

2021 ICM

Problem D: The Influence of Music

Music has been part of human societies since the beginning of time as an essential component of cultural heritage. As part of an effort to understand the role music has played in the collective human experience, we have been asked to develop a method to quantify musical evolution. There are many factors that can influence artists when they create a new piece of music, including their innate ingenuity, current social or political events, access to new instruments or tools, or other personal experiences. Our goal is to understand and measure the influence of previously produced music on new music and musical artists.

Some artists can list a dozen or more other artists who they say influenced their own musical work. It has also been suggested that influence can be measured by the degree of similarity between song characteristics, such as structure, rhythm, or lyrics. There are sometimes revolutionary shifts in music, offering new sounds or tempos, such as when a new genre emerges, or there is a reinvention of an existing genre (e.g. classical, pop/rock, jazz, etc.). This can be due to a sequence of small changes, a cooperative effort of artists, a series of influential artists, or a shift within society.

Many songs have similar sounds, and many artists have contributed to major shifts in a musical genre. Sometimes these shifts are due to one artist influencing another. Sometimes it is a change that emerges in response to external events (such as major world events or technological advances). By considering networks of songs and their musical characteristics, we can begin to capture the influence that musical artists have on each other. And, perhaps, we can also gain a better understanding of how music evolves through societies over time.

Your team has been identified by the **Integrative Collective Music (ICM) Society** to develop a model that measures musical influence. This problem asks you to examine evolutionary and revolutionary trends of artists and genres. To do this, your team has been given several data sets by the ICM:

- 1) “*influence_data*”¹ represents musical influencers and followers, as reported by the artists themselves, as well as the opinions of industry experts. These data contains influencers and followers for 5,854 artists in the last 90 years.
- 2) “*full_music_data*”² provides 16 variable entries, including musical features such as *danceability*, *tempo*, *loudness*, and *key*, along with *artist_name* and *artist_id* for each of 98,340 songs. These data are used to create two summary data sets, including:
 - a. mean values by artist “*data_by_artist*”,
 - b. means across years “*data_by_year*”.

¹ These data were scraped from AllMusic.com

² These data were obtained from Spotify’s API

Note: DATA provided in these files are a subset of larger data sets. These files **CONTAIN THE ONLY DATA YOU SHOULD USE FOR THIS PROBLEM.**

To carry out this challenging project, the ICM Society asks your teams to explore the evolution of music through the influence across musical artists over time, by doing the following:

- Use the *influence_data* data set or portions of it to create a (multiple) directed network(s) of musical influence, where influencers are connected to followers. Develop parameters that capture ‘*music influence*’ in this network. Explore a subset of musical influence by creating a subnetwork of your directed influencer network. Describe this subnetwork. What do your ‘*music influence*’ measures reveal in this subnetwork?
- Use *full_music_data* and/or the two summary data sets (with artists and years) of music characteristics, to develop measures of music similarity. Using your measure, are artists within genre more similar than artists between genres?
- Compare similarities and influences between and within genres. What distinguishes a genre and how do genres change over time? Are some genres related to others?
- Indicate whether the similarity data, as reported in the *data_influence* data set, suggest that the identified influencers in fact influence the respective artists. Do the ‘influencers’ actually affect the music created by the followers? Are some music characteristics more ‘contagious’ than others, or do they all have similar roles in influencing a particular artist’s music?
- Identify if there are characteristics that might signify revolutions (major leaps) in musical evolution from these data? What artists represent revolutionaries (influencers of major change) in your network?
- Analyze the influence processes of musical evolution that occurred over time in one genre. Can your team identify indicators that reveal the dynamic influencers, and explain how the genre(s) or artist(s) changed over time?
- How does your work express information about cultural influence of music in time or circumstances? Alternatively, how can the effects of social, political or technological changes (such as the internet) be identified within the network?

Write a **one-page document** to the ICM Society about the value of using your approach to understanding the influence of music through networks. Considering the two problem data sets were limited to only some genres, and subsequently to those artists common to both data sets, how would your work or solutions change with more or richer data? Recommend further study of music and its effect on culture.

The ICM Society, an interdisciplinary and diverse group from the fields of music, history, social science, technology, and mathematics, looks forward to your final report.

Your PDF solution of no more than 25 total pages should include:

- One-page Summary Sheet.
- Table of Contents.
- Your complete solution.
- One-page document to ICM society.
- References list.

Note: New for 2021! The ICM Contest now has a **25-page limit**. All aspects of your submission count toward the 25-page limit: Summary Sheet, Table of Contents, Main Body of Solution, Images and Tables, One-page Document, Reference List, and any Appendices.

Attachments

We provide the following four data files for this problem. THE DATA FILES PROVIDED CONTAIN THE ONLY DATA YOU SHOULD USE FOR THIS PROBLEM.

1. [influence_data.csv](#)
2. [full_music_data.csv](#)
3. [data_by_artist.csv](#)
4. [data_by_year.csv](#)

Data Descriptions

1. [influence_data.csv](#)

(Data is encoded in utf-8 to allow for handling of special characters):

- **influencer_id:** A unique identification number given to the person listed as influencer. (string of digits)
- **influencer_name:** The name of the influencing artist as given by the follower or industry experts. (string)
- **influencer_main_genre:** The genre that best describes the bulk of the music produced by the influencing artist. (if available) (string)
- **influencer_active_start:** The decade that the influencing artist began their music career. (integer)
- **follower_id:** A unique identification number given to the artist listed as follower. (string of digits)
- **follower_name:** The name of the artist following an influencing artist. (string)
- **follower_main_genre:** The genre that best describes the bulk of the music produced by the following artist. (if available) (string)
- **follower_active_start:** The decade that the following artist began their music career. (integer)

2. [full_music_data.csv](#)

3. [data_by_artist.csv](#)

4. [data_by_year.csv](#)

Spotify audio features from the “**full_music_data**”, “**data_by_artist**”, “**data_by_year**”:

- **artist_name**: The artist who performed the track. (array)
- **artist_id**: The same unique identification number given in the influence_data.csv file. (string of digits)

Characteristics of the music:

- **danceability**: A measure of how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable. (float)
- **energy**: A measure representing a perception of intensity and activity. A value of 0.0 is least intense/energetic and 1.0 is most intense/energetic. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy. (float)
- **valence**: A measure describing the musical positiveness conveyed by a track. A value of 0.0 is most negative and 1.0 is most positive. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry). (float)
- **tempo**: The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration. (float)
- **loudness**: The overall loudness of a track in decibels (dB). Values typical range between -60 and 0 db. Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). (float)
- **mode**: An indication of modality (major or minor), the type of scale from which its melodic content is derived, of a track. Major is represented by 1 and minor is 0.
- **key**: The estimated overall key of the track. Integers map to pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C#/D♭, 2 = D, and so on. If no key was detected, the value for key is -1. (integer)

Type of vocals:

- **acousticness**: A confidence measure of whether the track is acoustic (without technology enhancements or electrical amplification). A value of 1.0 represents high confidence the track is acoustic. (float)
- **instrumentalness**: Predicts whether a track contains no vocals. “Ooh” and “aah” sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly “vocal”. The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0. (float)
- **liveness**: Detects the presence of an audience in a track. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live. (float)

- **speechiness**: Detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks. (float)
- **explicit**: Detects explicit lyrics in a track (true (1) = yes it does; false (0) = no it does not OR unknown). (Boolean)

Description:

- **duration_ms**: The duration of the track in milliseconds. (integer)
- **popularity**: The popularity of the track. The value will be between 0 and 100, with 100 being the most popular. The popularity is calculated by algorithm and is based, in the most part, on the total number of plays the track has had and how recent those plays are. Generally speaking, songs that are being played more frequently now will have a higher popularity than songs that were played more frequently in the past. Duplicate tracks (e.g. the same track from a single and an album) are rated independently. Artist and album popularity are derived mathematically from track popularity. (integer)
- **year**: The year of release of a track. (integer from 1921 to 2020)
- **release_date**: The calendar date of release of a track mostly in yyyy-mm-dd format, however precision of date may vary and some just given as yyyy.
- **song_title (censored)**: The name of the track. (string) Software was run to remove any potential explicit words in the song title.
- **count**: The number of songs a particular artist is represented in the full_music_data.csv file. (integer)