Music Network Data Analysis Based on ISOMAP Algorithm Model

Kai Zhong¹, Shangqian Liu², Yue Li³ and Yanling Xu¹,*
¹College of Sciences, Huazhong Agricultural University, Wuhan 430000, China
²College of Public Administration, Huazhong Agricultural University, Wuhan 430000, China
³College of Economic & Management, Huazhong Agricultural University, Wuhan 430000, China
*Corresponding author e-mail: yanling_xu@mail.hzau.edu.cn

Abstract. The development of music is a tortuous process, and the network relationship between each genre and each artist is intricate. In order to have a better understanding of the history of music, this paper tells the stories hidden in the history of music by means of data processing. Firstly, this paper establishes a model to evaluate the similarity between music by using ISOMAP algorithm. At the same time, the forest evolution model was established to mark the most revolutionary musical characters. Finally, using the Page-Rank algorithm, we get the founders of several music genres. It turns out that the figures who led the development of music don't coincide with the figures who revolutionized music. Through the analysis of this paper, we can more clearly understand the development of music and the evolution of genres.

Keywords: Social Network, Visual Analysis, Forest Evolution.

1. Introduction
From politics to commerce, the influence of music on behaviour is acknowledged [1]. Music provided a powerful means of shaping Americans’ perceptions of the past to urge them to action in the present for outcomes envisioned in the future, compellingly anchored in terms of gratitude, virtue, and memory [2]. Though it makes sense to study the development of music.

Weiße C, et al used tone-based recognition to cluster and analyze the characteristics of classical music [3]. Nakamura E used statistical methods to analyze the evolution of music [4]. Klement B emphasized the importance of “extra-local knowledge” in musical creation [5]. At the same time, the evolution of music and the process of culture are linked together. Patrick E (2019) discuss the Cultural evolution of music in terms of educational and copyright [6]. Ben Lambert used musical evolution to describe the pace of modern culture [7]. However, no research has been done to visualize the evolution of music from a visual perspective.

For this study, we develop a model combining the ISOMAP and the social network analyze method to describe the changes in the evolution of music and find the key artist in the history of music. By the
process of dimensionality reduction and visualization, the connection and the evolution pattern are displayed directly.

2. Datasets Description
‘influence_data’ (These data were scraped from http://AllMusic.com) represents musical influencers and followers, as reported by the artists themselves, as well as the opinions of industry experts. These data contain influencers and followers for 5,854 artists in the last 90 years.

‘full_music_data’ (These data were obtained from Spotify’s API) 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 a summary data sets which is mean values by artist ‘data_by_artist’.

3. Measure of Different Music Genres Similarity
To better visualize and describe the similarity between different genres, ISOMAP algorithm is applied to the measurement of musical genres.

Nonlinear manifold learning is a popular dimension reduction method that determines large and high dimensional datasets’ structures [8]. ISOMAP is a classical algorithm of nonlinear dimensionality reduction (NLDR) or manifold learning (ML) [9]. For this question, the similarity between any two artists needs to be represented by the distance between them. But every artist in the data_by_artists are described by danceability, energy, valence, tempo and other variables with a total of 16. In the case of high latitude data, direct calculation of straight-line distance is not appropriate. By ISOMAP method, the distance between samples can be well approximated based on the nearest neighbor distance approximation.

Extract the common entry artist ID in data_by_artist and influence_data, and then match them. There are 291 artists whose genre is not identifiable, so they are excluded. Take 14 attributes of the remaining artists as input except name and id.

The original 14-dimensional attribute dataset D is mapped to a three-dimensional dataset Z by using ISOMAP. After obtaining Z, let Z = ZT, set two distance parameters as follows:

\[ D^{(l)} = Z(label = L), D^{(r)} = Z(label = R) \]

Among them, L and R are the two genres to be selected. Set the total evaluation value of genre similarity as the index to measure similarity, which is used as the measurement model of music similarity:

\[
\begin{align*}
    sim_{11} &= \frac{\sum_{a_1=1}^{k_1} \sum_{a_2=1}^{k_1} (d_{a_1}^{(l)} - d_{a_2}^{(l)})^2}{k_1^2} \\
    sim_{12} &= \frac{\sum_{a_1=1}^{k_1} \sum_{a_2=1}^{k_2} (d_{a_1}^{(l)} - d_{a_2}^{(r)})^2}{k_1 \cdot k_2}
\end{align*}
\]

Where, \( sim_{11} \) is the total evaluation value of the similarity of artists in the same genre, \( sim_{12} \) is the total evaluation value of the similarity of artists in different genres, \( k_1 \) is the number of \( D(l) \), and \( k_2 \) is the number of \( D(R) \).

\[
    D_{a_1}^{(l)} - D_{a_2}^{(r)} = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2}
\]

(a) Jazz & R&B  (b) Latin & R&B  (c) Vocal & Electronic

Figure 1. Visualization of the relative distance between the two genres
Use the music similarity measurement model to compare the similarity between and within genres. In order to improve the accuracy of the results, choose to use the data set full_music_data that contains more samples. In the first place, remove artists whose works lack genre information. Then the artists’ works of the same genre are sorted and grouped. Three pairs of genres are selected to be analyzed. The results are as Figure 1.

The total evaluation values of similarity within artists of the same genre and artists of different genres are as follows:

(a) \( \text{sim}_{11} = 0.7509, \text{sim}_{12} = 1.5267 \);
(b) \( \text{sim}_{11} = 1.0758, \text{sim}_{12} = 1.1249 \);
(c) \( \text{sim}_{11} = 0.5340, \text{sim}_{12} = 1.7685 \);

The results show that the similarity within the same genre is always greater than the similarity between different genres. Analyzing the graph in (b), we reach the conclusion that Latin and R&B music characteristics are similar.

4. Identify the Star Artists

4.1. Identify the Most Revolutionary Artists

In order to find the most revolutionary artists, every artist’s work should be considered. And each of music contains 12 attributes. What’s more, variations in each artist’s music should be taken into account. Considering the changes in influence over time and the relationship among followers and influencers, the "Forest Evolution Model" is innovatively proposed. In this model, take the artist as the root node, his direct followers as branch node and his followers’ work as leaf node. Though get a tree. All artists together form a complete forest. The forest contains the above 5854 artists.

Take an artist for example, his followers should be found first to form a subnet. After traversal, the ID of all leaf nodes and the number of layers of leaf nodes are obtained. Then the distance of nodes should be calculated to describe the similarities among the artists’ work by using ISOMAP algorithm. Then a three-dimensional coordinate table will be obtained. One element of the table represents the three-dimensional coordinate of a node. \( m_{uij} \) is used to represent the distance between the \( j^{th} \) leaf node of \( i^{th} \) branch node and the root node. Each branch node has its own weight, which is the reciprocal of the number of layers it belongs to. Considering that the same artist can have multiple identities, this results in them being at different levels. \( \text{lay}_{ik} \) is used to represent the \( k^{th} \) identity of \( i^{th} \) artist (branch node). Then add up the ownership of each artist to get a new weight \( W_i \). Select all the branch node whose weight is greater than \( \delta \) for analysis. Weight it through nodes and finally get the total score of each artist. And calculate the total score for each artist using the following formula:

\[
W_i = \sum_{j=1}^{m} \frac{1}{\text{lay}_{ij}}
\]

\[
Score = \sum_{i=1}^{a} W_i \cdot \sum_{j=1}^{b-1} (m_{uij} - m_{ui(j+1)})
\]

Among them, \( a \) is the number of works of each branch node, \( b \) is the number of branch nodes of each root node, and \( m \) is the number of each artist’s number of identities.

In this case, consider influencers with large followings to be stars in the field of music, like Bob Dylan and the Beatles. Only artists who are stars have the right to be influential.

According to Price Law, define that the number of stars is approximately equal to the square root of the total artists (artists + followers). In the whole data set, 75 artists are identified as stars. Calculate each artist's in-degree to make the top 75 people with the entry value to be stars. The relationship between the 75 people is plotted as a directed graph, where each color represents a different genre and the entry degree of each node determines the label size. The results are shown in the figure below:
Figure 2. The relationship between different stars

It can be found that The Beatles, Bob Dylan and The Rolling Stones not only exerted great influence within Pop/Rock genre, but also influenced the most influential influencers in other genres. We select the top 10 artists of 75 artists and put the data into the Forest Evolution algorithm to calculate. The result shows that the top three artists with total scores are John Coltrane, The Rolling Stones and Pink Floyd. And their corresponding scores are 8355.0517, 4377.0232 and 3926.2191.

4.2. Identify the Originators of Music Genres

This question uses PageRank algorithm to trace the source of influence. Brin and Page introduced the PageRank [10] algorithm, which is solely based on the graph structure of the web pages. We apply to this case: If artist A is influenced by artist B, it indicates that A thinks B is more important, so as to assign part of A's importance score to B, and so on. The score value of importance is expressed as follows:

\[ PR(A) = \frac{PR(B)}{L(A)} \]  

PR(A) is the PageRank value of A, and L(A) is the outdegree of A. This formula further explains: the PageRank value of A decreases, and the PageRank value of B increases accordingly. We can get the PageRank value of each artist's influence and the size relationship is as follows:

Figure 3. Social network graph under PageRank algorithm

It can be calculated that the average value of all artists PageRank is 0.00018, the sum of PageRank is 1. Genre and PageRank values of each artist are shown in the table below:
Table 1. Genre and PageRank values of each artist

| Artist Name         | Genre     | PageRank |
|---------------------|-----------|----------|
| Cab Calloway        | Jazz      | 0.0208   |
| Billie Holiday      | Vocal     | 0.0197   |
| Lester Young        | Jazz      | 0.0169   |
| Louis Jordan        | Jazz      | 0.0136   |
| T-Bone Walker       | Blues     | 0.0100   |
| Sister Rosetta Tharpe | Religious   | 0.0093   |
| The Beatles         | Pop/Rock  | 0.0090   |

The table shows that the influence of Cab Calloway, Lester Young, Billie Holiday are top3. This result is a little different from the stars we are familiar with. A strong explanation is that well-known artists such as The Beatles followed the artists who were active in the 1940s and divided their influence points among those artists. In addition, it can be found that Cab Calloway, Lester Young and Louis Jordan all belong to Jazz genre, which also shows that famous influencers active in 1960s were deeply influenced by Jazz. It also shows that the traditional American Jazz genre has a profound influence on the musicians in the datasets.

5. Conclusion
The development of music is influenced by different musical traditions. This paper makes a data mining of music development history. We define the index to describe the similarity of music genres, and select three pairs of genres for analysis. Finally, some important figures in the history of music development are obtained through the analysis of music attributes.

Acknowledgments
This research was financially supported by the general course of innovation and entrepreneurship construction of Huazhong Agricultural University “Data investigation, analysis and presentation” and Huazhong Agricultural University teaching research reform project (Grants No. 2019050).

References
[1] LOHMAN LAURA. Singing “Past, Present and Future”: Music in Early American Commemoration[J]. Journal of the Society for American Music, 2021, 15(2): 192-223.
[2] Philip Emanuel Rodney. Prison and the Law in Modern American and British Popular Music[J]. International Journal for the Semiotics of Law - Revue internationale de Sémiotique juridique, 2020: 1-26.
[3] Weiß C, et al. Investigating style evolution of Western classical music: A computational approach[J]. Musicae Scientiae, 2018, 23(4): 486-507.
[4] Nakamura E , Kaneko K . Statistical Evolutionary Laws in Music Styles[J]. Scientific Reports, 2019, 9(1).
[5] Klement B , Strambach S . How do new music genres emerge? Diversification processes in symbolic knowledge bases[J]. Regional Studies, 2019, 53(10): 1447-1458.
[6] Savage, P.E. Cultural evolution of music[J]. Palgrave Commun 2019, 5(16).
[7] Lambert, B., Kontonatsios, G., Mauch, M. et al. The pace of modern culture[J]. Nat Hum Behav, 2020, 4: 352–360.
[8] Mahwish Yousaf et al. NRIC: A Noise Removal Approach for Nonlinear Isomap Method[J]. Neural Processing Letters, 2021: 1-28.
[9] Mahwish Yousaf et al. NRIC: A Noise Removal Approach for Nonlinear Isomap Method[J]. Neural Processing Letters, 2021: 1-28.
[10] Honggui Li and Dimitri Galayko. Deep reconstruction of 1D ISOMAP representations[J]. Multimedia Systems, 2021: 1-16.
[11] Tang Xia et al. Anderson Acceleration of the Arnoldi-Inout Method for Computing PageRank[J]. Symmetry, 2021, 13(4) : 636-636.