Group members:
Edoardo Bastianello: 2053077
Stefano Binotto: 2052421
Gionata Grotto: 2052418

Finding Significant Statistics And Relationships In Music Networks With Node Features

MOTIVATION:

- Figuring out which artists are most relevant, based on year and geographic area.
- Figuring out which music genres are most relevant, based on year and geographic area.
- Figuring out which songs are the most relevant, based on music genre, year, geographic area and other factors.

DATA:

For this project, we are going to use the following 2 datasets:

1. This dataset contains data of 896 music genres, 13880 songs and 3611 artists. In particular, this dataset contains:

- Artist Collaboration Networks: networks of collaborating artists obtained from Spotify Charts for each considered market and year. Artists are linked when they collaborated on a hit song (i.e. this song appears on a Spotify Top 200 Chart). These networks are subdivided by geographic areas (US, Japan, France, …) and years (2017/2019). Each network has about 1000 edges, while the number of nodes is up to 3000. The graph is undirected, and each edge has a weight corresponding to the number of hit songs in which both artists collaborated.

- Info about the artists and charts: data about the top artists and the hit songs according to Spotify Charts in all considered markets (2017-19).

- Genre Collaboration Networks: networks of music genres obtained from Spotify Charts for each considered market (US, Japan, ...) and year (2017-2018-2019). There is an edge between two genres when artists belonging to these genres collaborated on a hit song. The graph is undirected, and each edge has a weight corresponding to the number of hit songs which are a collaboration between artists from these genres. These graphs have between 1600 and 3000 edges each, while the number of nodes is around 850.

- Hit songs: useful statistics about all hit songs who made into Weekly Top 200 Spotify Charts in all considered markets (2017-19) (time signature, loudness, tempo, energy, key, release date). The total number of songs is 13880.

Link to the dataset: https://opgabriel.github.io/ISMIR2020/

2. As the second dataset, we are going to use MusicOSet, which is a dataset of musical elements (artists, songs and albums) based on musical popularity classification. This dataset includes:

- metadata: this dataset consists of textual and numeric information about songs, artists and albums (album ID, album name, total track of the album, artists, followers popularity, song type, …)
- popularity: this dataset consists of information about the popularity of each song, album and artist (the position of a song or album in the billboard or a score of popularity of an artist).
- song features: this dataset consists of acoustic fingerprints (duration, key, tempo, time_signature, …) collected directly from spotify and lyrics resources.

We are going to use this dataset in order to obtain additional information with respect to the first dataset.

Link to the dataset: https://marianaossilva.github.io/DSW2019/#tables

METHOD:

Problem: compute the centrality scores for all nodes in the graphs specified below.

Algorithms: we are going to try to use an exact algorithm for the centrality scores.

INTENDED EXPERIMENTS:
We are going to use the NetworkX implementation of the closeness centrality and betweenness centrality algorithms. These are exact implementations of the methods. If that doesn't not work, we will use the approximated methods.

Machine for experiments:
- CPU = i7-8750H 2.20GHz - 6 cores
- NVIDIA GeForce GTX 1080 Max-Q
- 16GB RAM

Experiments: compute the centrality scores of the following networks:

- Artist Collaboration Networks (described above), for different years and geographic areas;
- Genre Collaboration Networks (described above), for different years and geographic areas;
- other possible networks that we are going to create from the data, like a network where each song is a node and there is an edge between 2 nodes if the 2 songs share at least 1 artist.

REFERENCES

Link to the first dataset: https://opgabriel.github.io/ISMIR2020/

Link to the second dataset: https://marianaossilva.github.io/DSW2019/#tables