ISSN 2618-1568 Vol. 3, Issue 2: 6-9, DOI: 10.25236/FAR.2021.030202

Research on Music influence based on Principal Component Analysis

Zhixiong Tao, Qiyang Pu, Gan Hu, Zheng Wang

Dalian Maritime University, Dalian, Liaoning, 116026

Abstract: *This paper mainly studies the data related to music in depth. The paper measure the impact of music by creating a directed network. The paper also develop measures of music similarity, analyze the influencing factors of music, and study the evolution of music. The paper have grouped the seven small questions into four large sections so that they can be studied in greater depth and clarity. In the first section, firstly, The paper create a directed network of music influence, and express the directed influence relationship by followers pointing to influencers, and then get the adjacency matrix. Secondly, the paper introduce the idea of between centrality and eigenvector centrality to quantify the musical influence index of artists with different meanings. The paper also come up with the interesting phrase enduring influence.*

Keywords: *Graph theory, Eigenvector centrality, Principal component analysis (PCA)*

1. Introduction

Music is the result of human's conscious cultural creation. Artists in the creation of music, the background of The Times, their own originality, personal experience and other factors affect their works. In the evolution of music, the creation of some artists has been deeply influenced and inspired by other artists and works of art [1].

Stephen P. Borgatti lays out in Centrality and network flow (2005) that a typology of network flows based on two dimensions of variation. Measures of centrality are then matched to the kinds of flows that they are appropriate for. Simulations are used to examine the relationship between type of flow and the differential importance of nodes with respect to key measurements such as speed of reception of traffic and frequency of receiving traffic. Zhong Yuan, Hong-mei Chen et al construct a hybrid feature outlier detection method based on fuzzy information entropy by using fuzzy approximate space with fuzzy similarity relation in Fuzzy information entropy-based adaptive approach for hybrid feature outlier detection (2020). Yin-Fu Huang [2], heng-Min Lin et al propose an automatic music genre-classification system based on a local feature-selection strategy by using a self-adaptive harmony search (SAHS) algorithm in Music genre classification based on local feature selection using a self-adaptive harmony search algorithm (2014).

The paper are going to develop a model to quantify the musical influence [3]. The model is designed to measure the impact of previously produced music on new music and artists. The paper also need to develop measures of music similarity and better understand how music evolves in society over time by examining the evolutionary and revolutionary trends of artists and genres.

2. Directed network of musical influence

Through studying the data set of influence_data, the paper can quantify the influence of artists on music. By creating a directed network of musical influence, the paper can analyze the influencers and followers of each artist to determine the level of influence of that artist [4].

The paper conduct data processing on the influence_data dataset and obtain 5603×5603 artist phalanx through the database full connection operation [5]. That is, every row or every column contains all followers and influencers, and there are no duplicating artists. The paper use a directed graph to represent all followers and influencers in the data set. Each node represents an artist, and if an artist influences another artist, there is a directed edge from that follower to the influencer. A network map can be formed. The paper extracted a small part of data from the data set to build a subnetwork, as shown in Figure 1.

ISSN 2618-1568 Vol. 3, Issue 2: 6-9, DOI: 10.25236/FAR.2021.030202

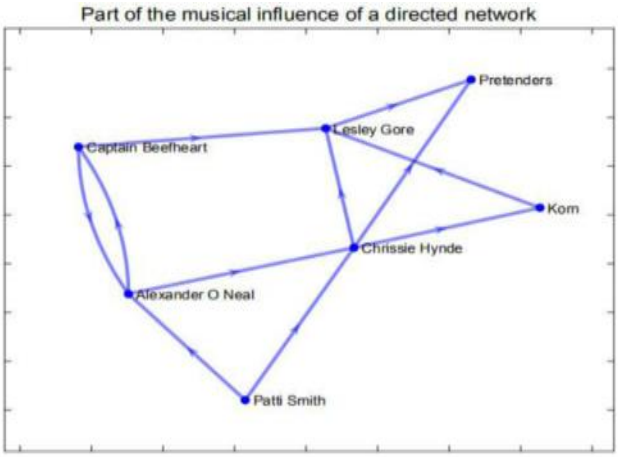

Figure 1: Music influence sub-network diagram

3. Analyzing the Network

The paper think that each node's between centrality is one of the music influence index. The between centrality indicates that in the process of mutual influence and inheritance of music, a certain node plays a relatively important role in connecting the preceding and the following. Between centrality is defined as how often a particular node acts as the midpoint of the shortest path between two other nodes. That is:

$$C_B(v) = \sum_{v \neq s, v \neq t, v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{1}$$

First, the paper construct the adjacency matrix. The paper through building the subnet adjacency phalanx A, as shown in the table 1, including line said followers, columns represent influencers. Aij equals 1 said the i line on behalf of the artist to follow the j columns represent artists, Aij is not equal to 1, said the current artists i and artists j distance (the distance to 0 or 999, of which 999 represent infinity).

The centrality of feature vectors is an important index to measure the influence of nodes in a network. The centrality of a node's eigenvector centrality is equal to the linear superposition of the centrality of its adjacent nodes. The difference between eigenvector centrality and between Centrality lies in that not every node connected to a particular node has the same contribution influence. Being adjacent to a node with high centrality makes a greater contribution to the improvement of the current node's own centrality than being adjacent to a node with low centrality. For example, if an artist is followed by only one very influential artist, his influence can not be underestimated. On the contrary, the influence of an artist is not necessarily greater if he is followed by a lot of lesser artists.

$$C_E(v) = x_v = \frac{1}{\lambda} \sum_{j \in N(i)} x_j = \frac{1}{\lambda} \sum_{j \in N} A_{ij} x_j \tag{2}$$

First, the paper construct the adjacency matrix. The paper through building the subnet adjacency phalanx A, which line represents the influencer, column represents the followers. Aij equals 1 said the i line on behalf of the artist to follow the j representative artists, on the contrary, Aij equals 0.

Table 1: Adjacency matrix for eigenvector centrality

	C.Beefheart	L.Gore	Pretenders.	Korn	Chrissie	A O.Neal	P.Smith
C.Beefheart	0	1	0	0	0	1	0
L.Gore	0	0	1	0	0	0	0
Pretenders.	0	0	0	0	0	0	0
Korn	0	1	0	0	0	0	0
Chrissie	0	1	0	0	0	0	0
A O.Neal	1	0	0	0	1	0	0
P.Smith	1	0	0	0	1	0	0

Then, the paper realized the solution of the eigenvector centrality through Python software, and the results are shown in the table 2.

One interesting thing the paper found when the paper captured music influence on the web was that

ISSN 2618-1568 Vol. 3, Issue 2: 6-9, DOI: 10.25236/FAR.2021.030202

influencers from the 1930s had followers from 2010, spanning 80 years. Focus on the present, the culture of nearly a hundred years ago can still remain in people's mind is undoubtedly the cultural classics and artistic treasures. Therefore, the paper think that the time span of followers and influencers can also be called one of the indicators of music influence.

Table 2 Centrality of matrix parameters

Name	C.Beefheart	L.Gore	Pretenders.	Korn	Chrissie	A O.Neal	P.Smith
Eigenvector Centrality	0.33333	0	0	0	0	0.33333	0.33333

4. Model result

The paper define the concept of "mean degree coefficient" in order to judge whether the above classification categories can reflect the genre. To illustrate this concept, let's make an analogy with a set of samples with two properties. As shown in Figure 2, this is a planar rectangular coordinate system. Where the abscissa X1 and ordinate X2 represent two attributes respectively, and each sample exists and has only a unique point in the plane Cartesian coordinate system. In the figure, the gray origin represents each sample, and the blue diamond represents the clustering center. As can be seen, the 11 samples are grouped into 3 categories. There is a new sample (represented by the red square). Clustering operation is performed again, and it is clustered into the leftmost category in the figure.

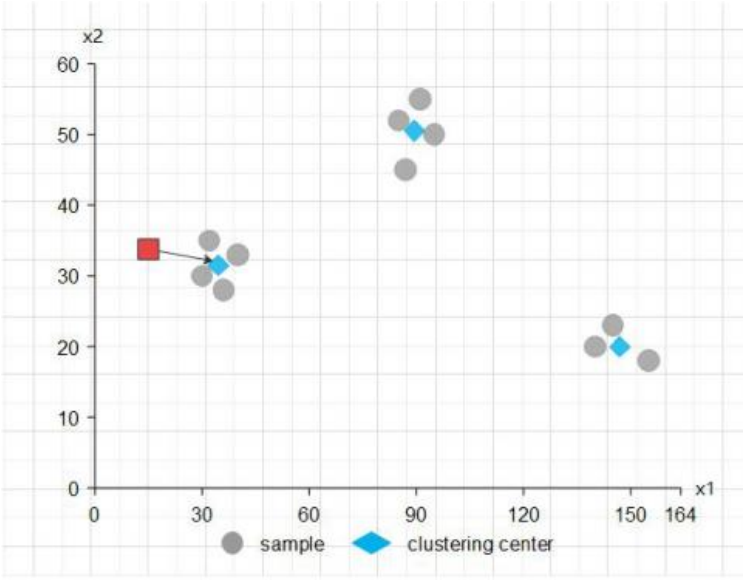

Figure 2: Concept of degree coefficient

Then, the degree coefficient of this category reflecting the new sample is calculated. The formula for calculating the degree coefficient is as follows:

$$L_{\alpha\beta} = \sum_{i=1}^{n} \left(\alpha_i - \beta_i \right)^2 \tag{3}$$

The smaller the degree coefficient is calculated, the more likely the category will cover the new category. In this paper, according to the pre-processed data, the paper know the mean value of each music attribute of each genre. Therefore, the paper can consider genres as artists, re-cluster them into the system, and calculate the degree coefficient from the results obtained. Genre itself is a set, which is analogous to sample individuals because of averaging processing. This is also the "mean" in the "mean degree coefficient". The paper analyze all artists who belong to category 2 after K-means clustering. Since each artist has its own genre in the original data, the paper present its genre in a pie chart in proportion to the number of artists, as shown in the left image of Figure 3.

ISSN 2618-1568 Vol. 3, Issue 2: 6-9, DOI: 10.25236/FAR.2021.030202

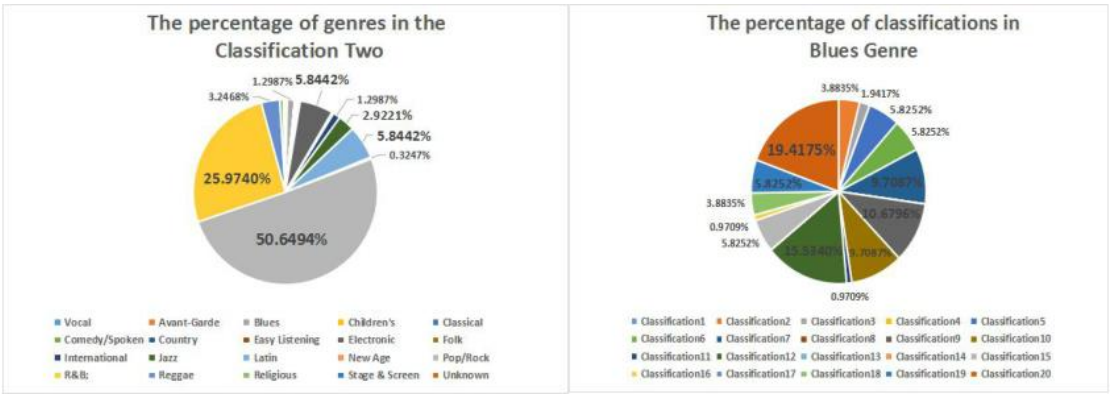

Figure 3: Pie chart

5. Conclusions

The arrival of the Internet has greatly promoted the exchange of information. The paper have found out the internal development law of music in the music data obtained through the Internet. This law will challenge the traditional understanding of music. Therefore, whether the paper can combine the traditional understanding of music and the use of big data to analyze music will be an urgent matter. In this paper, principal component analysis leads to information loss, which is measured by variance or eigenvalue. Linear fitting trend lines include error, random interference, SSR (sum of squares of regression) and SSE (sum of squares of errors).

This paper tries to use quantitative methods to measure the object relationship as far as possible. Centrality is used to measure genetic influence, and eigenvector centrality is used to express direct influence. The method of judging revolutionary breakthrough is put forward.

References

[1] J Q Chen, Y G Xi, Z J Zhang. A clustering algorithm for fuzzy model identification. Fuzzy Sets and Systems 98 (1998) 319-329.
[2] Stephen P. Borgatti.Centrality and network flow.Social Networks 27 (2005) 55–71.
[3] K Naveed, C Watanabe, P Neittaanmaki.Co-evolution between streaming and live music leads a way to the sustainable growth of music industry e Lessons from the US Experiences. Technology in Society 50 (2017) 1-19
[4] A J H Goulart, R C Guido, C D Maciel. Exploring different approaches for music genre classification. Cairo University Egyptian Informatics Journal (2012) 13, 59–63.
[5] Z Yuan, H M Chen, T R Li, J Liu, S Wang. Fuzzy information entropy-based adaptive approach for hybrid feature outlier detection. ISSN0165-0114.