132 user profiles do not contain any country information, we exclude those from our experiments since they do not contribute to defining a country's mainstreaminess. For each user in the remaining user set, we calculate the proposed mainstreaminess measures according to Table 3. Note that using the LFM-1b dataset, the global population is in our case the Last.fm users in the dataset.

4.2 Experimental Setup

While we are aware that a truly user-centric evaluation would be beneficial for this kind of research, conducting a user study on tens of thousands of users (or even only a representative subset of the users) is beyond the scope of this paper. We therefore stick to the common approach of quantifying the performance of a recommender system by conducting a rating prediction task. To this end, we normalize and scale the playcount values in the UAM to the range [0, 1000] for each user individually, assuming that higher numbers of playcounts indicate higher user preference for an artist.

We apply the common singular value decomposition (SVD) method according to [27] to factorize the UAM and in turn effect rating prediction. In 5-fold cross-validation experiments, we use root mean square error (RMSE) and mean absolute error (MAE) as performance measures.

To obtain a baseline, we first run the rating prediction experiment on the global group of 65,132 users and report results of the error measures in the first row of Table 4. To study the influence of both, the different mainstreaminess *definitions* and mainstreaminess *levels* on recommendation performance, we then create subsets of users for each combination of mainstreaminess measure and country with at least 1,000 users.[1] To this end, we split the users in each country into three (almost) equally sized subsets according to their mainstreaminess value: *low* corresponds to users in the lower 3-quantile (tertile) w.r.t. the respective mainstreaminess definition, *mid* and *high*, respectively, to the mid and upper tertile. In the individual experiments, *all* refers to the group of all users in each considered country, *low* only to the users in the lower 3-quantile (tertile) w.r.t. the respective mainstreaminess definition, *mid* and *high* defined analogously. Further, conducting the same experiment on all users in each country (user set *all*) allows for a comparison of a pure mainstreaminess filtering approach versus a combination of mainstreaminess filtering and demographic (country) filtering.

4.3 Results and Discussion

Table 4 shows the error measures (RMSE and MAE) for different *definitions* and *levels* of mainstreaminess, averaged over all considered countries (cf. Subsection 4.1), RMSE and MAE weighted by the number of users in the respective country. In the following

[1]The restriction to countries with at least 1,000 users was made to allow for a meaningful analysis, as performed in [31].

discussion, we concentrate on RMSE since it is more common and considers larger differences between predicted and true ratings disproportionately more severe than smaller ones.

As a general finding, our results show that tailoring the recommendations to a user's mainstreaminess level (*low*, *mid*, *high*) leads to substantial error reductions, irrespective of the applied mainstreaminess measure. More specifically, $C_{c:AF,u:AF}$ outperforms the other measures in four regards: First, it leads to the lowest overall RMSE of 14.349 (*all*). Second, the errors realized by $C_{c:AF,u:AF}$ are also the lowest for each of the three user sets (*low*, *mid*, *high*). If better performance is achieved on a set with another measure, the difference is just in the third position after the decimal point. Third, $C_{c:AF,u:AF}$ performs on each of the three user sets (*low*, *mid*, *high*) in a balanced way (weighted RMSE amounts to respectively 3.692, 4.270, and 3.687), whereas the other mainstreaminess measures yield a rather unbalanced picture since each of them performs on at least one set far worse than on the other(s), e.g., $C_{g:AF,u:AF}$ with 19.183, 7.443, and 3.681, respectively, for *low*, *mid*, and *high*. Fourth, $C_{c:AF,u:AF}$ performs well also on the low mainstreaminess user set (*low*), which is a user segment that is typically difficult to satisfy.

The fraction-based approaches $F_{g:AF,u:AF}$, $F_{c:AF,u:AF}$, and $F_{g:AF,u:AF \cdot ILF}$ have in common that they perform far better in the high mainstreaminess segment than in the mid and the low one. This could indicate that these measures still privilege globally popular items too much and, thus, produce more errors in the mid and low segments.

Interestingly, the approaches based on symmetrized Kullback-Leibler divergence (*D*) perform worse when tailored towards a user's country ($D_{c:AF,u:AF}$), compared to their application on a global level ($D_{g:AF,u:AF}$). Combining the country-specific tailoring with the AF-ILF weighting allows for better results compared to applying both separately.

While our results do not suggest a general superiority of mainstreaminess measures that incorporate AF-ILF, first results of our deeper analysis on the country level indicate that these measures seem to perform particularly well for countries far away from the global mainstream, such as Finland (RMSE of $D_{c:AF \cdot ILF,u:AF \cdot ILF}$ for *all*=5.985, *high*=1.346, *mid*=1.365, *low*=1.418), but worse for high mainstream countries, such as the USA (RMSE of $D_{c:AF \cdot ILF,u:AF \cdot ILF}$ for *all*=57.489, *high*=4.071, *mid*=4.077, *low*=55.968). In the presented example, the low mainstream country Finland is small, and the respective weighted error measures in Table 4 do not reflect this country's users to the same extent as the large and high mainstream United States. As part of our ongoing large-scale analysis, delving into detail on country-specific aspects, we will investigate as a next step what factors influence the performance differences between countries for a given mainstreaminess measure.

A direct comparison of the RMSE achieved by our approach with the RMSE reported in [37], the work closest to ours, is unfortunately impossible since Vigliensoni and Fujinaga quantized playcounts into a 5-point Likert rating scale: [1,5]. Still, in a rough estimation, our results suggest that the accuracy of our best $C_{c:AF,u:AF}$ approach delivers a new benchmark in the combination of demographic (country) filtering and mainstreaminess filtering, with a RMSE of 14.3 on a [0,1000] scale. The best RMSE reported in [37] when considering mainstreamness and country information is approximately 0.9 on the much narrower [1,5] scale (cf. approach *u.c.m.* in Figure 2 of [37]).

5 CONCLUSIONS AND OUTLOOK

We proposed 11 novel measures to quantify the music mainstreaminess of a user, a country, and an entire population. Those are based on fractional (*F*), divergence (*D*), and rank correlation (*C*) functions. Considering that music mainstream may be defined from a global but also a country-specific perspective, we particularly studied how the combination of a user's mainstreaminess and demographic (country) filtering influences the quality of music recommendations. Based on the LFM-1b dataset [30], we investigated the performance of the proposed measures in a rating prediction task, employing matrix factorization. To quantify performance, we computed country-averaged, weighted RMSE and MAE figures for all mainstreaminess definitions and various mainstreaminess levels, and compared these with a global baseline. Overall, our results suggest that incorporating any kind of mainstreaminess information outperforms the baseline. Our best approach combines demographic filtering (based on a user profile's country) and mainstreaminess filtering based on Kendall's τ (variant $C_{c:AF,u:AF}$) and outperforms applying these filtering approaches separately. While our results do not hint at a general superiority of mainstreaminess measures that incorporate AF-ILF, they do show that such measures perform much better than others for countries whose preference profiles are far away from the global taste (e.g., Finland).

As part of future work, we will take an in-depth look at the differences between countries, i.e. analyze in which countries which mainstreaminess functions perform particularly well or poorly. Additionally, we plan to analyze how well our results generalize to other datasets providing demographic user information, e.g., the Spotify playlists dataset [23] or the Million Musical Tweets Dataset [15]. We further plan a user study to qualitatively investigate whether incorporating mainstreaminess information improves the perceived satisfaction with recommendations.

6 ACKNOWLEDGMENTS

This research is supported by the Austrian Science Fund (FWF): V579.

REFERENCES

[1] Hyung Jun Ahn. 2006. Utilizing Popularity Characteristics for Product Recommendation. *International Journal of Electronic Commerce* 11, 2 (2006), 59–80. http://www.jstor.org/stable/27751212
[2] Christian Anderson. 2006. *The long tail: Why the future of business is selling less of more.* Hachette Books, New York, NY, USA.
[3] Sarah Baker, Andy Bennett, and Jodie Taylor. 2013. *Redefining mainstream popular music.* Routledge.
[4] Chumki Basu, Haym Hirsh, and William Cohen. 1998. Recommendation as classification: Using social and content-based information in recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence.* American Association Intelligence, 714–720.
[5] Christine Bauer, Marta Kholodylo, and Christine Strauss. 2017. Music Recommender Systems: Challenges and Opportunities for Non-Superstar Artists. In *Proceedings of 30th Bled eConference*, Andreja Pucihar, Mirjana Kljajić Borštnar, Christian Kittl, Pascal Ravesteijn, Roger Clarke, and Roger Bons (Eds.). University of Maribor Press, Bled, Slovenia, 21–32.
[6] Erik Brynjolfsson, Yu Hu, and Duncan Simester. 2011. Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product

Sales. *Management Science* 57, 8 (2011), 1373–1386. http://www.jstor.org/stable/25835786

[7] Oliver Budzinski and Julia Pannicke. 2017. Do preferences for pop music converge across countries?:Empirical evidence from the Eurovision Song Contest. *Creative Industries Journal* (2017), 1–20.

[8] Òscar Celma. 2010. *Music Recommendation and Discovery.* Springer, Berlin, Heidelberg, Germany.

[9] Òscar Celma and Pedro Cano. 2008. From Hits to Niches?: Or How Popular Artists Can Bias Music Recommendation and Discovery. In *Proceedings of the v2Nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition (NETFLIX 2008).* ACM, New York, NY, USA, Article 5, 8 pages. https://doi.org/10.1145/1722149.1722154

[10] Zhiyong Cheng and Jialie Shen. 2014. Just-for-Me: An Adaptive Personalization System for Location-Aware Social Music Recommendation. In *Proceedings of the International Conference on Multimedia Retrieval (ICMR 2014).* ACM, New York, NY, USA, 185:185–185:192. https://doi.org/10.1145/2578726.2578751

[11] Charles L.A. Clarke, Maheedhar Kolla, Gordon V. Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. 2008. Novelty and Diversity in Information Retrieval Evaluation. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008).* ACM, New York, NY, USA, 659–666. https://doi.org/10.1145/1390334.1390446

[12] Paolo Cremonesi, Franca Garzotto, Roberto Pagano, and Massimo Quadrana. 2014. Recommending without short head. In *Proceedings of the 23rd International Conference on World Wide Web (WWW 2014).* ACM, 245–246.

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

[14] Bruce Ferwerda. 2016. *Improving the User Experience of Music Recommender Systems Through Personality and Cultural Information.* PhD. Johannes Kepler University Linz, Linz, Austria.

[15] David Hauger, Markus Schedl, Andrej Košir, and Marko Tkalčič. 2013. The Million Musical Tweets Dataset: What Can We Learn From Microblogs. In *Proceedings of the 14th International Society for Music Information Retrieval Conference (ISMIR 2013).*

[16] Xiao Hu, Jin Ha Lee, Kahyun Choi, and J.S. Downie. 2014. A Cross-cultural Study of Mood in K-pop Songs. In *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014).* 217–238.

[17] Shinobu Kitayama and Hyekyung Park. 2007. Cultural shaping of self, emotion, and well-being: How does it work? *Social and Personality Psychology Compass* 1, 1 (2007), 202–222.

[18] Rajeev Kumar, B.K. Verma, and Shyam Sunder Rastogi. 2014. Social popularity based SVD++ recommender system. *International Journal of Computer Applications* 87, 14 (2014), 33–37.

[19] Audrey Laplante. 2014. Improving music recommender systems: what can we learn from research on music tags?. In *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR 2014).* 451–456.

[20] Greg Linden, Brent Smith, and Jeremy York. 2003. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing* 4, 1 (2003), 76–80.

[21] Brian McFee, Luke Barrington, and Gert Lanckriet. 2012. Learning content similarity for music recommendation. *IEEE transactions on audio, speech, and language processing* 20, 8 (2012), 2207–2218.

[22] Miquel Montaner, Beatriz López, and Josep Lluís de la Rosa. 2003. A taxonomy of recommender agents on the Internet. *Artificial Intelligence Review* 19, 4 (2003), 285–330.

[23] Martin Pichl, Eva Zangerle, and Günther Specht. 2015. Towards a Context-Aware Music Recommendation Approach: What is Hidden in the Playlist Name?. In *Proceedings of the 15th IEEE International Conference on Data Mining Workshop (ICDM 15).* IEEE, Piscataway, NJ, USA, 1360–1365.

[24] Tim Pohle, Markus Schedl, Peter Knees, and Gerhard Widmer. 2006. Automatically Adapting the Structure of Audio Similarity Spaces. In *Proceedings of the 1st Workshop on Learning the Semantics of Audio Signals (LSAS 2006).*

[25] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2015. *Recommender Systems Handbook* (2nd ed.). Springer, New York, NY.

[26] Paul Rutten. 1991. Local popular music on the national and international markets. *Cultural Studies* 5, 3 (1991), 294–305.

[27] Ruslan Salakhutdinov and Andriy Mnih. 2007. Probabilistic Matrix Factorization. In *Proceedings of the 20th International Conference on Neural Information Processing Systems (NIPS 2007).* Curran Associates Inc., 1257–1264. http://dl.acm.org/citation.cfm?id=2981562.2981720

[28] Gerard Salton, Andrew Wong, and Chung-Shu Yang. 1975. A Vector Space Model for Automatic Indexing. *Commun. ACM* 18, 11 (1975), 613–620. https://doi.org/10.1145/361219.361220

[29] Markus Schedl. 2013. Ameliorating Music Recommendation: Integrating Music Content, Music Context, and User Context for Improved Music Retrieval and Recommendation. In *Proceedings of International Conference on Advances in Mobile Computing & Multimedia (MoMM) (MoMM '13).* ACM, New York, NY, USA, 3:3–3:9. https://doi.org/10.1145/2536853.2536856

[30] Markus Schedl. 2016. The LFM-1b Dataset for Music Retrieval and Recommendation. In *Proceedings of the 17th ACM International Conference on Multimedia Retrieval (ICMR 2016).* New York, NY, USA.

[31] Markus Schedl. 2017. Investigating country-specific music preferences and music recommendation algorithms with the LFM-1b dataset. *International Journal of Multimedia Information Retrieval* 6, 1 (2017), 71–84. https://doi.org/10.1007/s13735-017-0118-y

[32] Markus Schedl and Christine Bauer. 2017. Distance- and Rank-based Music Mainstreaminess Measurement. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (July 9-12, 2017) *(UMAP 2017).* ACM, New York, NY, USA, 364–367. https://doi.org/10.1145/3099023.3099098

[33] Markus Schedl, Emilia Gómez, and Julián Urbano. 2014. Music Information Retrieval: Recent Developments and Applications. *Foundations and Trends in Information Retrieval* 8, 2-3 (2014), 127–261. https://doi.org/10.1561/1500000042

[34] Markus Schedl and David Hauger. 2015. Tailoring Music Recommendations to Users by Considering Diversity, Mainstreaminess, and Novelty. In *Proceedings of the 38th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2015).*

[35] Markus Schedl, Peter Knees, Brian McFee, Dmitry Bogdanov, and Marius Kaminskas. 2015. *Recommender Systems Handbook* (2nd ed.). Springer, New York, NY, USA, Chapter Music Recommender Systems, 453–492.

[36] Catherine J. Stevens. 2012. Music Perception and Cognition: A Review of Recent Cross-Cultural Research. *Topics in Cognitive Science* 4, 4 (2012), 653–667.

[37] 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* (August 7-11, 2016) *(ISMIR 2016).* 94–100.

[38] Linxing Xiao, Lie Lu, Frank Seide, and Jie Zhou. 2009. Learning a music similarity measure on automatic annotations with application to playlist generation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (April 19-24, 2009) *(ICASSP 2009).* IEEE, 1885–1888.

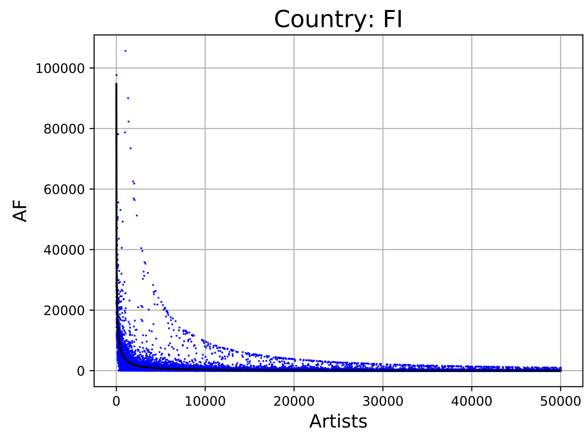
[39] Yan Yan, Tianlong Liu, and Zhenyu Wang. 2015. A Music Recommendation Algorithm Based on Hybrid Collaborative Filtering Technique. In *4th National Conference on Social Media Processing* (November 16-17, 2015) *(SMP 2015)*, Xichun Zhang, Maosong Sun, Zhenyu Wang, and Xuanjing Huang (Eds.). Springer, Singapore, 233–240. https://doi.org/10.1007/978-981-10-0080-5_23

[40] JungAe Yang. 2016. Effects of Popularity-Based News Recommendations ("Most-Viewed") on Users' Exposure to Online News. *Media Psychology* 19, 2 (2016), 243–271. https://doi.org/10.1080/15213269.2015.1006333

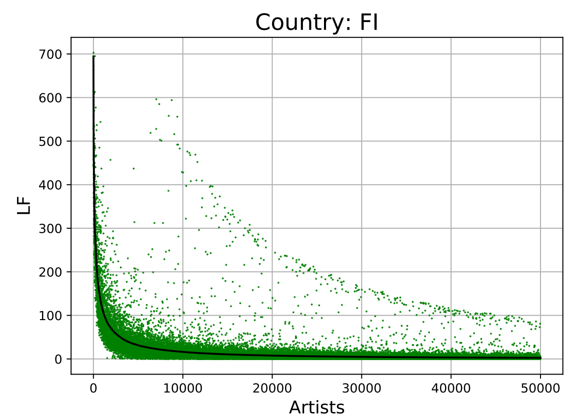
[41] Yuan Cao Zhang, Diarmuid Ó. Seaghdha, Daniele Quercia, and Tamas Jambor. 2012. Auralist: Introducing Serendipity into Music Recommendation. In *Proceedings of the 5th ACM International Conference on Web Search and Data Mining (WSDM 2012).* Seattle, WA, USA.

Mainstreaminess	user set	w.RMSE	w.MAE
Baseline (global UAM)		29.105	25.202
$F_{g:AF,u:AF}$	*all*	26.377	24.050
	high	3.714	1.308
	mid	12.574	9.887
	low	14.186	11.625
$F_{g:AF,u:AF\cdot ILF}$	*all*	21.137	18.617
	high	3.681	1.299
	mid	11.035	8.191
	low	14.426	11.868
$F_{g:AF\cdot ILF,u:AF\cdot ILF}$	*all*	19.140	16.769
	high	11.777	9.121
	mid	13.396	10.833
	low	8.708	5.806
$F_{c:AF,u:AF}$	*all*	14.465	11.958
	high	3.723	1.309
	mid	8.681	6.112
	low	12.706	9.952
$F_{c:AF\cdot ILF,u:AF\cdot ILF}$	*all*	17.615	15.301
	high	9.237	6.648
	mid	3.686	1.305
	low	10.122	7.610
$D_{g:AF,u:AF}$	*all*	24.026	21.705
	high	10.561	8.024
	mid	9.854	7.299
	low	5.365	2.909
$D_{c:AF,u:AF}$	*all*	28.021	25.746
	high	5.365	2.912
	mid	13.510	10.840
	low	25.923	22.621
$D_{c:AF\cdot ILF,u:AF\cdot ILF}$	*all*	14.628	11.624
	high	3.656	1.281
	mid	7.035	4.515
	low	8.589	5.670
$C_{g:AF,u:AF}$	*all*	15.906	13.525
	high	3.680	1.291
	mid	7.443	4.472
	low	19.183	16.373
$C_{c:AF,u:AF}$	*all*	14.349	12.032
	high	3.687	1.290
	mid	4.270	1.833
	low	3.692	1.308
$C_{c:AF\cdot ILF,u:AF\cdot ILF}$	*all*	30.827	28.535
	high	7.680	5.187
	mid	4.825	2.340
	low	10.785	8.1084

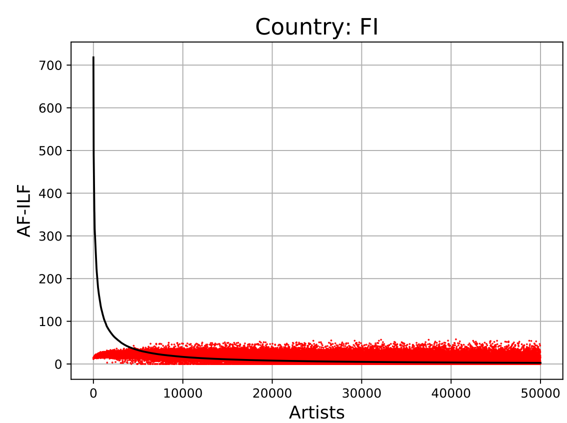
Table 4: Weighted root mean square error (RMSE) and weighted mean absolute error (MAE) for various mainstreaminess definitions and levels, i.e. user sets. Rating values are scaled to [0, 1000]. Experiments are conducted on the country level (except for first row using the complete UAM irrespective of country) and error measures are averaged (arithmetic mean) over all countries with more than 1,000 users and weighted by number of users in the respective country. In the individual experiments, *all* refers to the group of all users in each considered country, *low* only to the users in the lower 3-quantile (tertile) w.r.t. the respective mainstreaminess definition, *mid* and *high* defined analogously.



(a) Artist frequency (AF)



(b) Listener frequency (LF)



(c) Artist frequency–inverse listener frequency (AF-ILF)

Figure 1: Artist frequency (AF), listener frequency (LF), and artist frequency–inverse listener frequency (AF-ILF) for the top 50,000 artists in Finland. Artist IDs (x-axis) are sorted by global AF, LF, or AF-ILF values, respectively. The black line indicates the global values.