

Music Influence Analysis based on Weighted Network

Yang Liu, Yiming Cai, Xuhui Zhang

Wuhan Institute of Technology, Wuhan, China.

Abstract

Music plays an important part in human history. It affects all aspects of people's life and work. Music has undergone many changes with the progress of the times. These changes have led to different genres and music styles. It's meaningful to build models to measure music influence. Before solving each problem, we analyzed and preprocessed the data. The data that does not in conformity with the reality are eliminated. The data is normalized processing according to the topic requirements. For the first problem, according to influence_data data set, we establish a weighted directed network between influencer and follower, and show it in Figure 1. The noticing, influence of edge weights and node activity are used as parameters to judge the influence of music. Three of the most influential individuals are selected in each year and a total of 27 individuals form a subnetwork. The specific network diagram is shown in Figure 2. The influence evaluation index of music individuals in the sub-network is calculated to construct a multiple influence matrix (table 2). The matrix reflects the size of music influence. We use full_music_data to build the measurement model of music similarity. In view of many indicators in the data set, multivariate factor analysis is used to reduce dimensionality. We get three factors after dimensionality reduction. We used songs as the research object, and then used cluster analysis to divide the data into 20 categories. The result is shown in Table 3 and Figure 7. According to the results, most of the works of artists in each category belong to the same genre. It shows that the measurement model is reasonable. By comparison, it can be concluded that artists of the same genre are more similar than artists of different genres. Principal component analysis and entropy weighting method were used to compare the similarity and influence of different schools. Each genre is ultimately given a different score, as shown in Table 5. You can distinguish different schools by the size of the score. Figure 8 and Figure 9 show how the different genres have changed over time. In view of different indexes, there is a certain similar relationship between different schools. On the basis of the first question, a subnet diagram centered on the core influencer is established. Combined with the scoring model established by the third question, the differences of music characteristics between influencers and followers are analyzed. Finally, one-way analysis of variance is used to find the fluctuation of music characteristics between the influencer and the follower, so as to judge the infectious music characteristics. From the data we find that there are features that mark major leaps in the development of music, and these features are clearly represented in the turning points in the image. Within the established network, some people have created new musical genres and musical characteristics. They represent revolutionaries. Then, we mainly analyzed the process of Pop/Rock genres over time, as shown in Figure 11. Furthermore, the main historical reasons for the change are analyzed. Finally, we take the development and change of the Internet as an example to analyze how to identify the influence of the Internet in the network relationship. In the network relationship, when the connection between artists becomes more close and the network becomes more diverse, we can think that the emergence of the Internet accelerates the spread and development of music.

Keywords

Weighted Directional Network; Multiple Influence Matrix; Multivariate Factor Analysis; Cluster Analysis; Single Factor Analysis of Variance.

1. Introduction

1.1. Restatement of the Problem

Music is an indispensable part of human life, no matter what era, music always exists in the long river of history. The topic gives several data sets, requiring us to conduct quantitative analysis on the influence of music according to these data sets, including studying the mutual influence between artists, music transformation, etc., so as to analyze the development and transformation of music.

1) Build a personally oriented network based on influence_data to reflect music influence. Then on the basis of the network, we need to set some parameters. Based on these parameters, a sub-network model is established to describe the model and analyze the relationship between parameters and the model.

2) Based on full_music_data and two summary data sets, a model is established to measure music similarity. And the model is used to compare similarity between artists. Then it is necessary to compare the relationship and changes between schools.

3) We need to take data_influence data set as the research object. Does the similar data indicate that the identified influencers have an impact on artist? What is the specific impact? In the data, determine if there are features that mark a big change in the music landscape. And on the basis of network relationships, find out which artists represent revolutionaries.

4) It is required to study the changing process of a certain genre of music over time. The factors of the dynamic influencers are judged to explain the reasons why genres or artists change over time. Then it further analyzes the influence of external factors (such as time, technology, etc.) on network relationship.

5) On the basis of the above questions, we need to write a one page document to reveal the value of the network model in analyzing the impact of music, and explore the impact of music on culture.

1.2. Problem Analysis

The question first gives the historical background and basic knowledge of the problem, which is the basis for solving the problem.

First of all, we need to preprocess the data and eliminate some of the data that do not conform to the actual situation. And carry on the data screening, obtains the final suitable topic request data. After establishing the individual-specific directional network, we can define the weight of edges and nodes to determine the parameters of music influence and establish the subnetwork model. Then through the evaluation index of the influence of individual music in the subnetwork, the multiple influence matrix is constructed, so that the size of the music influence can be obtained objectively.

Through the multivariate factor analysis, the dimensionality reduction of multiple indexes is easy to establish the music similarity model. Then a similar model can be established by clustering analysis. And analyze the corresponding problems through the model.

For the difference of different music genres, its fundamental purpose is to find the differences in music characteristics and music types of different genres. Firstly, principal component analysis can be used to reduce the dimension of the data to obtain effective few factors, and then a discrimination model based on entropy weight can be established. Then, weighting can be given according to the proportion of discrimination to establish a scoring model with high

discrimination degree. Finally, schools can be divided into different ranges according to their specific scores.

In view of the trend of genre change over time, we can first visualize the musicians and the time when they started music, and initially intuitively analyze the trend of music time change. Then the music released by each genre is corresponding to the time, and the change of music genre over time is further analyzed according to the number of albums released by each genre in each year.

2. Assumptions and Justifications

In order to facilitate the establishment of the model, based on the analysis of the topic, we make some reasonable assumptions:

2.1. When follower_active_start is less than influencer_active_start, cull the data.

The meaning of followers is to follow influential artists. If there is no influence of the former, there will be no follow of the latter. For some data, the time of followers is earlier than that of influencer, which obviously does not conform to the actual situation, so they are removed.

2.2. The proportion of influential schools is roughly equal to that of followers

This assumption can simplify the calculation and optimization model. We can also balance the network when building models without causing too much deviation.

3. Notations

The key mathematical notations used in this paper are listed in Table 1.

Table 1. Notations used in this paper

Symbol	Description
Y_{ij}	The weight of different music genres in different years
$R(m/n)$	the weight of edge
r_m	'm' total number of interactions received and sent
mean	The mean of all interactions
β	regulatory factor
α	Adjusting parameters (a rational number greater than 1)
S_i^c	Cross strength of node i
S_i^{sin}	The input strength of node i
S_i^{cout}	Outgoing strength of node i
X	data matrix
Y	Standardized matrix

Note: Symbols not represented in the table are explained in the text

4. Directional Network and Subnetwork Model

4.1. Establishment and Analysis of Directional Network Relationship

4.1.1. Establishment of directional network

According to the topic requirements, the directional network relationship between influencer and follower is established by using influence_data data set, as shown in Figure 1.

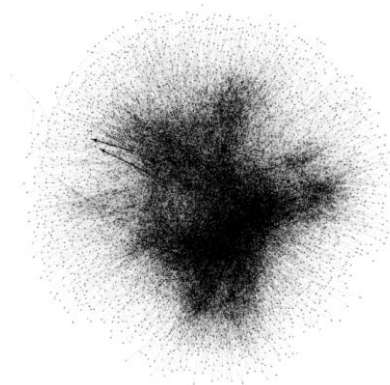


Figure 1. Directional network diagram

It can be seen from this network diagram that the relationship between the influencer and the follower is complex. An influencer can have multiple followers, and a follower will influence multiple people. Eventually form an interwoven network. According to this directed network graph, we establish a directed weighted model. The network diagram can be quantified. The edges and points in the large network are given weights as parameters of music influence.

4.1.2. Definition of network edge rights

The interaction frequency between individuals in the network is used as the weight index of their relationship, that is, the weight is directly reflected in the weight of the music network model. The intensity of interaction in different music networks varies according to the type of interaction between individuals (the same genre or different genres) and the number of times. The interaction between musician individual 'm' and 'n' is denoted as $n \rightarrow m$. Because the interaction between individual musicians has the characteristics of pointing, the edges in the music oriented network model should be directed. In addition, if the number of interactions

$m \rightarrow n$ is 'rmn', then $\sum_{u \in \text{om}} r_{mu}$ shows that the number of interactions 'm' sends out, where 'om' is the collective of all musicians whose network individual m sends out interactions. Identically, $\sum_{u \in \text{qm}} r_{um}$ can be used to represent the number of interactions received by an individual, where 'qm' represents the collective of all musicians who receive the interaction of an individual 'm'. Similarly, we define the number of interactions sent and received by 'n'.

When measuring the relationship strength of network individual $m \rightarrow n$, we mainly consider two aspects: The first is the proportion of 'n' in all interactions issued by musician individual 'm'. If the value is larger, it indicates that 'm' has higher attention to 'n'. Secondly, What is the proportion of interactions from 'n' in all interactions received by individual musicians 'm'. This ratio indicates the influence of musician individual 'n' on 'm'. If this value is higher, then the influence of musician individual 'n' on 'm' is correspondingly higher.

Based on the above ideas, the following two ratios are used to measure the relationship strength of individual musicians from 'm' to 'n'.

1) Define the musician individual 'm' to 'n' attention is:

$$f_{mn} = \frac{r_{mn}}{\sum_{u \in \text{om}} r_{mu}} \quad (1)$$

$\sum_{u \in \text{om}} r_{mu}$ represents the sum of the outgoing degrees of musicians in the data.

2) The influence of musician individual m on n is defined as:

$$d_{mn} = \frac{r_{mn}}{\sum_{u \in qm} r_{um}} \quad (2)$$

$\sum_{u \in qm} r_{um}$ represents the sum of the musicians in the data.

Then the relationship degree of musician individual $m \rightarrow n$ can be defined as:

$$R(m/n) = \frac{f_{mn} \times d_{mn}}{f_{mn} + d_{mn}} \times \frac{r_m}{\text{mean}} \times \beta \quad (3)$$

Similarly, the degree of relation of $n \rightarrow m$ is:

$$R(n/m) = \frac{f_{nm} \times d_{nm}}{f_{nm} + d_{nm}} \times \frac{r_n}{\text{mean}} \times \beta \quad (4)$$

1) If there is a one-way relationship (that is, one musician 's only worship of another), then

$$\frac{d_{mn}}{f_{mn}} = 0 \Rightarrow R = 0 \quad (5)$$

So there 's:

$$R(m/n) = \left(\frac{f_{mn} \times d_{mn}}{\alpha} \right) \times \frac{r_m}{\text{mean}} \times \beta \quad (6)$$

2) Finally get the weight of the edge:

$$R(m/n) = \begin{cases} \frac{f_{mn} \times d_{mn}}{f_{mn} + d_{mn}} \times \frac{r_m}{\text{mean}} \times \beta \\ \left(\frac{f_{mn} \times d_{mn}}{\alpha} \right) \times \frac{r_m}{\text{mean}} \times \beta \end{cases} \quad (7)$$

4.1.3. Definition of network node weights

In the definition of node weight in double weighted network, considering that it is a weighted network, it needs the reputation and activity of nodes in the influence age. The more active a node in the music network is, the more influential it can lead to more nodes and participate in the interaction of the network. Therefore, the node weight index in this music network can be measured by the activity of its nodes in a certain era.

Activity is based on the influence-data data set. We can calculate the proportion of the number of musicians in each era as influencers, the popularity of music genres in the development process of the times. By normalizing these data, we can get the node weights in the weighted music network we want.

Based on the above analysis, we give a general definition of node weights in music networks. Let the number of interactions between individual musicians in the network 's' in a certain era be M_s^i (i represents the times). The individual activity of the musician can be defined in this

network as $a_i = \frac{M_i}{\sum M_i}$ ($\sum M_i$ represents the total number of musicians interacting in an era).

However, although the above definition can describe the activity of individual musicians in the network to a certain extent, it is likely that individual musicians appear in a short period of time in the music network and interact in the network in a short period of time. At the same time, most musicians are often kept at a low level of activity, and the activity described by the definition can be seen, which cannot well show the real activity of a node.

For these reasons, we can propose a new weighted music network node weight definition method, the definition formula is as follows:

$$A_s^i = \lambda + \frac{M_s^i}{\ln \sum_{s \in sm} M_s^i} \quad (8)$$

' M_s^i ' is the interactive influence of musician individual 's' in the era 'i'. ' λ ' is a regulator of individual activity. 'sm' represents all musicians in the era 'i'. The reason for this definition is mentioned above. However, the main reason is that when we actually deal with influence-data dataset, we encounter many abnormal musicians, which are not enough to reflect the role and play the characteristics in the big network. In order to make all the network individuals can have a description index to show their activity, and can not cause great fluctuation due to the difference of musicians' individual influence. Therefore, combined with relevant knowledge, we add a regulatory factor to keep the activity between 0 and 1, so as to carry out the next sub-network analysis for an impactor.

4.2. Establishment and analysis of sub-networks

4.2.1. Establishment of subnetwork

According to the directional network model established above, we select the sub-networks. According to the above calculation method of edge and node weight, the influence (out-degree) of each musician is calculated in the directional network in the year. Then select some people in the directional network as the research object. The basis of choosing these people is: according to 1930-2010, each year musician influence (out) selected the first three, nine years a total of 27. Finally, these 27 musicians are nodes in the sub-network relationship. According to the selected musicians, establish a sub-network as shown in Figure 2. Different colors in the picture represent different music genres.

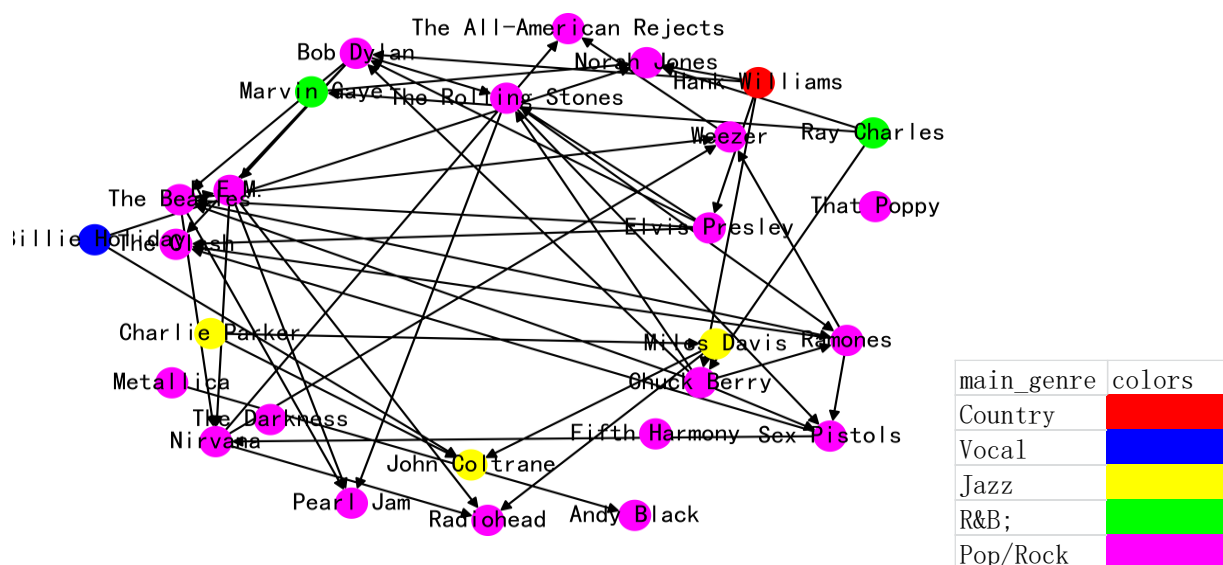


Figure 2. Subnetwork diagram

This subnet relationship reveals the relationship between artists with high influence in each year. It can also be seen from the graph that Pop / Rock has the largest number of people, indicating that even in different years, the influence of the music genre is higher.

4.2.2. The influence evaluation index algorithm of music individual in sub-network and the construction principle of multiple influence matrix.

The parameters (attention, influence, activity, etc.) obtained from the influence-data dataset serve as the basis for the basic edge and node weights in music networks.

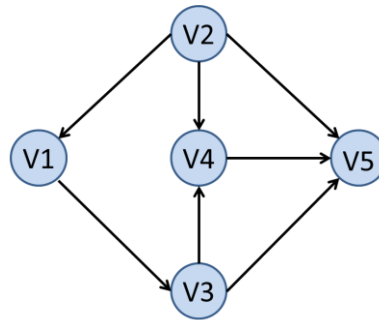


Figure 3. Topography

In this algorithm, for any directed graph, the influence degree between the two connected nodes V_2 and V_5 can be defined as $I(V_2|V_5)$, and $I(V_2|V_5)=R(V_2|V_5)$. For any two non-adjacent nodes, such as node V_1 and node V_5 , it can be seen that there are many paths from V_1 to V_5 , $\{V_1 \rightarrow V_3 \rightarrow V_5, V_1 \rightarrow V_3 \rightarrow V_4 \rightarrow V_5\}$. And the relative influence of V_1 to V_5 is:

$$I(V_1|V_5) = \max\{P(V_1|V_5) \times A_{V_i}\} \quad (9)$$

A_{V_i} denotes the activity of a node directly adjacent to V_5 , that is, the node weight.

$$P(V_1|V_5) = \frac{I(V_1|V_i) \times I(V_i|V_5)}{I(V_1|V_i) + I(V_i|V_5)} \quad (10)$$

V_i and V_5 are direct adjacent nodes.

In summary, we can define the relative influence degree of any node V_a on V_b as :

$$I(V_a|V_b) = \begin{cases} R(V_a|V_b) \\ \max\left\{\frac{I(V_a|V_i) \times I(V_i|V_b)}{I(V_a|V_i) + I(V_i|V_b)} \times A_{V_i}\right\} \end{cases} \quad \textcircled{2} \quad (11)$$

Equation ① means that V_a is directly adjacent to V_b . ② indicates that V_a and V_b are not directly adjacent, and V_i is directly adjacent to V_b .

Analyze the definition of relative impact between nodes as follows :

Firstly, the relational intensity between the two nodes is defined as the relative influence degree. Because the relational intensity is the harmonic mean of the attention and influence degree between the two musicians. When the difference between the two is large, the harmonic average is closer to a smaller value than the arithmetic average. When the node V_a interacts with the node V_b less, V_b interacts with V_a more, it will produce one-way high strength relationship. The harmonic average can effectively avoid this situation.

Secondly, in all paths between two non-adjacent nodes, the reason for taking the maximum path value is based on the degree of influence between two musicians, which is determined by the most direct connection. Each musician on the path has the greatest influence combined. As for the two nodes that are not directly adjacent, the harmonic mean value of importance between nodes is explained in the first point. This is to avoid unilateral influence, that is, individual worship.

Multiple influence matrix can be constructed by calculating the relative influence degree between nodes. Then the dependence of each individual musician on the remaining nodes in the subnet can be represented by matrix P :

$$P = (P_{ij})_{n \times n} = \begin{pmatrix} 0 & I_{12} & \cdots & I_{1n} \\ I_{21} & 0 & \cdots & I_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ I_{n1} & I_{n2} & \cdots & 0 \end{pmatrix} \quad (12)$$

In matrix P, $P_{ij}=I_{ij}$ represents the relative influence degree value of node V_i on V_j . On the basis of considering the relative influence degree of the whole network node to be evaluated, the important influence degree B_i of node V_i can be defined by combining the local importance (cross strength value) of the node to be evaluated.

First, the cross strength of node i is:

$$S_i^c = \lambda S_i^{\text{sin}} + (1 - \lambda) S_i^{\text{cout}} \quad 0 < \lambda < 1 \quad (13)$$

$\begin{cases} \lambda > 0.5, & \text{The input strength of the node has a greater effect} \\ \lambda < 0.5, & \text{The output strength of the node has a greater effect} \end{cases}$

Then there is:

$$B_i = S_i^c \sum_{j=1, j \neq i}^n I_{ji} \quad (14)$$

$\sum_{j=1, j \neq i}^n I_{ji}$ represents the relative importance of node V_i in the whole network, which is the sum

of all nodes' influence on node V_i , denoted as P_i .

After the final normalization, the important influence degree of node V_i is:

$$B_i^1 = \frac{B_i}{\sum_{k=1}^n B_k} \quad (15)$$

4.2.3. The construction of sub-network multiple influence matrix

Using the above data and formula, the multiple influence matrix of the subnet is calculated by python. The matrix is 27×27 , the table is too large to display in the article, so here only show some matrix.

Table 2. A partial matrix of multiple influences

	Hank Williams	Billie Holiday	Charlie Parker	Miles Davis	John Coltrane	Ray Charles	Marvin Gaye	Elvis Presley	Chuck Berry
Hank Williams	0	0	0	0	0	0	0	0	0
Billie Holiday	0	0	0	0	0	0	0	0	0
Charlie Parker	0	0	0	0	0	0	0	0	0
Miles Davis	0	0	2.08E-05	0	0	0	0	0	0
John Coltrane	0	2.57E-05	2.3E-05	5.22E-06	0	0	0	0	0
Ray Charles	0	0	0	0	0	0	0	0	0
Marvin Gaye	0	0	0	0	0	9.77E-06	0	0	0
Elvis Presley	7.69E-06	0	0	0	0	0	0	0	0
Chuck Berry	8.21E-06	0	0	0	0	9.95E-06	0	0	0
The Beatles	0	0	0	0	0	0	0	2.61E-05	1.88E-05
Bob Dylan	8.15E-06	0	0	0	0	0	0	2.63E-05	1.89E-05
The Rolling Stones	0	0	0	0	0	0	0	2.37E-05	1.75E-05
Sex Pistols	0	0	0	0	0	0	0	0	2.04E-05
The Clash	0	0	0	0	0	0	0	2.79E-05	0
Ramones	0	0	0	0	0	0	0	0	1.98E-05

The degree of influence between different musicians in the matrix reflects the size of their influence. The larger the value is, the greater the influence is. When the value is 0, there is no effect. Through this matrix, we can intuitively and objectively see the interaction between artists. According to this result, it can be found that some of the 27 artists have great influence. The network lines around these people are also dense in the subnet diagram.

The following conclusions can be drawn by sub-network diagram and multiple influence matrix. First, it can be found that the Pop /Rock genre has the largest number of musicians, indicating that the genre is influential. Secondly, by observing the trend of network lines, it can be found that there is a close relationship between the same schools. The interaction is also more.

5. Model of music similarity and comparative analysis of related genres

5.1. Music similarity measurement model and analysis

5.1.1. Data preprocessing

Taking full_music_data as the research object, 16 music indicators are found. According to the requirements of the topic, we find that year, release_date, these two indicators have little relationship with the topic. So we eliminate them. Finally, 14 indicators were taken as the research object. After further data preprocessing, it is found that 26 out of the danceability column are 0, which does not meet the research requirements and is deleted. Then, 3σ criterion is used to delete some singular data. There are 109 columns of energy. Liveness is 2936. Speechiness is 1896. The column duration_ms has 1467.

All indicators are then normalized. The normalization method has two forms. One is to change the number into the decimal between (0,1), and the other is to change the dimensional expression into the dimensionless expression. Mainly for the convenience of data processing, the data mapping to 0~1 range processing, more convenient and fast[1].

5.1.2. Dimension Reduction by Multivariate Factor Analysis

There is a correlation between the indicators between the data, so factor analysis can be done. Factor analysis is to convert multiple measured variables into a few comprehensive indicators (or latent variables). It reflects a dimension reduction idea. By reducing the dimension, the variables with high correlation are clustered together to reduce the number of variables to be analyzed and the complexity of problem analysis[2]. There are many indicators to be processed here, so factor analysis is needed to reduce the dimension. Because of the mixed different types of data indicators, multivariate factor analysis is used.

We use the prince library in Python to calculate it. Prince is a library for doing factor analysis. This includes a variety of methods including principal component analysis (PCA) and correspondence analysis (CA). The goal is to provide an efficient implementation for each algorithm along with a scikit-learn API. Call this library to complete mixed multivariate analysis. Finally, three factors after inductive dimension reduction are obtained. Factor 1: Basic characteristics of music; Factor2: external influence of music; Factor3: musical emotional characteristics.

Bar3D-Label Distribution

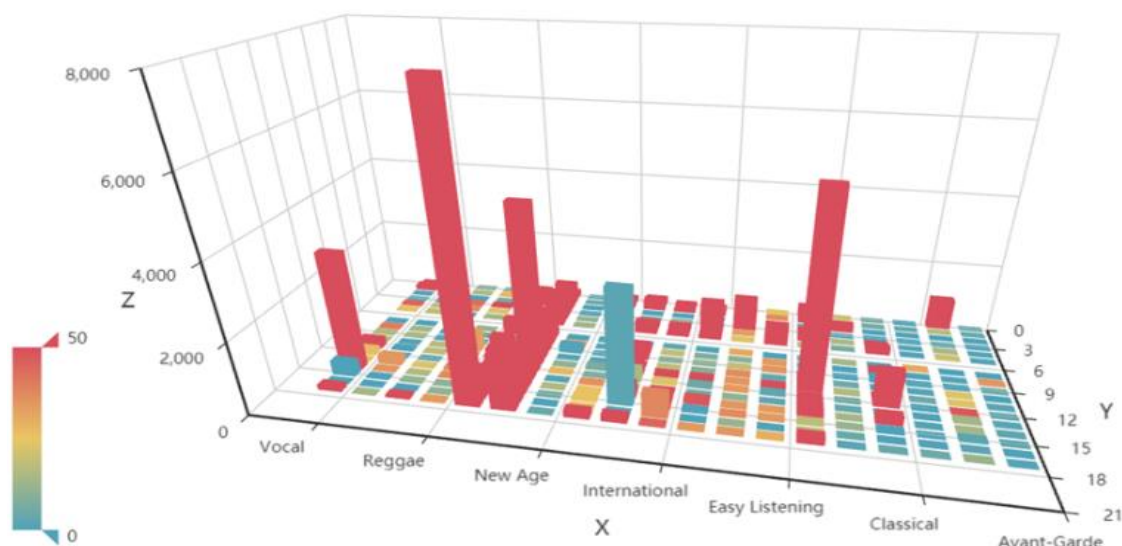


Figure 4. Distribution of songs in pop genres

5.1.3. Establishing Music Similarity Measurement Model by Cluster Analysis

The K-means algorithm in Python uses distance as the evaluation index of similarity. That is to say, the closer the distance between the two objects is, the greater the similarity is. Using this algorithm and three factors after dimension reduction, clustering analysis can be done well. We take each song in full_music_data as the research object. Because there are a total of 20 schools, so we divide the songs into 20 categories according to cluster analysis. Statistics the distribution of songs in 20 labels. The distribution of songs in pop genres is illustrated as Figure 4:

Through this diagram, it can be seen that pop genre music is mainly distributed in the label, that is, the characteristics of pop genre artists ' songs are similar and can be divided into labels. Thus we make20labels represent the collection of high similarity of artists in each genre, which can be divided into the following table:

Table 3. Genre clustering results

label1	label2	label3	label4	label5
Avant-Garde	Unknown	Comedy/Spoken	Electronic	Folk
label6	label7	label8	label9	label10
Easy Listening	Latin	Children's	Avant-Garde	New Age
label11	label12	label13	label14	label15
Reggae	Country	Classical	Stage & Screen	Vocal
label16	label17	label18	label19	label20
Jazz	International	R&B	Religious	Pop/Rock

The 20 labels represent a collection of highly similar works by artists in each genre. A three-dimensional cluster diagram as shown in Figure (6)? is made. This figure can directly reflect the distribution of 20 categories. Different colors represent different categories.

At the same time, we can extract two schools for comparison:



Figure 6. Folk and Latin comparison chart

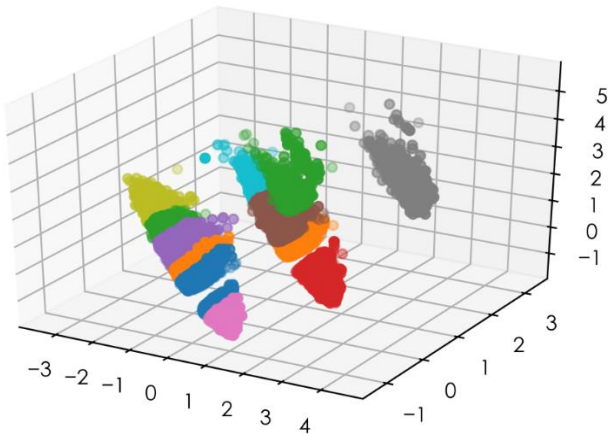


Figure 7. Clustering analysis diagram of songs

At the same time, we can extract two schools for comparison:

The picture above shows that artists from the same genre are more similar in style, while there is less connection between works from different genres.

5.2. Comparative analysis of different schools

5.2.1. Problem analysis

For the difference of different music genres, its fundamental purpose is to find the differences in music characteristics and music types of different genres. Firstly, principal component

analysis was used to reduce the dimension of the data, and three effective factors were obtained. Then, a discrimination model based on entropy weight was established, and a scoring model with high discrimination degree was established according to the proportion of discrimination degree. Finally, the schools were divided into sections according to their specific scores

In view of the trend of genre change over time, firstly, the musicians and the beginning of music were visualized to initially intuitively analyze the trend of music time change. Then the music released by each genre is corresponding to the time, and the change of music genre over time is further analyzed according to the number of albums released by each genre in each year.

The relationship between schools is further analyzed on the proportion data based on the clustering model of problem two, based on the similar characteristics between schools. After filtering the data, data_by_artist removes 243 artists that do not appear on the influence_data list. The full_music_data removes 7781 artists that do not appear on the influence_data list.

5.2.2. Model preparation

Due to the different number of artists in each genre, the rationality of the data is considered. In this paper, the index of all artists is weighted by the proportion of the number of music created by the musician in all music of the same genre. The weighted average processing is used to obtain the feature vector of each genre, and an eigenvalue matrix is obtained. The data is forwarded and normalized.

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \quad (16)$$

$$\text{Among them, } r_{ij} = \frac{1}{n-1} \sum_{k=1}^n (X_{ki} - \bar{X}_i)(X_{kj} - \bar{X}_j) = \frac{1}{n-1} \sum_{k=1}^n X_{ki} X_{kj}$$

Using PCA, based on the principle of information maximization, three factors with the largest contribution are obtained, which are the rhythm factor z1, the type factor z2, and the emotional factor z3.

$$\begin{cases} z_1 = l_{11}x_1 + l_{12}x_2 + \cdots + l_{1p}x_p \\ z_2 = l_{21}x_1 + l_{22}x_2 + \cdots + l_{2p}x_p \\ \vdots \\ z_m = l_{m1}x_1 + l_{m2}x_2 + \cdots + l_{mp}x_p \end{cases} \quad (17)$$

Then the entropy weighting method is used to calculate. It is a widely used weighting method, and the calculated discrimination has the characteristics of objectivity and robustness. If 'n' evaluation objects and m indicators are set, the initial information matrix $X' = (x_{ij}')_{n \times m}$ can be obtained, where X_{ij}' is the value of the j index of the i evaluation object. Since the permanent deformation and penetration displacement indexes are inverse indexes, the indexes are first homogenized according to Formula (18) to obtain the data matrix $X = (x_{ij})_{n \times m}$. Secondly, dimensionless is carried out according to Equation (19) to obtain the standardized matrix $Y = (y_{ij})_{n \times m}$. Thirdly, the entropy method is weighted according to Equations (20) and (21), and the entropy of the jth evaluation index is defined as, and the entropy weight of the jth evaluation index is defined as. Finally, the discrimination is calculated according to formula (22).^[3]

$$x_{ij} = \frac{1}{x_{ij}'} \quad (i = 1, 2, \cdots, n; j = 1, 2, \cdots, m) \quad (18)$$

$$y_{ij} = \frac{x_{ij}}{\sum_{k=1}^n x_{kj}} \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, m) \quad (19)$$

$$H_j = -k \sum_{i=1}^n y_{ij} \ln y_{ij} \quad \left(j = 1, 2, \dots, m; k = \frac{1}{\ln n} \right) \quad (20)$$

$$\omega_j \in [0, 1] \quad \text{and} \quad \sum_{j=1}^m \omega_j = 1 \quad (21)$$

$$DI_j = \left(\max_{i=1}^n y_{ij} - \min_{i=1}^n y_{ij} \right) \omega_j \quad (22)$$

5.2.3. Establishment and Solution of Model

Factor discrimination and weight are respectively:

Table 4. Factor discrimination and weight

	Z_1	Z_2	Z_3
Entropy (H)	0.92	0.95	0.99
Entropy weight (ω)	5.71	0.312	0.117
Differentiation (DI)	0.28	0.06	0.01

According to the weight data of maximum discrimination, the maximum discrimination scoring model based on entropy weight is established.

$$Y = 0.571 \times z_1 + 0.312 \times z_2 + 0.117 \times z_3 \quad (23)$$

By calculating the scores of each genre, such as table 3, the score shows a certain difference, so the music genre can be distinguished according to the difference.

Table 5. Factor scores of different schools

Genre	score	genre	score	genre	score
Country	-1.07	Folk	1.85	Stage&Screen	3.25
New Age	0.33	Blues	2.03	Reggae	3.47
Pop/Rock	0.66	Jazz	2.20	Children's	3.65
Electronic	0.87	Vocal	2.40	Comedy/Spoken	3.81
Classical	1.12	Avant-Garde	2.61	R&B	4.60
Religious	1.38	Latin	2.79		
Easy Listening	1.64	International	2.92		

According to the scores above, the scores of Country School and R & B School are less than 0 and greater than 4, respectively. As the extreme value, the interval span is large and easy to distinguish. For the schools with intermediate scores, the score gap between the two neighboring schools is about 0.2, and it is easy to distinguish between the two schools.

After being tested by some artists, the error is less than 30 %, so the model is reasonable.

According to The observation of images and data, The genre of Avant-Garde, Stage & Screen, Classical, Comedy/ Spoken has a very small number of people, and its distribution is relatively average, with no significant development and no disappearance. The genre of Children's began in 1950 and had little development in The following years; The genre of New Age started in 1960, with some development in The following years, but few artists were involved. The genre

of Easy Listening had a few artists during The period 1930-1950, and then declined; The genre of International, Folk and Blues has only a few artists, both of them reached their peak in 1960 and then declined; The genre of Pop/Rock R&B originated from 1930 to 1940 and developed rapidly in The following decades, becoming The main force of The art world, among which Pop/Rock developed The most rapidly and reached The peak number of 665 in 1990.

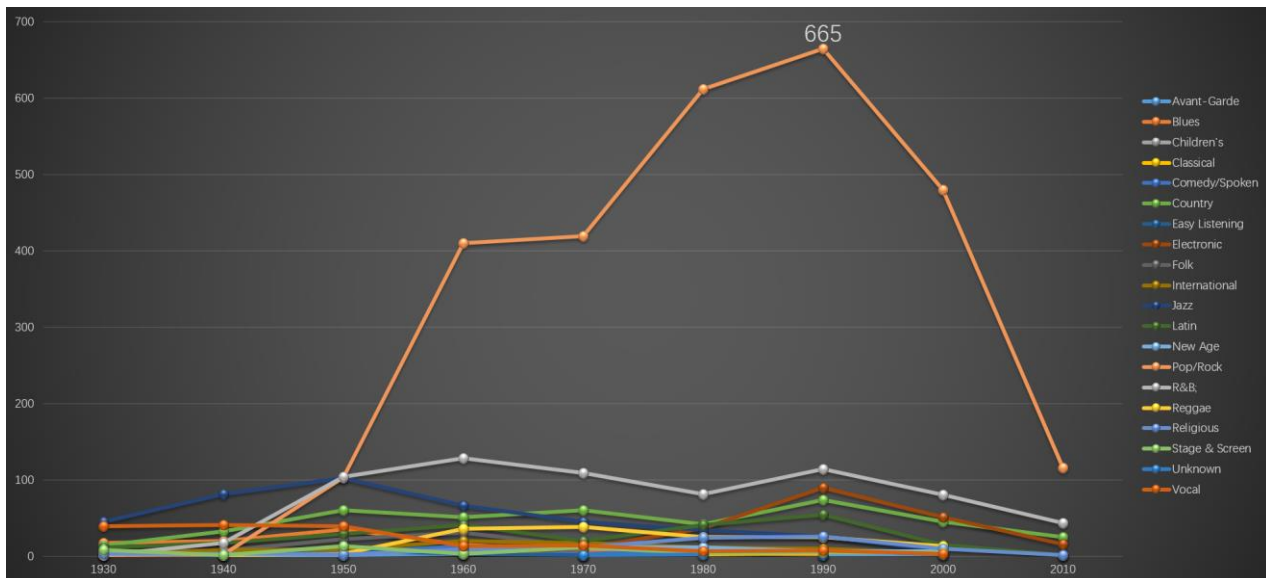


Figure 8. Allocation Year Map Based on Exit and Ingress

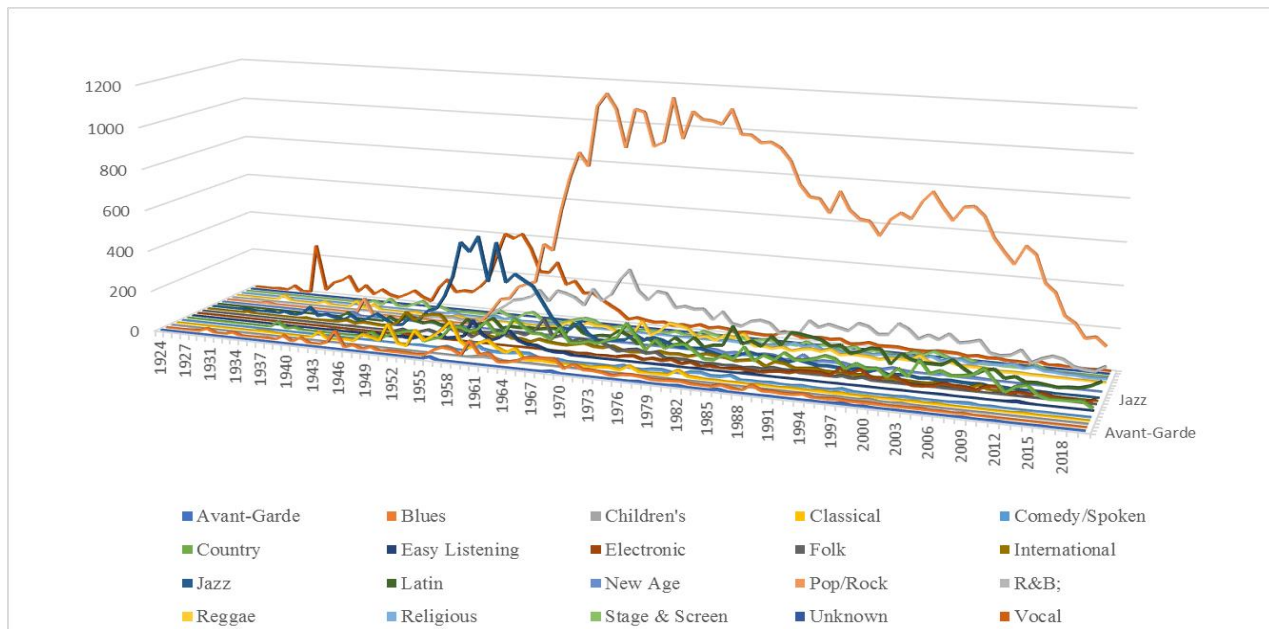


Figure 9. Chart of the amount of music produced by genre over time

Since the total amount of music created by Avant-Garde and Children's is relatively small (60 and 53 respectively), it is not analyzed. Blues composed 1,179 music pieces, most of which were released between 1956 and 1969. The number of songs released gradually decreased, but they continue to be released every year. Classical genre produced 1332 pieces of Music during the 20 years from 1945 to 1964, and then disappeared rapidly. After 1984, the genre almost completely disappeared. Comedy/Spoken composed a total of 319 pieces of Music. The genre first appeared in 1943 and experienced two periods of prosperity, 1959-1965 and 1978-1986,

before disappearing completely. Country genre appeared in 1936 and began to develop in 1949. Up to now, there are still 100 works issued every year. It has a long history and a good development situation. The genre of Easy Listening flourished from 1952 to 1968, peaked in 1958, and disappeared after 1970, producing 784 pieces. Electronic music genre developed well after 1988. Electronic music benefited greatly from the development of Electronic technology. Its boom period has not come yet, so it is predicted that this genre will continue to develop. Folk genre had a certain development during 1951-1977, but due to the limitations of its music, it could not be changed in time in the era of ever-changing trends, and it completely declined after 1979. The Pop/Rock genre is the largest in the art world, accounting for more than half of the music industry in terms of both number of people and number of works. It was started in 1957 and then developed rapidly, accounting for more than 70% of the total number of works in the period 1965-2009. Latin and R&B schools are similar, both of which are larger schools, which developed and prospered after 1955 and still have a certain creative power and have been circulated for a long time, but the number of R&B and the number of creation is higher than that of Latin. Jazz flourished during the period 1949-1968 and then declined more rapidly, as its musical characteristics and conveyed emotions were consistent with the ideology of the time. VOCAL began in 1933, developed steadily for a period of time, prospered from 1954 to 1963, and then declined.

From the perspective of the continuity of popularity among genres, VOCAL was the largest group in the period from 1933 until 1951, when Jazz developed to equal its share. By 1963, Pop/Rock had become a major force in the music industry. At the same time, the R&B and Country genres occupied the number two spot in the music industry. So far, the music industry has not changed the pattern of one dragon and two tigers.

6. The influence of influencer

6.1. The problem of four

6.1.1. Data preprocessing

Take data_by_artists as the research object. Delete the musicians not in the influence-data data set, because to analyze the influence of music, it involves relationships of multiple directed networks between followers, and 243 musicians not in the multiple directed networks need to be deleted.

6.1.2. Influence of influencers on music created by followers

Through the weighted directional subnetwork, a group of music works of influencers and followers are found in the influence-data, and the musical characteristics between their works are analyzed. The evaluation model of the third question is used to score, and the size of the score is analyzed to reflect whether it has an impact. The closer the score is, the greater the effect. Through the analysis of the data, it can be concluded that when the influencer and followers are in the same genre, the similarity of the music works is large, which proves that the influence is great. When influencers and followers are in different genres, the similarity of musical works is in the middle or high, which proves that only partial influence exists.

6.1.3. Analyse the role of musical characteristics in the directional network of influencers and followers

Firstly, take the core influencer as the core to generate a subnet. Then the changes in music characteristics between the core influencers and followers in the subnet are analyzed. The music features of musicians in each subnet are extracted from the data_by_artists dataset.

Core influencer selection criteria: select the most influential follower genre musicians as the core influencer. This is the most representative.

Core influencers: Hank Williams

Table 6. Hank Williams normalized data

artist_name	genre	artist_id	danceability	energy	valence	tempo	loudness	mode	key
Hank Williams	Country	0.58	0.14	0.60	0.48	0.46	1	0.36	4

Analysis process:

1. The data of Hank Williams and his followers are normalized.
2. Excluding the maximum and minimum values in each musical feature, excluding the extreme values resulting from normalization, and then making box diagrams

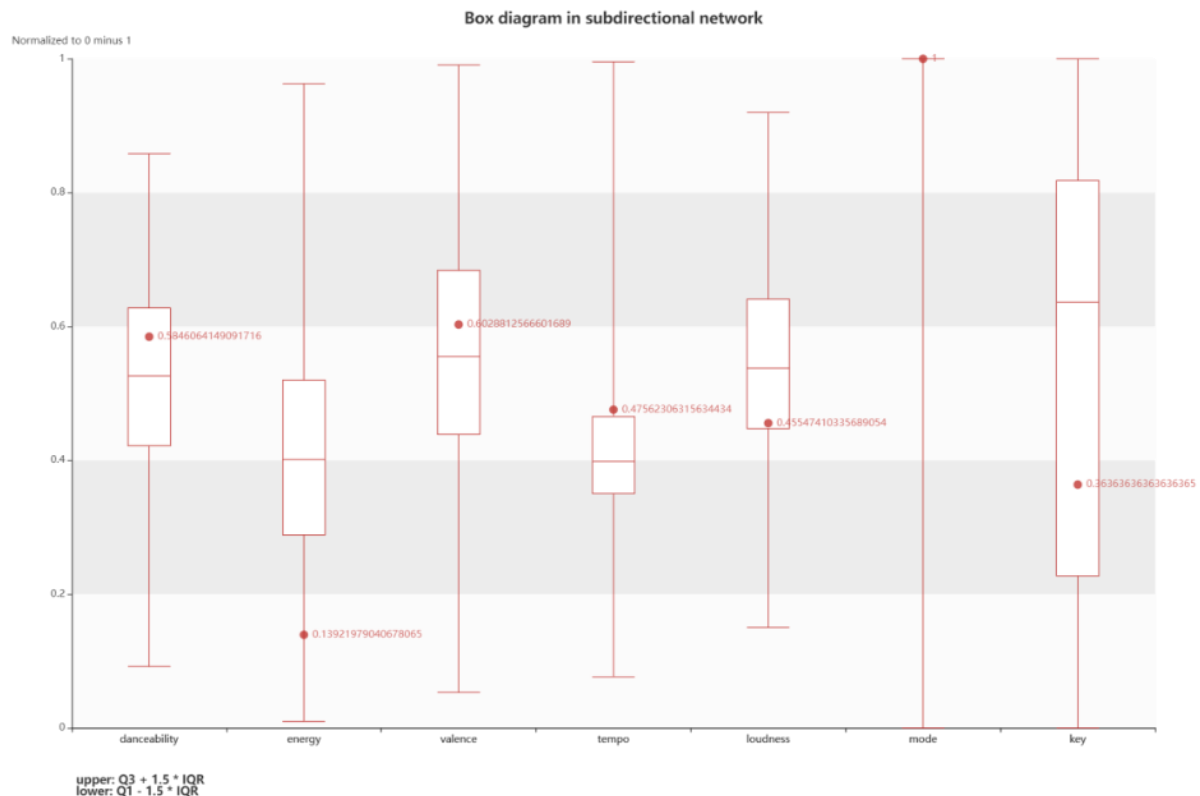


Figure 10. Box plot

In the box diagram above, the red dots represent seven music eigenvalues of the core influencer, and the rectangular box represents the music eigenvalues of the followers. It can be seen that the core influencer Hank William and the followers have a large similarity in the music characteristics 'tempo' and 'danceability', which further indicates that these two characteristics have a greater role in the sub-network with Hank William as the core, namely, tempo and danceability more 'embodiment' than others.

Table 7. Multiple comparison

Multiple Comparisons						
Variable: tempo						
LSD						
Category		Mean difference (I-J)	Standard error	Significance	95% Confidence Interval	
					Lower limit	Up limit
Hank William (Country)	Pop/Rock Genre	(0.00324)	0.00377	0.39077	0.01067	0.00419
	Folk Genre	(0.00026)	0.00855	0.97587	0.01713	0.01661
Pop/Rock Genre	Hank William(Country)	0.00324	0.00377	0.39077	0.00419	0.01067
	Folk Genre	0.00298	0.00866	0.73107	0.01410	0.02006
Folk Genre	Pop/Rock Genre	0.00026	0.00855	0.97587	0.01661	0.01713
	Hank William(Country)	(0.00298)	0.00866	0.73107	0.02006	0.01410

3. Variance analysis

Through the single factor analysis of variance of this table, we can see through the F test, significant $0.000 < 0.05$, which shows that for a music feature, there are at least two musicians between different musicians are significant differences, which have significant differences in most groups are influencers and followers in different schools, and the significant difference between musicians is not big, indicating that the music feature does play an important role between influencers and followers.

By observing the LSD results of different music features, it is concluded that there is no significant difference among the musicians in a certain music feature (conclusion), and by observing the mean figure below, we can also see that in this music feature, the trend of music features of different musicians tends to be stable, which means that the music feature is more 'contagious' than other music features.

6.2. Features of Revolution in Music Development and Artists Representing Revolutionaries

6.2.1. Problem analysis

Each index is analyzed separately in data _by _year, and further analysis is made on the combination of data with large fluctuations. Since the emergence and popularity of schools will affect the overall music indicators of the year, this phenomenon can be regarded as a music revolution. Music revolutionaries can be compared according to the data of their influencers and followers. If the musician has a large difference from his influencers and a small difference from his followers, he can be defined as a revolutionist.

6.2.2. Problem solving

By analyzing the data of danceability, energy, valence, tempo, loudness, it is found that all the data fluctuated greatly in 1946, and the number of music records issued in that year decreased sharply. According to the time distribution of music creation, Latin schools issued some records in the year. According to the analysis of records, it conforms to the data of the year. According to historical reasons, 1946 was the worst year in the U. S. economy in nearly a hundred years. Therefore, its music tends to be low and sad, corresponding to ideology.

The second is based on the results of factor analysis in question two, the basic characteristics of music, music emotional characteristics of the two factors in 1964, 1983, 1991 between the three years changed greatly.

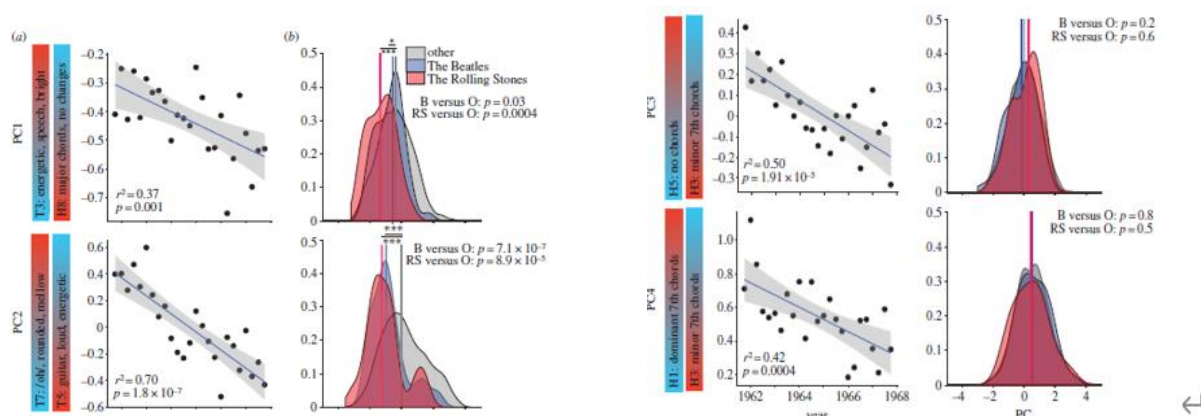


Figure10. Factor clustering analysis in 1964

According to the figure 10 in 1964, music feature data fluctuated greatly. Combined with music data, 1964 belongs to the pop /rock genre boom period, in the history of music and The Beatles

music impact on the United States. Therefore the music characteristic factor can act as the music development revolution influence characteristic[4].

7. The change of schools over time and the influence of external environment

7.1. Genre music varies over time

7.1.1. Changes of music of a genre over time (choose rock and roll)

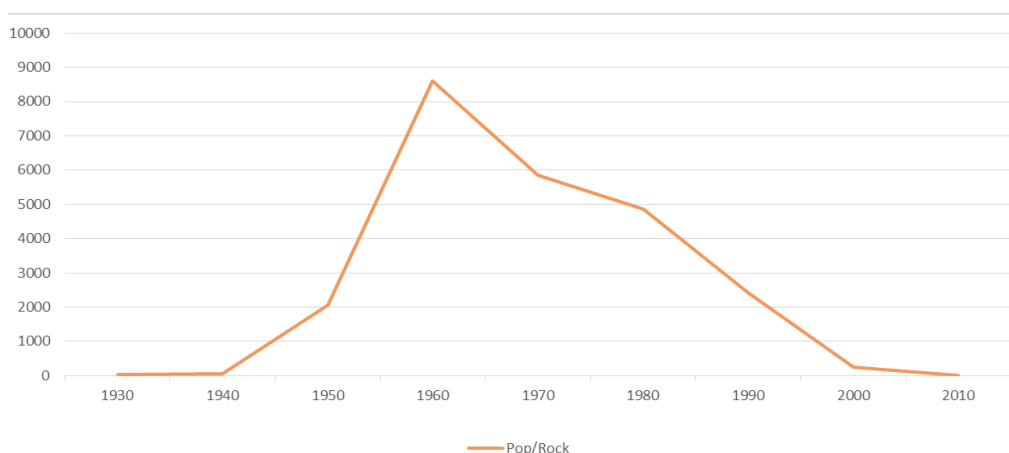


Figure 11. Pop / Rock music genre changes over time

According to Figure XX, music such as Pop /Rock grew significantly in 1940, turning again in 1950 and growing further. By 1960 reached the peak, then the number began to decrease.

7.1.2. Reasons for these changes

Through the search of data, it is found that the main reason is historical. In the 1950s, there was a significant increase in the number of various genres, based on the historical background and social environment at that time. In the long history of the United States, rock music as an important part of the genre, its prosperity and development is closely related to The Times.

The 1950s during World War II was a relatively stable and prosperous era in the history of the United States. At this time, young people lived in a very superior environment. They did not like the sentimental pop songs of their parents and needed to find new pop music of their own. And the rhythm of the "rock" is derived from black blues and jazz, country music has a strong rhythm, form, and free music type with the characteristics of the young and active, energetic, disposition, and rock music subject matters involved in the issue of concern is closely linked with them, it's the music characteristic determines its relevant characteristics of mass culture, cater to the young generation of music appreciation requirement, so it would appear however, attracting thousands of teenagers[4].

In the 1960s, rock music in the United States went into a low ebb. During this period, the United States was in dire straits, the social situation was turbulent, the domestic economy declined, the unemployment rate increased; Abroad, the Vietnam War is getting deeper and deeper. In this situation, the once fanatical rock music began to fall silent[4]. many popular music groups have also started to see a decline in numbers.

7.2. Identify other changes in network relationships (the impact of the Internet on music)

Many factors will affect the development and dissemination of music, such as time, environment, social changes and other external factors. In the network relationship model

established above, the changes of music genres are different from those in other situations, and sudden changes or anomalies occur. In combination with years, 2000 is just the time of the rise of the Internet, and it can be considered that the Internet the Internet on music. With the popularization of the Internet, the carrier of music works has undergone essential changes. In the record industry era, the carrier of music is black tape, wire recording, and then developed to tape, CD, and then developed to digital MP3. It is the Internet that subverts everything and makes music become a pure digital communication signal. The carrier of music has become a free resource download website, audio-visual platform website and various APPs adapted to intelligent terminals[5]. The development of Internet has changed the way of music transmission and promoted the spread and influence of music.

7.3. Sensitivity Analysis

Problem 3 establishes a maximum discrimination scoring model based on entropy weight. The weight of the principal component is modified. The weight of the first factor is added and subtracted by 1% -5%, and the remaining weight is based on the principle of sum 1. According to the comparative analysis of new data and original data, the data change trend is limited, indicating that the model has good robustness and strong stability.

8. Strengths and Weaknesses

8.1. Strengths

- 1) Use multiple factor analysis to reorganize the information of the original variables, find out the common factors affecting the variables and simplify the data.
- 2) Principal component analysis uses dimension reduction technology to replace the original variables with a few comprehensive variables, which gather most of the information of the original variables. Secondly, it calculates the score of the comprehensive principal component function.
- 3) Robustness of scoring model based on entropy weight.

8.2. Weaknesses

- 1) Models are difficult to test.

9. A document to the ICM Society



Figure 12. Notes and curve spectrum

In the above research process, we have established a number of network relationships, and analyzed the principles and relationships. It can be seen from the network relationship that the

influence of music reflects the spread and development trend of music to a large extent. Reflects the influence of a particular musician or genre, and can be further inferred about changes in the music. Online relationships allow us to intuitively and objectively understand the value of music influence. Music influence and network relation complement each other and influence each other. The network relationship reflects the music influence, and different music influence will lead to different network relationship. Using our method to establish a network model helps to understand the value of music influence.

In this problem, 20 music genres are taken as the research object and analyzed based on the data given by the question. This will make the final network and results have certain limitations and particularity. As more data becomes available, the directional network becomes more complex. There will be more contact between influencers and followers, and greater influence between genres and within the same genre. The measurement of musical similarity should be further improved. The increase in the amount of data will lead to more information between music and music. In order to obtain similar information, it is necessary to develop more detailed classification standards for data. When the research object is a genre, the increase of data volume will increase the number of schools. The analysis of genre information should be further planned. If the time line of the study is extended, the correlation between the data and time is even greater. The changes over time will be more pronounced.

Concerts have a visible effect on culture. Culture also influences music in many ways. Music is a functional part of culture. The meaning of music is inferred from the perspective of culture, and it is inextricably linked with culture. Music's function is given by culture. In today's society, the trend of diversification is becoming more and more obvious. The proposition of multicultural music education and the music culture of every nation in the world are worth learning. The cultural value of music is equal and there is no distinction between advanced and backward music. [6] The development of music and culture is influenced by the external environment. When science and technology gradually developed, people's living standard improved. There will be more positive and lively music, which in turn will promote the prosperity of culture.

References

- [1] 360 encyclopedia. Normalization method. <https://baike.so.com/doc/5331658-5566896.html>
- [2] 360 Q & A. What is the use of factor analysis ? <https://wenda.so.com/q/1383524221061452?src=150&q=%E5%9B%A0%E5%AD%90%E5%88%86%E6%9E%90w>
- [3] ZhangHuaizhi,WangDi,YangYanhai. Differentiation Investigation of High Temperature Performance Evaluation Indices of Asphalt Mixtures [J/OL]. Journal of Building Materials:1-12[2021-02-08]. <http://kns.cnki.net/kcms/detail/31.1764.TU.20201109.1548.002.html>.
- [4] Matthias Mauch, Robert M. MacCallum, The evolution of popular music: USA 1960–2010[J]. 01 May 2015.
- [5] Guo Jinliang. Development and Characteristics of American Popular Music [J]. Young Writers, 2012(06):92.
- [6] Liu Yangjun. The development and reform of modern music under the mode of 'Internet+' [J]. Journal of Xinjiang Art College, 2018,16(01):29-33.
- [7] Gu Yingying. Comparative study on Sino-US music education from the perspective of culture [D]. Qufu Normal University, 2008.