Clustering My Spotify Playlist by Genre Using Network Analysis

Have you ever wondered what kinds of music you frequently listen to? I did, so I pulled data from one of my favorite Spotify playlists and clustered the tracks based on genre similarities. This project gave me new insight into my listening habits, and showed how network clustering methods can uncover patterns in music taste.

Research Question

What genre clusters exist within a user's playlist, and how can we visualize these genre clusters?

Stakeholder: Music enjoyers, musicians, or Spotify's recommendation team.
Use Case: Understanding listener profiles for playlist curation or recommendation algorithms.

The Data: My Spotify Playlist

I used the playlist with ID: "79mXvSd3YpbxeAztsIBz7z", which includes a diverse selection of my most played songs. From each track, I collected:

- track_name
- artist_name
- artist_id
- genres (via the primary artist)

Each track is tied to an artist, and each artist is associated with a set of genres on Spotify. These genres are the basis for clustering.

Data Collection

I used the Spotipy Python library and the Spotify API.

1. Fetched all track IDs from the playlist using sp.playlist_items().
2. For each track:

- Fetched the track details
- Fetched the primary artist's genre. (genre was not available for each song)

Measuring Similarity

Genres are multi-label categorical data, so I used MultiLabelBinarizer to convert genre lists into a binary feature matrix. Then I computed cosine distances between each track based on their genre vectors. This became the foundation for clustering the songs together.

Clustering Method & k Selection

I used Agglomerative Clustering from scikit-learn, which works with the precomputed distance matrices and can detect hierarchies in data. Since this was exploratory, I started with k = 4 to see if any meaningful genre groupings would emerge, which they did!

What the Clusters Revealed

Each cluster captured a distinct style or vibe. For example:

Cluster 0: Pop

- "Lose yourself to dance" - daft punk
- "Decaptacon" - Le tigre

Cluster 1: Hip-Hop

- "PRIDE - Kendrick Lamar
- "C.R.E.A.M" - Wu-Tang Clan

Cluster 2: Indie

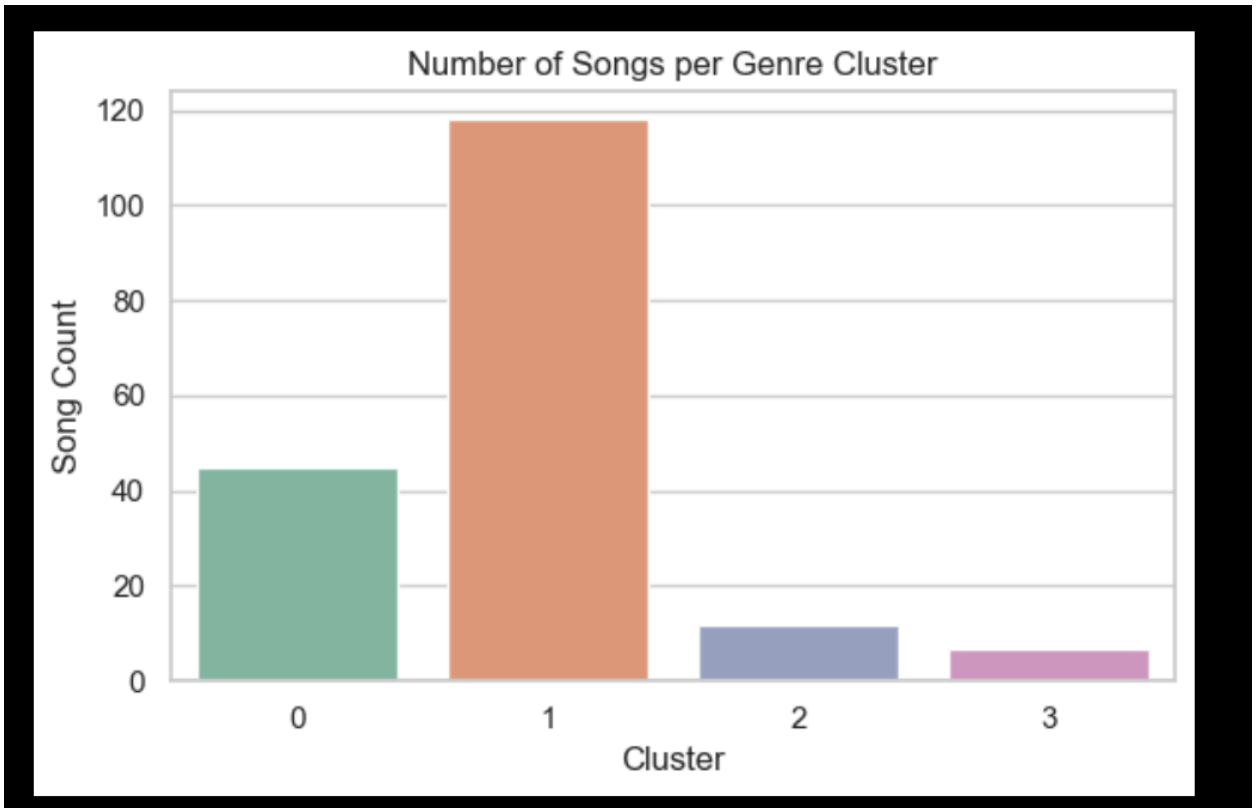
- "Eventually" - Tame Impala
- "My Kind of Woman" - Mac DeMarco

Cluster 3: Raggae

- “No Guns Allowed” - Snoop Lion
- “Could You Be Loved” - Bob Marley & the Wailers

Each cluster reflected real genre communities. This speaks to the validity of Spotify’s genre tags and the power of cosine similarity.

Visualizing the Results



Data Cleaning & Bugs

Common issues I ran into:

- Some tracks had no genres (typically for new or small artists).
- API limits: I used time.sleep(0.1) to avoid issues.
- Duplicate tracks had to be filtered using set().

Limitations

- Genres are only as good as Spotify's tagging. Many artists make music under multiple genres, and this can mess up the tagging.
- Clustering with a fixed k can oversimplify the diversity in music.
- Tracks with no genre were excluded.

Code & Reproducibility

This small project showed how network data + clustering can reveal hidden structure in something as personal as your music taste. With just a playlist and a few lines of code, you can uncover genre patterns or even recommend songs based on similar styles.

You can run the full project here