ISSN 2618-1568 Vol. 3, Issue 2: 45-50, DOI: 10.25236/FAR.2021.030209

The Evaluation of Musical Influence and Evolution

Liuyuan Jiang

Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, 215000, China

Abstract: *Music, which has evolved for thousands of years, plays an important role in social culture. To understand musical evolution, we need to measure the influence of previous music on new music and examine the revolutionary shift over the past years. We prepare the sets of artists, songs and genres as 3 basic types of music carriers. Together with music features and time, we have several 3-dimentional attribute matrices describing the musical properties during different time spans for artists and songs. Artists, songs and genres also have pairwise embeddings and they have internal one-to-one influencer-to-follower relationships.*

Keywords: *similarity measure, classification model, K-means clustering algorithm*

1. Background

Music occupies a pivotal position in the cultural heritage. Many factors can affect the process of making music such as artists' innate ingenuity, current social or political environment or other artists. Music is constantly changing because of both internal and external influence. We will develop a model to measure musical influence and also study the evolutionary and revolutionary trends of artists and genres.

2. Influence measurement model

2.1 Direction network and quantitative influence indicators

To conduct quantitative analysis of musical influence, we first need to figure out the influence network. In "influence_data.csv", there are 42765 entries of influencer- follower data with respective year and genre. To measure the influence, we need to (i) land the directed inter-artist relationship network(see figure 1), (ii) find some indicators to quantitatively measure the influence at each node, (iii) take time into consideration.

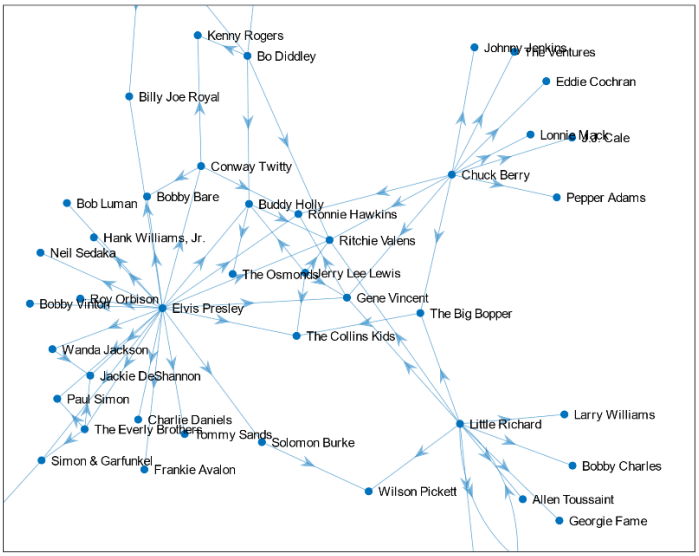

Figure 1: partial sum of direction network

We apply MATLAB coding (see Appendix) thus landing a influencer-to-follower relationship matrix

ISSN 2618-1568 Vol. 3, Issue 2: 45-50, DOI: 10.25236/FAR.2021.030209

$R \in \mathbb{R}^{4810\times4810\times9}$ where $r_{ijk} = n$ representing artist $i \in A$ at time period $k \in K_{decade}$ influenced artist $j \in A$ at $n-1$ decades later ($n = 1$ means follower and influencer are contemporaries), $r_{ijk} = 0$ if no connection.

In this way, we can land the (i) direction network by detecting $r_{ijk} \neq 0$, (ii) **Number of Followers** during period k as indicator by

$$NOF_{ik} = \Sigma_{j\in J}\tilde{r}_{ijk}$$

Where $\forall r_{ijk} \neq 0,\ \tilde{r}_{ijk} = 1$ and **Influence Existing Time** as indicator by

$$IET_{ik} = 10 \times max_{j\in A}(r_{ijk} - 1).$$

Note that (iii) time is taken considered in since both indicators present influencer's influence during specific time period k. This enable us to compare the influence of times in later pages.

To simplify the record by ignoring (iii), we have $NOF_i = NOF_{ik_{k\in K_{decade}}}$, $IET_{ik} = IET_{ik_{k\in K_{decade}}}$.

Part of the NOF ranking is shown in Table 1.

Table 1: The rank of the numbers of the followers

rank	artist_name	followers	Rank	artist_name	followers
1	The Beatles	614	9	Hank Williams	184
2	Bob Dylan	389	10	The Velvet Underground	181
3	The Rolling Stones	319	11	Black Sabbath	171
4	David Bowie	238	12	Marvin Gaye	169
5	Led Zeppelin	221	13	Elvis Presley	166
6	Jimi Hendrix	201	14	Miles Davis	160
7	The Kinks	191	15	Chuck Berry	159
8	The Beach Boys	185	16	The Byrds	158

2.2 Influencer: The Beatles

With regards to these 2 indicators, we find *The Beatles* tops both ranks with $NOF = 610$ and $IET = 60$.

The Beatles was an English rock band was officially formed in 1960 by John Lennon, Paul McCartney, George Harrison and Ringo Starr. [6] They had their great success during the 1960s and 1970s with unmatched influence their in the 60's. As illustrated by its influence over time measured by IET_{ik}, we can see that the Beatles had greatest followers in the 1960s.

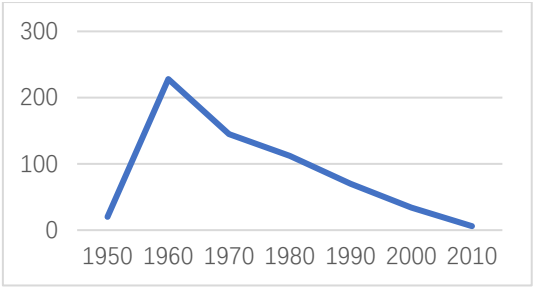

Figure 2: Number of followers during decade

This band's musical style is derived from 1950s rock music. Elvis Presley, as illustrated in direction Figure 2, is one typical influencer to the Beatles from the 1950s. He is the 13th in the grand rank and he tops the $NOF_{ik=3}$ rank as the most influential artist in the 1950s. The Beatles is also influenced by his contemporary Bob Dyla, another seminal artist who in the grand rank second only to the Beatles. However, the band's followers has limited proportion of big, yet its influence can never be taken lightly given the number 610 in total and its influence till now.

ISSN 2618-1568 Vol. 3, Issue 2: 45-50, DOI: 10.25236/FAR.2021.030209

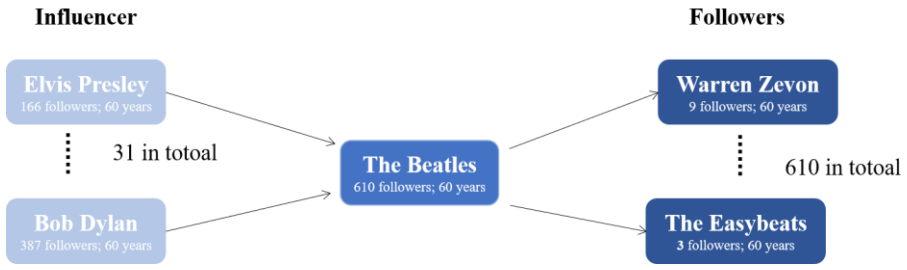

Figure 3: Subnetwork

3. Similarity between and within genres

To compare the music between or within music carriers, say artists, genres or songs, we need to clarify what represent such carrier. Spotify audio features can be used as quantitative pivotal measurement for presentation. We thus take Spotify audio features as critical quantifications. Matrices $S^{I\times J}$ and $Q^{I\times J}$ are thus prepared for similarity analysis.

However, in the available dataset "data_by_artist.csv" and "full_music_data.csv", genres are not concluded. Although "influence_data.csv" provides some embedding between artists and genres, we still don't have complete information. Thus, we use multiple logistic regression based on record for known-genre artists and their music features, predicting the genre for those artists with unknown genres. Before running regression, we conduct factor analysis to reduce number of variables to reduce the negative impact caused by imperfect multi-co-linearity between variables.

3.1 Factor analysis

Factor analysis reduces the complex relationship between variables into a few comprehensive factors by studying the correlation coefficient matrix between variables. This is a typical method for dimensionality reduction, through which we can avoid multi-co-linearity between variables. It is credit for its interpretability compared with other akin method such as principal component analysis.

We can find 13 factors/characteristics of music from "full_music_data.csv". Since "duration" and "popularity" is not very suitable to measure the similarity, then we only choose the leaving 11 factors. SPSS is used for all the following procedures.

Before conducting factor analysis, we need to test whether our data is suitable to this method. **KMO test** and **Bartlett spherical test** are conducted.

KMO test tests the relative sizes of simple correlation coefficients and partial correlation coefficients between original variables, and is mainly applied to factor analysis of multivariate statistics like our data. Bartlett spherical test is a test method to test the degree of correlation between various variables.

We take Kaiser's KMO test standard as in table 2 and we land both our test outcomes in table. We can see both p-value for Matlett's test are nearly 0.000 which means factor analysis is suitable and the KMO=0.773 means suitability is general.

Table 2: KMO test standard

KMO > 0.9	**0.8 < KMO < 0.9**	**0.7 < KMO < 0.8**	**0.6 < KMO < 0.7**	**KMO < 0.5**
Very suitable	suitable	general	Not very suitable	Not suitable

Table 3: KMO and Matlett's test for sphericity

KMO sampling fitness measure		.773
Matlett's test for sphericity	The approximate chi-square	23492.291
	Degrees of freedom	78
	significant	.000

Now we conduct the test. Then, we can obtain the following discriminant condition.

Assume there are n samples and p factors, then we can construct an $n\times p$ sample matrix

ISSN 2618-1568 Vol. 3, Issue 2: 45-50, DOI: 10.25236/FAR.2021.030209

$x = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix}$. We also assume the mean of the matrix x is $u = (u_1, u_2, \cdots, u_p)'$. Therefore, the general model of factor analysis is

$$\begin{cases} x_1 = u_1 + a_{11} f_1 + a_{12} f_2 + \cdots + a_{1m} f_m + \varepsilon_1 \\ x_2 = u_2 + a_{21} f_1 + a_{22} f_2 + \cdots + a_{2m} f_m + \varepsilon_2 \\ \vdots \\ x_p = u_p + a_{p1} f_1 + a_{p2} f_2 + \cdots + a_{pm} f_m + \varepsilon_p \end{cases} \quad (1)$$

Where $f_1, f_2, \ldots, f_m$ are common factors and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ are special factors.

Table 4: Factor loading matrix

	1	2	3	4	5
danceability	.047	.875	-.078	.217	-.064
energy	.884	.070	.184	-.096	.079
valence	.067	.797	-.008	-.203	.120
tempo	.395	-.161	.005	-.403	.259
loudness	.848	.181	.091	-.135	.010
mode	-.071	-.067	-.024	-.643	-.257
key	.013	.033	-.010	.087	.908
acousticness	-.873	-.054	.031	-.055	-.055
instrumentalness	-.275	-.468	-.140	.800	.061
liveness	.054	-.138	.803	-.120	.012
speechiness	.003	.129	.798	.172	-.020

From Table 4, we can lower the dimension of 11 factors to 5 factors. We use the most related original factor to explain the new factor and name the 5 factors energy2, danceability2, liveness2, instrumentalness2 and key2 respectively. Additionally, SPSS gives us the score of the 5 new factors for each artist, and thus we create a new artist attribute matrix Q_{ij}' with i denoting the artist and j denoting 5 new features.

3.2 Classification model

Logistic regression is such a process: in the face of a regression or classification problem, the cost function is established, and then the optimal model parameters are solved iteratively through the optimization method, and then the quality of the model we solve is tested and verified. Logistic method is mainly used to study the probability of occurrence of certain events.

We move the genre of each artist in “influence_data.csv” into “data_by_artist.csv”. To deal with artists without genre information, we use **multiple logistic regression** to classify these artists into the genres that they are most likely to belong to. The prediction accuracy of our model is 85.5%.

We also use **neural network** to calculate, and the results are basically similar to the results obtained by multiple logistic regression, and the accuracy is 82.8%.

The above two methods both confirm our results are relatively accurate.

We then obtain tables with all artists and songs have corresponding genres.

3.3 Genre-genre analysis: Feature similarity model

Distance is used to measure the distance existing in space between individuals. The longer the distance, the greater the difference between individuals. [3]

Euclidean distance is the most common measure of distance, which measures the absolute distance between points in multidimensional space. The formula is as follows:

ISSN 2618-1568 Vol. 3, Issue 2: 45-50, DOI: 10.25236/FAR.2021.030209

$$D = dist(X, Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{2}$$

We first need to prepare some matrices.

●We have an attribute matrix $S^{I\times J}$, with s_{ij} represents /the i^{th} song's j^{th} feature, j=1,…,11 is the subset of *J*(|*J*|=13) which takes *danceability, energy, valence, tempo, loudness, mode, key, acousticness, instrumentalness, liveness, speechiness.*

●We also have a matrix $Q^{I\times J}$,with q_{ij} means the i^{th} artist's j^{th} factor.

●We then have a matrix $P^{I\times J}$,with p_{ij} means the i^{th} genre's j^{th} factor. Each value of J is obtained from average of song records within the genre *i*.

●After **standardization** and **factor analysis**, 11 factors were reduced to 5 factors. Therefore, the above two matrix become $Q_{ij}^{'}$ and $J_{ij}^{'}$ with $j = 5$.

We choose an artist a and assume his corresponding genre is b. Then, Q_{aj} is the attribute matrix of the artist a and P_{bj} is the attribute matrix of genre b.

Then, we calculate the distance within genre, denoted as $D_1 = dist(Q_{aj}, P_{bj})$. We also calculate the distance between genres, denoted as

$$D_2 = \frac{\sum_{i=1}^{18} dist(Q_{aj}, P_{bj})}{17}, i \neq b \tag{3}$$

Using the data above, we can measure the similarity of artists within and between genres, and get the following Figure 4.

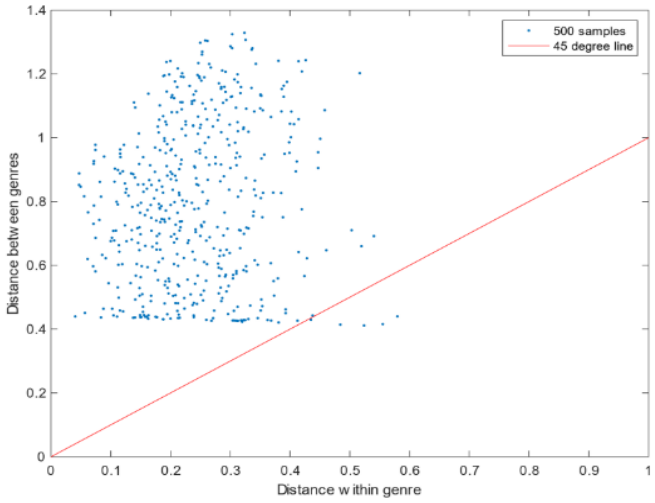

Figure 4: Distance within and between genres

We can clearly see that the most points lie above the 45° line. Then the distance between genres is smaller than distance within genre. It is clear that artists within genre are more similar than artists between genres.

4. Sensitivity Analysis

In the previous calculation, we chose the most common Euclidean distance to measure the similarity. We now change the measurement method to Chebyshev distance to explore the differences in the results. We plot the similarity between the schools using two methods respectively (see figure 5). We compare and find that the results are pretty much the same. This indicates that our model is not sensitive to the change of the parameter of distance,

ISSN 2618-1568 Vol. 3, Issue 2: 45-50, DOI: 10.25236/FAR.2021.030209

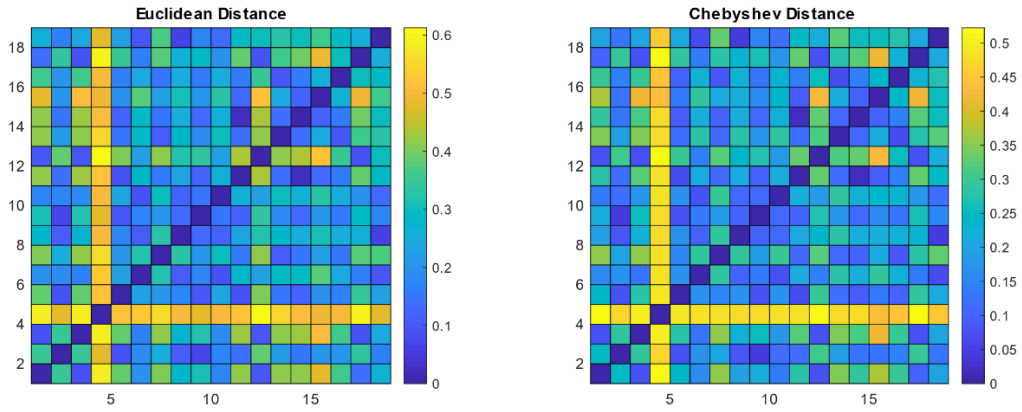

Figure 5: Comparison between Euclidean and Chebyshev Distance

5. Strengths and Weaknesses

The three dimensional matrix we build clearly integrate the data together, and it is easy to retrieve the data in the later calculation. We map all the associated matrices one by one to build a directed network. We quantify our results and present them visually in part. But There is a lack of detailed analysis and research on specific musical attributes.

References

[1] Le Wu, Peijie Sun, Yanjie Fu, Richang Hong, Xiting Wang, and Meng Wang, July 21-25 2019. A Neural Influence Diffusion Model for Social Recommendation. https://arxiv.org/abs/1904.10322
[2] Dan Shi, 2013. Research on Music Genre Similarity Detection Algorithm.
[3] Dong Wang, 2018. Euclidean distance and similarity in machine learning. https://blog.csdn.net/ wangdong2017/ article/details/81302799
[4] James H.Stock, Mark W.Waston, 2020. Introduction to Econometrics.
[5] Sitan Yang, 2018. The rise of American rock and roll and its social influence.
[6] Yi Wang, 2017. The Beatles. https://baike.so.com/doc/2493243-2634820.html.