Music Recommender Systems

JOHANNES KEPLER
UNIVERSITY LINZ

INSTITUTE OF COMPUTATIONAL PERCEPTION

MULTIMEDIA MINING AND SEARCH GROUP

Markus Schedl

www.cp.jku.at, www.mschedl.eu
markus.schedl@jku.at
@m_schedl

**JOHANNES KEPLER
UNIVERSITY LINZ**
Altenberger Straße 69
4040 Linz, Austria
jku.at

Markus Schedl

Full Professor

Johannes Kepler University Linz, Austria

Institute of Computational Perception, Head of Multimedia Mining and Search Group

Linz Institute of Technology, AI Lab, Head of Human-centered AI Group

Contact: markus.schedl@jku.at | www.mschedl.eu | www.hcai.at | @m_schedl

Outline

- Introduction to Recommender Systems: Motivation, Domains, Tasks, Flavors
- Problems of Recommender Systems
- Large-scale Music Exploration (Emotional Music Tower Blocks - EmoMTB)
- Popularity Bias in Recommendations (Black Holes of Popularity - BHP)
- Gender Unfairness of Recommender Systems

Recommender Systems

What Are Recommender Systems?

"Recommender systems are tools and techniques that provide suggestions for items that are most likely of interest to a particular user, ...what items to buy, what music to listen to, or what online news to read."

[Ricci et al., 2022]

For products

amazon

For fashion

STITCH FIX

dressipi

For jokes

Jester 5.0

Jokes for *your* sense of humor

For travel

trivago

For music

DEEZER

Spotify

For movies/series

NETFLIX

For videos

YouTube

Tik Tok

For users and user-generated content

facebook

twitter

For books

WHAT SHOULD I READ NEXT?

Why Do We Need Recommender Systems?

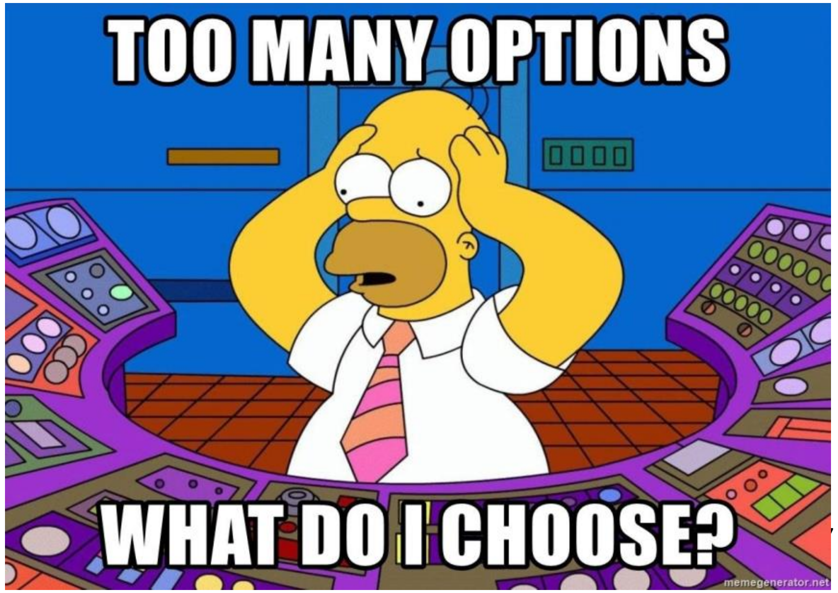
*Obtaining a **personalized** selection of items from a large catalog to support decision making*

Ex.: Spotify (https://newsroom.spotify.com/company-info/)

- 70M+ music tracks
- 381M users

Ex.: Youtube (https://www.globalmediainsight.com/blog/youtube-users-statistics)

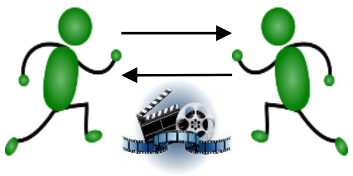
- 5B+ videos watched per day
- 1B hours of video watched every day
- 500 hours of video uploaded per minute
- 2B users



Main Flavors of Recommender Systems

Collaborative filtering (CF):

Recommend to target user items that
other *similar users* liked in the past



Content-based filtering (CBF):

Recommend to target user *content similar*
to what he or she liked in the past

Context-aware recommender system (CARS):

Recommend to target user items that he, she,
or other users liked in a given *context or situation*

Hybrid RS: Any *combination* of the above

Common Recommendation Tasks

(Missing) rating prediction:
Which rating would a user give to an unseen item?

	Item 1	*Item 2*	*Item 3*	*Item 4*	*Item 5*
User 1	3		2		3
User 2	4	3	4	3	
User 3		2		5	4
User a		5	4	3	**?**

Explicit ratings

Common Recommendation Tasks

Interaction prediction:
Will a user interact with an unseen item?

	Item 1	*Item 2*	*Item 3*	*Item 4*	*Item 5*
User 1	1	0	0	1	0
User 2	0	0	1	1	0
User 3	1	0	1	1	1
User a	0	1	1	1	**?**

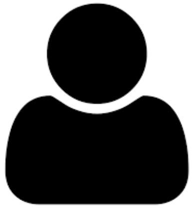
Often implicit feedback (e.g., clicks, views, listening events, skipping behavior)

Common Recommendation Tasks

	Item 1	*Item 2*	*Item 3*	*Item 4*	*Item 5*
User 1	3		2		3
User 2	4	3	4	3	
User 3		2		5	4
User a		5	4	3	**?**

Explicit ratings

Recommender



Ranked list of items

$< i_7, i_9, i_{12}, i_5, i_{42} >$

determined by a ranking algorithm (based on predicted ratings, degree of similarity/matching, probability, confidence score, etc.)

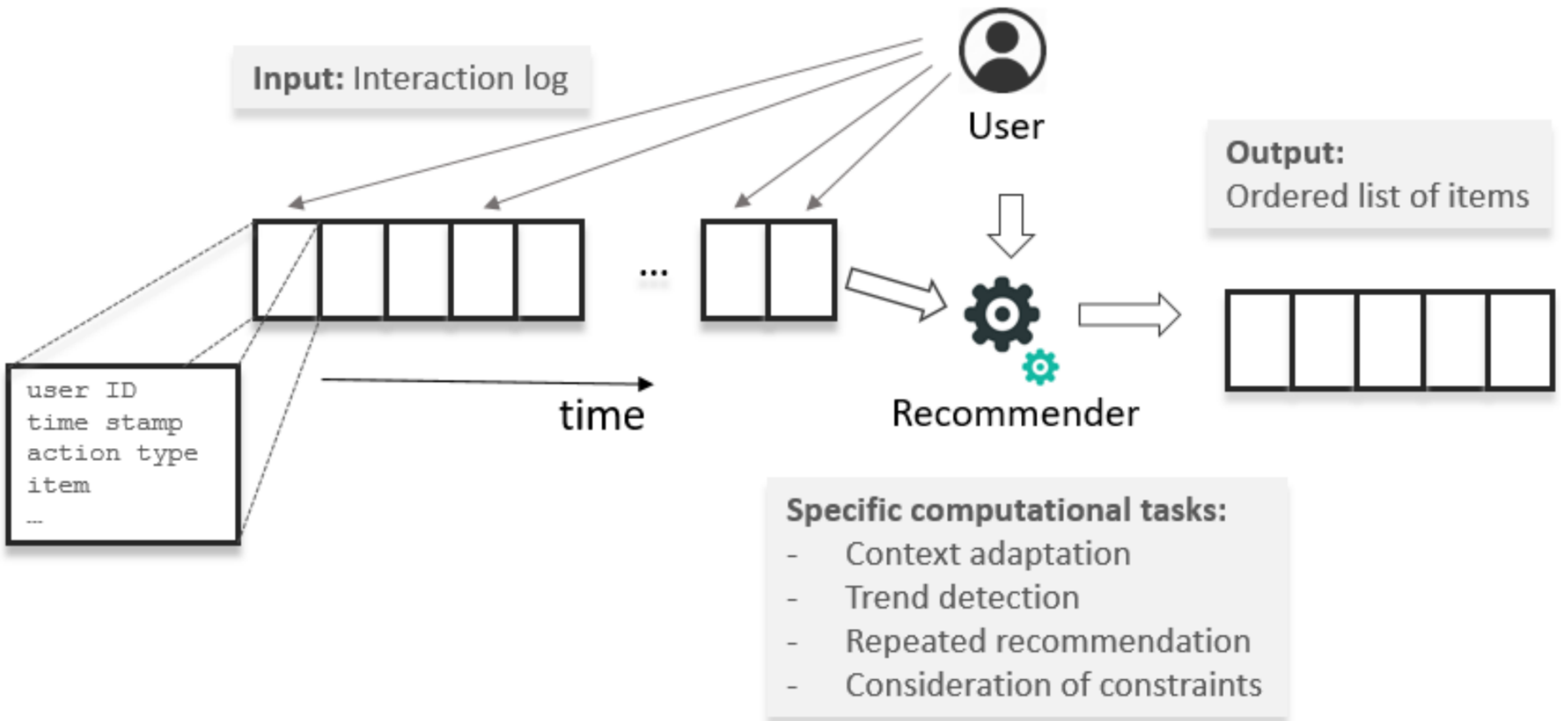
	Item 1	*Item 2*	*Item 3*	*Item 4*	*Item 5*
User 1	1	0	0	1	0
User 2	0	0	1	1	0
User 3	1	0	1	1	1
User a	0	1	1	1	**?**

Implicit feedback (e.g., clicks, views, listening events, skipping behavior)

Recommender

Common Recommendation Tasks

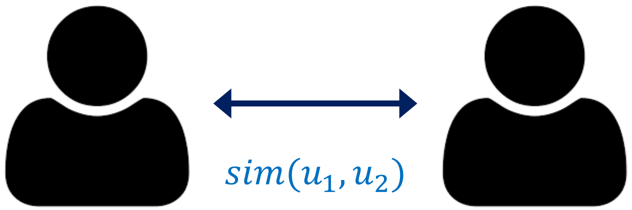
Sequential recommendation
(e.g., automatic playlist generation)



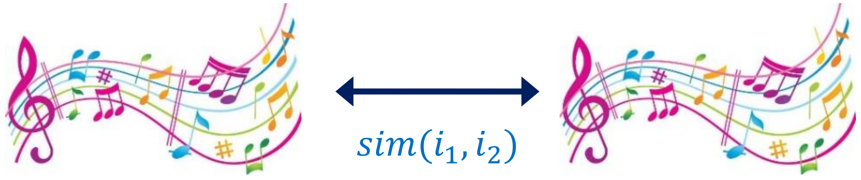
Cremonesi et al., *Sequence-Aware Recommender Systems*, Tutorial at the Web Conference 2019

Important Concept in Recommender Systems

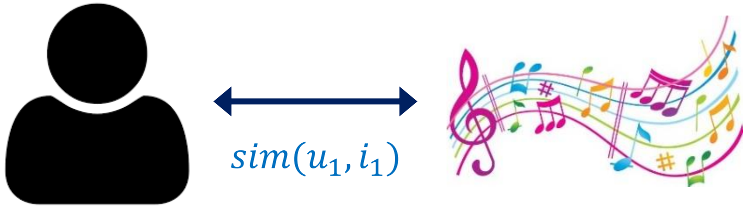
Similarity



(e.g., memory-based CF, UserKNN)



(e.g., content-based filtering)



(e.g., latent factor model)

Problems of Recommender Systems

Despite all their benefits, recommender systems suffer from shortcomings and problems, including…

- Cold-start problem (new users or new items unknown to the system)
- Privacy and security concerns
- Lack of transparency and explainability
- **Over-personalization → users get stuck in a “filter bubble” → less diverse results**
- **Exploration vs. exploitation**
- **Biases (e.g., popularity or demographic)**
- **Fairness and discrimination**

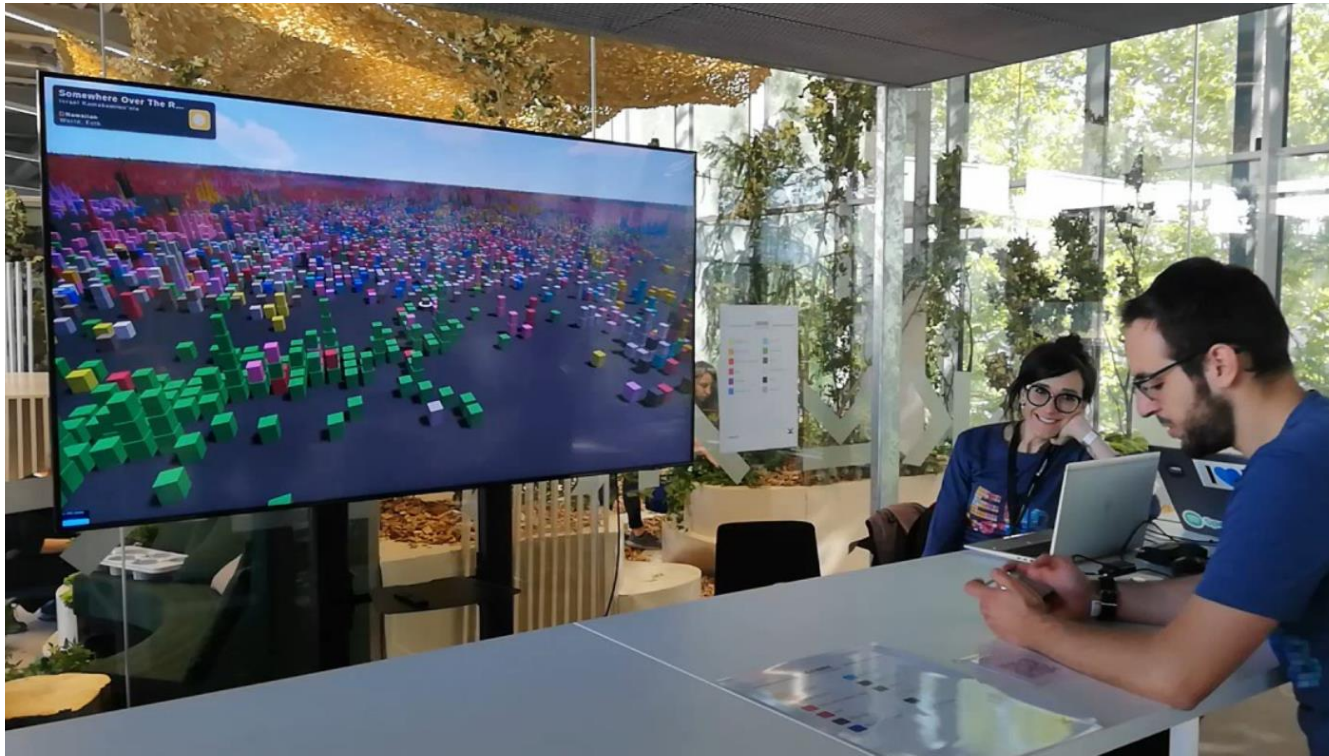
Large-scale Music Exploration (EmoMTB)

https://ars.electronica.art/newdigitaldeal/en/music-tower-blocks

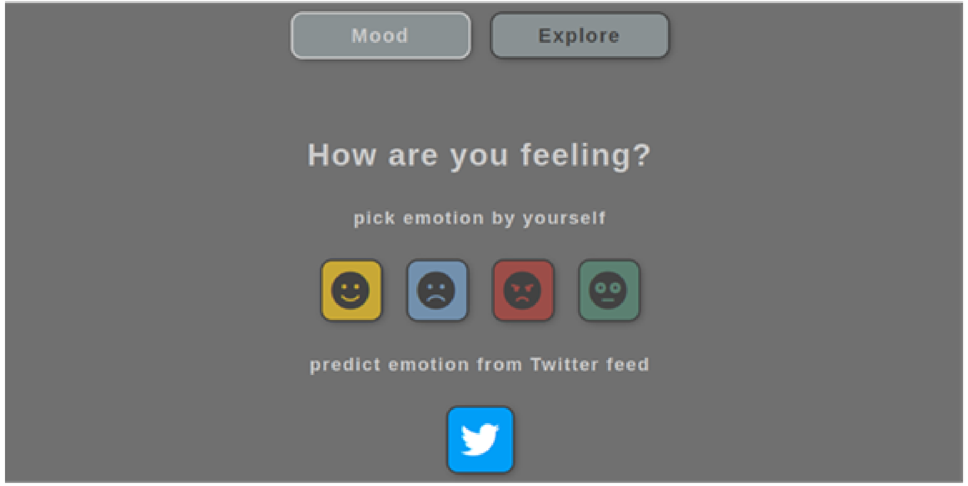
Emotional Music Tower Blocks (EmoMTB)
Emotion-aware Music Recommendation and Exploration

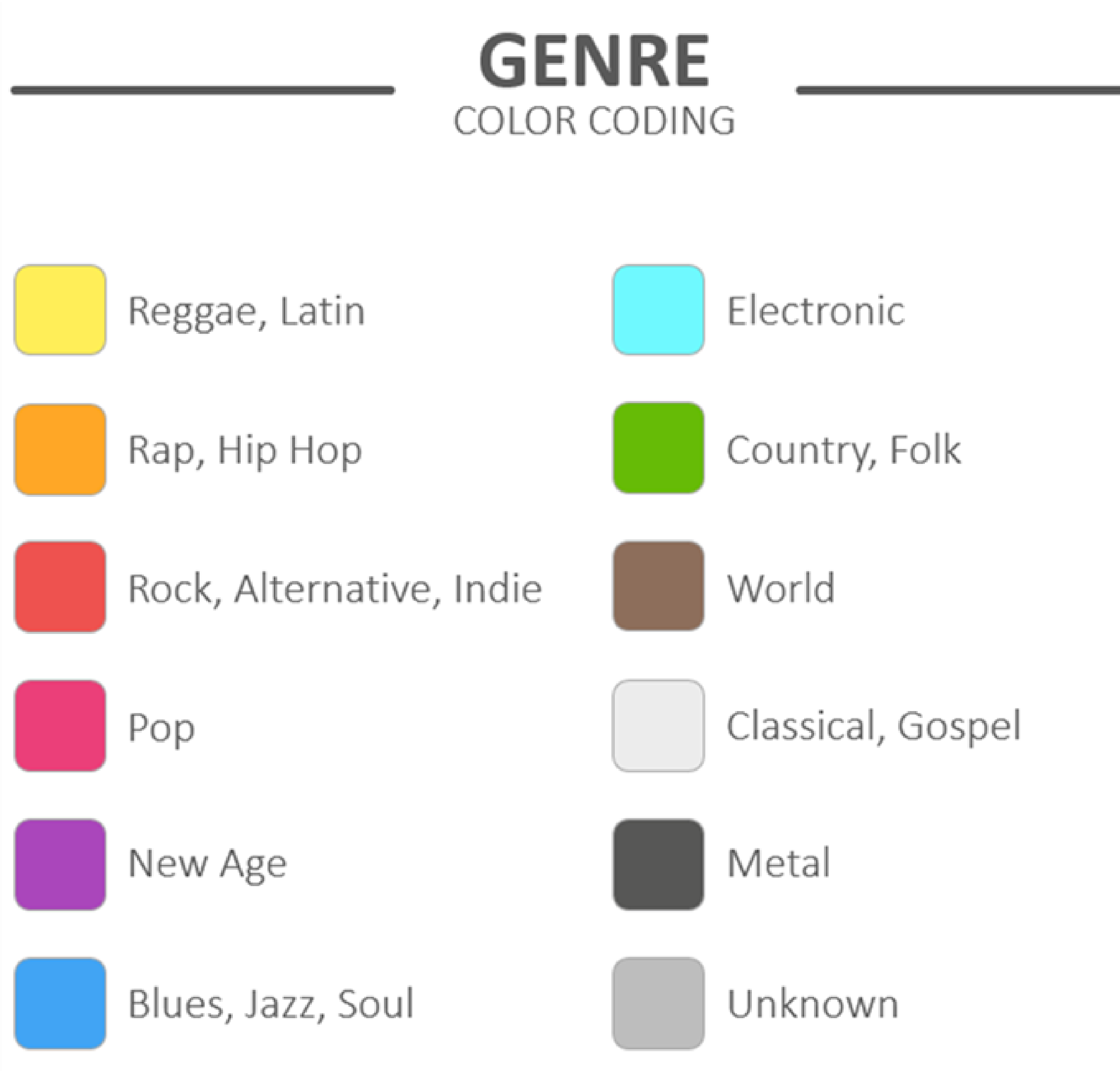
Addressed problems: Exploration vs. exploitation, diversification

- Artistic/scientific project presented at Ars Electronica Festival of Media Arts 2021
- Audiovisual exploration of a music collection (~500K tracks) using metaphor of city
- Tracks are clustered based on (very fine-grained) genre information and audio features
- Visualized as blocks; very similar ones are stacked to form buildings
- Nearby buildings form neighborhoods of similar genres (genres are color-coded)
- Each track is assigned an emotion (predicted from Last.fm tags)
- User selects an emotion → recommendations and visualizations update accordingly
- Explanatory video: https://bit.ly/3hfVH1S

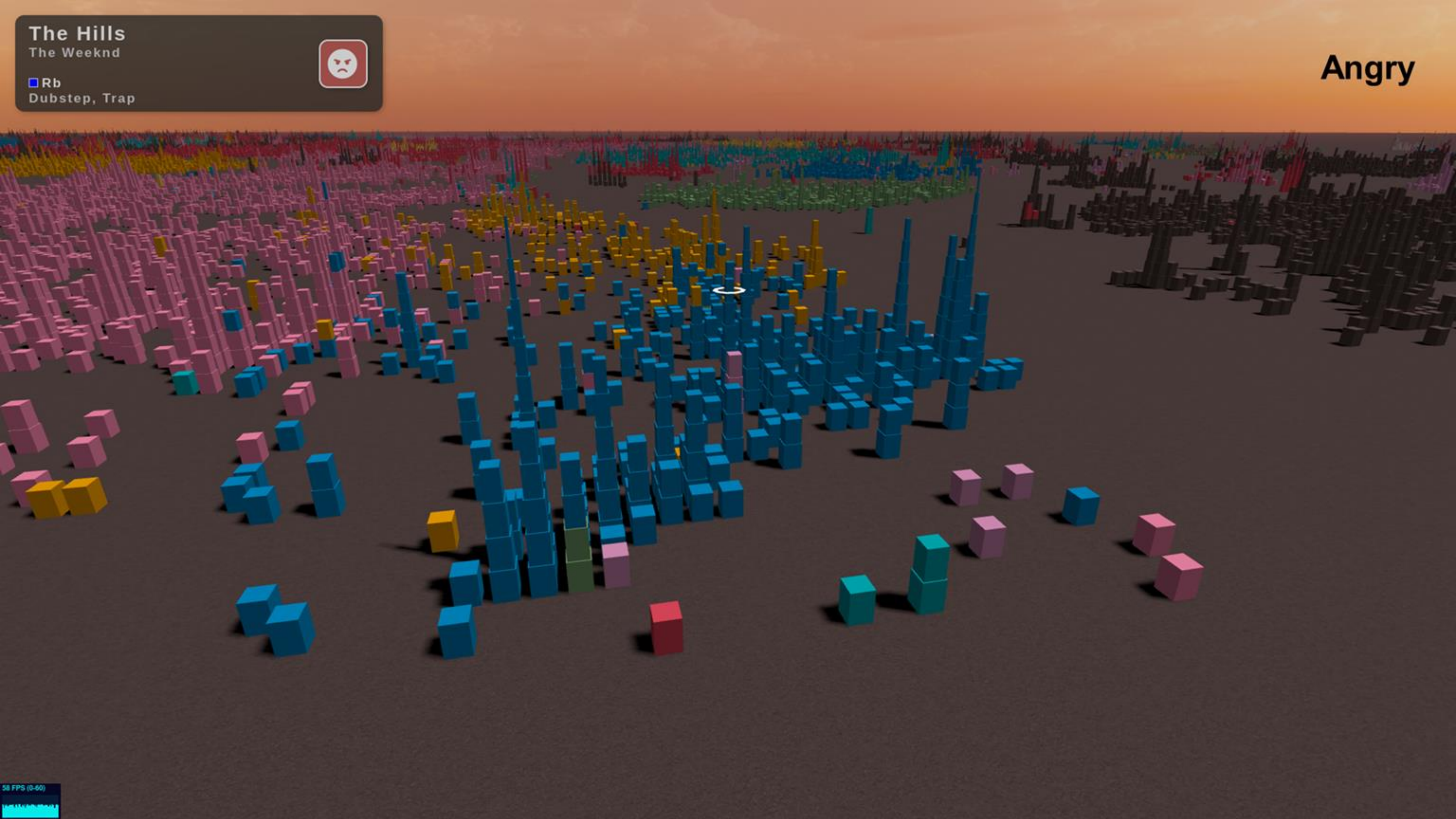


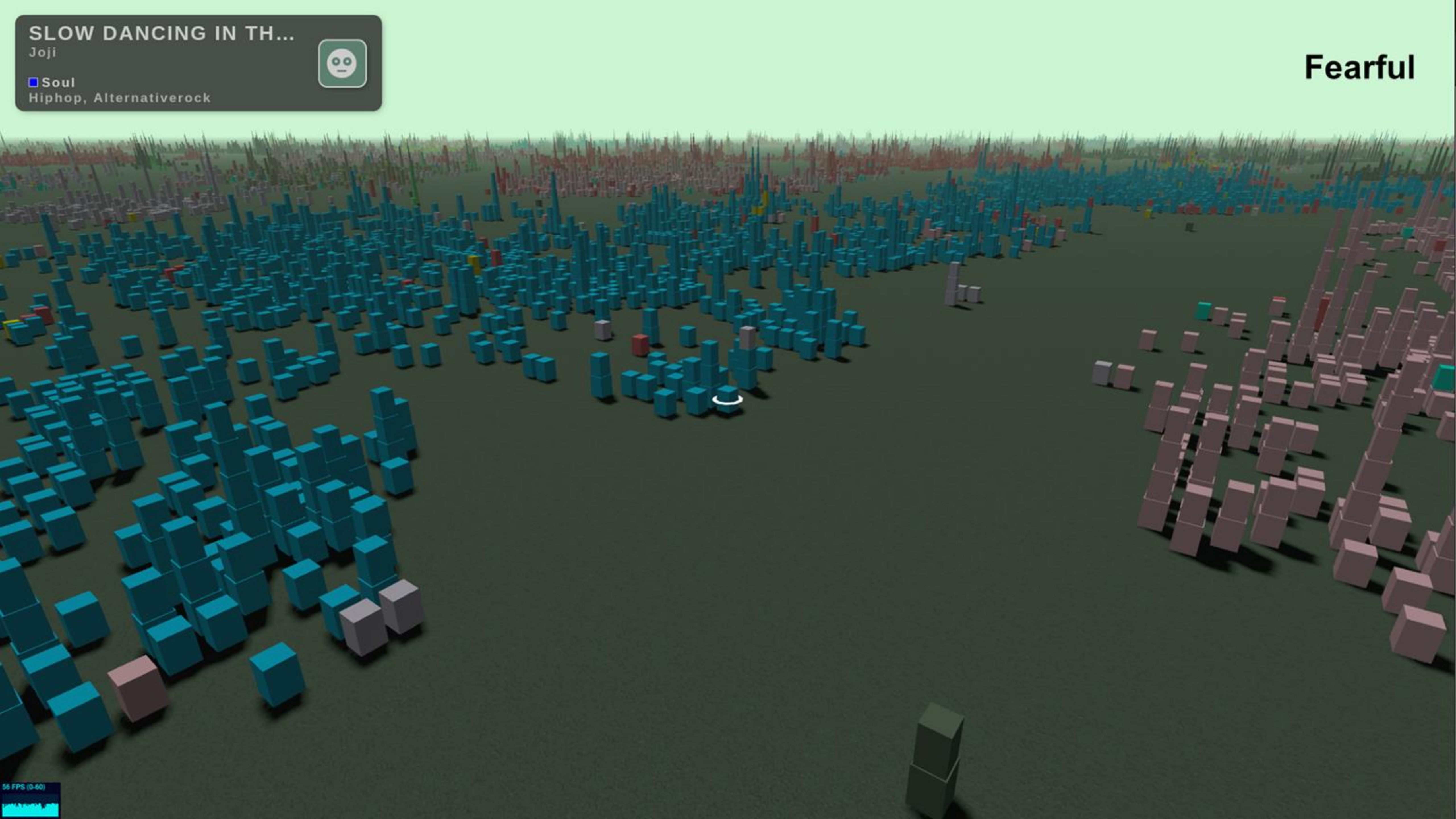
EmoMTB: User Controls

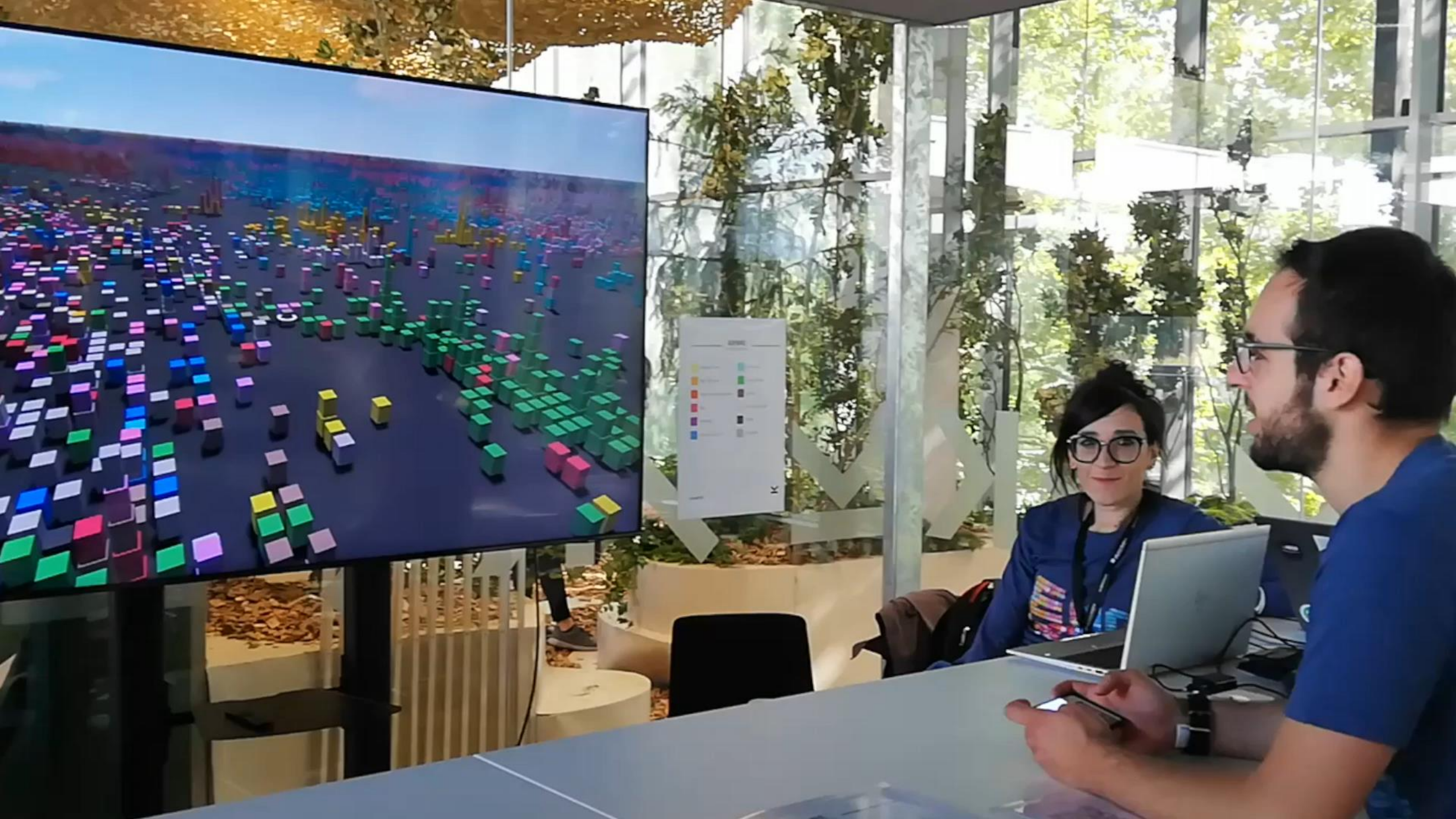





Hips Don't Lie (feat. W...
Shakira, Wyclef Jean
Pop
Latin, Reggaeton
Happy
55 FPS (0-60)


The Hills
The Weeknd
Rb
Dubstep, Trap
Angry
58 FPS (0-60)


SLOW DANCING IN TH...
Joji
Soul
Hiphop, Alternativerock
Fearful
56 FPS (0-60)

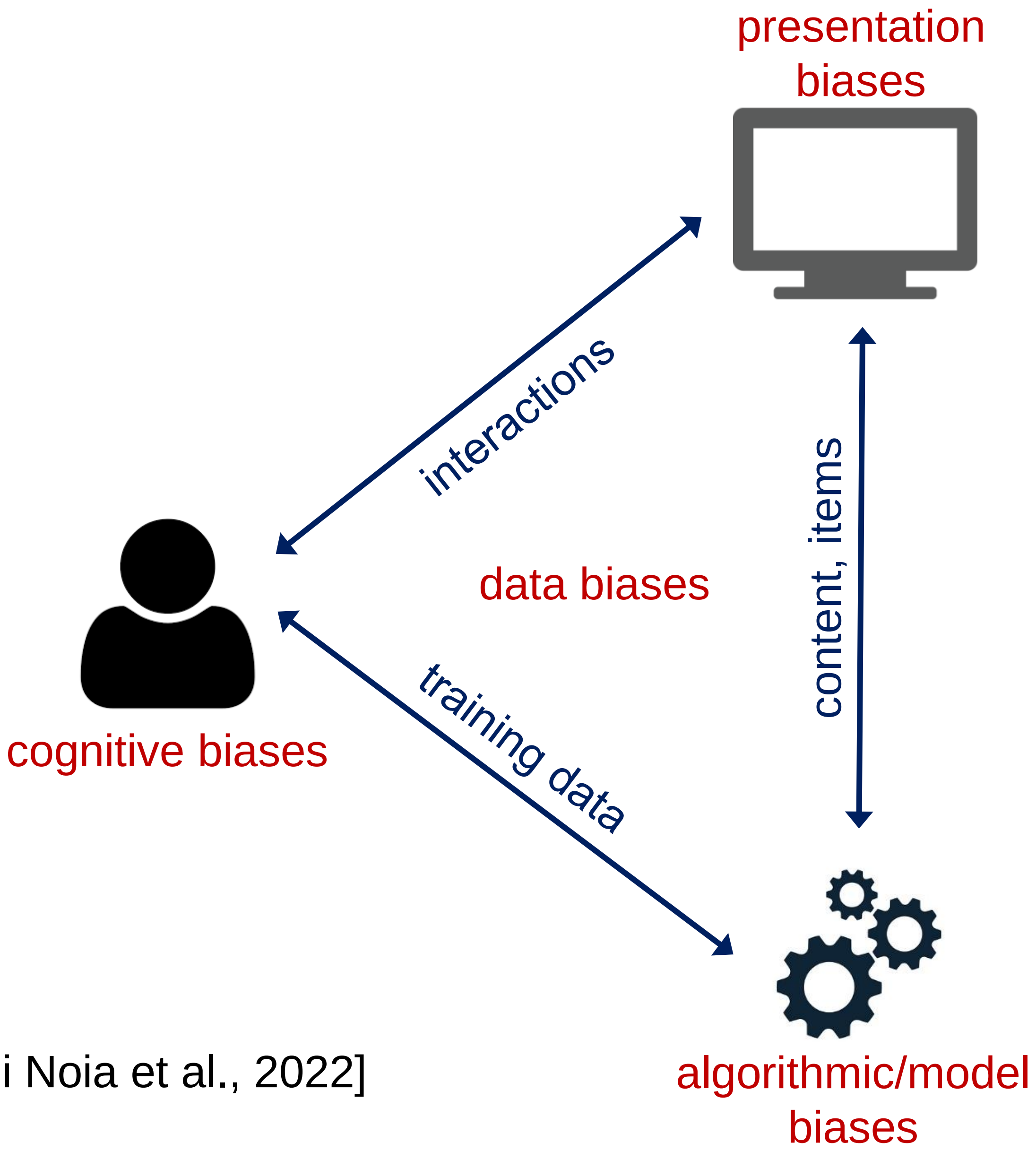


Biases in Recommendations (Black Holes of Popularity)

Biases in Recommender Systems

Decisions made by RSs are affected by various biases (influencing each other), originating from:

- *Data:* e.g., unbalanced dataset w.r.t. group of users → demographic bias, community bias
- *Algorithms:* e.g., reinforcing stereotypes or amplify already popular content ("rich get richer" effect) → popularity bias
- *Presentation:* e.g., positions of recommended items on screen
- *User cognition or perception:* e.g., serial position effect, confirmation bias



[Di Noia et al., 2022]

When Are Biases Problematic?

Biases can result in different treatment of users or groups of users

“The system systematically and unfairly discriminates against certain individuals or groups of individuals in favor of others.” [Friedman and Nissenbaum, 1996]

In case of **popularity bias**: reinforcing already popular items/artists, while limiting exposure of less popular ones (bad for artists and users ☹)

Setting a optimal level of popularity in recommendations is tricky, though!

What is the desired level of popularity in recommendations?

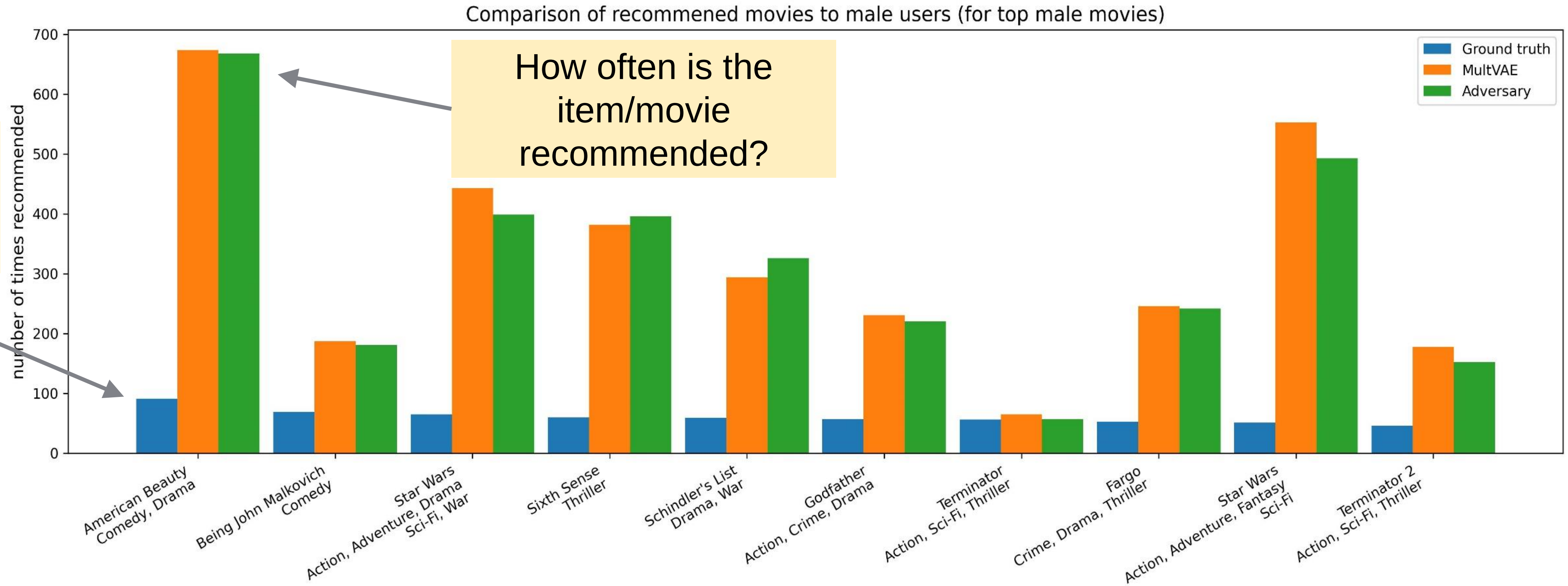
- Should a system recommend *all* items with the same probability?
- Should the popularity of items in the recommendation list match the popularity of items in the user’s consumption history (“calibration”)?
- Should the popularity of items in the recommendation list match with the item popularity in the consumption history of all (or a subset of) users of the RS?

Popularity Bias: Simple Example

[Lesota et al., 2021]

Metric: Difference between an item’s recommendation frequency and consumption frequency in user profiles

How often is the item/movie consumed?



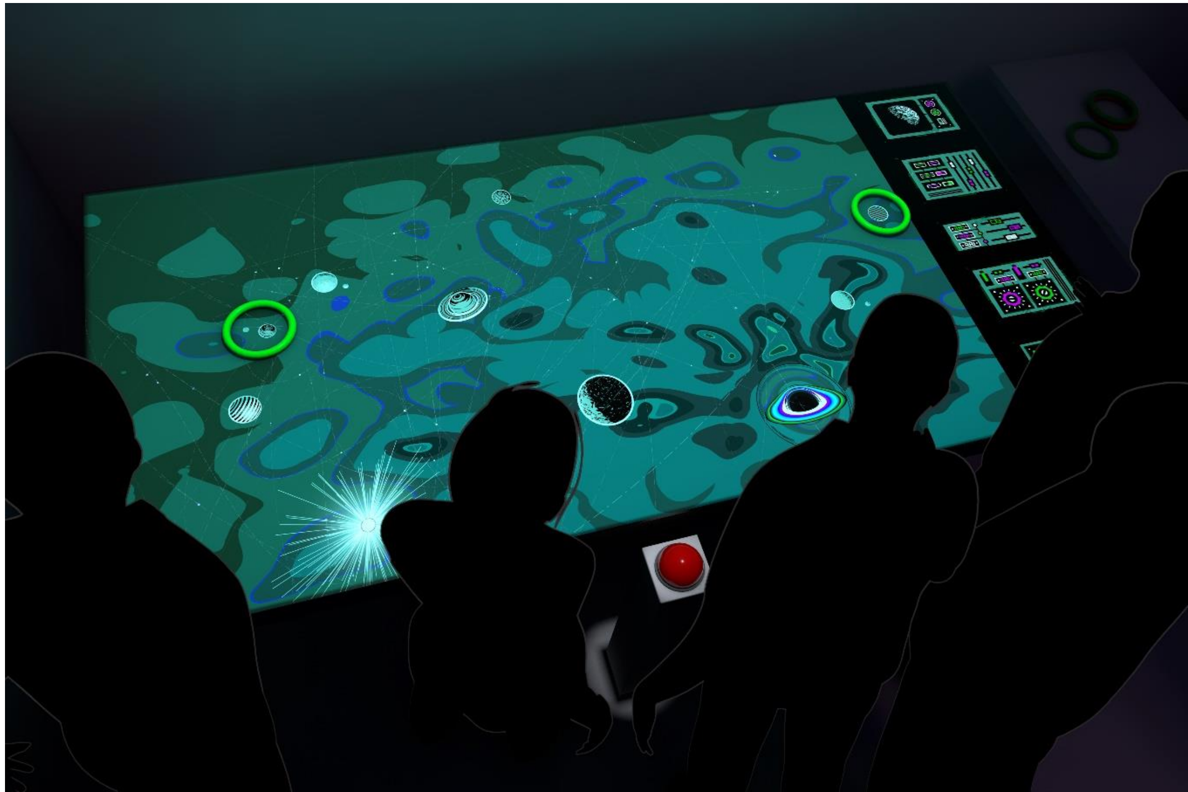
How often is the item/movie recommended?

https://ars.electronica.art/planetb/en/black-holes-of-popularity

Black Holes of Popularity

Addressed problems: Popularity bias

- Artistic/scientific project presented at Ars Electronica Festival of Media Arts 2022
- Raising awareness of artist popularity bias in music recommendation
- Exploration of music via genre, using metaphor of a universe
- Cosmic bodies represent songs with varying levels of popularity (planets, stars, black holes)
- User interacts by means of a lifebuoy with planets and stars, selecting which ones to save from being eaten by the black hole
- Influence of user's song saving activities is computed by in/decrease of fairness score, shown to the user
- Explanatory video: https://bit.ly/3VBAbqT




SELECTED REGION
COUNTRY & FOLK
LATIN & REGGAE
BLUES & JAZZ
PUNK & ROCK
POP

Gender Unfairness of Recommender Systems

User Gender Bias

[Melchiorre et al., 2021]

Metric: *RecGap* measures performance difference of system for different user groups

Model	Scenario	All	M/F	*RecGap*
POP	STANDARD	.046	.045/.049	.004 (f)
	RESAMPLED	.045	.044/.051	.007 (f) †
ItemKNN	STANDARD	.301	.313/.259	.054 (m) †
	RESAMPLED	.292	.304/.250	.054 (m) †
BPR	STANDARD	.127	.129/.117	.012 (m) †
	RESAMPLED	.123	.124/.116	.008 (m)
ALS	STANDARD	.241	.251/.205	.046 (m) †
	RESAMPLED	.238	.248/.204	.044 (m) †
SLIM	STANDARD	.364	.378/.315	**.063** (m) †
	RESAMPLED	.359	.372/.312	**.060** (m) †
MultiVAE	STANDARD	.192	.197/.173	.024 (m) †
	RESAMPLED	.183	.188/.166	.023 (m) †

Female users often receive worse recommendations than male users!

Artist Gender Bias

[Ferraro et al., 2021]

Debiasing: Penalize/downrank content by the majority group (male artists) by λ positions in the recommendation list, created with ALS CF approach

	Algo	Avg position 1st female	Avg position 1st male	% females rec.
LFM-1b	ALS	6.7717	0.6142	25.44
	POP	0.1325	1.7299	32.44
	RND	3.3015	0.3046	23.30
LFM-360k	ALS	8.3165	0.7136	26.27
	POP	0.9191	0.2713	29.31
	RND	3.3973	0.2951	22.77

Female artists tend to occur further down in the recommendation lists!

Personality Bias

[Melchiorre et al., 2020]

RQ: Do music recommender algorithms treat users with different personality traits equally?

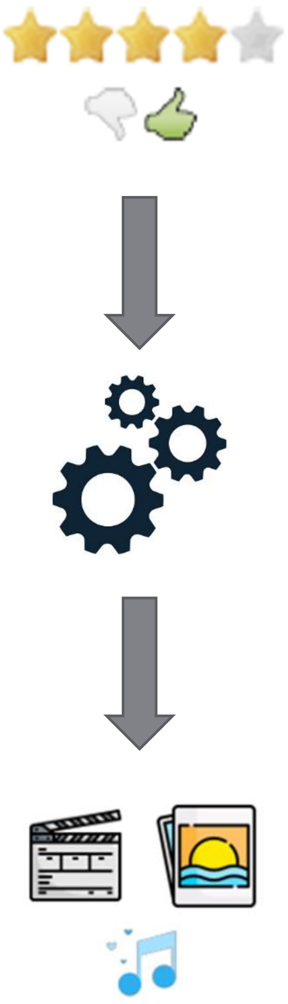
Open minded and extravert users receive worse recommendations, while neurotics receive better ones!

Trait	Algorithm	@5 All	@5 High	@5 Low
Agr.	EASE	0.0311	0.0295	0.0327
	SLIM	0.0279	0.0263	0.0295
	Mult-VAE	0.0380	**0.0385***	**0.0374***
Con.	EASE	0.0311	**0.0274***	**0.0349***
	SLIM	0.0279	**0.0241*****	**0.0319*****
	Mult-VAE	0.0380	0.0353	0.0407
Ext.	EASE	0.0311	**0.0266****	**0.0355****
	SLIM	0.0279	**0.0242****	**0.0317****
	Mult-VAE	0.0380	**0.0340****	**0.0417****
Neu.	EASE	0.0311	**0.0366*****	**0.0257*****
	SLIM	0.0279	**0.0335*****	**0.0224*****
	Mult-VAE	0.0380	**0.0436*****	**0.0324*****
Ope.	EASE	0.0311	**0.0221*****	**0.0400*****
	SLIM	0.0279	**0.0196*****	**0.0363*****
	Mult-VAE	0.0380	**0.0285*****	**0.0473*****

Not All Hope Is Lost

Harmful Biases Can Be Mitigated

Strategies to Mitigating Harmful Biases



Pre-processing strategies

- *Data rebalancing* (e.g., upsample minority group, subsample majority group)

In-processing strategies

- *Regularization* (e.g., include bias correction term/bias metric in loss function used to train a model)
- *Adversarial learning* (e.g., train a classifier that predicts the sensitive attribute and adapt model parameters to minimize performance of this classifier)

Post-processing strategies

- *Reweigh/Rerank* items in recommendation list
- *Filter* items (e.g., remove items from overrepresented groups)

Open Questions and Challenges (You May Solve ☺)

- Which novel methods, algorithms, architectures do we need to debias state-of-the-art RS algorithms? How to address the trade-off personalization vs. fairness?

- How to leveraging multimedia data? And how to beneficially integrate it with collaborative data?

- How can a machine understand user intent (purpose why they want to listen to music now)? And how should intent be integrated into RSs?

- How should requirements and aims of various stakeholders (e.g., content creator, consumer, provider, policymakers, etc.) be accounted for?

- Do computational bias metrics really capture how users perceive fairness?

- What are the economic, social, and legal consequences of biases resulting from RS technology adopted in high-risk areas (e.g., in recruitment)?

Thank You!

Markus Schedl
Johannes Kepler University Linz, Austria
Linz Institute of Technology, Austria
markus.schedl@jku.at | www.mschedl.eu | @m_schedl

JOHANNES KEPLER UNIVERSITY LINZ

References

[Chen et al., 2020]: *Bias and Debias in Recommender System: A Survey and Future Directions*, CoRR abs/2010.03240, https://arxiv.org/abs/2010.03240, 2020.

[Di Noia et al.m 2022]: *Recommender systems under European AI regulations*. Communications of the ACM 65(4): 69-73, 2022.

[Ekstrand et al., 2021]: *Fairness and Discrimination in Information Access Systems,* CoRR abs/2105.05779, 2021.

[Ferraro et al., 2021]: *Break the Loop: Gender Imbalance in Music Recommenders*, Proceedings of the ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR), Canberra, Australia, 2021.

[Friedman and Nissenbaum, 1996]: *Bias in Computer Systems*, ACM Transactions on Information Systems 14(3):330-347, 1996.

[Ganhör et al., 2022]: *Mitigating Consumer Biases in Recommendations with Adversarial Training*, Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2022), Madrid, Spain, 2022.

[Kowald et al., 2020]: *The Unfairness of Popularity Bias in Music Recommendation: A Reproducibility Study*, Proceedings of the 42nd European Conference on Information Retrieval (ECIR 2020), Lisbon, Portugal, 2020.

[Lesota et al., 2021]: *Analyzing Item Popularity Bias of Music Recommender Systems: Are Different Genders Equally Affected?*, Proceedings of the 15th ACM Conference on Recommender Systems (RecSys 2021), Amsterdam, the Netherlands, 2021.

[Melchiorre et al., 2020]: *Personality Bias of Music Recommendation Algorithms*, Proceedings of the 14th ACM Conference on Recommender Systems (RecSys 2020), Virtual, 2020.

[Melchiorre et al., 2021]: *Investigating Gender Fairness of Recommendation Algorithms in the Music Domain*, Information Processing & Management, 58(5), 2021.

[Ricci et al., 2022]: *Recommender Systems Handbook*, 3rd edition, Springer, 2022.