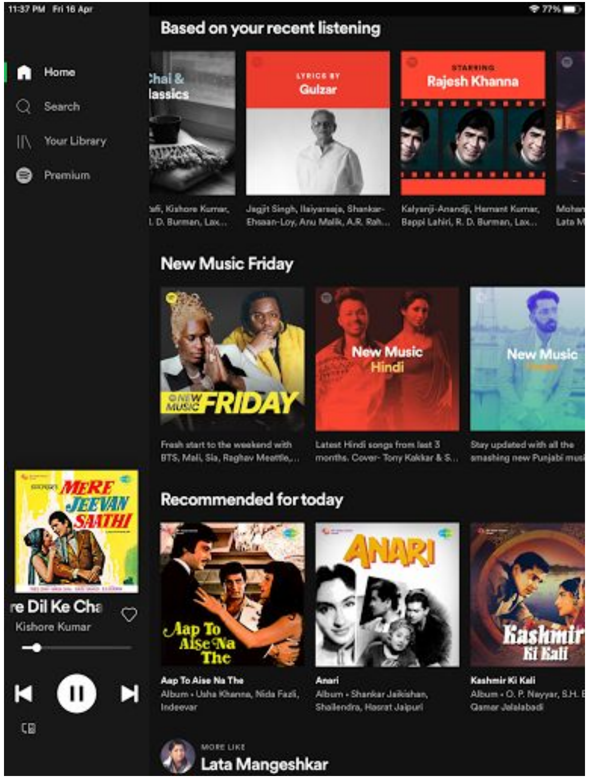
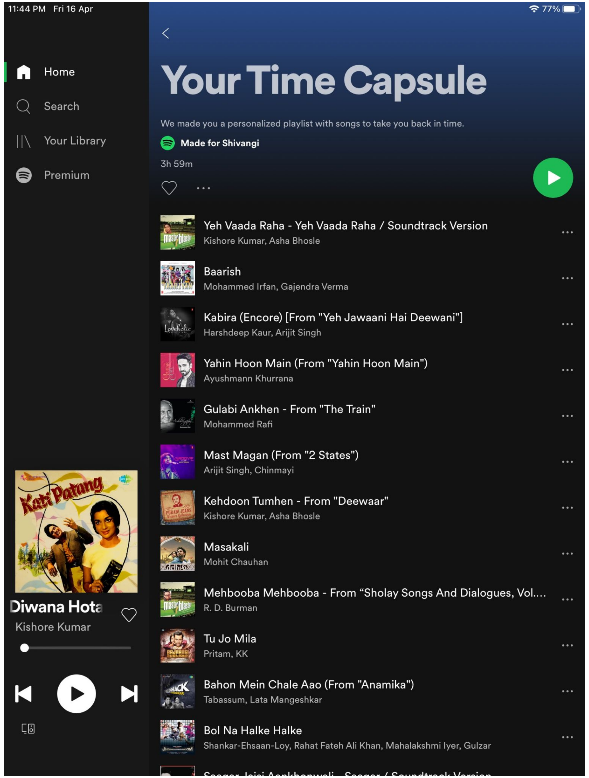
Algorithmic Effects on the Diversity of Consumption on Spotify

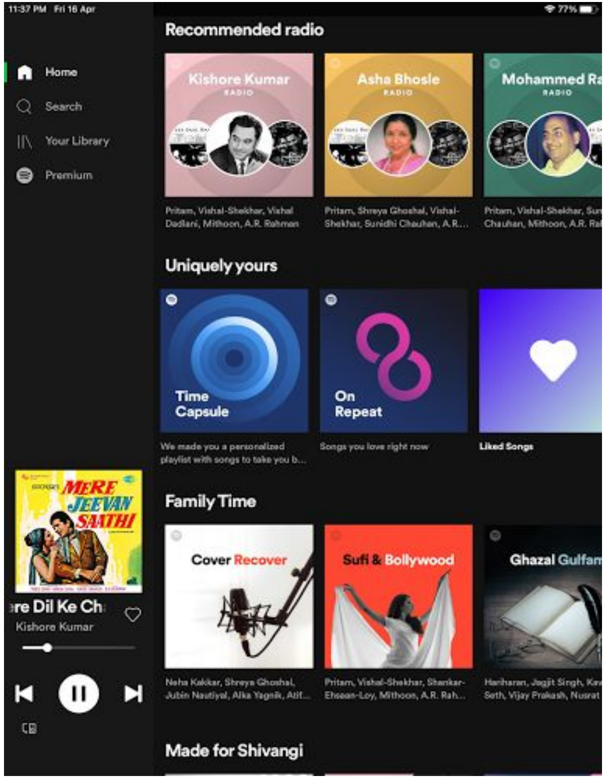
Ashton A., Lucas M., Rishabh M., Ian A., Mounia L.
University of Toronto & Spotify

WWW '20

Spotify Recommendations







Research Focus & Contribution

1. Analysis of diversity in user behaviour in recommender systems and online platforms
2. Study of music listening and streaming services
3. Study of trade off between short-term and long-term engagement in online platforms

Some Related Knowledge

1. Filter Bubble: A **filter bubble** is a state of intellectual isolation that can result from personalized searches when a website algorithm selectively guesses what information a user would like to see based on information about the user, such as location, past click-behavior and search history.
2. Diversity: In this paper, Diversity is referred to as how similar a piece of music is to the type of music the user has historically streamed
    a. It can facilitate exploration by helping users discover new content or inculcate new tastes.
    b. It can help the platform spread consumption across artists and facilitate consumption of less popular content.

Data

- 100M distinct premium users data
- Time period of data collection: 28 days of July 2018 + 28 days more for temporal analysis
- Data divided into
  - User-driven data collection
  - Algorithm-driven data collection

Music Embeddings

- Embeddings used to encode latent representations between users and content.
- Bag-of-words Word2vec model
- User generated playlist - documents
- Songs in the playlist- terms
- 850M playlists used with some filtration
- User: average of the song embeddings that they listen
- Vectors - 40 dimensional vectors
- Cosine similarity

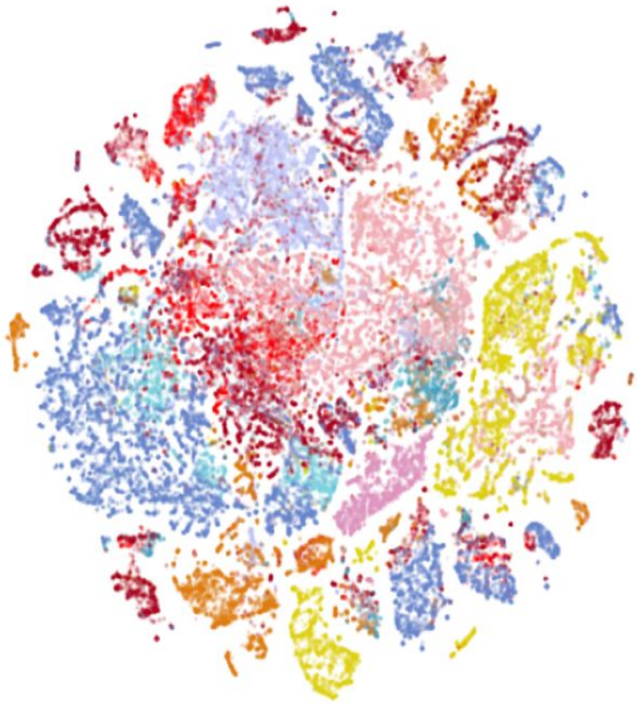


Figure 1: Two-dimensional *t*-SNE projection of 100,000 songs from our song embedding. Songs are represented as points, where two songs are close together if they have high usage overlap in our data. Colors represent genres of music.

Generalist - Specialist Score

- Used to quantify musical diversity of user
- Specialist - if listens to very similar songs
- Generalist - if listens to a diverse set of songs
- GS-score measures the average cosine similarity between a song vector and the average of the user's song vectors.

$$\vec{\mu_i} = \frac{1}{\sum w_j} \cdot \sum_j w_j \vec{s_j}$$

$$GS(u_i) = \frac{1}{\sum w_j} \sum_j w_j \frac{\vec{s_j} \cdot \vec{\mu_i}}{\|\vec{s_j}\| \cdot \|\vec{\mu_i}\|}$$

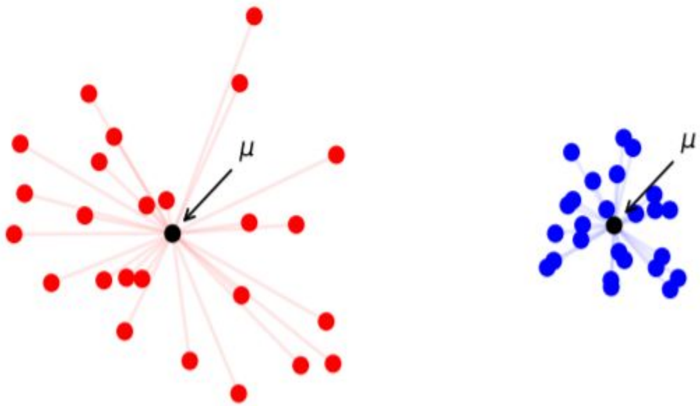


Figure 1: A schematic depicting the vector representations of communities contributed to by a generalist (left) and a specialist (right). The generalist's communities are spread out, and the specialist's communities are clustered together.

Diversity by Activity Level

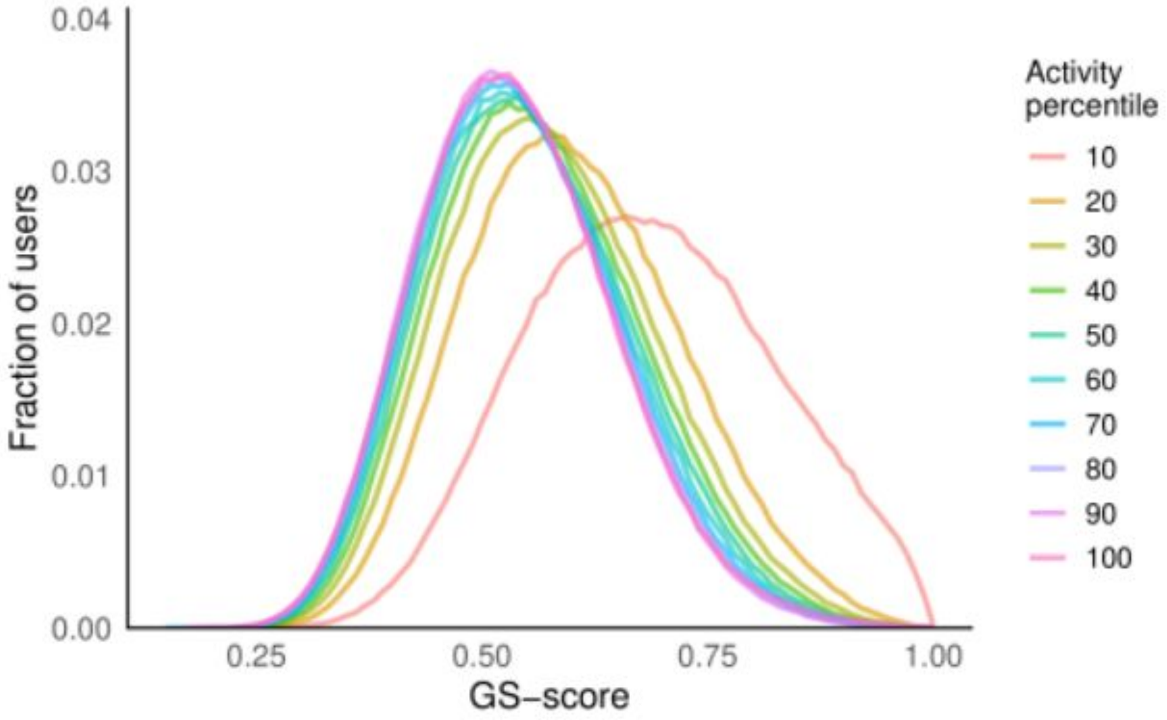
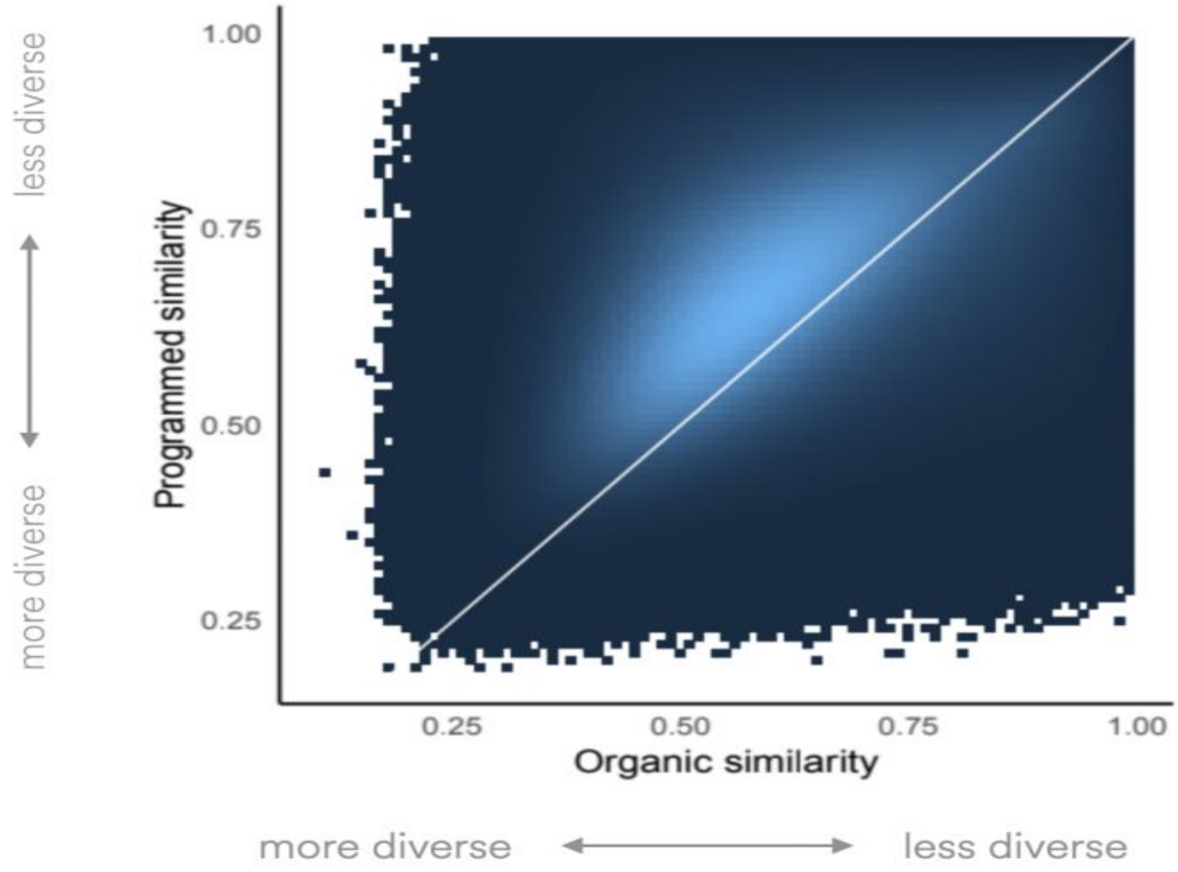


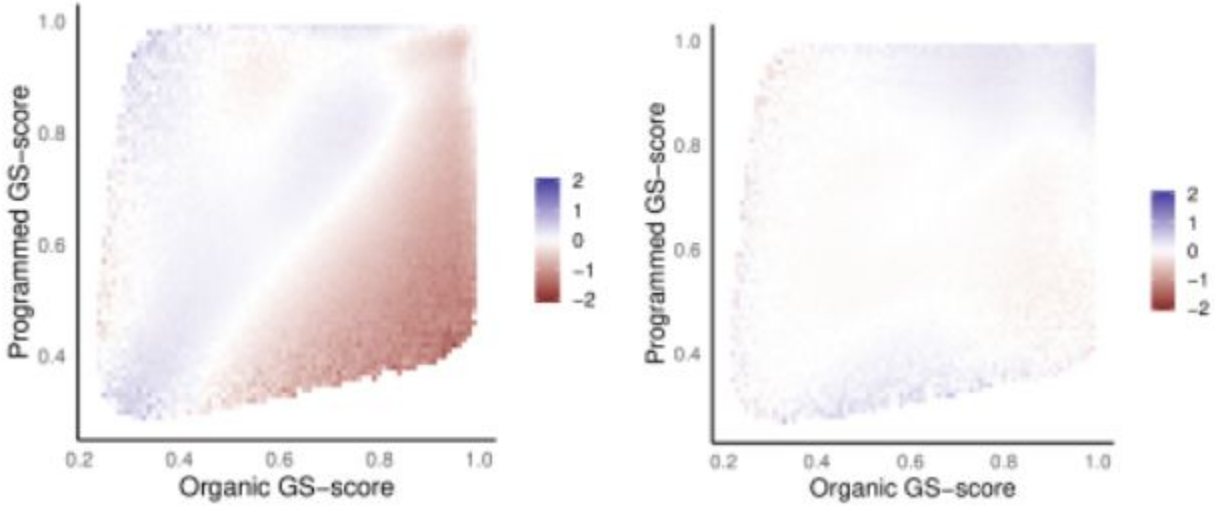
Figure 2: Distributions of diversity controlling for activity.

Organic Vs Programmed Diversity

The *y* = *x* line is shown in white. The vast majority of users are above this line, indicating that their programmed listening is less diverse than their organic listening.

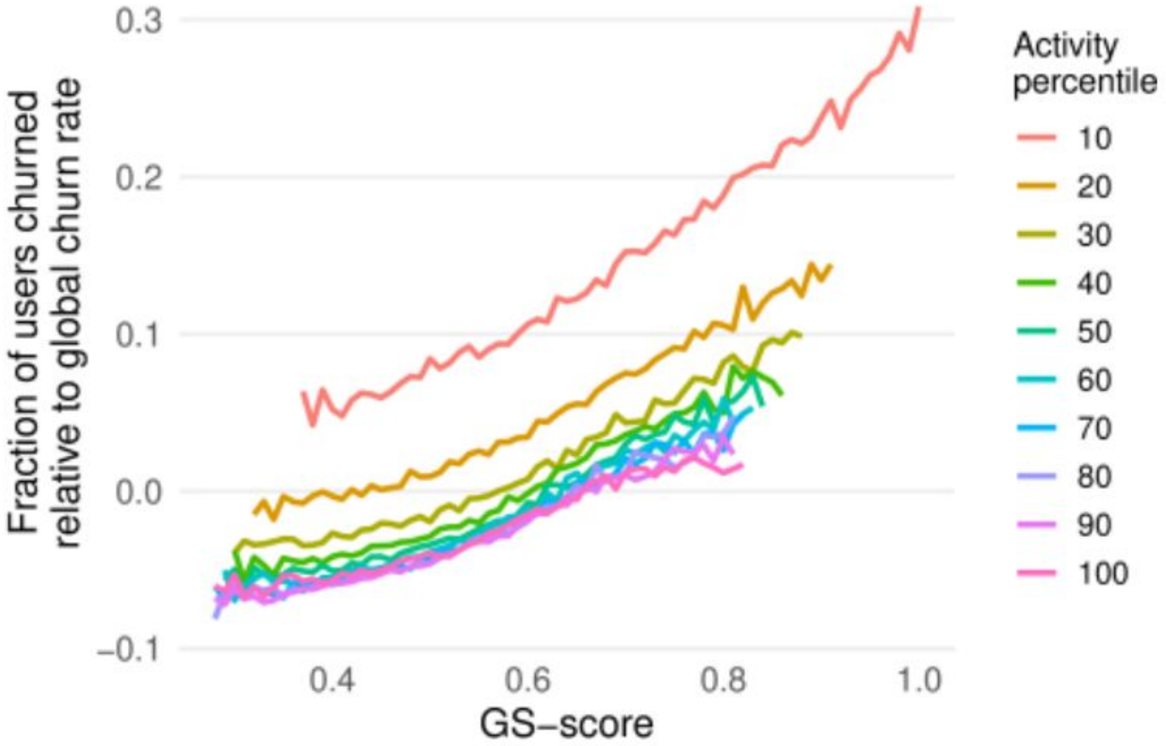


Demographic variation in Musical Diversity

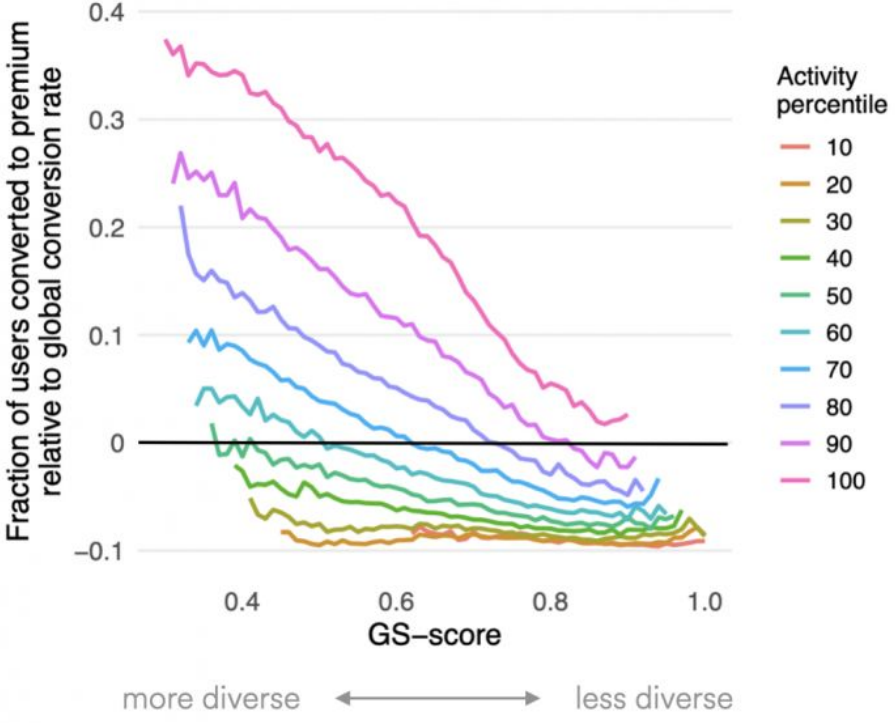


Log odds ratios of diversity distributions for (a) younger and older users and (b) male and female users.

Retention: Are users remaining on the platform?



Conversion: Are generalists or specialists more likely to become premium members?



Conclusion 1

- Conversion and retention are associated with greater content diversity.
- Recommendation Algorithms of spotify, though derives a significant amount of engagement, are associated with lower content diversity.
- Recommendation Algorithm can be effective in short-term while they are counterproductive in long-term.

My Experiment

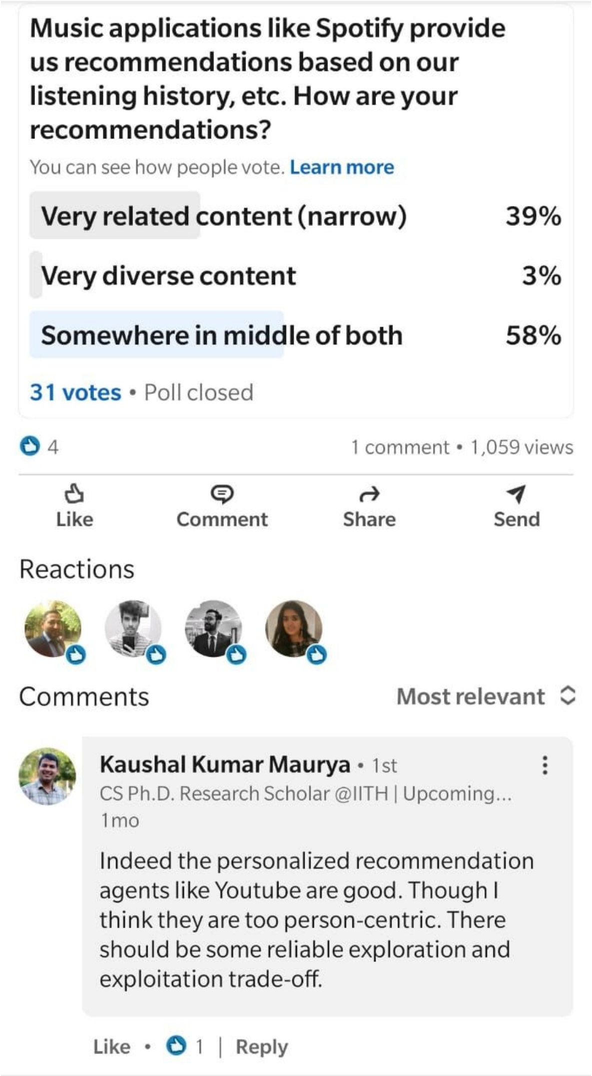
I created a poll after reading this paper, to know how people in my linkedin connection feel about their Spotify Recommendations

- 39% - Less Diverse
- 3% - More Diverse
- 58% - Mixture of both

Total people voted = 31

Total views on poll = 1059

(PS: Not an influencer yet to get more people see my post and vote)

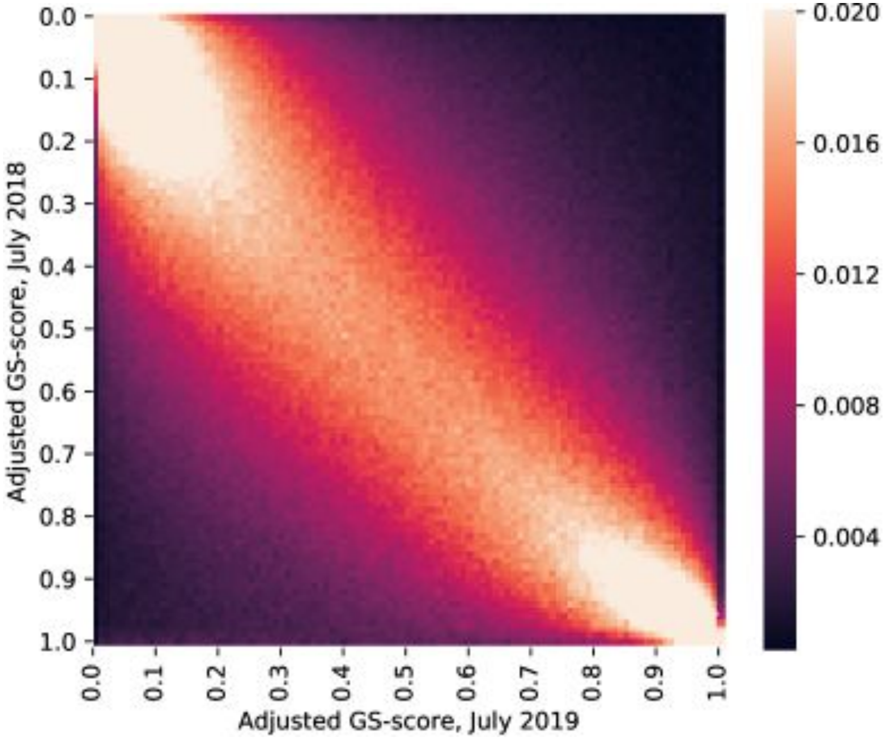


Musical Diversity Over Time

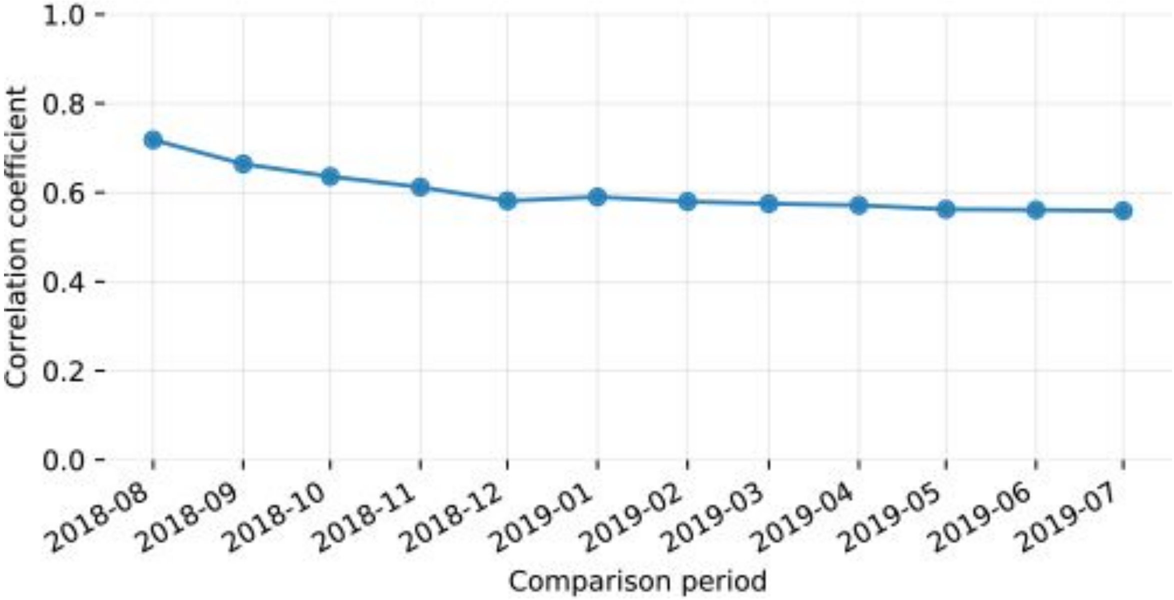
Activity Adjusted GS Score

For a given time period T, a user activity-adjusted GS SCore is the percentile rank of their GS SCore relative to all users in the same activity bucket at time T. This enables comparing GS Scores across time, controlling for any activity biases.

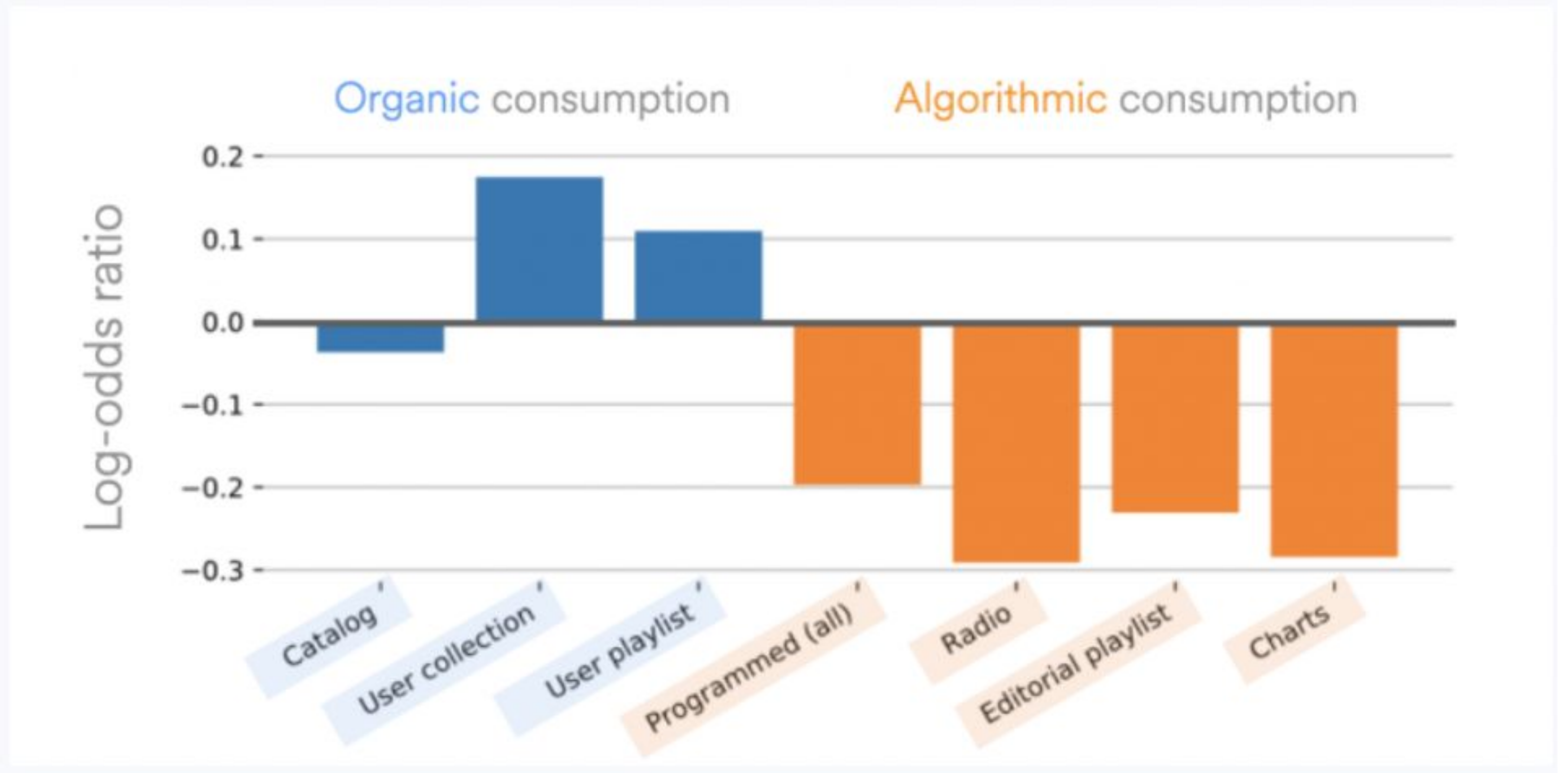
The Stability of Musical Diversity



The Stability of Musical Diversity



Mechanism of Change



Conclusion 2

- The GS-score is quite stable over time, clarifying that it captures an inherent behavioral characteristic of users.
- On explicitly analyzing users who increased their diversity over time, author found that these changes are accompanied by increases in their organic streaming and decreases in their algorithmically-influenced consumption

Impact of Recommendations for Generalist & Specialist

Experimental Setup

- Author Considered 7 popular algorithmic playlists
- Randomly assign a ranking algorithm in these playlists
  - Popularity Ranker - popularity in descending order
  - Relevance Ranker - relevance to user
  - Learned Ranker - based on neural network model with user-level, song-level and interaction level features.
- Online A/B Test on 5,40,000 users using free version of spotify

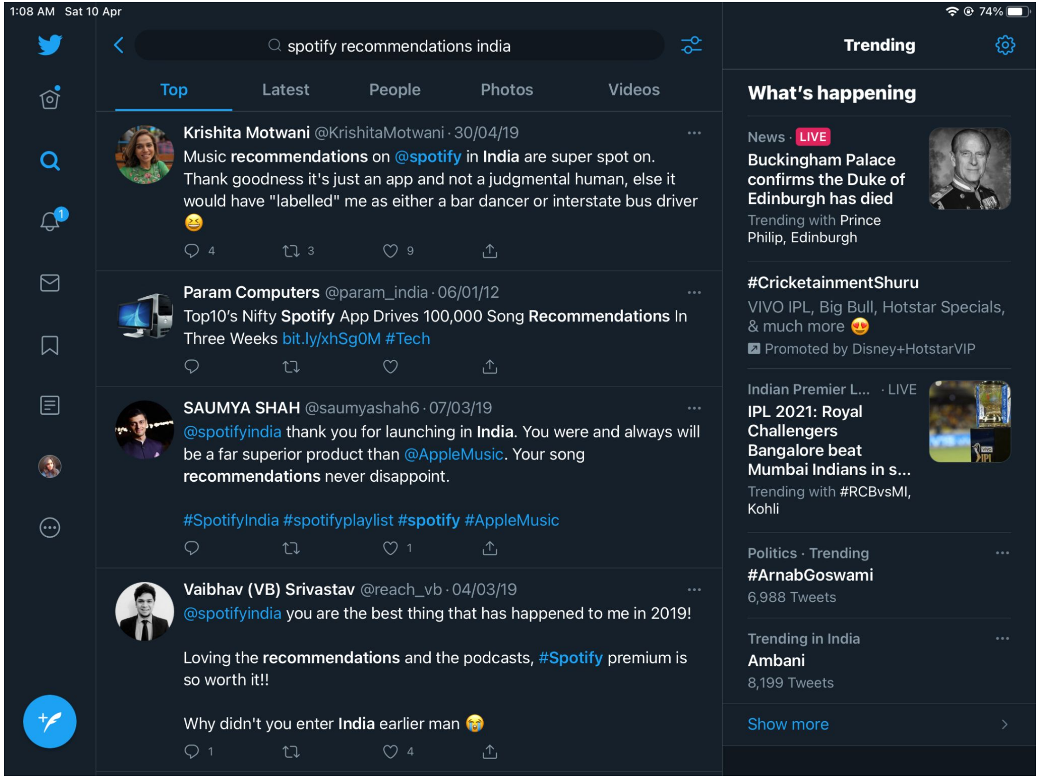
Relative Performance of Different Ranking Algorithms

Comparison	User Type	Song Streams	Song Skips
Relevance over Popularity	Generalists	+10.03%	+4.71%
	Specialists	+25.66%	+2.89%
Learned over Relevance	Generalists	+1.82%	+0.90%
	Specialists	+1.30%	−9.76%

Conclusion 3

- Consumption diversity is a useful signal in understanding how users will respond to recommendations.
- Classical recommendation models perform much better for specialists than for generalists.
- There is a need of more diversity-aware ranking methods for Generalist users.
- Author hints that there are risks to algorithmic over-specialization in online platforms, and to measuring the effectiveness of recommender systems too narrowly.

Different opinions from people on these recommendations



Limitations of paper

- Cold-start based recommendations
- Challenges in generating and using the dataset.
- Need to explore more diversity - aware algorithms.
- The importance of diverse recommendations for artists and content creators.

Resources

- Algorithmic Effects on the Diversity of Consumption on Spotify
- Generalists and Specialists: Using Community Embeddings to Quantify Activity Diversity in Online Platforms
- Algorithmic Effects on the Diversity of Consumption on Spotify

THANK YOU :)