

# The Influence of Music

## Summary

Music is an important part of human society and culture, which has extraordinary significance and weight in human history. Nowadays, with the development of computer, it is possible to quantify the influence of music and find out the relationship between music and artists. In this paper, through the establishment of quantitative music and artist model, we explore the similarity and mutual influence between music genres and artists. Based on this, we can have a more accurate understanding of the development of music, the evolution and revolutionary trend of artists and genres.

For Problem 1, based on influence\_data set, we establish an artists directed influence network. In the network, we take artists as nodes and the following relationship between artists as edges. Then, based on the established network, we propose an improved VPR model to measure artist's "music influence", and clarify the performance and significance of "music influence" in the network. "Music influence" in the network is represented by the size of the node and the direction of the edge, in which the size of the node represents the influence of the individual artist, and the direction of the edge represents the transmission direction of the influence.

For Problem 2, we take cosine similarity as the measure of music similarity, and quantify artist according to it, so as to compare the similarity between artists of the same genre and different genres. We find that compared with different genres, there is a higher similarity between artists of the same genre. Finally, we use Mann Whitney U test to verify the reliability of our conclusion.

For Problem 3, we first establish the scheme of genre quantification, that is, we first carry out principal component analysis on 15 characteristics of music, extract four principal components, and describe the significance of the four principal components according to the linear relationship between the principal components and the original indicators. According to the difference and connection of the four principal components in each genre, this paper compares the similarity and mutual influence between genres, and obtains the difference of distinguishing each genre. Then, it shows the changes of each genre with time in the form of Thermodynamic diagram.

For Problem 4, based on the quantified results of the artist in Problem 2, we verified that the influencer really influenced the followers, and identified six characteristics, such as energy and loudness, which are more "infectious" than other characteristics of music.

For Problem 5, we first define the revolution in the process of music development, that is, we use (gene, cascade, characteristic) triple to describe it. We have established innovative indicators to measure the size of changes in the development of music and other metrics, combined with model 1, to identify the revolutionaries in the process of music development.

For Problem 6, we first establish the measurement index of the change size of music in the evolution process. This paper selects a representative artist as an example, manually selects 5 artists who have influenced him. Then we get these 6 artists' characteristics based on

full\_music\_data set. 12 music characteristics are reduced by principal component analysis, and the evolution process of the six artists' works is shown in the form of scatter diagram. For Problem 7, we infer the effects of social, political or technical changes that might have existed by analyzing the changes in the characteristics of various kinds of music in different periods and the popularity of music in society as a whole.

**Keywords:** Principal Component Analysis , Quantification of music , Complex network.

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Problem Background . . . . .	3
1.2	Problem Restatement . . . . .	3
1.3	Our Approach . . . . .	4
<b>2</b>	<b>Assumptions and Symbols</b>	<b>4</b>
2.1	Assumptions . . . . .	4
2.2	Symbols . . . . .	5
<b>3</b>	<b>Quantitative Model of Music Influence</b>	<b>5</b>
3.1	Music Influence Evaluation Index . . . . .	5
3.1.1	The Degree of Preference . . . . .	6
3.1.2	Music Influence . . . . .	6
3.2	Model Implementation and Result Analysis . . . . .	6
<b>4</b>	<b>The Measurement of Music Similarity</b>	<b>8</b>
4.1	Data Cleaning . . . . .	8
4.2	Data Standardization . . . . .	8
4.3	Similarity Calculation . . . . .	9
<b>5</b>	<b>Correlation between Different Genres</b>	<b>9</b>
5.1	Principal Component Analysis . . . . .	10
5.1.1	Data Standardization . . . . .	10
5.1.2	Feasibility Test . . . . .	11
5.1.3	Principal Component Extraction . . . . .	11
5.1.4	Calculate the Principal Component Value . . . . .	12
5.2	Result Analysis . . . . .	12
<b>6</b>	<b>The Influence of Influencers on Followers</b>	<b>13</b>
6.1	The Definition of Music_Influence . . . . .	13
6.2	The Relevance between Music_Influence(influencers,followers) and Similarity .	14
6.3	Discussion on the Appeal of Characteristics . . . . .	14

6.4	The Influence of Influencers on Followers' Styles . . . . .	15
<b>7</b>	<b>The Insight into Evolutionary and Revolutionary Trends of Artists and Genres</b>	<b>16</b>
7.1	Main work . . . . .	16
7.1.1	The Definition of Revolutionary Characteristics . . . . .	16
7.1.2	The Definition of Significant Change . . . . .	17
7.2	Establishment of Indicators to Measure the Reformer . . . . .	17
7.2.1	The Definition of Innovative . . . . .	17
<b>8</b>	<b>Indicators that Reveal the Dynamic Influencers</b>	<b>18</b>
8.1	Define the Influence in the Process of Music Evolution . . . . .	18
8.2	The Evolution of Artists . . . . .	18
<b>9</b>	<b>Influence of Music in Time and Circumstances</b>	<b>19</b>
<b>10</b>	<b>Sensitivity Analysis</b>	<b>19</b>
<b>11</b>	<b>Model Evaluation</b>	<b>20</b>
11.1	Strengths . . . . .	20
11.2	Weaknesses . . . . .	21
<b>12</b>	<b>Conclusion and Further Discussion</b>	<b>21</b>
	<b>Appendices</b>	<b>23</b>
	<b>Appendix A Tools and software</b>	<b>23</b>
	<b>Appendix B The Codes</b>	<b>23</b>
B.1	Task4 Contagious . . . . .	23

# 1 Introduction

## 1.1 Problem Background

In the long river of history, the twentieth century is an era of upheaval. In the field of culture and art in this period, while negating the past forms and concepts, artists pursued and explored from a wide range of angles, so that various ideological trends filled the atmosphere of the art field. It can be said that the 20th century society, no matter from which point of view, presents a situation of coexistence of different styles. The music of the 20th century itself is a complex and diversified culture, and the composers of different genres of this period are constantly innovating from the starting point of inheriting the tradition, showing an extremely individualized tendency. At the same time, they are trying to explore the musical language and means of expression that accord with the established artistic thought and aesthetic taste. Composers in this period enjoy greater freedom to write and create than ever before. In the 20th century, music developed to multipolarization, but its development always had a clear context. Various styles coexist and tend to be integrated. Composers try everything possible to create by all means available, making the field of music wider, although some music is not accepted.

With the development of computers, people can quantify music in 13 aspects, such as danceability, music energy, valence, tempo, loudness, mode, key, acousticness, instrumentality, liveness. Based on information about music and its authors, And the quantitative data, We can find the interaction between music artists, And the development of music in society over time. The development of music is more and more clear in front of us.

## 1.2 Problem Restatement

In order to understand and measure the impact of previously produced music on new music and music artists, develop a model to measure music influence, through which we can check the evolution and revolutionary trend of artists and genres.

Task1: Use the influence\_data set to build a digraph that can reflect the influence of music. Develop a parameter that reflects the music influence and explain the realization and significance of music influence in the network diagram.

Task2: Develop a measurement method of music similarity and use this method to quantitative artists. Then calculate and compare the similarity of artists in the same genre and between different genres

Task3: Compare the similarity, influence and relevance between genres and within genres. Develop a method to differentiate genres. Explore the relationship between genres and time.

Task4: According to the similarity, verify whether the influencer affects the followers. Identify whether some musical features are relatively more "contagious.". Judge whether the characteristics of influencers and followers are related.

Task5: Establish the measurement method that can measure the revolution in music evolution, and identify the music revolutionaries

Task6: Establish an indicator to reflect the dynamic influence factors of music development, and explain how music genres or artists change over time.

Task7: Using the established model, analyze how to express the information about the cultural influence of music in time or environment. How to identify the impact of social,

political or technological changes (such as the Internet) in the network.

### 1.3 Our Approach

Firstly, we process and clean the data set. Remove the artists of Unknown genre, and use the given data set to match the corresponding relationship between songs and genre, artists and genre, and standardize the data for specific problems.

After that, we propose an improved VPR model to measure the music influence of artists, and select artists between 2000 and 2010 as data to implement the model. We draw an influencer-follower directed network node graph, and take Kings of Leon as an example to calculate the music influence.

For music similarity, we use cosine similarity to establish the calculation method of similarity index. Using this method we can get the similarity among genres. Then, we use the method of principal component analysis to make a simple analysis of the characteristics of different genres. For the influence of influencers on followers, we calculate the Pearson correlation coefficient of music influence and similarity. At the same time, taking James Taylor as an example, we define a new index to indicate the similarity of various characteristics between influencers and followers, and finally come to the conclusion that influencers do have an impact on the music created by followers.

In addition, by calculating the Pearson correlation coefficient between popularity and other characteristics, we get several more infectious characteristics. Through the data analysis in the time dimension, we get the changes of music genres over time, and give a detailed definition of revolutionary characteristics, major changes and reformers. Through the analysis of statistical results and the improved VPR model, we find that the Beatles is the reformer of Pop / Rock. At the same time, we define a new index to reflect the influence of music evolution, and illustrate its specific meaning with examples. Finally, we analyze the sensitivity, advantages and disadvantages of the model and method.

So far, we have developed a complete program to understand and measure the impact of previously produced music on new music and music artists.

## 2 Assumptions and Symbols

### 2.1 Assumptions

To simplify our model and eliminate the complexity, we make the following main assumptions in this literature. All assumptions will be re-emphasized once they are used in the construction of our model.

**Assumption 1.** *Each artist belongs to only one music genre and has the characteristics of that genre.*

**Assumption 2.** *Artists don't change their music genres.*

**Assumption 3.** *Artists can only create their own style of music, that is to say, for music created by one person, music and creators belong to the same genre.*

## 2.2 Symbols

In this work, we use the nomenclature in Table 1 in the model construction. Other nonefrequent-used symbols will be introduced once they are used.

Table 1: Notations used in this literature

Symbols	Description
$Pre_i$	Followers' preference for influencers
$MI(i)$	music influence
$w(i, j)$	The weight factor among artists
$LS(t)$	The popularity of the artist's genre
$f(i, j)$	the time interval between the decade that the artists began their music career
$Sim_i$	The alienation degree of followers and influencers on the i-th character
$c_{artist^i}$	The i-th character of an artist named artistname
$C^*$	Comment Quantification Value
$number_f$	the total number of the followers

## 3 Quantitative Model of Music Influence

An artist is influenced by influencers, and as an influencer he influences other artists. If an artist is followed by many artists, his music influence may be great. If an artist is not followed by many artists, his music influence may be great. If an artist is followed by one artist with great influence, his music influence may be great. We also consider that an artist's music influence is influenced by time. In different periods of time, an artist is followed by different numbers of artists with different influences, and that artist's music influence will change accordingly. At the same time, if a genre is popular, the music influence of its artists increases. To sum up, we evaluate the influence of music by considering the influence size of followers, the time interval between the decade that the artists began their music career s, the popularity of the genre in the period of followers, and the attention of followers. In this section, we propose an improved VPR model to measure the music influence of artists[1].

### 3.1 Music Influence Evaluation Index

Music influence is used to measure the influence of a musician on the music industry, which is divided into two dimensions: the influence on artists of the same genre and the influence on artists of different genres. The influence on artists in the same genre is that it causes more artists to enter the genre and start their music careers. The influence on artists in different genres means that artists in another genre are involved in the creation of works in the field.

The essence of an artist following another artist is the process of the spread of the artist's music influence. The concept of time heterogeneity is involved here, and factors such as the different prevalence of different schools in different time spans should also be considered to reduce the impact of time heterogeneity on the evaluation results. Therefore, in order to minimize the impact of the temporal heterogeneity of the evaluation index of music influence on the analysis results, different weights should be set for the communication index of music influence composed by artists of different genres in different years.

### 3.1.1 The Degree of Preference

An artist may have many influencers, and the degree of preference for different influencers is different, so we define the degree of preference as

$$pre_i = \begin{cases} a * \frac{1}{number_f} & \text{the influencer belongs to the same genre as the artist} \\ (1 - a) * \frac{1}{number_f} & \text{the influencer is not in the same genre as the artist} \end{cases} \quad (1)$$

If the influencer belongs to the same genre as the artist, then the artist prefers the influencer, and the weight is  $a$ . If the influencer is not in the same genre as the artist, then the weight is  $1-a$ .  $a \in (0.5, 1)$

### 3.1.2 Music Influence

If he is followed by many artists who have great influence in recent years, his music influence will also be high. Therefore, for each artist, his  $MI(i)$  depends on the music influence of all artists who follow him and the popularity of his genre in recent years.

The formula is as follows:

$$MI(i) = d \sum_{j \rightarrow i} MI(j) * \omega(i, j) + \frac{1 - d}{|V_p|} \quad (2)$$

where  $w(i,j)$  is the probability function of jumping from  $j$  to  $i$

$w(i,j)$  is the probability function of jumping from  $j$  to  $i$

$$\omega(i, j) = \alpha * LS(t) + (1 - \alpha) * f(i, j) \quad (3)$$

where  $LS(t)$  refers to the popularity of the genre of the artist, and  $f(i,j)$  refers to the time interval between the decade that the artists began their music career.

$$f(i, j) = e^{-\sigma(T_j - T_i)} \quad (4)$$

where  $\sigma$  is a decline factor,  $T_i$  is the decade that the artist  $i$  began his music career.

We believe that there is a linear relationship between the influence of a genre and the proportion of the genre to the total number of artists in a given year. So:

$$LS(t) = \frac{\text{the increase in number of the genre}}{\text{the number of the genre not include the increase}} \quad (5)$$

## 3.2 Model Implementation and Result Analysis

We implement our proposed model with the data given. The modified VPR model has four parameters in total:  $a, \alpha, \sigma, d$ .  $d$  is set to 0.85, learned from the research of Professor Zhang [1].  $\sigma$  is set as 0.137. Parameter  $\alpha$  is used to adjust the time factor. So  $\alpha \in [0.1, 1]$ . We set it as 0.3

Here, we only select artists beginning their music careers in 2000 and 2010 as data for model implementation. The relationship between Influencers and followers is shown in Figure 1. Set the follower as target and the Influencer as source. As shown in Figure 1, Kings of Leon's different music influences on Saving Abel, Sheppard, Young the Giant, The Lumineers, George



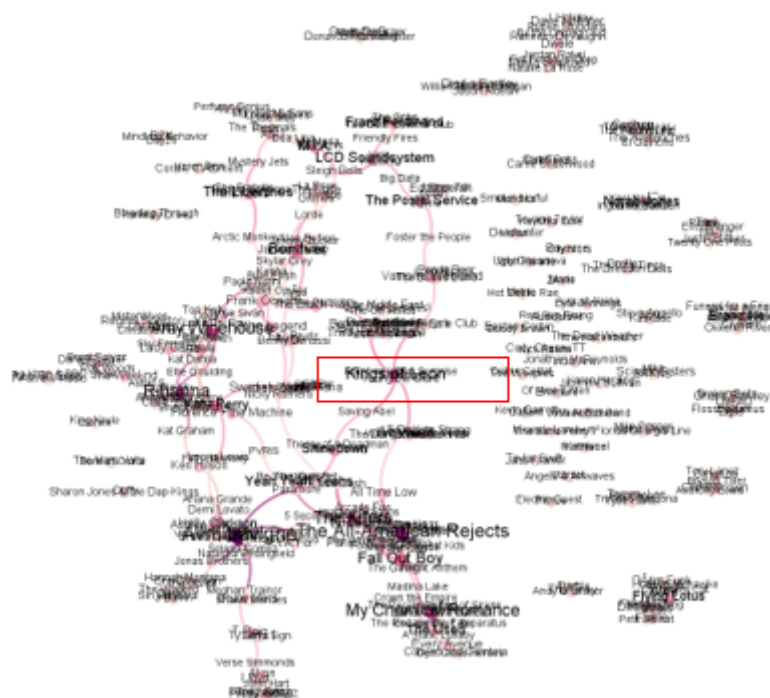


Figure 1: Directed influence network based on the influence association between artists.



Figure 2: Directed diagram based on the interaction between artists.

Ezra, Rainbow Kittens Surprise and Rainbow Kittens Surprise are shown. The size of the circle indicates the amount of influence.

We can find that this model can well represent the music influence. Combined with the data\_influence file, it can be seen that Kings of Leon influences Saving Abel, Sheppard, Young the Giant, The Lumineers, George Ezra, Rainbow Kittens Surprise, Rainbow Kittens Surprise, all of which are pop/rock genre, Saving Abel, Sheppard, Young the Giant started their musical career in 2000, but by 2010 no one had followed them. The Lumineers started Pop/Rock in 2000 and had a following The Oh Hellos, who started in 2010. As we can see from the percentage of genres new additions, Pop/Rock was more popular in 2000, so MI (Kings of Leon) was the biggest of the bunch.

## 4 The Measurement of Music Similarity

### 4.1 Data Cleaning

In order to obtain real and credible results, we first conduct relevant processing on the data set. We found that in the full\_music\_data data set, there are some songs composed by two or more artists, and according to the statistics, there are 4410 such songs. There may be artists of different genres among the artists of these songs, leading to the unknown genre of songs, which may have a certain impact on the calculation results. Therefore, all these data are eliminated. At the same time, we combined the influence\_data data set, matched the genre of corresponding artists with the unique identifier artists\_id, and filled the result back into full\_music\_data data set to form a one-on-one correspondence relationship with the song, and eliminated the data that failed to match successfully. For songs whose genre is Unknown, we should also disregard this part of data and eliminate all of them. By analyzing the characteristics contained in the music, we find that release\_data is the concrete expression of year, so we delete the release\_data characteristics that are not easy to quantify. At this point, the data cleansing is complete, and we end up with a new data set of 79,704 data items named full\_music\_data\_sortedtemp.

### 4.2 Data Standardization

Each song contains 15 traits in the newly generated data set, all of which are used for similarity calculations. We found that the dimensionality of these variables is not consistent,

tempo	mode	year
energy	key	explicit
valence	liveness	duration_ms
loudness	acousticness	popularity
danceability	instrumentalness	speechiness

Figure 3: All 15 characteristics of music.

and to eliminate the effect of dimensionality on the final result, we process the data through data normalization. Here we use the min-max normalization process. For sequences  $x_1, x_2, \dots, x_n$ , we use formula followed to calculate:

$$y_i = \frac{x_i - \min_{1 \leq j \leq n} \{x_j\}}{\max_{1 \leq j \leq n} \{x_j\} - \min_{1 \leq j \leq n} \{x_j\}} \quad (6)$$

New sequence  $y_1, y_2, \dots, y_n \in [0, 1]$  can be obtained and there's no dimension. Using this method, the 15 variables are processed successively, and their values are normalized to the interval of 0-1, and the influence of dimension on the final result is eliminated.

### 4.3 Similarity Calculation

For music similarity, we calculate by using cosine similarity [2]. Cosine similarity is a method of computing similarity. This method first maps the individual index data to vector space, and then measures the inner product between two individuals vector space. The closer the Angle between two individual vectors is to 0 degree, that is, the greater the cosine value of the Angle is, the higher the similarity between two individuals is. The closer the Angle between two individuals is to 180 degrees, the smaller the cosine value of the Angle is, indicating the lower the similarity. As can be seen from Figure 1, compared with vector B, vector A and vector C have higher similarity.  $\cos \theta$  is defined as the similarity between two vectors, and the

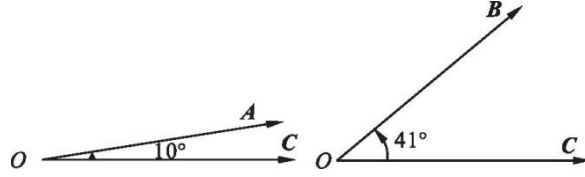


Figure 4: Schematic illustration of cosine similarity

value range is  $[-1, 1]$ . The cosine similarity of the two individuals is calculated by the following formula:

$$\cos \theta = \frac{Y_i \cdot Y_s}{|Y_i| \cdot |Y_s|} = \frac{(y_{i1}, y_{i2}, \dots, y_{in}) \cdot (y_{s1}, y_{s2}, \dots, y_{sn})}{\sqrt{\sum_{j=1}^n (y_{ij})^2} \cdot \sqrt{\sum_{j=1}^n (y_{sj})^2}} \quad (7)$$

$Y_i$  and  $Y_s$  represent the  $i^{th}$  and  $s^{th}$  different vector respectively.

Using this method, the 15 characteristics of each song are taken as a vector and substituted into the formula to calculate the similarity of the two songs.

For the degree of similarity between genres, songs are classified according to their genres. Due to the large amount of data, its not realistic nor necessary to do all the calculation. Therefore, our team extracted 10 songs from each genre randomly and made it as the sample for the further calculation. Compare the data and type calculating results into an Excel file ( named sim.xlsx.) Then average data are calculated in order to compute the similarity of same and different genres. The results depicting similarity among various genres are as follows:

As shown in the figure, the value of each point is the similarity between the two genres marked by the horizontal and vertical coordinates of the point. The diagonal represents the similarity of artists of the same type. The closer its value is to 1, the greater the similarity is. It is not hard to find that all genres, except Pop/Rock, are slightly more similar to themselves than to other genres, consistent with the conventional wisdom. It can be argued that each genre has its own characteristics, used to distinguish each genre. Take Reggae as an example: It can be found that the self-similarity within the Reggae genre is 0.94, higher than the similarity between Reggae and other genres, where Reggae and Stage&Screen has the lowest similarity, which is 0.79.

## 5 Correlation between Different Genres

In order to analyze the music styles of different genres, we first classify songs according to genres, and calculate the average value of each characteristic of songs in the same genres as the

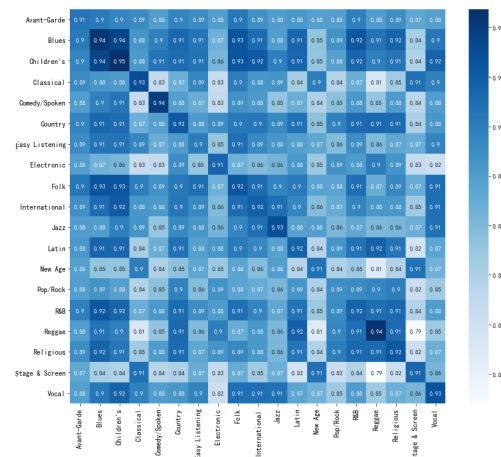


Figure 5: Similarity between genres.

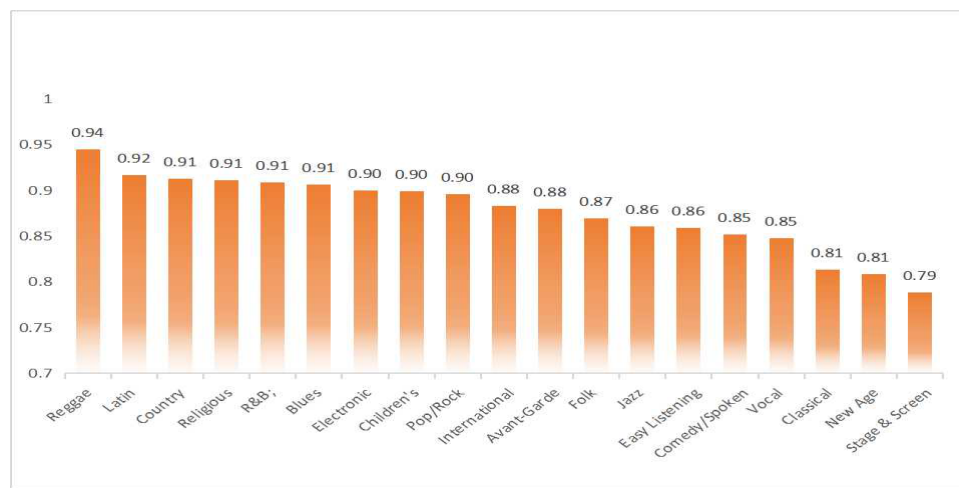


Figure 6: The similarity between reggae and other genres.

characteristics of genres. But considering that there are 15 characteristics, the data dimension is too high for further analysis, so we adopt the principal component analysis method to reduce the dimension of these 15 variables.

## 5.1 Principal Component Analysis

Principal component analysis (PCA) is a common method for comprehensive evaluation. The principle is to transform multiple index problems into a few comprehensive indexes by dimension reduction, and simplify high-dimensional space problems into low dimensional space problems[3]. The comprehensive index formed by dimension reduction is a linear combination of original variables. They are not related to each other and can explain most of the information contained in the original index. The basic process is as follows:

### 5.1.1 Data Standardization

In question 2, the data has been standardized,so we skip this step.

### 5.1.2 Feasibility Test

KMO and Bartlett test were carried out for each variable. The closer KMO value is to 1, the better the correlation of variables is, and the more suitable this method is; the closer the significance value  $p$  is to 0, the stronger its significance is, and the more suitable this method is. The KMO test value of these 15 variables is  $0.535 > 0.5$ ,  $P = 0 < 0.01$ , which indicates that the data can be analyzed by principal component analysis, but the performance may be poor.

### 5.1.3 Principal Component Extraction

the number of principal components is usually determined according to the principle that the eigenvalue is greater than or equal to 1. The analysis results of matrix eigenvalue and cumulative contribution rate are obtained by using SPSS software, as shown in the following figure: The

factor	Characteristic root	Variance percentage	Cumulative contribution rate
Z1	5.824	38.829%	38.829%
Z2	3.376	22.509%	61.338%
Z3	2.727	18.182%	79.520%
Z4	1.226	8.174%	87.694%
Z5	0.708	4.717%	92.411%
Z6	0.486	3.237%	95.648%
Z7	0.244	1.623%	97.271%
Z8	0.186	1.241%	98.512%
Z9	0.081	0.537%	99.049%
Z10	0.062	0.412%	99.461%
Z11	0.042	0.277%	99.738%
Z12	0.024	0.158%	99.896%
Z13	0.009	0.062%	99.958%
Z14	0.004	0.029%	99.987%
Z15	0.002	0.013%	100.000%

Figure 7: Results of principal component analysis.

results show that the eigenvalues of the first four principal components are greater than 1, and the cumulative contribution rate is 87.694%, so we choose the first four factors as principal components.

The slope of the first four principal component eigenvalues is larger than that of the later ones, which proves the rationality of extracting the four principal components. In order to analyze the meaning of the four principal components, the new load of each variable on the four principal components is obtained by the method of quartic maximum rotation. The thermodynamic diagram of factor load matrix is drawn as follows:

The higher the load of the variable in the factor, that is, the closer it is to 1 or - 1, the more the factor can reflect the correlation of the variable. By analyzing this graph, we can get the characteristics mainly reflected by four principal components

Factor1: valence,danceability,loudness,tempo,energy,instrumentaness

Factor2: year,populrity,acousticness

Factor3: speechiness,explicit,liveness

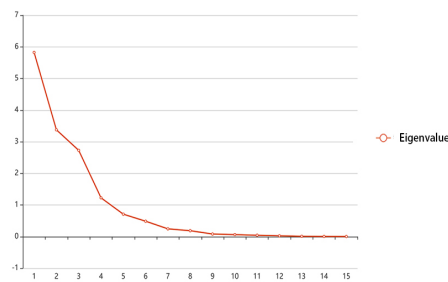


Figure 8: Gravel map



Figure 9: Thermodynamic diagram of factor loading matrix.

Factor4: duration\_ms, key, mode

### 5.1.4 Calculate the Principal Component Value

The eigenvector  $a_j$  of each principal component is obtained by using the following formula

$$a_j = \frac{Y_{ij}}{\sqrt{Z_j}} \quad (8)$$

$Y_{ij}$  represents the load coefficient of each principal component (i represents the i-th principal component, j represents the j-th index), and  $Z_j$  represents the eigenvalue. Then the standardized evaluation matrix elements are multiplied by the corresponding eigenvalues to obtain four principal component scores F1, F2, F3 and F4.

## 5.2 Result Analysis

We then substitute 15 variables for each genre into the formula to get 4 principal component scores for each genre.

Through the above steps, we reduced the dimension of 15 variables of each genre to 4 variables, and the results are shown in the figure below:

F1		F2		F3		F4	
Reggae	3.29	Comedy/Spoken	1.55	Comedy/Spoken	0.35	New Age	0.63
Latin	3.21	Children's	0.86	Electronic	0.25	Religious	0.59
R&B	3.14	International	0.65	New Age	0.02	Stage & Screen	0.54
Pop/Rock	3.14	Vocal	0.59	Stage & Screen	-0.06	Pop/Rock	0.51
Religious	3.09	Blues	0.50	Pop/Rock	-0.10	Country	0.51
Electronic	3.00	Folk	0.50	Religious	-0.11	Electronic	0.50
Country	2.92	Classical	0.44	R&B	-0.13	Comedy/Spoken	0.47
Children's	2.76	Country	0.32	Avant-Garde	-0.15	Avant-Garde	0.41
Comedy/Spoken	2.57	Jazz	0.28	Reggae	-0.19	Latin	0.39
Blues	2.57	Easy Listening	0.26	Latin	-0.21	Classical	0.37
Folk	2.16	Reggae	0.20	Jazz	-0.31	Folk	0.36
International	2.12	R&B	0.17	Easy Listening	-0.33	R&B	0.35
Avant-Garde	1.95	Religious	0.16	Classical	-0.35	Easy Listening	0.35
Vocal	1.92	Latin	0.13	Country	-0.39	Children's	0.32
Jazz	1.86	Avant-Garde	0.11	Vocal	-0.46	Blues	0.30
Easy Listening	1.63	Pop/Rock	0.07	Folk	-0.47	Vocal	0.29
Stage & Screen	1.27	Stage & Screen	0.06	Blues	-0.48	Reggae	0.25
New Age	1.18	New Age	-0.15	International	-0.51	Jazz	0.23
Classical	1.04	Electronic	-0.33	Children's	-0.68	International	0.19

Figure 10: Results of principal component analysis.

- for F1, genres value the more extreme, the genre may in valence and danceability, loudness, tempo, energy, instrumentaness has distinctive features in one or more characteristics.

We can see from the picture F1 larger Reggae, Latin, R&B, Pop/Rock, Religious, Electronic, these genres have a strong sense of rhythm, high strength, high loudness etc, thus has the high valence, danceability, loudness, tempo, energy value, in line with the actual situation. Genres with less F1 are Classical, New Age, Stage& Screen, all of which have the characteristics of serenity, ease and leisure, and are mainly composed of pure music with few vocals, so these genres have high instrumentanness value.

- for F2, genres value the more extreme, this genre may in the year, popularity, acoustiness has distinctive features in one or more characteristics.

Analyzing the above figure, Comedy/Spoken usually directly adopts by the artist's original sound, so it has a high acoustiness value. However, with the development of time in recent years, Electronic has become more and more popular among people, especially young people, and thus has a higher popularity. In addition, Electronic music is created by Electronic instruments, and most of them have technology or enhancements and electrical amplification by computers and other devices, so they have low acoustiness.

- for F3, it reflects the main speechiness, explicit, liveness values.

Referring to the Figure above, we have an insight of Comedy/Spoken. Among them, Comedy/Spoken is mainly made up of dialogue (high speechiness), most of which include lyrics (high explicit value), and some songs include audience reactions. In most comedies, audience laughter (high liveness) is added to enhance the appealingness (which is in line with the calculated results).

- For F4, duration\_ms, Key, Mode values are reflected. These three values are not related to the style of music, so there is little difference in F4 values between different genres.

Through the analysis above, we can draw a general way to distinguish genres, that is, to distinguish them through some characteristics. However, since the KMO test value calculated in the previous principal component analysis method is 0.535, the result may be poor. Therefore, there may be some deviation in the result for the music genres with no obvious relevant characteristics, that is, the genres in the middle position in the above table.

## 6 The Influence of Influencers on Followers

In this section, to determine if the similarity in the DATA\_influence data set indicates that the identified influencers actually influence their respective artists. We did a correlation analysis between music\_influence and similarity.

### 6.1 The Definition of Music\_Influence

We simplified the 1.2 model to define music\_influence(influencers, followers). Suppose that in the constructed influence network, the influencer directly connected has the greatest influence on the follower. The influence of indirectly connected influencers on follower decreases with the increase of interval, time and other factors. For two unconnected artists, it is considered that they have no influence on each other.

## 6.2 The Relevance between Music\_Influence(influencers,followers) and Similarity

In this section, we verify the theoretical results through the actual experimental results, that is, we use the similarity between influencers and followers to verify the influence of the former on the latter. If the similarity between followers and influencers can show that influencers affect followers, we can study it through the correlation between similarity and influence. If the correlation is high, influencers do affect followers.

Using music\_influence(influencers,followers) in section 6.1, We group the data according to the interval size to explore the correlation. We use Pearson correlation coefficient to calculate the correlation between "influencer's influence on follower" and "similarity of the two"[4].

Pearson correlation coefficient formula is as follows

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X}) (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (9)$$

Through verification, we find that the average similarity of influencer and follower gradually

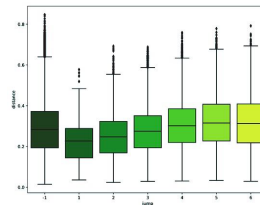


Figure 11: Similarities between genres.

decreases with the increase of interval, and converges to the similarity between the two artists without mutual influence. At the same time, it also shows that influencer does have an impact on follower.

## 6.3 Discussion on the Appeal of Characteristics

In order to explore whether some musical features are more "infectious" than others, we turn this question into which musical features will affect the popularity of music. For this reason, we calculate the correlation coefficient between other characteristics and music popularity in turn, here we use Pearson correlation coefficient.

60000 songs were randomly selected from the standardized data set. The Pearson correlation coefficient between popularity and other characteristics of these songs was calculated. In order to better analyze the results, 1000 songs were randomly selected from 60000 songs, and the scatter plot of some characteristics and popularity was drawn.

Based on the above results, the closer Pearson correlation coefficient is to 1 or - 1, the more influence this characteristics has on the popularity of songs. It can be found from the figure that popularity is positively correlated with year, energy and loudness, that is, the larger these indicators are, the more popular music is.

For the year characteristics, with the increase of years and the popularity of the Internet, songs have a wider spread than before. People have more and more ways to obtain and listen to music. And excellent songs can be found by more people, so the popularity is higher.



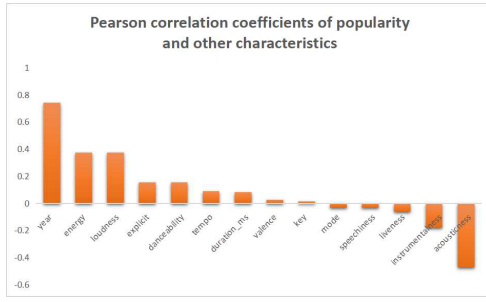


Figure 12: Pearson correlation coefficients of popularity and other characteristics.

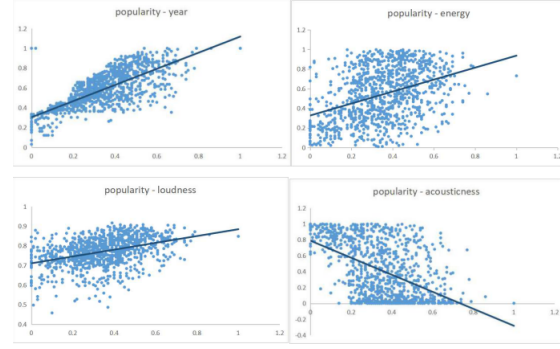


Figure 13: Relationship between musical characteristics and popularity.

Combined with the previous data, we can find that Electronic, Pop / Rock, Religious and Latin all have high energy and loudness values, and they also have high popularity. Secondly, there is a negative correlation between the popularity and the acousticness characteristics, that is, the lower acousticness is, the more people like the music. It is embodied in the Electronic genre, which has the lowest acoustinccness among all genres, and at the same time, it is also the most popular among all genres.

In conclusion, we speculate that the following characteristics may be more "infectious":

- energy, loudness, explicit, danceability, instrumentalness, acousticness.

## 6.4 The Influence of Influencers on Followers' Styles

In order to explore whether "influencers" really affect the music created by followers, we select James Taylor with ID 204386 as an example to discuss. James Taylor is a Pop/Rock artist. He has 52 followers, including 41 Pop / Rock, 7 Country and 4 Folk. The network node diagram is as follows:

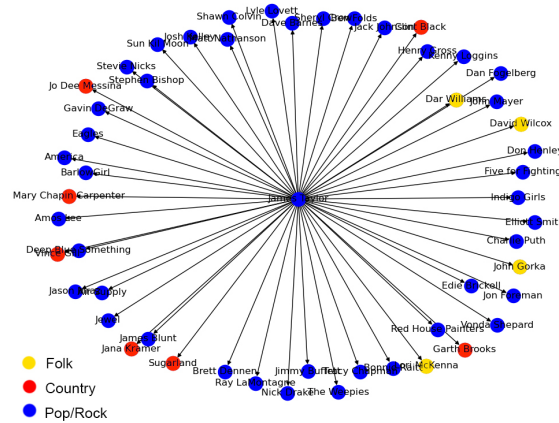


Figure 14: Artists' network

In order to judge the influence of James Taylor on his followers, we establish the following formula to calculate:

$$sim_i = \frac{\sum \frac{|c_{JamesTaylor}^i - c_{Follower_i}^i|}{c_{JamesTaylor}^i}}{n} \quad (10)$$

where  $c_{JamesTaylor}^i$  represents James Taylors i-th characteristics.  $c_{Follower_i}^j$  represents his j-th followers i-th characteristics.  $n$  represents the number of James Taylors followers.  $sim_i$  represents the alienation degree of followers and influencers on the i-th index. The larger the value is, the greater the difference between them in this characteristics is. The smaller the value is, the more similar they are in this characteristics.

We combine data\_ by\_ Artist and influence\_ Data set to get the relevant indicators of James Taylor and his 52 followers. We choose to remove the mode, key, count that are not related to music style, and leave 11 indicators as the values we want to calculate. The results are as follows: This shows that James Taylor and his followers have a high similarity in the indicators

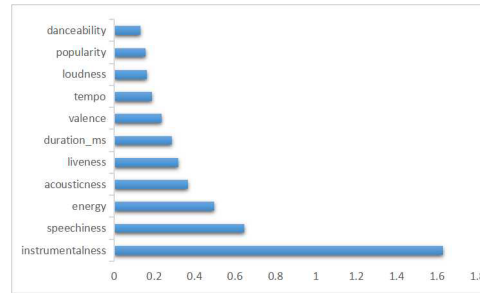


Figure 15: Similarity of music characteristics.

of danceability, popularity, loudness, tempo and validity, and have a great difference in the indicators of instrumentalness, speech and energy. We know that James Taylor belongs to Pop / Rock genre, and most of his followers also belong to Pop / Rock genre. Combining with the results of the principal component analysis in question 3, we can know that danceability, loudness, tempo and validity are the main characteristics to distinguish Pop / Rock genre. Therefore, we can think that James Taylor, as an influencer, influenced the above characteristics of his followers and made them join the Pop / Rock genre with James Taylor to a large extent. So to sum up, the influencer will influence the music created by the followers to a certain extent.

## 7 The Insight into Evlutionary and Revolutionary Trends of Artists and Genres

### 7.1 Main work

Due to the rapid changes in politics, economy and culture, music is also developing rapidly. The evolution process of music is also the significant change process of music characteristics of different genres. So in this section, we propose revolutionary features to indicate the major changes in the evolution of music.

#### 7.1.1 The Definition of Revolutionary Characteristics

The evolution of music is actually the result of the change of different genres. The changes of different genres are reflected in the great changes of characteristics. Therefore, we define the revolutionary characteristics of music evolution as atriple as follows:

$$(characteristic, genre, decade)$$

Among them, genre refers to a certain music genre, decade refers to the period of significant changes in the music genre, and characteristic refers to the indicator of significant changes in the music genre during the period.

### 7.1.2 The Definition of Significant Change

We quantify the large change of features as  $\left| \frac{Currentvalue - Lastyear'svalue}{Lastyear'svalue} \right|$ . If it is greater than 0.7, we can think that a significant change has taken place.

Through the given data, we demonstrate the possible revolution in the evolution of music. As shown in figure (a), during the period of 1930-1940, energy declined significantly in most genres. It can be found that there were significant changes in the music world during this period. As shown in figure (b), during the period of 1930-1940, each genre had significant changes in key. In fact, it was true that music tended to be stable after 1940. From 1930 to 1940, the tempo of each school had an obvious downward trend. In 1930-1940, the values of various genres experienced a sharp decline.

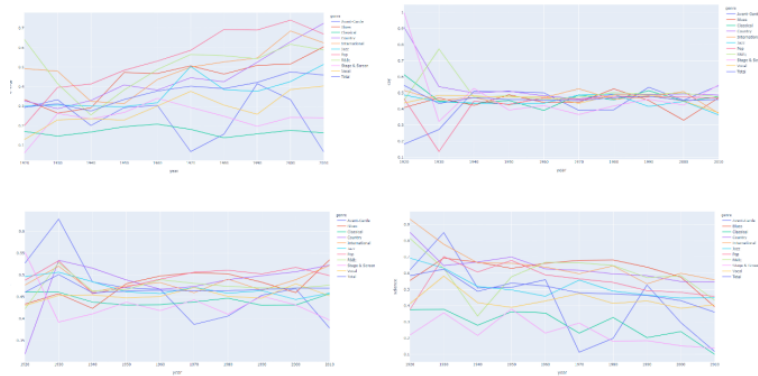


Figure 16: Music characteristics change over time.

## 7.2 Establishment of Indicators to Measure the Reformer

In the development of music, some artists, as models of genres, create music that is very different from other artists in the genres. In this section, we described the reformers, and found the qualified reformers according to the established indicators. Through the verification of one of the reformers, it is found that the established indicators can find the reformer well.

### 7.2.1 The Definition of Innovative

Firstly, the innovator should be an artist with great musical influence, which has been discussed in model 1; secondly, the innovator should be highly innovative.

$$revo(i) = \left| \frac{1}{similarity(i, influencer)} \right| \quad (11)$$

If the similarity between the reformer and his influencer is very high, it does not have strong innovation, on the contrary, it has strong innovation. In addition, the change trend of each index of the reformer in a certain period is the same as that of the genre. We use the line chart to see if the candidate has a period of high consistency with the trend of the school's characteristics. We can see that Ben Webster is an artist with great influence and strong innovation.

Then we continue to compare the trend of genre parameters with that of the artist.

It can be found that the characteristic change of ID 7554032 is consistent with Pop / Rock, which shows that the Beatles is a reformer of Pop / Rock.

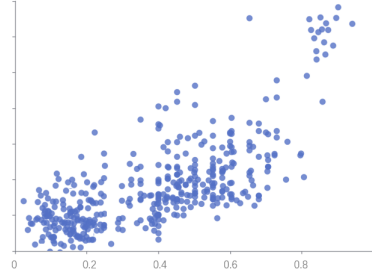


Figure 17: Relationship between creativity and popularity.

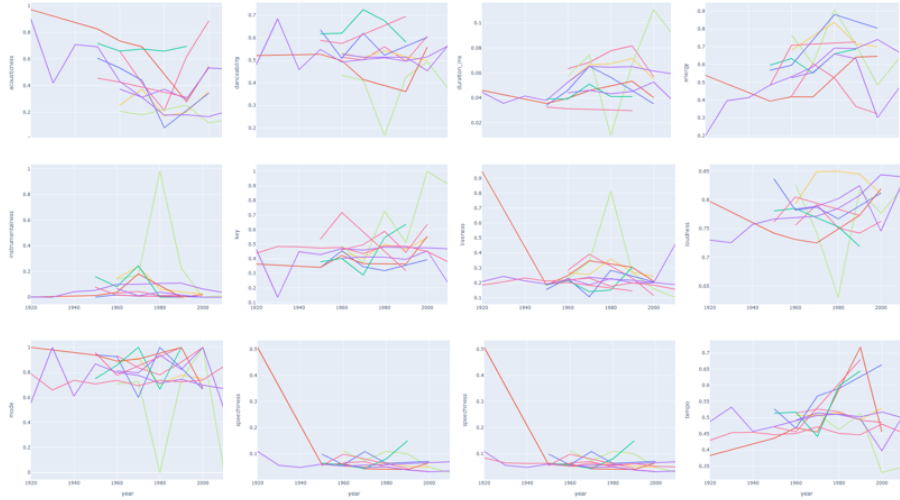


Figure 18: Changes in the characteristics of music over time.

## 8 Indicators that Reveal the Dynamic Influencers

### 8.1 Define the Influence in the Process of Music Evolution

Artists are more or less influenced by other artists. As for the influence on artists, we now define such an index to measure it:

$$influence = \frac{\delta \cdot \sigma}{|\sigma|} \quad (12)$$

$$\delta = artist_{years2} - artist_{years1} \quad (13)$$

$$\sigma = influencer_{years2} - influencer_{years1} \quad (14)$$

$artist_{years1}$  and  $influencer_{years1}$  are the average vectors of artist and influencer in the first time period (such as the 1990s).  $artist_{years2}$  is the average vector of artist in the second time period (such as the 2000s). Therefore,  $\delta$  indicates the difference of artist's song style in these two periods,  $\sigma$  expresses the difference between influencer and artist in the first period.  $influence$  is the module length of the projection of  $\delta$  in the direction of  $\sigma$ . Because  $\sigma$  is transformed into a unit vector,  $influence$  is the degree of deviation of artist to influencer in fact.

### 8.2 The Evolution of Artists

In order to fully test the effectiveness of the index, we select an artist who is influenced by artists in many fields as an example. An artist whose id is 55939 and whose name is Azymuth

is influenced by many artists. Select the influencers who are active during the active period of Azymuth (1970s-2000s), select 5 influencers manually with Excel, and calculate their influence on Azymuth in different years according to the following table.

The changes of influencer and azymuth from 1980 s to 2010s are shown in the figure below. Because the data feature has 15 dimensions, it is not easy to display, so we use principal component analysis to reduce the dimension to two dimensions before displaying.

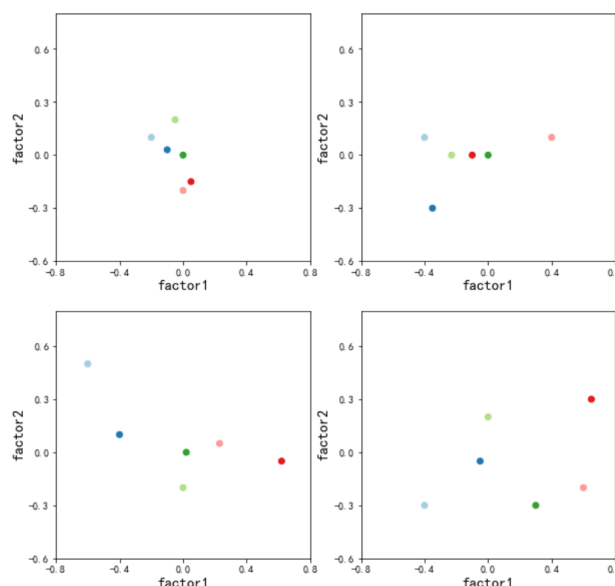


Figure 19: Azymuth and his influencers' changes in style from 1980s to 2010s.

Factor1 mainly represents the rhythm of music, Factor2 mainly represents the trendiness of music. The red dots in the picture indicate the style of Azymuth. It is not difficult to see that his music style has increased in both factor 1 (rhythm) and factor 2 (trendiness). In other words, during the active period of Azymuth, his music works became more rhythmic and trendy with the passage of time.

## 9 Influence of Music in Time and Circumstances

Society, economy and culture are influencing the development of music, and music is also influencing them imperceptibly.

Music has different popularity in different periods. Since 1950, the number of new artists has increased significantly, and the energy in music has also increased significantly. Music has brought people strength and vitality, inspired people to come out of the shadow of World War II and actively build society. The number of Pop / Rock also began to surge. At that time, Pop / Rock preached beliefs and doctrines, and many artists hoped to appeal to society and oppose fascism through music.

## 10 Sensitivity Analysis

As the construction of our model in 4.1, there are four parameters i.e.,  $\sigma$ ,  $d$ ,  $\alpha$ ,  $a$ . We show the sensitivity of our model on these four parameters.

Figure 26 depicts the change of music influence (MI) calculated by the model when the four parameters change respectively when the other three parameters are optimal. As can be seen

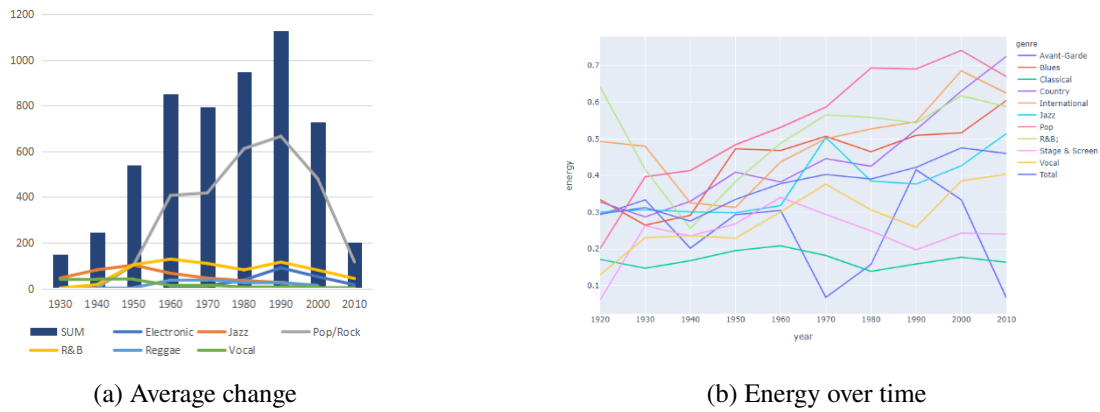


Figure 20: Total number of major music genres over time.

from the figure, with the gradual increase of  $\sigma$ , MI shows a downward trend; with the gradual increase of  $d, \alpha, a$ , MI shows an upward trend. When these parameters change, MI does not fluctuate sharply, but has a gentle upward or downward trend. Because these four parameters restrict each other, when one of them changes, the final MI does not change much. For this data set, the ideal values of  $\sigma, d, \alpha, a$  are:  $[0.1, 0.2]$ ,  $[0.8, 0.9]$ ,  $[0.3, 0.4]$ ,  $[0.6, 0.7]$ .

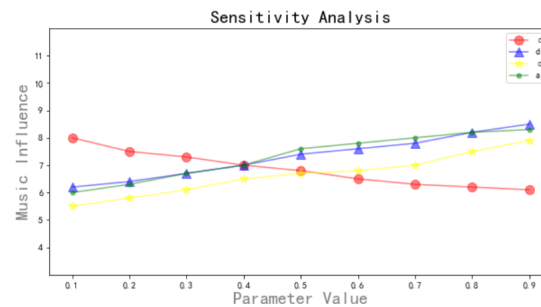


Figure 21: Sensitivity Analysis

## 11 Model Evaluation

### 11.1 Strengths

When defining music influence, the music influence of each artist not only considers the number of followers, but also takes into account different degrees of indirect followers, and adds factors such as the time interval of debut and the popularity of the genre. Only in this way can the influence be more reasonable and comprehensive.

When solving problem 4, we use similarity to verify whether influencers really affect followers. We consider the Pearson correlation coefficient of similarity and influence to quantify the relationship between them; at the same time, we also classify the relationship according to the interval number between influencers and followers, so that influencers really affect followers

When solving problem 5, when looking for a reformer, we consider not only his influence on the school, but also his "innovation", and quantify the latter. After verification, the reformer is very representative.

Pay attention to the validation process, and test the results in many steps to prove the

accuracy of our model. For example, we use Mann Whitney U test to verify the reliability of our conclusion after we get the conclusion that the similarity between genres is lower than within genres.

## 11.2 Weaknesses

When analyzing evolution process, only one pop / rock with the largest number of people is selected for analysis.

When calculating genre similarity, we did not consider the changes of each genre's characteristics over time in the calculation process, which may cause certain errors.

## 12 Conclusion and Further Discussion

Music is playing an increasingly important role. In order to explore the influence of music, we propose a series of novel models to solve the sub problems from the influence of artists to genres, similarity between artists and music evolution. These models have high accuracy and robustness.

In this project, we do detailed data cleaning for the given data set. After that, we visualize our analysis visually and diversely, and explore the correlation between the data from various directions. We defined music\_ Influence, similarity and other indicators to measure the artist's influence, genre and genre, artist and artist similarity and other music indicators. Using the methods of principal component analysis and Pearson correlation coefficient, this paper explores the stable relationship among music evolution, artists and similarity, and provides valuable conclusions about music influence for ICM Association

## References

- [1] Zhang Fang. Evaluation method of academic influence based on Journal influence and time factor [J]. Journal of Nanyang Institute of technology, 2020, 12(06): 46-52+68.
- [2] Chen Dali, Shen Yantao, Xie Bingzhu, Ma Yingyi. Best coach selection algorithm based on cosine similarity model [J]. Journal of Northeast University (NATURAL SCIENCE EDITION), 2014, 35 (12): 1697-1700
- [3] Xiao Jianzhong, Xiao Yutong, Li jiastrontium. High quality development evaluation of China's natural gas market based on principal component analysis [J]. Journal of China University of Petroleum (SOCIAL SCIENCE EDITION), 2021 (01): 20-27
- [4] Edelmann Dominic, Móri Tamás F., Székely Gábor J.. On relationships between the Pearson and the distance correlation coefficients. 2021, 169:108960-.

Dear, ICM Society:

An artist's music influence is restricted by many factors, such as his age, region, personal style, personal repertoire preference theme, popularity and so on. These factors will directly or indirectly reflect the influence of the artist in the whole music history. After the relevant data processing of the data set, our team selected artists who appeared in the 2000s and 2010s to create an influencer and follower network, and proposed an improved VPR model to quantitatively measure the music influence of artists. The value reflects the contribution rate of the artist to the development of music style. Some musicians with high music influence will also have a significant change in the history of music evolution. Music influence is directly reflected in the number of followers of the musician in the network diagram. After adding a time dimension to the network, combined with the influence of music, it is easy to find the great musicians who have played a crucial role in the history of music development. Through the study of these musicians, we can get the general direction of change of relevant music genres in the whole history of music evolution. It is of great significance to trace the origin of music in the past and predict the development of music in the future.

At present, our model is based on the given data. For more abundant and comprehensive data, our model is still applicable, but it may have the disadvantages of too much calculation and tedious calculation process. To solve this problem, we can first classify the data according to the genre, establish a network diagram for each genre to calculate the music influence, and finally integrate the results to get the final analysis results. For the problems of similarity analysis, characteristics change with time, our results will be more accurate with the abundance of data. If the type of music index is more abundant than the given data, the principal component analysis can also be used to reduce the dimension, reduce the number of variables and reduce the difficulty of calculation. To sum up, our solution has high applicability in the face of richer and more comprehensive data.

Through our calculation and consulting relevant information, we find that music has certain epochal character. After the Second World War, blues, jazz and other music became more popular, attracted countless audiences, and these music genres developed rapidly. After that, with the change of public psychology, especially the growing rebellious spirit of teenagers, rock and pop music has gradually become the most popular genre. In the new era, with the development of computer technology and the increase of route of transmission, music gradually presents the trend of diversified development. Various genres exchange and fusion, forming our prosperous music world today. Therefore, the study of music should not only focus on the excavation of the style of music, but also make a correlation analysis with the background of the times and the origin of music. Specific background of the times and specific social culture produce specific music style. Music, in turn, influences social culture. It can be said that music originates from culture but is higher than culture. Good music can not only make people happy physically and mentally, but also promote people to pursue a better life. Some music can even represent the spirit of a nation or even a country. Therefore, the study of music should be multi-dimensional. We should explore the connotation of music and reveal the social and cultural phenomenon.

Sincerely yours,

Your friends



# Appendices

## Appendix A Tools and software

Paper written and generated via  $\text{\LaTeX}$ , free distribution.

Graph generated and calculation using MATLAB R2019a.

## Appendix B The Codes

Here are simulation programmes we used in our model as follow.

### B.1 Task4 Contagious

---

```
def multipl(a,b):
    sumofab=0.0
    for i in range(len(a)):
        temp=a[i]*b[i]
        sumofab+=temp
    return sumofab

def corrcoeff(x,y):
    n=len(x)
    sum1=sum(x)
    sum2=sum(y)
    sumofxy=multipl(x,y)
    sumofx2 = sum([pow(i,2) for i in x])
    sumofy2 = sum([pow(j,2) for j in y])
    num=sumofxy-(float(sum1)*float(sum2)/n)
    den=sqrt((sumofx2-float(sum1**2)/n)*(sumofy2-float(sum2**2)/n))
    return num/den
```

---