Art and Performance Letters (2021) 2: 52-57
Clausius Scientific Press, Canada

DOI: 10.23977/artpl.2021.22009
ISSN 2523-5877

Research on Evaluation of Music-about Artist Influence, Genres and Characteristics

Qianqian Xu

School of Management, Hainan University, Haikou, Hainan, 570228

Keywords: Graph association network, Music characteristics, K-means cluster analysis

Abstract: Music evolves among the development of human society. In this paper, we first transformed the influencing information to a reflection association mapping network and determined two basic indicators of the network. Then we established a TOPSIS music influence evaluation model based on the entropy method for the sub-network. We compared influencers' influences respectively and revealed people's preference for music characteristics of the era. Second, we provided methods to measure music similarity. Dimensionality of music feature indicators are reduced with PCA. Then cosine similarity method was used to establish a multi-index music similarity measurement model. To compare the similarity of artists within and between genres, K-means cluster analysis was used to re-divide artists into 19 music styles we did cross comparison with the original genres and compared different artists within the genre. It leads to the conclusion that artists within genres are more similar than artists between genres.

1. Introduction

In all ages, music has been an indispensable part of people's life, but also has a pivotal position in the cultural field. It has been constantly evolving, forming many different genres and styles, such as pop, rock and roll, electronic music, etc [1]. The genres have their own characteristics in the structure, rhythm, lyrics and other aspects of the song.

2. Directional Music Influence Network

2.1 Generating Graph Network

The given data set *influence_data* shows the influencer and follower information of 5854 artists from 1930 to 2010. The following processing is based on the data after data cleaning. They reflect the influence of the artists and their followers but the relationship is ambiguous. Thus, we trace the source of the data, that is, establish a connection relationship of "influencer-follower " to form a tree diagram with multiple levels, and the Initial Nodes (artists who only affect others but are not affected) are arranged in chronological order. Due to the large amount, only part of the data are shown in Figure 1.

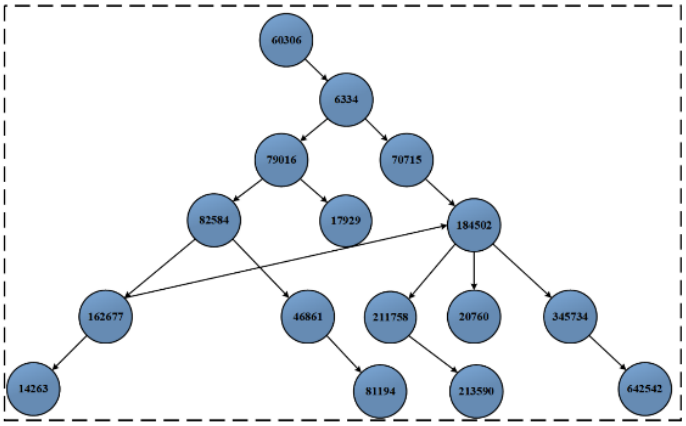

Figure 1. Part of the graph network

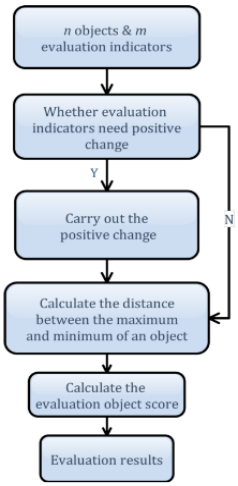

Figure 3. The flow chart of TOPSIS

2.2 Subnetwork Parameters

Figure 1 reflects the influence relationship between artists. We take Jazz as and example to show association relationship of artists that we provide a Figure 2, where the numbers on the circles are IDs. In this figure, the more direct followers and influencers an artist have, the darker the circle is. Through analysis, we contract two sub-network parameters: the total number of people in each genre in a certain year and the number of influential artists of each genre in a certain year.

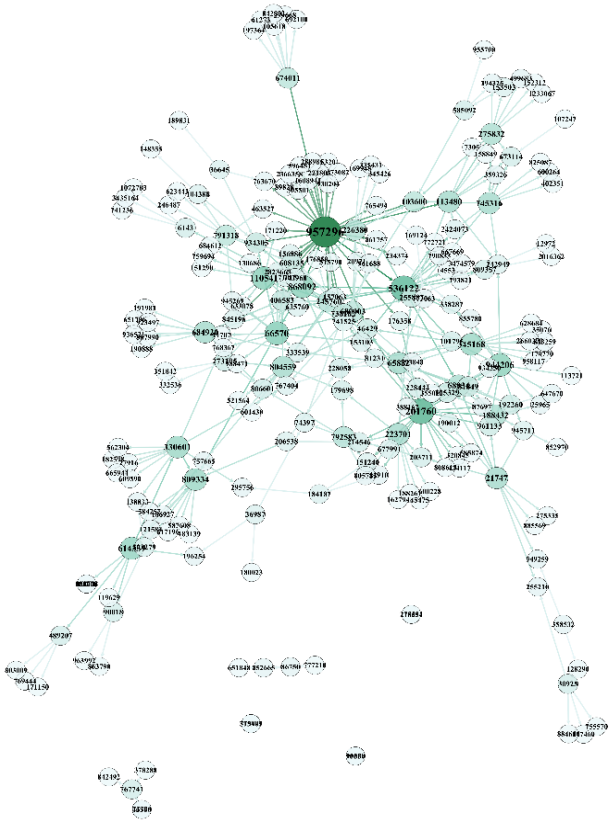

Figure 2. Association Degree of Jazz Artists

2.3 Music Influence Model

Now we are going to develop "Music Influence" model by using entropy method to determine the weight coefficient of each indicator through entropy weight method and then evaluate them with TOPSIS. Finally, rank the score of all artist to reflect their influence.

2.3.1 Determine the Weight of Each Indicator

Entropy method is an objective method of value assignment which has a wide range of applications in engineering technology, social economy and other fields. The basic idea is to determine the objective weight based on the variability of indicators. Its main steps are listed below:

1) Data Standardization: $\boldsymbol{Y_{ij}} = \frac{\boldsymbol{X_{ij} - min(X_i)}}{\boldsymbol{max(X_i) - min(X_i)}}$.

2) Solve the information entropy of each index: $\boldsymbol{E_j} = -\mathbf{ln}(\boldsymbol{n})^{-1} \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{p_{ij}} \mathbf{ln}\, \boldsymbol{p_{ij}} (\boldsymbol{p_{ij}} = \frac{\boldsymbol{Y_{ij}}}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{Y_{ij}}})$.

3) Get the Weight of Each Indicator: $\boldsymbol{W_i} = \frac{\boldsymbol{1 - E_i}}{\boldsymbol{k - \sum E_i}}$.

2.3.2 Confirm the Indicators of Entropy Weight Method

According to the data and information given, we construct five indicators shown in Table 1.

Table 1. The indicators and corresponding meanings

Serial	The indicator	The meaning of the indicator
1	The number of followers of an artist	The number of followers of artists in a certain field in a certain year
2	The total number of people in the genre	The number of all artists in the record range
3	Number of people in a certain year in the field	The total number of artists in a certain field in a certain year
4	The total follower number of an artist	The number of followers of an artist from 1930 to 2010
5	Competitiveness of the year	The total number of artists in all fields in this year

Indicator 5, namely "Competitiveness of the year" remains the same for all artists, so we only calculate indicators 1-4. After calculating with MATLAB, the results, namely entropy method index weights of indexes 1-4 are 0.0671, 0.1728, 0.3627 and 0.3974.

2.3.3 Evaluation through TOPSIS

TOPSIS ranks a limited number of evaluators based on how close they are to the desired goal. It is a commonly used effective method in multi-objective decision analysis, and it is also called the distance method of superior and inferior solutions. The processes of TOPSIS are shown in Figure 2.

2.3.4 The Result of Music Influence Model

Provided with the index weight and five indicators constructed, we get the result of TOPSIS through MATLAB and we present it by ploting a comprehensive coefficient heat map(Figure 4). The horizontal axis represents the IDs of 18 artists, who are randomly taken from 18 genres as representatives.

From the graph, we can see an obvious difference of influence between those genres. For exam- ple, ID "66915" (representing POP&Rock) has the highest influence at 1, followed by ID"614339"(representing Jazz) and"774956" (representing R&B), whose indexs are both 0.98.

In this way, our model is able to clearly declare the level of music influence of various genres.

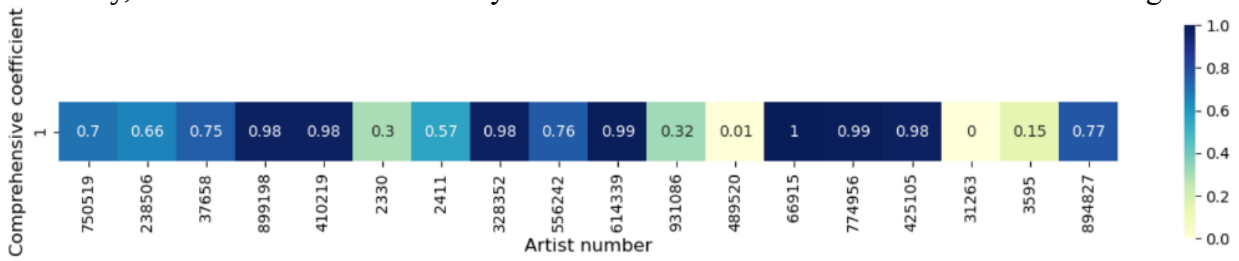

Figure 4. Comprehensive coefficients of genre representitives

3. Music Similarity Measurement Model

3.1 Data Processing

In this part, we use data from table *influence_data* and *full_music_data* we first process the data. We need adequate information for analysis and notice that not all artists in *full_music_data* appear in *influence_data*, so we compare the two tables to remove those artists with Python and SQL.

3.2 Principal Component Analysis

The data provided include 16 musical characteristic indicators, which makes the problem relatively complex. To measure the similarity between music, we need to reduce dimensionality and describe musical features with fewer indicators. Dimension reduction is to retain some of the most important features of high-dimensional data and remove noise and unimportant features. Principal component analysis (PCA) is used here. It is a statistical analysis method that divides the original variables into a few comprehensive indicators, which can be used for dimensionality reduction.

The indicators we get after dimension reduction are: danceability, energy, valence, tempo, loudness, mode, key and acousticness. The number of indicators reduced from sixteen to eight.

3.3 Similarity Evaluation with Cosine Similarity

We choose to evaluate the level of music similarity with cosine similarity.

Cosine similarity, also known as cosine similarity, evaluates the similarity of two vectors by calculating the cosine of the angle between them. The smaller the value of the similarity measure, the smaller the similarity between individuals, and the larger the similarity value, the greater the individual difference.

Multidimensional space cosine function formula is: $\boldsymbol{\cos(\theta)} = \frac{\sum_{i=1}^{n}(x_i \times y_i)}{\sqrt{\sum_{i=1}^{n}(x_j)^2} \times \sqrt{\sum_{i=1}^{n}(y_i)^2}} = \frac{\boldsymbol{a \cdot b}}{\|\boldsymbol{a}\|\boldsymbol{x}\|\boldsymbol{b}\|}$.

Using the fomula, we are able to find the similarity index between every two songs, ranging from 0 to 1. We illustrate part of them in figure 5.

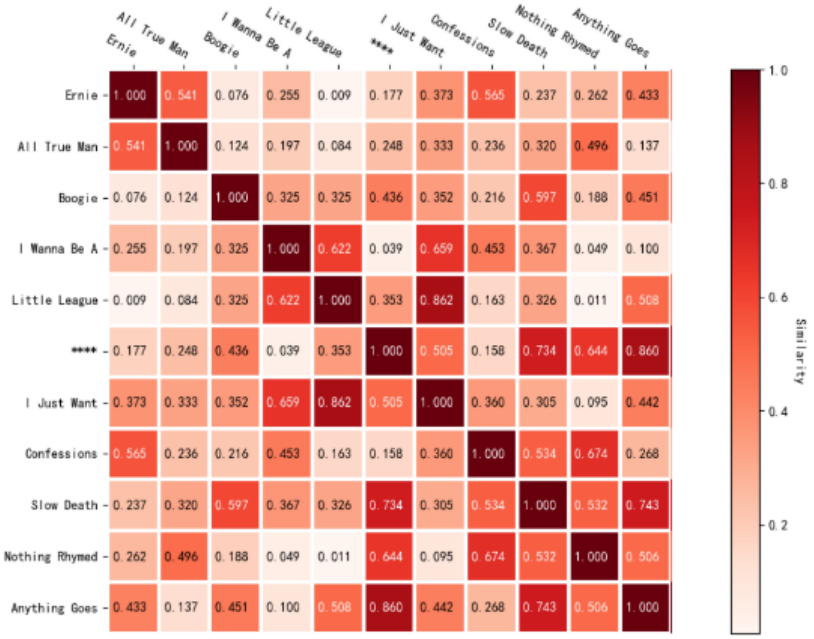

Figure 5. Music similarity between songs

3.4 Multi-scale K-Means Clustering

The K-Means algorithm is an unsupervised clustering algorithm. Its basic idea is: given a data set with n elements or records, the split method will construct k groups, and each group represents a cluster, k < n, and these k groups meet the following conditions: each group has at least one data record: each record belongs to only one group.

In this way, we defined a new set of classification consists of 19 music styles and we display part of the style information here:

Table 2. 19 Music Styles

Style	The Content(artists are represented by IDs)
Avant-Garde	9 artists including 186927, 496525, 475362, etc
Blues	12 artists including 59156, 608701, 1006782, etc
Children's	4 artists including 744969, 784597, 808626,etc
Classical	26 artists including 803752, 239859, 740576, etc
...	...

That is a new set of classification of music genres, and we measure the similarity difference between and in genres based on it. We took samples of artists (1) in all songs and (2) within a genre (Jazz) respectively and polt both of them (Figure 6 and 7) in radar graphs. From both graphs, we can see that the plots in figure 6 are more dispersive and those in figure 7 are more concentrate, which indicate that the artists within genres are more similar to some extent.

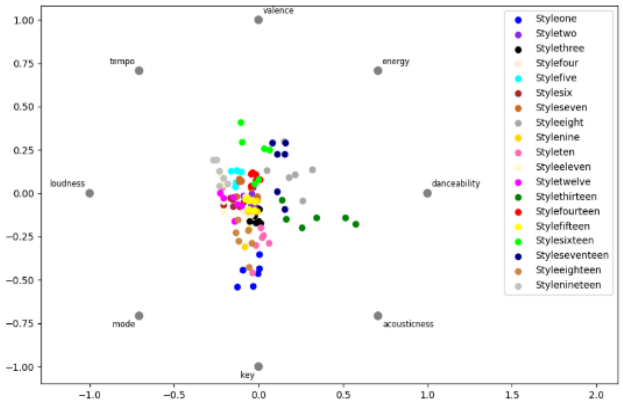

Figure 6. Similarity between All Artists

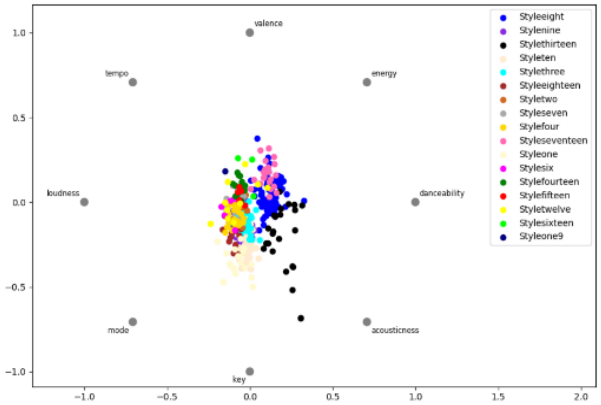

Figure 7. Similarity between Jazz Artists

4. Conclusion

In this article, we created a directed network of musical influence, where influencers are connected to followers. We developed parameters that capture 'music influence' in this network and explored a subset of musical influence by creating a subnetwork of the directed influencer network. We also developed measures of music similarity, through which we proved that artists within genre are more similar than artists between genres. Our work has strengths including: (1) We clarified the connection and influence between influencer and follower through the construction of graph association network. (2) The data visualization done through graphical methods has a significant effect on feature comparison. (3) Cosine similarity is used as a music similarity measurement model, which has a better behavioral scoring mechanism and a better judgment effect. Future work can also study the deeper cross-influence between various musical characteristics and the one-sided solution to the effect of similarity.

References

[1] Orpen K S, Huron D. Measurement of Similarity in Music: A Quantitative Approach for Non-parametric Representations Similarity: Qualitative and Quantitative Aspects [J]. Computers in Music Research, 1992.