

User Insights on Diversity in Music Recommendation Lists



Interactive Version

Kyle Robinson¹, Dan Brown¹, Markus Schedl²

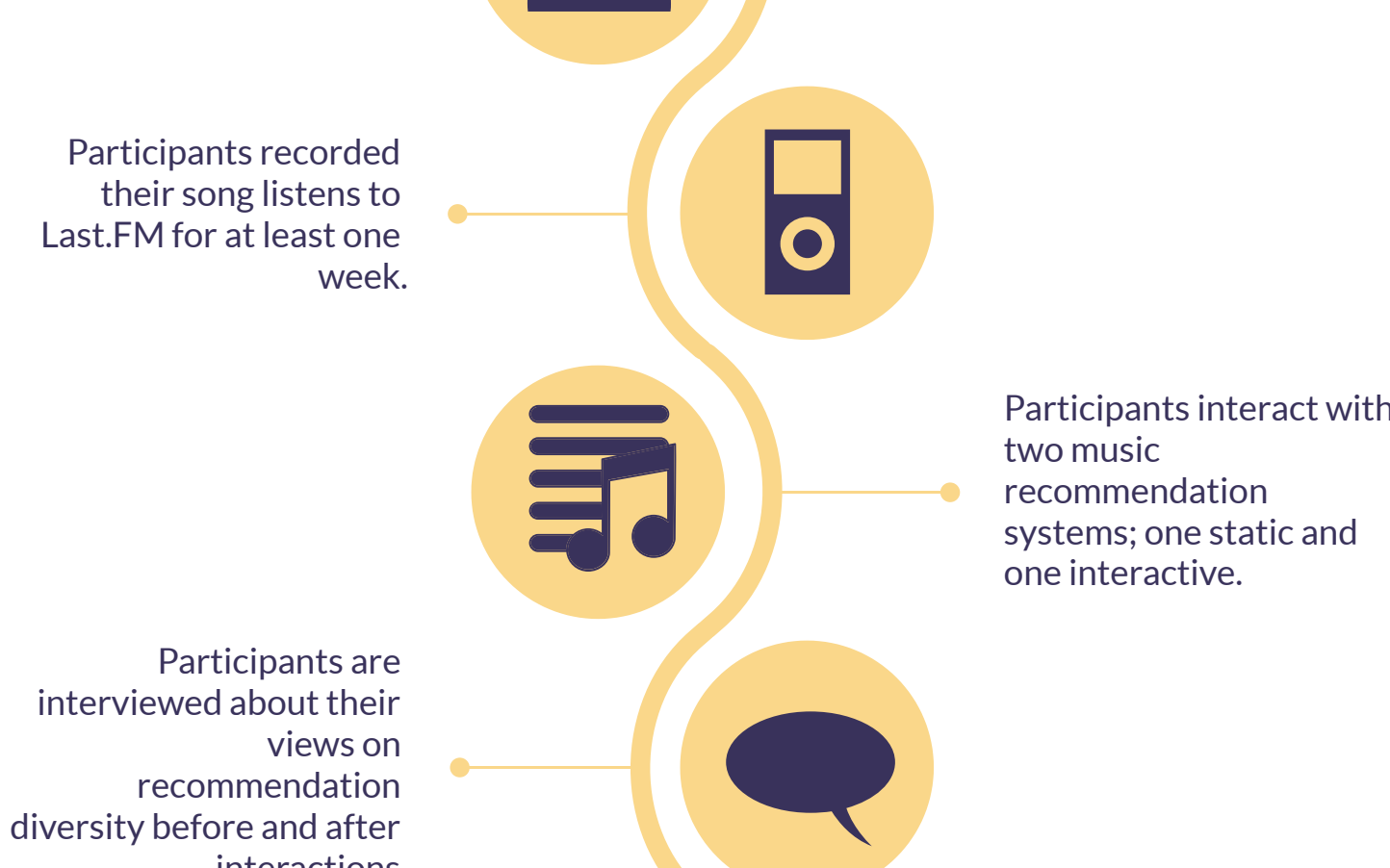
¹David R. Cheriton School of Computer Science, University of Waterloo, Canada

²Institute of Computational Perception, Johannes Kepler University Linz, Austria

Research Questions

- 1 What are users' **feelings about diversity** in personalized music recommendation lists?
- 2 How might users **optimize the level of diversity** in personalized recommendation lists?

Study Overview



DIVERSITY?

Diversity is a beyond-accuracy metric for recommender systems that is often defined as the opposite of similarity.

Rooted in information retrieval, and often used to prevent over-personalization of recommendations.

INTRA-LIST DIVERSITY (ILD)

One commonly used measure of diversity based on pairwise (dis)similarity.

$$d_i = \sum_{\substack{j=1 \\ j \neq i}}^n ||v_i - v_j||$$

A diverse list has songs (d_i) whose **features** (γ) differ from all other songs **features** (v_j).

Note: Total ILD is a normalized sum of all d_i .

Last.FM Dataset



341,764,569

Total scrobbles (listening events)

51,669

Unique usernames

1 Year

Scrobbles recorded from Jan 2019 to Feb 2020

Recommendation Lists

Four unique top-100 recommendation lists were generated for each participant using implicit matrix factorization (Alternating Least Squares) trained on the above dataset and their own listening events.

Each participant was then shown the top-7 recommendations from each top-100 list displayed as Spotify song previews first using system 1 and then using system 2.

System 1

Static top-7 recommendations based on recommendation relevance (R_i) from matrix factorization.

$$F_i = R_i$$

- 1) Well Known: 10% KAYTRANADA
- 2) Well Known: Risk FKJ, Bas
- 3) Well Known: Sofia Clairo
- 4) Well Known: In Your Eyes The Weeknd
- 5) Well Known: A Pale ROSALIA
- 6) Well Known: At The Door The Strokes
- 7) Well Known: Never Come Back Caribou

System 2

Interactive re-ranking of top-7 recommendations incorporating ILD (D_i) using a slider. Beta (B) is a value between 0 and 1 controlled by the slider.

$$F_i = (1 - B) * R_i + B * D_i$$

- Trial: 5/8
- 1) Well Known: 10% KAYTRANADA
 - 2) Well Known: Risk FKJ, Bas
 - 3) Well Known: Sofia Clairo
 - 4) Well Known: In Your Eyes The Weeknd
 - 5) Well Known: A Pale ROSALIA
 - 6) Well Known: At The Door The Strokes
 - 7) Well Known: Never Come Back Caribou

Interviews

Pre-Interview

- When you look for diverse music recommendations, what do you mean by that?

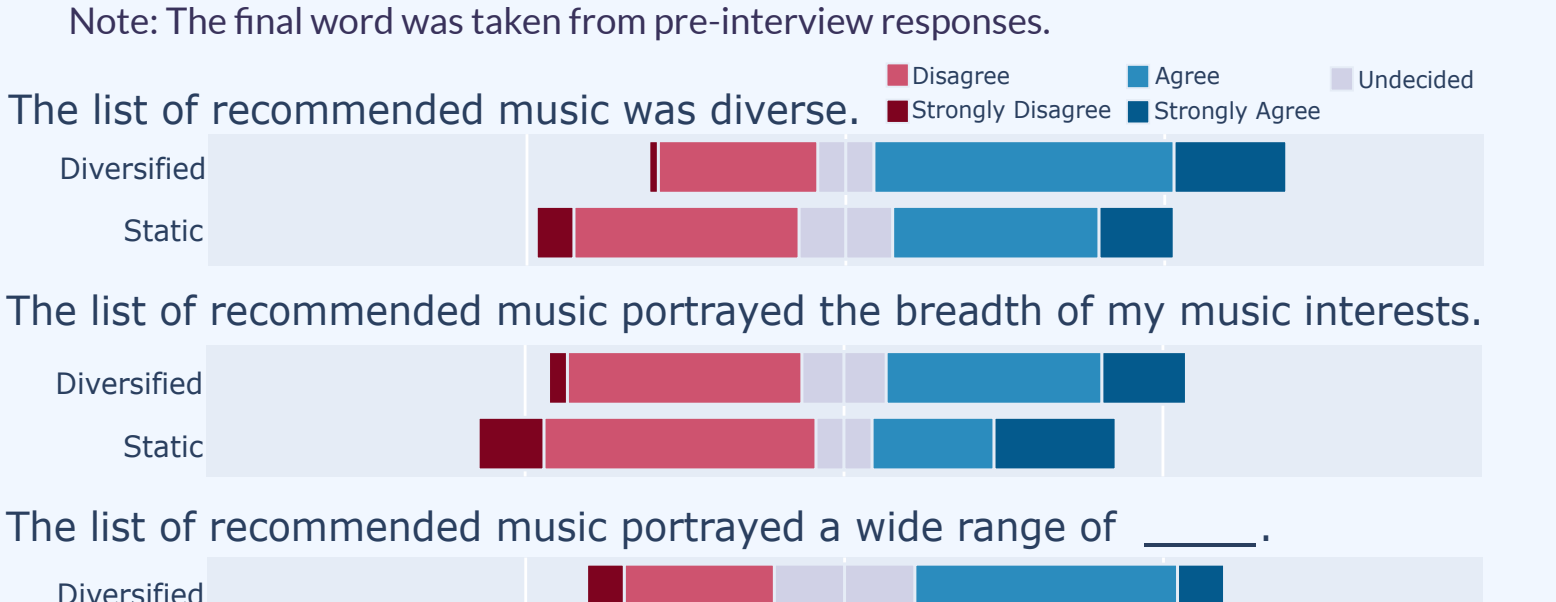
Post-Interview

- Which system did you prefer, and why?
- How clear was the effect of the slider, and what did it seem to change?
- What are some positives and negatives of more and less diversity in recommendations?

Quantitative Results

Likert responses on diversity of recommendations after selecting a final slider position.

Note: The final word was taken from pre-interview responses.



Key Qualitative Insights

Inner Diversity

Diversity within the bounds of existing preferences.

"Diverse... within the boundaries of the things that I like."

Outer Diversity

Diversity outside of existing preferences.

"Something that I'm not used to, like I've never heard before."

Mood

Commonly referenced as an important factor in how much diversity participants wanted.

"[I] like a piece of music right now because of the mood that I am in..."

"You just want to listen to sad songs. You just want a playlist that has a sad song. You don't want diversity."

Diversity in music recommendations should have at least as solid a foundation in user perception as in information retrieval.