INTRODUCING GLOBAL AND REGIONAL MAINSTREAMINESS FOR IMPROVING PERSONALIZED MUSIC RECOMMENDATION

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

Markus Schedl & Christine Bauer
Johannes Kepler University Linz
markus.schedl@jku.at / christine.bauer@jku.at
http://www.cp.jku.at

15th International Conference on Advances in Mobile Computing & Multimedia (MoMM2017), 4 December 2017, Salzburg, Austria

BACKGROUND

- information overload
- recommender systems important
- example: music recordings on Spotify or YouTube → music recommender systems

MUSIC DOMAIN

hit-driven domain

popularity-based approaches

- approach assumes that a random user is more likely to like a very popular music item than one of the far less popular items
- helpful in cold-start situations

one specific approach in the music domain: describing music listeners in terms of the degree to which they prefer music items that are currently popular or rather ignore such trends

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).

A FRACTION-BASED APPROACH TO QUANTIFY A USER'S MUSIC MAINSTREAMINESS

- quantifies the extent to which a user's listening preferences correspond to those of the population at large
- general: overlap between a user's and the global listening profile
- listening profile computed for user u and globally (g)
- compute artist listening frequency for all artists A in dataset (considering *g* or *u*):
  ⟦AF⟧_a and ⟦AF⟧_(a,u), respectively

$$F_u \quad = \quad 1 - \frac{1}{|A|} \cdot \sum_{a \in A} \frac{|\widehat{AF_{a,u}} - \widehat{AF_a}|}{\max\left(\widehat{AF_{a,u}}, \widehat{AF_a}\right)}$$

A	set of artists
$\widehat{AF_a}$	normalized artist frequency (sum-to-unity)
$AF_{a,u}$	artist frequency of artist *a* listened to by user *u*

higher values indicate closer to the mainstream, whereas lower ones indicate farther away from the mainstream

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.

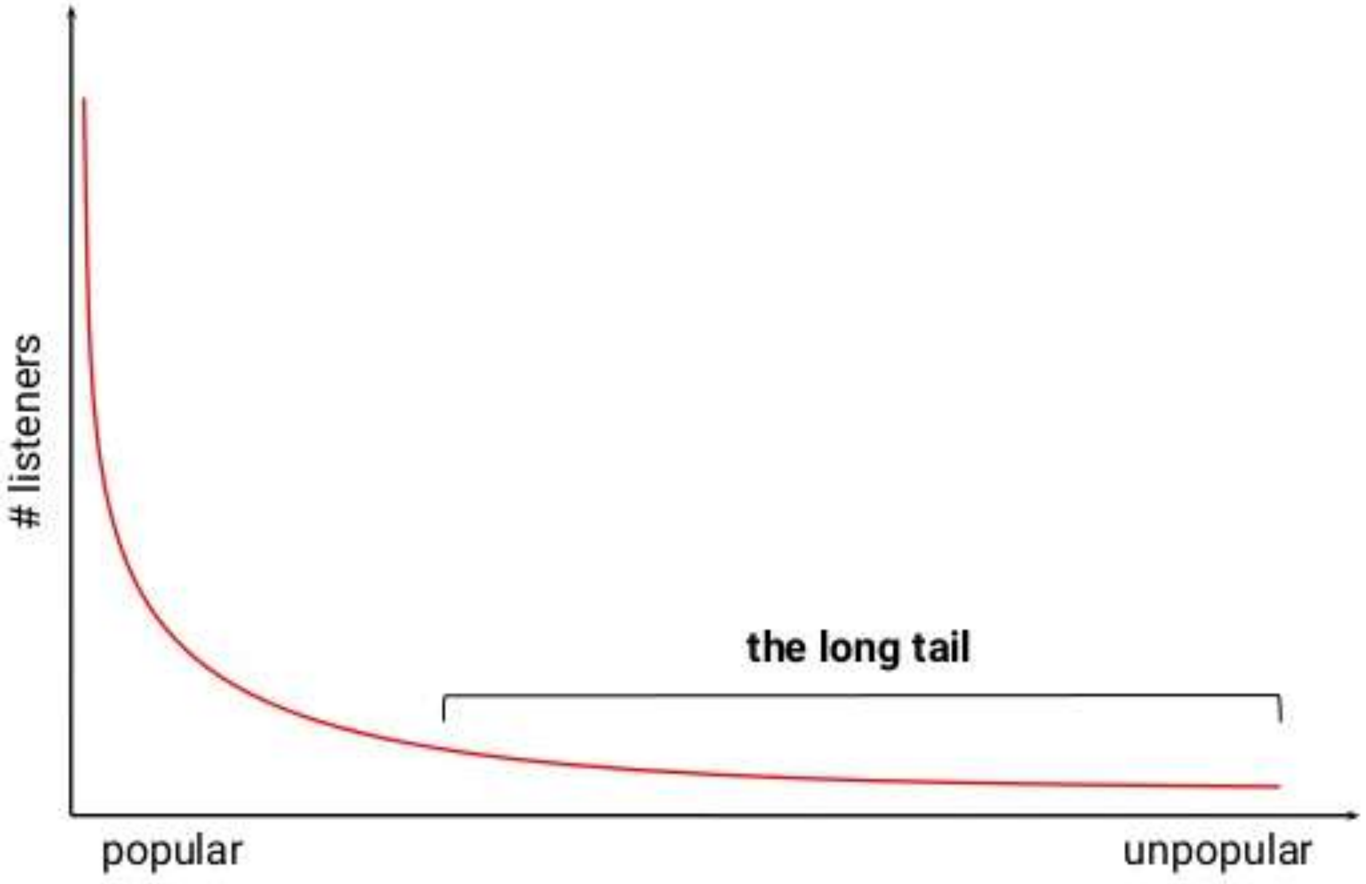
Markus Schedl and David Hauger. 2015. Tailoring Music Recommendations to Users by Considering Diversity, Mainstreaminess, and Novelty. In Proceedings of the 38th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2015).

PROBLEM – “SUPERSTAR” PHENOMENON

also known as “long- tail” or “‘hit-driven” phenomenon

phenomenon describes that relatively small numbers of items (the head) dominate the market, while there is a considerable long tail of less popular items

disproportionately higher influence of absolute top hits (the head) in fraction-based definitions of mainstreaminess



DISTANCE- AND RANK-BASED APPROACHES TO QUANTIFY A USER'S MUSIC MAINSTREAMINESS

- Distance-based (D_u): symmetrized Kullback-Leibler (KL) divergence between global and user's artist frequency
- Rank-based (C_u): rank-order correlation according to Kendall's **τ** between global and user's preference profiles
- Fraction-based (F_u): baseline; average difference between user's artist frequency and global artist frequency

higher values indicate closer to the mainstream, whereas lower ones indicate farther away from the mainstream

$$\mathrm{D_u} = \frac{1}{\frac{1}{2}\cdot\left(\sum_{a\in A}\widehat{AF_{a,u}}\cdot\log\frac{\widehat{AF_{a,u}}}{\widehat{AF_a}} + \sum_{a\in A}\widehat{AF_a}\cdot\log\frac{\widehat{AF_a}}{\widehat{AF_{a,u}}}\right)} \quad (1)$$

$$\mathrm{C_u} = \tau\left(ranks\left(PP_u\right), ranks\left(PP_g\right)\right) \quad (2)$$

$$\mathrm{F_u} = 1 - \frac{1}{|A|}\cdot\sum_{a\in A}\frac{|\widehat{AF_{a,u}} - \widehat{AF_a}|}{\max\left(\widehat{AF_{a,u}}, \widehat{AF_a}\right)} \quad (3)$$

where A is the set of artists in the dataset, $\widehat{AF_a}$ denotes the normalized artist frequency AF_a (sum-to-unity over all artist frequencies), $\widehat{AF_{a,u}}$ defined accordingly; $ranks(PP_u)$ denotes a function that converts the real-valued preference profile (vector over artist frequencies) of user u to ranks, $ranks(PP_g)$ accordingly on the global level, i.e. considering all users.

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

PROBLEM
COUNTRY-SPECIFIC MAINSTREAM

Global (53,258 users)

Artist	LF
Radiohead	24,829
Nirvana	24,249
Coldplay	23,714
Daft Punk	23,661
Red Hot Chili Peppers	22,609
Muse	22,429
Queen	21,778
The Beatles	21,738
Pink Floyd	21,129
David Bowie	20,602

Finland (1,407 users)

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

Italy (972 users)

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

Turkey (479 users)

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 FREQUENCY–INVERSE LISTENER FREQUENCY (AF-ILF) APPROACH TO QUANTIFY A USER'S MUSIC MAINSTREAMINESS

- what is considered mainstream depends on the selection of a population;
  we define it globally and on a country-specific level
- our approach is inspired by the well-established monotonicity assumptions in text processing and information retrieval: the TF-IDF (term frequency–inverse document frequency) weighting
- → artist frequency–inverse listener frequency (AF-ILF)

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

$AF_{a,U}$ sum of the number of tracks by artist *a* listened to by a set of users *U*

$LF_{a,U}$ number of listeners of artist *a* within a user population *U*

U_1 and U_2 may represent a single user, all users in the same country, or all users in the dataset
(allows to formalize the **global** and the **regional** definitions of mainstreaminess, by varying U_1 and U_2)

DISTILLING COUNTRY-SPECIFIC MAINSTREAM BY TF-IDF-LIKE WEIGHTING

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)

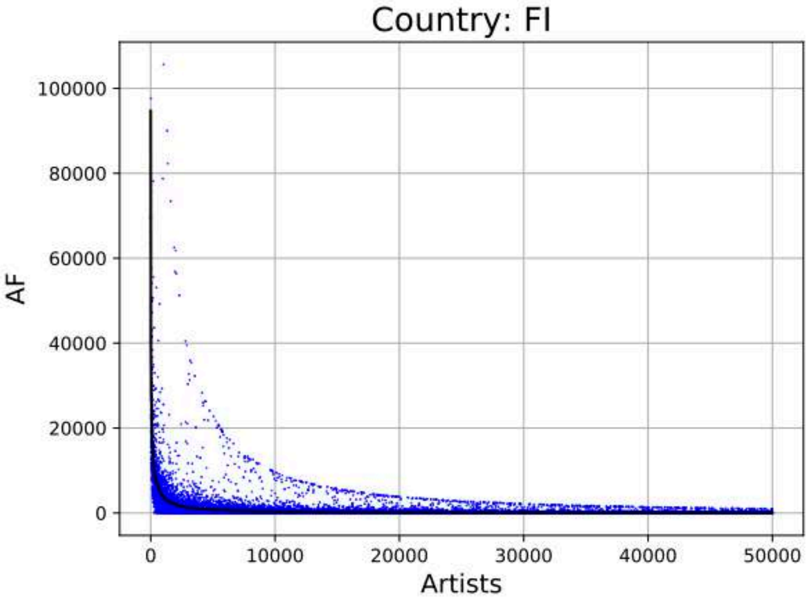
Artist	LF
Radiohead	556
Pink Floyd	539
The Beatles	505
David Bowie	500
Muse	500
Nirvana	497
Coldplay	475
The Cure	466
Depeche Mode	459
Daft Punk	457
Artist	**AF-ILF**
CaneSecco	68.451
DSA Commando	66.049
Veronica Marchi	65.864
Train To Roots	65.459
Alessandro Raina	64.228
Machete Empire	63.915
Danti	62.958
Dargen D'Amico	62.453
宝塚歌劇団・宙組	62.228
Aquefrigide	61.663

(b) Italy (972 users)

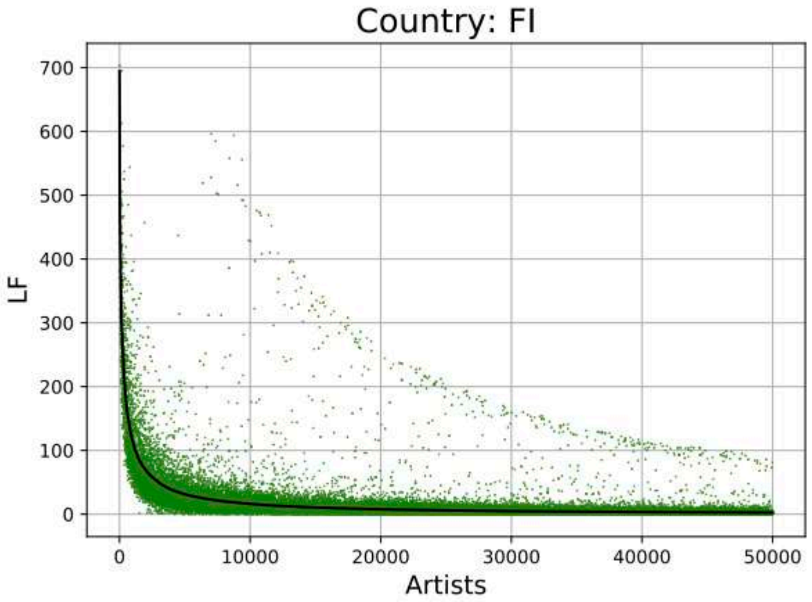
Artist	LF
Pink Floyd	292
Radiohead	289
Metallica	268
Coldplay	261
Nirvana	251
Massive Attack	249
The Beatles	240
Red Hot Chili Peppers	240
Queen	238
Led Zeppelin	236
Artist	**AF-ILF**
Cüneyt Ergün	64.473
Floyd Red Crow Westerman	61.955
Fırat Tanış	58.666
Acil Servis	58.439
Taste (Rory Gallager)	58.366
Mezarkabul	57.799
Rachmaninoff Sergey	57.733
Mabel Matiz	57.619
Grup Yorum	56.855
Yüzyüzeyken Konuşuruz	56.748

(c) Turkey (479 users)

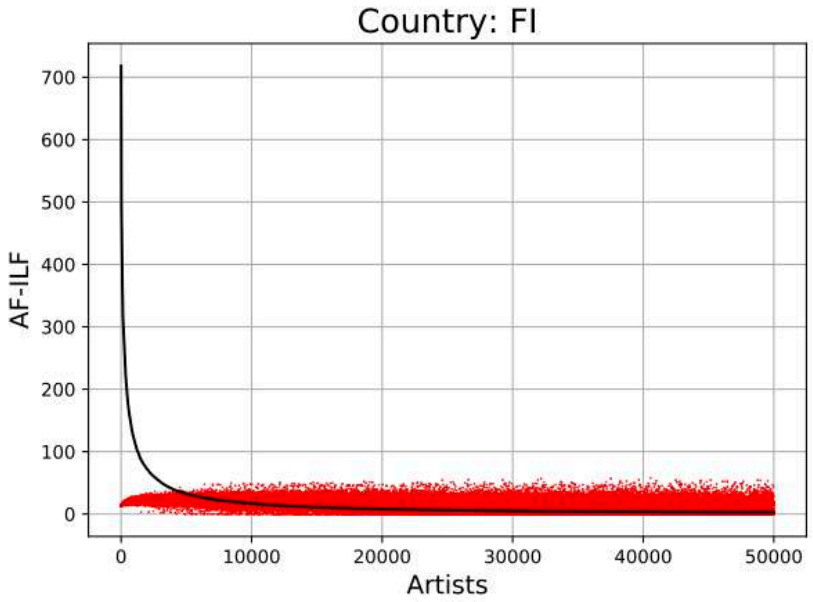
THE DIFFERENT WEIGHTINGS ON THE EXAMPLE FINLAND



AF for FI sorted, Top50k



LF for FI sorted, Top50k



AF-ILF for FI sorted, Top50k

11 VARIATIONS OF QUANTIFYING A USER'S MUSIC MAINSTREAMINESS

- for distance-based (D_u), rank-based (C_u), and fraction-based (F_u):
- combinations of (c,u) and (g,u) with AF and AF-ILF weighting

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

DATA FOR EVALUATION

LFM-1b dataset of listening histories

LFM-1B: OVERVIEW

lost.fm

- > 1b listening events (LE)
- > 120k users
- LE = <user, artist, album, track, timestamp>
- LEs covering Jan 2005 – Aug 2014
- Seed list of 250 top tags → fetch top fans → 465k active users → random subset of 120k users → fetch their listening histories
- Demographic information of (anonymized) listeners
- Data cleaning: remove users/artists with < 10 unique artists/users

120k x 585k user-artist-playcount matrix

Markus Schedl. 2016. The LFM-1b Dataset for Music Retrieval and Recommendation, Proceedings of the ACM International Conference on Multimedia Retrieval (ICMR), New York, USA, April 2016

LFM-1B: DISTRIBUTION AMONG COUNTRIES

Country	No. of users	Pct. in dataset
US	10255	18.581 %
RU	5024	9.103 %
DE	4578	8.295 %
UK	4534	8.215 %
PL	4408	7.987 %
BR	3886	7.041 %
FI	1409	2.553 %
NL	1375	2.491 %
ES	1243	2.252 %
SE	1231	2.230 %
UA	1143	2.071 %
CA	1077	1.951 %
FR	1055	1.912 %
N/A	65132	54.131 %

EXPERIMENTS

MUSIC RECOMMENDATION TAILORED TO USER MAINSTREAMINESS

EVALUATION APPROACH FOR MUSIC RECOMMENDATION TAILORED TO USER MAINSTREAMINESS

evaluation method	• rating prediction on playcounts scaled to [0, 1000]
algorithm	• model-based collaborative filtering (SVD)
analysis	• different definitions and levels of mainstreaminess
definitions	• distance-based, rank-based, fraction-based
levels	• user tertiles w.r.t. mainstreaminess (lower, mid, upper 1/3)
performance measures	• root mean square error (RMSE) and mean average error (MAE)

Mainstreaminess	user set	w.RMSE	w.MAE
Baseline (global UAM)		29.105	25.202
$C_{g:AF,u:AF}$	*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
$D_{g:AF,u:AF}$	*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
$F_{g:AF,u:AF}$	*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

FINDINGS (1/2)

- tailoring the recommendations to a user's mainstreaminess level (low, mid, high) leads to substantial error reductions
- $C_{c:AF,u:AF}$ outperforms other measures in 4 regards:
  - lowest overall RMSE of 14.349 (all)
  - errors also the lowest for each of the three user sets (low, mid, high)
    - if better performance on a set with other measure, difference just 0.00x
  - performs on each of the 3 user sets (low, mid, high) in a balanced way (weighted RMSE: respectively 3.692, 4.270, and 3.687)
    - other measures: on at least one set very low performance
  - performs well also on the low mainstreaminess user set (low), which is a user segment that is typically difficult to satisfy
- the 3 fraction-based approaches: perform far better in the high mainstreaminess segment (high)
  - still privileges globally popular items too much?

Mainstreaminess	user set	w.RMSE	w.MA
Baseline (global UAM)		29.105	25.20
$F_{g:AF,u:AF}$	*all*	26.377	24.05
	high	3.714	1.30
	mid	12.574	9.88
	low	14.186	11.62
$F_{g:AF,u:AF\cdot ILF}$	*all*	21.137	18.61
	high	3.681	1.29
	mid	11.035	8.19
	low	14.426	11.86
$F_{g:AF\cdot ILF,u:AF\cdot ILF}$	*all*	19.140	16.76
	high	11.777	9.12
	mid	13.396	10.83
	low	8.708	5.80
$F_{c:AF,u:AF}$	*all*	14.465	11.95
	high	3.723	1.30
	mid	8.681	6.11
	low	12.706	9.95
$F_{c:AF\cdot ILF,u:AF\cdot ILF}$	*all*	17.615	15.30
	high	9.237	6.64
	mid	3.686	1.30
	low	10.122	7.61
$D_{g:AF,u:AF}$	*all*	24.026	21.70
	high	10.561	8.02
	mid	9.854	7.29
	low	5.365	2.90
$D_{c:AF,u:AF}$	*all*	28.021	25.74
	high	5.365	2.91
	mid	13.510	10.84
	low	25.923	22.62
$D_{c:AF\cdot ILF,u:AF\cdot ILF}$	*all*	14.628	11.62
	high	3.656	1.28
	mid	7.035	4.51
	low	8.589	5.67
$C_{g:AF,u:AF}$	*all*	15.906	13.52
	high	3.680	1.29
	mid	7.443	4.47
	low	19.183	16.37
$C_{c:AF,u:AF}$	*all*	14.349	12.03
	high	3.687	1.29
	mid	4.270	1.83
	low	3.692	1.30
$C_{c:AF\cdot ILF,u:AF\cdot ILF}$	*all*	30.827	28.53
	high	7.680	5.18
	mid	4.825	2.34
	low	10.785	8.108

FINDINGS (2/2)

- symmetrized Kullback-Leibler divergence (D) perform worse when tailored towards a user’s country (Dc:AF,u:AF), compared to their application on a global level (Dg:AF,u:AF)
- combining the country-specific tailoring with the AF-ILF weighting allows for better results compared to applying both separately
- on first sight: no general superiority of AF-ILF measures, but deeper analysis on the country level indicates that these measures seem to:
  - perform particularly well for countries far away from the global mainstream, e.g., Finland (RMSE of Dc:AF ·ILF,u:AF ·ILF for all=5.985,high=1.346,mid=1.365,low=1.418)
  - but worse for high mainstream countries, e.g., USA (RMSE of Dc:AF·ILF,u:AF·ILF for all=57.489,high=4.071, mid=4.077, low=55.968)

Mainstreaminess	user set	w.RMSE	w.MA
Baseline (global UAM)		29.105	25.20
$F_{g:AF,u:AF}$	*all*	26.377	24.05
	high	3.714	1.30
	mid	12.574	9.88
	low	14.186	11.62
$F_{g:AF,u:AF\cdot ILF}$	*all*	21.137	18.61
	high	3.681	1.29
	mid	11.035	8.19
	low	14.426	11.86
$F_{g:AF\cdot ILF,u:AF\cdot ILF}$	*all*	19.140	16.76
	high	11.777	9.12
	mid	13.396	10.83
	low	8.708	5.80
$F_{c:AF,u:AF}$	*all*	14.465	11.95
	high	3.723	1.30
	mid	8.681	6.11
	low	12.706	9.95
$F_{c:AF\cdot ILF,u:AF\cdot ILF}$	*all*	17.615	15.30
	high	9.237	6.64
	mid	3.686	1.30
	low	10.122	7.61
$D_{g:AF,u:AF}$	*all*	24.026	21.70
	high	10.561	8.02
	mid	9.854	7.29
	low	5.365	2.90
$D_{c:AF,u:AF}$	*all*	28.021	25.74
	high	5.365	2.91
	mid	13.510	10.84
	low	25.923	22.62
$D_{c:AF\cdot ILF,u:AF\cdot ILF}$	*all*	14.628	11.62
	high	3.656	1.28
	mid	7.035	4.51
	low	8.589	5.67
$C_{g:AF,u:AF}$	*all*	15.906	13.52
	high	3.680	1.29
	mid	7.443	4.47
	low	19.183	16.37
$C_{c:AF,u:AF}$	*all*	14.349	12.03
	high	3.687	1.29
	mid	4.270	1.83
	low	3.692	1.30
$C_{c:AF\cdot ILF,u:AF\cdot ILF}$	*all*	30.827	28.53
	high	7.680	5.18
	mid	4.825	2.34
	low	10.785	8.108

FUTURE AVENUES OF RESEARCH

- considering highly varying “music listening culture” in different countries
- integration of more data sources
- deployment of additional research instruments (e.g., surveys)

TAKE AWAY…

- 11 novel measures to quantify the music mainstreaminess of a user, a country, and an entire population
- based on fractional (F), divergence (D), and rank correlation (C) functions
- combination of a user's mainstreaminess and demographic (country) filtering
- CF enhanced by grouping users according to any kind of mainstreaminess category outperforms non-personalized approach
- best approach combines demographic filtering (based on a user profile's country) and mainstreaminess filtering based on Kendall's τ $C_{c:AF,u:AF}$
- AF-ILF perform much better than others for countries whose preference profiles are far away from the global taste (e.g., Finland)

JOHANNES KEPLER UNIVERSITY LINZ

Department of Computational Perception

Markus Schedl & Christine Bauer
Johannes Kepler University Linz
markus.schedl@jku.at / christine.bauer@jku.at
http://www.cp.jku.at

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