

Proposal_Lange_Sofia

October 31, 2019

1 The Group

Student Name: **Sofia Lange**

Group Members: **Srirachan Reddi Varra**

2 The Data

The dataset has been derived from: [EveryNoise](#)

We have attached the notebook which generated the dataset, it is titled EveryNoiseScraper.ipynb

```
In [4]: import pandas as pd
        df_all_genres = pd.read_csv('all_genres.csv')
        df_school_genres = pd.read_csv('school_genres.csv')
        df_all_genres.head()
```

```
Out[4]:
```

	Unnamed: 0	GENRE \		SIM_GENRES \		SIM_WEIGHTS \
0	0	backgroundmusic	0	['reiki', 'calminginstrumental', 'yoga', 'back...	0	['103', '120', '101', '240', '102', '125', '11...
1	1	classicmoroccanpop	1	['bluesrock', 'andalusianclassical', 'chaabima...	1	['103', '107', '109', '100', '102', '103', '10...
2	2	orthodoxchant	2	['universitychoir', 'polishchoir', 'ukrainianc...	2	['102', '105', '107', '104', '106', '240', '10...
3	3	tar	3	['oud', 'persiantraditional', 'bansuri', 'tar'...	3	['106', '120', '101', '240', '103', '113', '10...
4	4	reggae	4	['dancehall', 'brazilianreggae', 'skarevival',...	4	['109', '105', '114', '101', '100', '127', '11...

```

                                OPP_GENRES \
0 ['mahraganat', 'divahouse', 'oldschoolukhiphop...
1 ['shiverpop', 'gravewave', 'tassieindie', 'est...
2 ['sambass', 'funkybreaks', 'hardbass', 'hardco...
3 ['happyhardcore', 'aggrotech', 'nightcore', 'b...
4 ['brazilianlo-firock', 'martialindustrial', 'p...

```

```

                                OPP_WEIGHTS \
0 ['160', '100', '121', '110', '135', '136', '12...
1 ['112', '112', '105', '102', '124', '100', '16...
2 ['107', '110', '158', '100', '119', '111', '10...
3 ['109', '107', '127', '114', '115', '120', '12...
4 ['106', '107', '155', '121', '160', '108', '14...

```

```

                                REL_ARTISTS \
0 ['NoemiNucci', 'JuditheAbelsen', 'FabienneWill...
1 ['HajHusseinToulali', 'AbdelhadiBelkheyat', 'H...
2 ['LegeArtis', 'PeterMichaelides', 'CapellaGreg...
3 ['KiyarashSaket', 'YusefForutan', 'HamidMoteba...
4 ['TheItals', 'MykalRose', 'JohnnieClark', 'Asw...

```

```

                                ARTIST_WEIGHTS \
0 ['113', '117', '106', '131', '109', '115', '11...
1 ['102', '103', '103', '108', '108', '102', '10...
2 ['114', '100', '160', '127', '115', '115', '10...
3 ['100', '100', '111', '100', '100', '100', '10...
4 ['101', '102', '100', '105', '101', '105', '10...

```

```

                                SPOTIFY_URL
0 https://open.spotify.com/user/thesoundsofspoti...
1 https://open.spotify.com/user/thesoundsofspoti...
2 https://open.spotify.com/user/thesoundsofspoti...
3 https://open.spotify.com/user/thesoundsofspoti...
4 https://open.spotify.com/user/thesoundsofspoti...

```

In [2]: df_school_genres.head()

```

Out[2]:   Unnamed: 0                                SCHOOL \
0         0  Texas A & M University-College Station
1         1    The University of Texas at Austin
2         2    Ohio State University-Main Campus
3         3    University of Central Florida
4         4    Arizona State University-Tempe

```

```

                                GENRES \
0 ['reddirt', 'outlawcountry', 'texascountry', '...
1 ['modernrock', 'indiepop', 'pop', 'indieoptim...

```

```

2 ['pop', 'indieoptimism', 'modernrock', 'indie...
3 ['pop', 'poprap', 'rap', 'hiphop', 'electrohou...
4 ['edm', 'electrohouse', 'pop', 'electronictrap...

                                PLAYLIST_LINK
0 https://embed.spotify.com/?uri=spotify:playlis...
1 https://embed.spotify.com/?uri=spotify:playlis...
2 https://embed.spotify.com/?uri=spotify:playlis...
3 https://embed.spotify.com/?uri=spotify:playlis...
4 https://embed.spotify.com/?uri=spotify:playlis...

```

3 The Project

3.1 Questions of Interest

1. What is the community structure of subgenres across Spotify?
2. How are communities assembled with various community detection / formation algorithms such as AGM, BigCLAM, Spectral Graph Partitioning, Graph Laplacian and Modularity Maximization?
3. In addition while community detection algorithms can show relationships between genres, how would other metric / metadata based methods find communities?
4. How are artists distributed across the subgenre graph?
5. What are their primary features that categorize a subgenre?

3.2 Description of everynoise.com, and the dataset we scraped from there.

The website is a collection of 3,652 (and increasing) subgenres of music across Spotify along with several other interesting lists such as one for the top genres of over 2000 universities across the globe. Each genre links to a page where the user can open a playlist for that genre through Spotify and listen to example tracks that characterize that genre. The website expands to Spotify playlists themselves, where their API can be used to gather information about the playlist and audio features of the songs themselves.

Each subgenre has a table of related subgenres, as well as a table of so-called opposite or distant subgenres. Also available are example artists whose work heavily fits into the genre.

We used python and BeautifulSoup to scrape our initial dataset from this website. It includes, for each subgenre, the related subgenres, the subgenres which are deemed the "farthest away" from them, as well as that subgenres related artists. Furthermore, we collected the size of the elements referring to these entities because the website sizes elements according to how related they are (or not related), to possibly use as weights for the edges between genres. This will provide an interesting weighted graph.

We also generated the dataset called "school_genres" using the every school add on at <http://everynoise.com/every-school.cgi>. We got all schools in the USA and the most popular genres for that school, and we also collected the links to the related spotify playlists for that school in order to gather data about music features.

3.3 Methods of Analysis

Finding community structure can be done with several algorithms such as AGM, BigCLAM, Spectral Graph Partitioning, Graph Laplacian, and Modularity Maximization. In addition to this each song has a set of audio features $\{key, mode, time - signature, acousticness, etc..\}$. With this set of features we can construct a vector for each song and get averages for the subgenre playlists. Genres could then be generated using clustering algorithms like as GMM and K-Means.

In addition to the exploration of music subgenres *everynoise* provides a list of the top genres for each over 2000 universities. This can be analyzed by using graphical orderings/embeddings to find the ordering of top genres of across universities in United States. This ordering can be constructed with page rank and the SpringRank Hamiltonian. The hypothesis is that the different community detection algorithms will in general find the majority of same communities with only little outliers and edge cases that will vary between each of the 5 methods. The analysis will use Python and specifically pre-constructed network packages such as NetworkX and ideally graph-tool (for it's performance).