Exploring and Visualizing Personal Music Trends with Spotify

Nithyashree Senguttuvan (011808333)
Simarjeet Singh (011870326)

March 31, 2024

Contents

1 Introduction

In this project, we examine our Spotify music listening patterns systematically. Utilizing Spotify's analytical tools, we gather our personal music data to explore our most frequently played songs and preferred artists. Our aim is to understand the relationships within our music preferences and observe how they change over time. Additionally, we enrich our analysis with insights such as music genres and artist collaborations. This project is focused on uncovering hidden patterns within our individualized soundtrack.

2 Intended Approach

Our approach aims to provide Spotify users with a comprehensive method to analyze and visualize their streaming activity using R and the Spotify API through the SpotifyR package. The process involves several distinct steps:

2.1 Data Collection

We utilize the user-specific data on streaming activity acquired which includes information such as track names, artists, timestamps, and user interactions such as likes or skips. By leveraging the functionalities offered by SpotifyR, we streamline the process of gathering this data, ensuring efficiency and accuracy.

2.2 Data Preprocessing

Before conducting any analysis, it's crucial to preprocess the raw data to make it suitable for exploration. This involves tasks such as handling missing values, standardizing data formats, and filtering out irrelevant information. By employing R's versatile data manipulation tools, we ensure that the dataset is clean and structured, ready for further analysis.

2.3 Exploratory Data Analysis (EDA)

EDA plays a pivotal role in understanding the underlying patterns and trends within the streaming data. Through various statistical and visualization techniques, we uncover insights such as most frequently streamed tracks, preferred genres, temporal patterns in listening behavior, and user engagement metrics. This step not only provides valuable insights into individual listening habits but also sets the stage for more advanced analyses.

2.4 Community Detection

Community detection algorithms are then applied to identify clusters or communities within the user's streaming activity network. In this context, a community refers to

a group of tracks or artists that are closely connected based on user interactions. By employing basic analysis techniques, we partition the streaming network into distinct communities, revealing underlying structures and preferences in the user's listening behavior.

2.5 Signed Networks

In the context of analyzing streaming activity, signed networks can provide deeper insights into user preferences and relationships between tracks or artists. A signed network incorporates positive and negative interactions, such as liking or disliking a track, into the analysis. By considering both positive and negative edges, we gain a more nuanced understanding of user preferences and the dynamics of their listening habits. This allows us to uncover patterns such as polarizing tracks or communities with strong positive/negative associations.

2.6 Visualization

Finally, we visualize the results of our analysis using interactive plots and dashboards. Through visually appealing and intuitive representations, users can explore their streaming activity, track community memberships, and understand the sentiments associated with different tracks or artists. This not only enhances the user experience but also facilitates interpretation and sharing of insights.

By following this intended approach, users can gain a deeper understanding of their Spotify streaming behavior, uncover hidden patterns, and explore the rich landscape of music preferences within their personal listening history.

3 Data Acquisition Strategy

The data acquisition strategy involves leveraging Spotify's Web API and the "Request My Data" feature for a detailed collection of personal streaming history, likes, and playlist contents. Scripts will be written to parse the obtained JSON or CSV files for relevant data like track IDs and artist names. Additional metadata, such as genre and collaborations, will be fetched using Spotify's API to enrich the dataset. The organized data will then be stored in a suitable database format, preparing it for comprehensive network analysis.

4 Background Review

Centrality Measures :

Utilize degree centrality to rank artists and tracks by frequency of appearance, indicating personal preference.

Community Detection :

Apply algorithms like the Louvain method for detecting communities within the network of genres or artists, revealing hidden patterns in music preferences.

Graph Embeddings :

Explore advanced techniques like node2vec to represent complex relationships and similarities between artists and genres in a multi-dimensional space.

Epidemic Models :

Using epidemic models, model listening behavior changes over time, simulating how new music interests spread and take hold.

Temporal Analysis :

Employ time-varying network analysis to track changes in music preferences over different periods, understanding how external events or discovery of new music influences listening habits.

Signed Networks :

Incorporate signed networks to capture positive and negative user interactions, providing a nuanced understanding of social dynamics and preferences within the music community.

5 Tentative Plan

5.1 Data Collection

- Obtain Spotify user data using the Spotify App and the Spotify API through the `spotifyr` package in R.
- Collect data related to user listening history, favorite artists, genres, and other relevant information.

5.2 Data Preprocessing

- Clean the collected data to remove any inconsistencies or irrelevant information.
- Transform the data into a suitable format for analysis, ensuring compatibility.

5.3 Community Detection

- Apply community detection algorithms to identify groups or clusters of users with similar listening preferences.
- Utilize techniques to partition the Spotify user network into cohesive communities.

5.4 Signed Network Analysis

- Construct a signed network representation of Spotify user interactions, considering both positive (e.g., liked songs, artists) and negative (e.g., artists least listened) connections.
- Explore metrics such as balance and tension within the signed network to understand the nature of relationships between users.

5.5 Integration of Community Detection and Signed Networks

- Investigate the relationship between community structure and signed network properties.
- Analyze how communities within the Spotify user network are influenced by both positive and negative interactions.
- Identify influential users or bridges between communities based on their interactions within the signed network.

5.6 Visualization and Interpretation

- Visualize the results of analysis using appropriate graphical representations.
- Interpret the findings to uncover patterns of user behavior, community dynamics, and network structure within the Spotify ecosystem.
- Provide insights into how these findings can inform recommendations for personalized music discovery and user engagement strategies on the Spotify platform.

5.7 Documentation and Reporting

- Document the entire analysis process, including data sources, preprocessing steps, algorithm implementations, and results interpretation.
- Prepare a comprehensive report detailing the methodology, findings, and implications of the analysis for Spotify users, researchers, and platform developers.

5.8 Conclusion and Future Directions

- Summarize the key findings and contributions of the analysis in understanding Spotify user behavior and network dynamics.
- Discuss potential avenues for future research, including refining community detection techniques, exploring additional network properties, and applying to other social music platforms.

This tentative plan outlines a structured approach for analyzing Spotify user activity using R and SpotifyR, with a focus on community detection and signed network analysis. By integrating these techniques, the analysis aims to provide valuable insights into the underlying structure and dynamics of the Spotify user network, facilitating personalized music recommendations and enhancing user engagement on the platform.

References

[1] Leskovec, Jure, Daniel P. Huttenlocher and Jon M. Kleinberg. "Signed networks in social media." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2010): n. pag.

[2] Kunegis, Jérôme & Lommatzsch, Andreas & Bauckhage, Christian. (2009). The Slashdot Zoo: Mining a Social Network with Negative Edges. WWW'09-Proceedings of the 18th International World Wide Web Conference. 10.1145/1526709.1526809.

[3] Adamic, Lada A. and Eytan Adar. "How to search a social network." *Soc. Networks* 27 (2003): 187-203.

[4] Fortunato, Santo. "Community detection in graphs." *ArXiv* abs/0906.0612 (2009): n. pag.

[5] Brandes, Ulrik and Thomas Wilhelm Erlebach. "Network Analysis: Methodological Foundations." (2010).