

# The Sound of Music

## Summary

Music is the pearl in the palm of human art and the treasure of the human spiritual world. The influence of music has been part of human society from the very beginning, as an important part of cultural heritage. In order to analyze the evolution of music, our team designed a series of models that studied the influence of musicians, the similarity of music, and the trend of musical characteristics over time.

For the measurement of musicians' influence, we selected a part of the data as an example, and built a network of musicians through the selected data. We adopt two algorithms, the point degree centrality algorithm and the eigenvector centrality algorithm, and combine them using the scale coefficient  $c$ . We finally calculated the influence coefficient of the musician, which measures the influence of the musician.

The similarity of musicians or genres is measured by the similarity of influencers and followers, and we use an algorithm that combines Euclidean distance similarity and cosine similarity. To combine the two algorithms, we define the scale coefficient  $\lambda$ , which measures the size of the similarity between the two genres or the similarity between the two musicians by calculating the size.

For the analysis of musical characteristics over time, we use the moving average method to predict a feature, and finally determine whether a major historical change has occurred in the year by calculating whether the size of the gap between the actual value and the predicted value meets the hypothesis test conditions we set. In the years we calculated for historical change, we counted the number of followers of musicians who released music before and after the change to determine who was most likely to be the revolutionaries who led the musical revolution.

In addition, our team combines the transformation of music with historical factors, summarizes the historical reasons for the change in music, and analyzes the response of music to these social or historical events.

In addition, we performed a sensitivity analysis of our model at the end of the paper, which ensured the stability and effectiveness of our model when dealing with problems.

**Keywords:**cosine similarity algorithm;moving average algorithm;

## Contents

<b>1 Introduction .....</b>	<b>3</b>
1.1 Problem Background and Restatement .....	3
1.2 Our Work .....	4
<b>2 Assumptions and Justifications .....</b>	<b>5</b>
<b>3 Notations .....</b>	<b>6</b>
<b>4 Network analysis of music impacts .....</b>	<b>7</b>
4.1 Problem analysis .....	7
4.2 Results and Analysis .....	9
<b>5. Musical similarity model .....</b>	<b>10</b>
5.1 Problem analysis .....	10
5.2 Global similarity measurement of music .....	11
5.3 Measures of local similarity of music .....	12
5.4 Results testing and analysis .....	13
<b>6 Analysis of similarity and influence of musical genres .....</b>	<b>14</b>
6.1 Analysis of similarity of musical genres .....	14
6.2 Influences between and within musical genres .....	15
<b>7 Changes in music genres over time .....</b>	<b>16</b>
<b>8 Impact Analysis .....</b>	<b>17</b>
8.1 Impact Evaluation .....	17
8.2 Contagion Analysis of Impact .....	18
<b>9 The music revolution .....</b>	<b>19</b>
9.1 An analysis of the musical revolution .....	19
9.2 Artists representing revolutionaries .....	21
<b>10 Dynamic indicators in the evolution of music genres .....</b>	<b>22</b>
<b>11 The influence of music .....</b>	<b>25</b>
11.1 Acousticness of music .....	25
11.2 Changes in prevalence of each category over time .....	26
11.3 Valence of music .....	27
<b>12 Model Evaluation and Further Discussion .....</b>	<b>28</b>
<b>13 Document .....</b>	<b>29</b>

# 1 Introduction

## 1.1 Problem Background and Restatement

Music is an important cultural heritage of human society and is highly influential. Musicians may be influenced by their predecessors or contemporaries to create similar styles of music, or they may be influenced by changes within society to create new genres of art. By studying the characteristic relationships between songs, we can understand the interplay between artists and also the development of music over time. With the data of artists and genres over the years, a model can be built to measure musical influences.

Taking the background information and restricted conditions identified in the problem statement into consideration, the following problems are required to be solved:

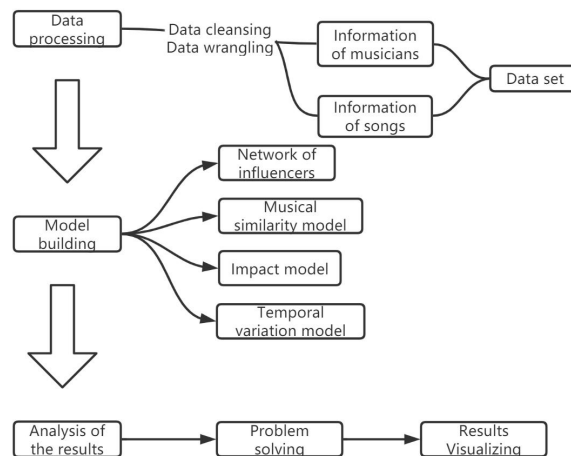
- Problem 1: Create a directed network connecting influencers and followers, and establish the parameters in this network that capture ‘musical influence’. Create and describe a sub-network of the directed influencer network, use this sub-network to explore a subset of musical influences and describe what the ‘musical influences’ in the sub-network reveal.
  - Problem 2 Develop measures of music similarity. And use the measure to judge whether artists of the same genre or artists of different genres are more similar.
  - Problem 3 Tell the difference between genres. How do genres change? Is there any similarity between genres?
  - Problem 4 Does the similarity data reported in the ‘data\_influence’ data set mean that the influencers actually influence other artists? Whether the influencers have more contagious characters or they play similar roles in influencing certain artists’ music?
  - Problem 5 Are there characteristics that might signify music revolution? Which artists are revolutionaries in the network?
  - Problem 6 Determine the indicators that reveal the dynamic influencers and explain, as time pass by, how the genre(s) or artist(s) change.
  - Problem 7 How can social, political, technological or other changes, like computers, be identified from the network?
- Music is an important cultural heritage of human society and is highly influential. Musicians may be influenced by their predecessors or contemporaries to create similar styles of music, or they may be

influenced by changes within society to create new genres of art. By studying the characteristic relationships between songs, we can understand the interplay between artists and also the development of music over time. With the data of artists and genres over the years, a model can be built to measure musical influences.

Taking the background information and restricted conditions identified in the problem statement into consideration, the following problems are required to be solved:

- Problem 1: Create a directed network connecting influencers and followers, and establish the parameters in this network that capture ‘musical influence’. Create and describe a sub-network of the directed influencer network, use this sub-network to explore a subset of musical influences and describe what the ‘musical influences’ in the sub-network reveal.
- Problem 2 Develop measures of music similarity. And use the measure to judge whether artists of the same genre or artists of different genres are more similar.
- Problem 3 Tell the difference between genres. How do genres change? Is there any similarity between genres?
- Problem 4 Does the similarity data reported in the ‘data\_influence’ data set mean that the influencers actually influence other artists? Whether the influencers have more contagious characters or they play similar roles in influencing certain artists’ music?
- Problem 5 Are there characteristics that might signify music revolution? Which artists are revolutionaries in the network?
- Problem 6 Determine the indicators that reveal the dynamic influencers and explain, as time passes, how the genre(s) or artist(s) change.
- Problem 7 How can social, political, technological or other changes, like computers, be identified from the network?

## **1.2 Our Work**



Picture 1:our work

As shown in Picture 1, we obtain the music influence score based on the weighting process and visualize it to obtain the weighting network. Meanwhile, we further analyze the similarities, differences, developments and connections, etc., of music genres to get the main content of this paper.

## 2 Assumptions and Justifications

In order for our work to continue well, we made the following assumptions:

1. Assume that the data is accurate and there is no falsified data: This means that our analysis is factual.
2. For music that is unknown to musicians, we did not include these data in the model for calculation due to its unknown genre.
3. Suppose the time of the musician's debut is the year they first released the song, because the exact year of the musician's premiere is unknown.
4. Suppose that there is only one genre of musicians, and according to known data, all music of musicians belongs to this one genre only.
5. Assuming that all musical features we choose to calculate similarity are of equal status, we assume that they have equal weight on the impact of the final result.
6. Assume that affected writers will be influenced by their influencers throughout their creations, that is, assume that all of their works are theoretically influenced by their influencers.

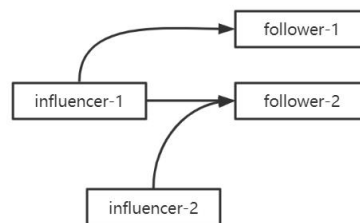
### 3 Notations

Symbol Name	Meaning of Symbols
$D_{C_i}$	The degree centrality of node i.
$E_{C_i}$	Importance measurement of node i
$d_{ij}$	Indicates whether there is a directed edge between node i and node j.
$w_{ij}$	Represents the weight of the edge.
c	Scale factor of the model.
$MI_i$	Artist i's influence indicator.
$d$	Euclidean distance of features of two musical compositions.
$S_E$	Euclidean similarity of two musical compositions.
$S_{cos}$	Cosine similarity of two musical compositions.
$\lambda$	Scalet factor of the model.
$S_{global}$	Global similarity of features of two pieces of music.
$m_i$	The attribute value of the I-th row of a music feature
$m_{min}$	The minimum value of the attribute value in the i-th row of a music feature
$m_{max}$	The maximum value of the attribute value in the ith row of a music feature
$w_i$	Data after normalizing the attribute value of the ith row of a music feature
$t_{ij}$	The difference between the i-th feature and the j-th song under the i-th feature.
$gnere_i$	The mean value of the difference degree of the music genre in the i-th feature.
$\bar{\mu}$	The mean value of musical characteristics over a ten-year period.
$\sigma_p$	The mean square error of musical features over ten years.
$MI_{test}$	Influence coefficient test value after changing the scale factor.
$MI_{fact}$	Influence coefficient value under the actual use parameter of task
$S_{global_{test}}$	The global similarity test value after changing the scale factor.
$\alpha$	Metrics in sensitivity analysis about MI
$\beta$	Metrics in sensitivity analysis about similarity

## 4 Network analysis of music impacts

### 4.1 Problem analysis

To analyze the influence of 5,854 artists on each other, our first step was to build an unweighted directed network in which the nodes represent an artist and the two nodes are directly connected to represent the influence relationship, starting with the influencer and pointing to his followers. Each artist has several followers and influencers.

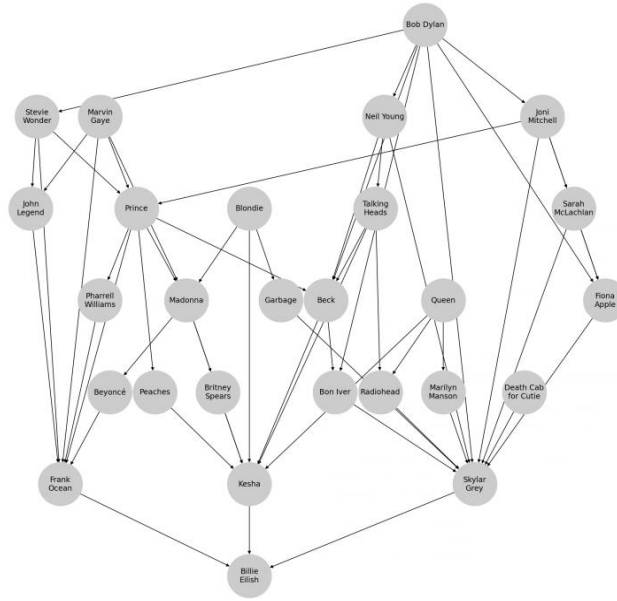


Picture 2: Influence relationships

When analyzing the influence of an artist, it is not difficult to find that whether it is the same genre or not, an artist is always influenced by several other artists, and its influencers are influenced by others, so in fact, an artist is often directly and indirectly influenced by many artists.

When analyzing the influence of an artist, it is not difficult to find that whether it is the same genre or not, an artist is always influenced by several other artists, and its influencers are influenced by others, so in fact, an artist is often directly and indirectly influenced by many artists.

We randomly selected an artist named Billie Eilish, selected some data, looked for her influencers, and built a network like this.



Picture 3: influence the network

After the graph is established, start calculating the node importance.

For this directed graph, we choose the outdegree of the node as an indicator to measure its importance. The more degrees of a point, the more important the point. The implication behind this is that the more followers an artist has, obviously the more influence that artist has.

Definition 1: The centrality of the point  $i$  is:

$$D_{C_i} = \sum_{j=1}^n w_{ij} d_{ij}$$

$N$  is the total number of nodes of the network, and  $d_{ij}$  is used to determine whether there is a directed edge between node  $i$  and node  $j$ . A value of 1 means that  $i$  and  $j$  have no edge directly, and a 0 means that there is an edge between  $i$  and  $j$ .

The eigenvector centrality algorithm measures the transmission of node influences, and the basic idea of centrality is that the centrality of a node is a function of the centrality of neighboring nodes. A relationship from a high-score node contributes more to a node's score than a relationship from a low-score node, and a node with a high score means that it is connected to many high-score nodes. Eigenvector centrality emphasizes the surrounding environment in which the node is located.

Definition 2: The importance measurement algorithm for node  $i$  is as follows:



$$E_{C_i} = c \sum_{j=1}^n a_{ij} x_j$$

C is the proportionality constant,  $x = [x_1, x_2, x_3, \dots, x_n]^T$ .

We combine the eigenvector centrality algorithm and the point centrality algorithm to comprehensively evaluate the nodes in the network, that is, the influence of an artist in the network is comprehensively evaluated by the number of followers and the importance of its followers.

The modified model  $E_{C_i}$  is:

$$E_{C_i} = \sum_{j=1}^n w_{ij} d_{ij} + c \sum_{j=1}^n a_{ij} x_j$$

In this,  $\sum_{j=1}^n w_{ij} d_{ij}$  is the point degree centrality  $D_{C_i}$  defined earlier. After iteration,  $E_{C_i}$  converges to a fixed value.

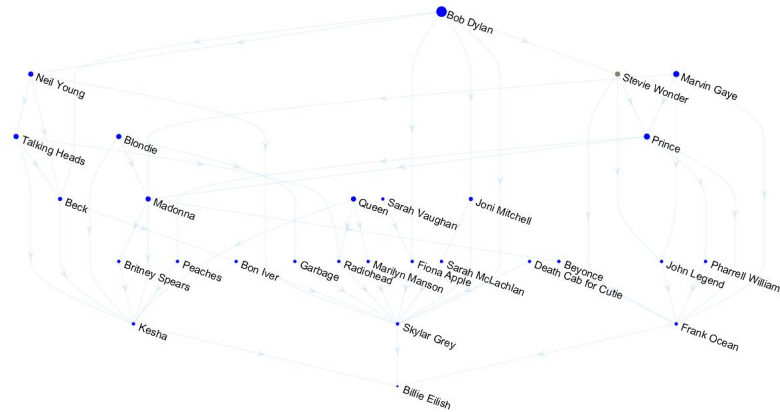
Definition 3: The influence of Artist i is:

$$MI_i = \frac{E_{C_i}}{\max\{E_{C_j}, j = 1, 2, \dots, n\}}$$

Use  $MI_i$  to measure an artist's impact.

## 4.2 Results and Analysis

The calculations can be done to get the impact of each of the artists in the network. We use the size of the point to indicate the importance of the point, the larger the point means the greater the influence of the point, and the smaller the point means the less influence of the point.

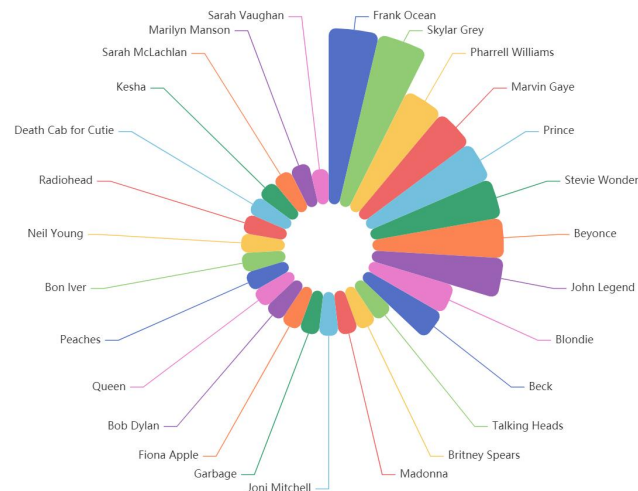


Picture 4: network reflecting influence

As can be seen in the figure, when an artist has more followers and the greater the influence of followers, the greater the point represented by the artist himself, that is, the greater the influence of the artist. For example, Neil Young in the picture,

in our grid, he has influenced many people, including Beck, and the people he influenced are also influential, and the reaction in the figure represents his point is larger.

For Billie Eilish, she was influenced by others in the grid as shown in the image:



Picture 5: influencers of Billie Eilish

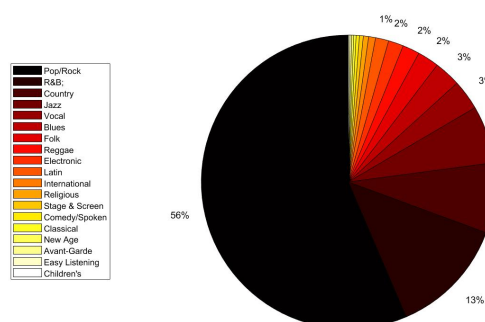
It can be seen that there are direct influences, such as Frank Ocean has a greater influence on Billie Eilish, while Kesha, who has an indirect influence and is less important in itself, has less influence on Billie Eilish.

## 5. Musical similarity model

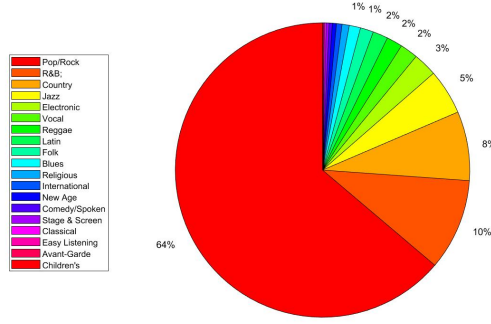
### 5.1 Problem analysis

We exclude some of the musicians in the influence\_data file whose genre belongs to unknown. full\_music\_data a small number of songs in the file are choral songs that do not fully represent the song and genre style of a single musician, and we exclude this data.

First of all, we calculate the proportions of the influencers and followers in the influence\_data, and after sorting, we can get the results as shown in the figure below:



Picture 6: influencer genre distribution



Picture 7: followers genre distribution

As can be seen from the pie chart drawn, more than half of the musicians belong to the Pop/Rock genre, followed by the most numerous genres: R&B, Country, Jazz, etc. Next, we take Pop/Rock, R&B, and Country as examples of the most popular genres for music similarity and genre similarity analysis.

## 5.2 Global similarity measurement of music

Suppose there are two pieces of music, use their features to construct the feature vectors A and B of the music, use the Euclidean similarity to measure the distance similarity of the two musical feature vectors, and use the cosine similarity to measure the angular direction difference of the two musical feature vectors

### Euclidean similarity:

Using the absolute distance between the two vectors, the absolute difference in the eigenvalues of the two songs is reflected, and d is defined as the absolute distance between the two eigenvectors A and B:

$$d = \sqrt{\sum_{i=1}^n (A_i - B_i)^2}$$

Defining the Euclidean similarity  $S_E(A, B) = \frac{1}{1+d}$ , the similarity is reduced to the interval (0,1], and the closer the two musical features are, the greater the Euclidean similarity  $S_E(A, B)$ .

### Cosine similarity:

Using the cosine value of the angle between the two vectors, the angle relationship between the eigenvectors of the two songs is reflected, and  $\cos\langle A, B \rangle$  are the angles between the two vectors

$$\cos\langle A, B \rangle = \frac{\vec{A} \cdot \vec{B}}{||A|| ||B||}$$

Define cosine similarity  $S_{cos}(A, B) = \frac{1}{2} + \frac{\vec{A} \cdot \vec{B}}{2||A|| ||B||} = \frac{1}{2} + \frac{1}{2} \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$ , and

reduce the value of similarity to the interval [0,1], the closer the two musical features, the greater the cosine similarity  $S_{cos}(A,B)$ .

**Global similarity of the two music:**

$$S_g(A,B) = \frac{S_E(A,B) + S_{cos}(A,B)}{2}$$

The interval is (0,1], and the greater the similarity, the closer the characteristics of the two music.

### 5.3 Measures of local similarity of music

Due to the high feature vector dimension of music, when comparing similarities between different music, it may be better to measure musical similarity by some features or a combination of some features. For example, the loudness and rhythm of the two pieces of music are similar, and although there are other similarities, they can still be considered similar songs. Within the same genre, there may be similarities in some characteristics, but there are many differences in others.

On this basis, by analyzing the characteristics of a certain genre of music, analyzing the comparison of the similarity of music and the global similarity under these characteristics or feature combinations, if this characteristic or feature combination measures better, it can be considered that these characteristics or feature combinations can better reflect the characteristics of this genre, and in these dimensions, the similarity of songs of the same genre is higher, and this genre is distinguished from other genres.

Suppose that a total of songs of a genre are selected, and the attribute value of each song in the  $i$ th dimension is defined as the difference between the song and the song of the same genre in the  $i$ th dimension

$$t_{ji} = \sum_{k=1}^s (y_i - y_k)^2$$

The mean of variance in the  $i$ -th dimension of the genre is defined as:

$$gene_i = \frac{\sum_{j=1}^s t_{ji}}{s}$$

This gives you the difference value of a genre in each dimension. From them, the dimensions with the smallest difference value are selected, and the number of selections depends on the difference between the difference values, so we select four dimensions as the common feature of the songs of this genre, that is, to measure the partial similarity of a genre that distinguishes it from other genres.

## 5.4 Results testing and analysis

Using two files, influence\_data and full\_music\_data, we selected four musicians from each of the three genres of Pop/Rock, R&B and Country, with a total of 12 songs. As shown in the figure below, next, these 12 songs are used to analyze the similarity between genres and within genres.

artist_names	song_title (censored)	genre
Bob Dylan	When I Got Troubles	Pop/Rock
P.O.D.	Goodbye for Now	Pop/Rock
Korn	Seed	Pop/Rock
Pixies	Trompe le Monde	Pop/Rock
TLC	No Scrubs	R&B
R. Kelly	**** For You	R&B
Joe	Another Used To Be	R&B
Syreeta	Black Maybe	R&B
Kenny Chesney	When I See This Bar	Country
Randy Travis	**** Tonk Moon	Country
Deana Carter	We Danced Anyway	Country
Brad Paisley	Anything Like Me	Country

Picture 8: Songs information

After that, the similarity results of the 12 songs calculated according to the global similarity are as follows:

1	0.693265416	0.592767957	0.5431718	0.748520212	0.703337878	0.644073157	0.721350092	0.658678165	0.732593402	0.73238667	0.692267646
0.693265416	1	0.936574138	0.770444709	0.902159889	0.812109205	0.912538872	0.806817767	0.916157374	0.854952196	0.915449838	0.874922109
0.592767957	0.936574138	1	0.794616025	0.760250273	0.689259797	0.846546978	0.83486878	0.815302175	0.710451099	0.790784461	0.829389112
0.5431718	0.770444709	0.794616025	1	0.602602687	0.540693389	0.593289629	0.766769281	0.618248024	0.668242623	0.726178428	0.704716344
0.748520212	0.902159889	0.760250273	0.602602687	1	0.924807744	0.866963886	0.660301605	0.91757629	0.85374451	0.86645182	0.798283594
0.703337878	0.812109205	0.689259797	0.540693389	0.924807744	1	0.906425168	0.641571445	0.927246308	0.849756885	0.854509712	0.878748556
0.644073157	0.912538872	0.846546978	0.593289629	0.866963886	0.906425168	1	0.743382483	0.955236525	0.873044931	0.901861605	0.93331067
0.721350092	0.806817767	0.83486878	0.766769281	0.660301605	0.641571445	0.743382483	1	0.716740851	0.694533448	0.751510173	0.816183328
0.658678165	0.916157374	0.815302175	0.618248024	0.91757629	0.927246308	0.955236525	0.716740851	1	0.886805077	0.908557573	0.877443422
0.732593402	0.854952196	0.710451099	0.668242623	0.85374451	0.849756885	0.873044931	0.694533448	0.886805077	1	0.971460145	0.806501113
0.73238667	0.915449838	0.790784461	0.726178428	0.86645182	0.854509712	0.901861605	0.751510173	0.908557573	0.971460145	1	0.865215966
0.692267646	0.874922109	0.829389112	0.704716344	0.798283594	0.878748556	0.93331067	0.816183328	0.877443422	0.806501113	0.865215966	1

Picture 9: the result of the similarity calculation

From the results of using local similarity, after using the similarity of songs in each genre and the similarity of songs within genres as the standard for measuring the musical similarity between genres and between genres, the similarity of songs within genres increases compared with the global similarity, and at the same time, when using local similarity, the genre can be clearly distinguished from other genres. From the examples we chose, the similarities of valence, danceability, key, and speechiness are what distinguish the Pop/Rock genre from the others. The similarities between loudness, instrumentalness, liveness, duration\_ms are what distinguish the R&B genre from the others. The similarities of danceability, loudness, instrumentalness, and speechiness are what distinguish the Country genre from the others.

## 6 Analysis of similarity and influence of musical genres

### 6.1 Analysis of similarity of musical genres

In order to analyze the characteristics of genres, we use the characteristics of genres without music. We use the full\_music\_data to perform similarity analysis on the description of the music.

We build matrices for each genre's data. First, the data is processed, removing useless data, error data, and blank values in the data, and then normalizing the data.

Definition 4: The normalization formula is calculated as follows

$$w_i = \frac{(m_i - m_{\min})}{(m_{\max} - m_{\min})}$$

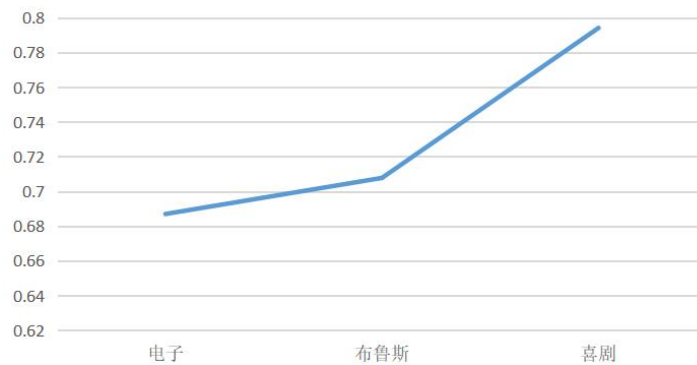
$m_{\max}$  and  $m_{\min}$  is the maximum and minimum values in the data, is the current calculated value, and the result obtained is the result of normalization processing. By normalizing the data to between 0~1, it is convenient for later calculations.

After processing the data, we compare the two genres to analyze their similarity. We use Euclidean similarity and cosine similarity to calculate similarities between genres synthetically. The calculation formula is as follows:

$$S_{global}(A, B) = \varphi S_E + (1 - \varphi) S_{Cos}$$

$$S_{Cos} = \frac{1 + \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} + \sqrt{\sum_{i=1}^n B_i^2}}}{2}$$

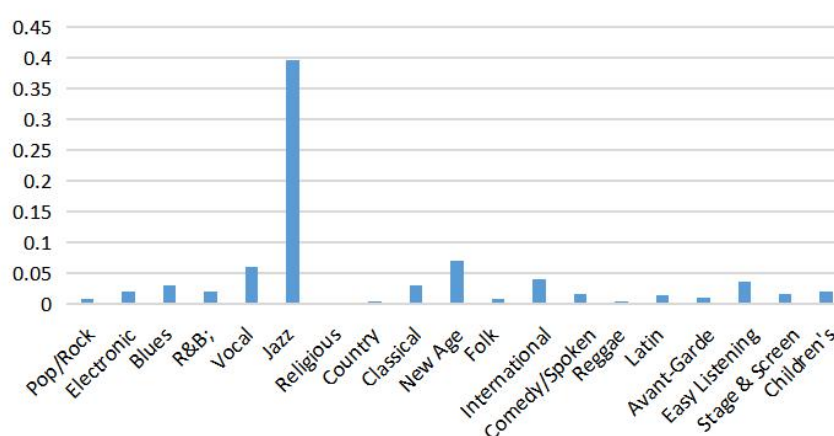
Where  $\varphi$  is a constant, we assign a value of 0.7, which is the Euclidean similarity. Based on this, we calculated that comedy music has a similarity of 0.794099, comedy music has a similarity of 0.686927 to electronic music, and comedy music has a similarity of 0.707721 to blues music.



Picture 10: Comedy music is similar to other music

## 6.2 Influences between and within musical genres

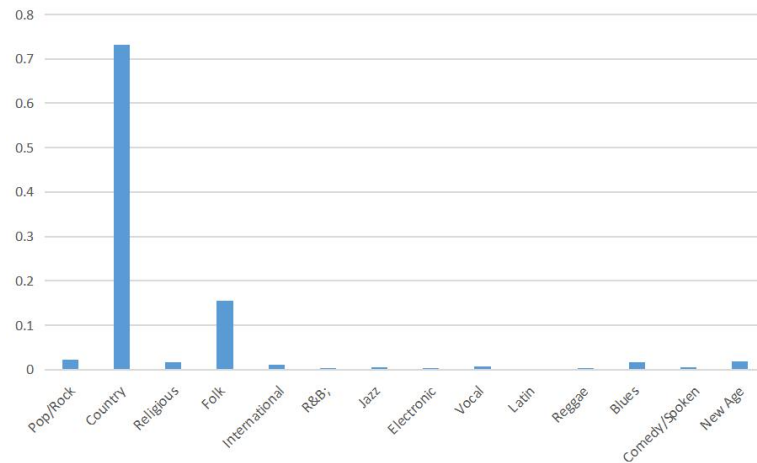
The development of art is always influenced by different influences, as we showed in the previous grid, a musician always has more or less other influences on other musicians, always has its followers, and a musical genre that develops will also have an impact on other musical genres or itself. Based on the data in the influence\_data, we counted the styles of followers of all musicians in each genre.



Picture 11: Jazz's influence

We counted the followers of musicians in the Jazz genre and got the number of followers of their different genres. Since the number of musicians varies from genre to genre, we don't directly judge Jazz's influence on different genres based on the number of followers. We count the total number of records affected by each genre, and then use the number of records of Jazz's impact on each genre and the total number of records affected by the genre to get the size of the genre affected by Jazz, which is represented in the form of a bar chart as shown in the figure. We can clearly see that jazz has influenced this genre much more than it has influenced other genres.

In the same way, we analyze the influence of the Country genre:

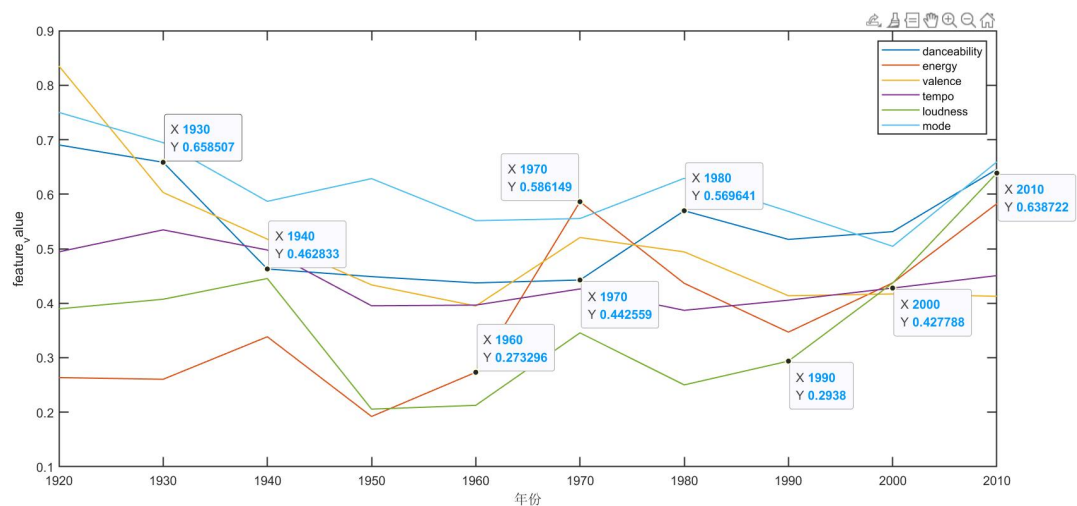


Picture 12: Country's influence

## 7 Changes in music genres over time

Music is not static. In different time periods, the characteristics of each music genre are different. With the development of time, the style and characteristics of each genre of songs have changed. We take time as the coordinate axis, count and average all the music characteristics of the music genre in that year, and get the change of music characteristics within the genre with the development of time, so as to analyze the change of music genre.

Taking Jazz music as an example, we counted the changes of Jazz-style songs over time with four indicators such as energy, and then normalized the data to draw the following table.



Picture 13: changes

It can be seen that the changes of tempo and danceability are relatively stable. From the 1920s to the 1960s, the valence value of Jazz music showed a downward trend. In the 1970s, energy and valence showed a large increase. Growth continued in the 1990s while valence gradually declined thereafter.

A musician influences other people's styles etc., and a follower usually has many influencers.

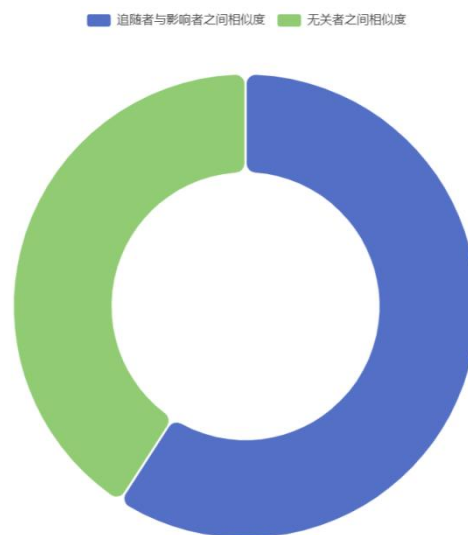


In order to judge whether a follower is really influenced by its influencer, we randomly selected several sets of data, from which, we calculated the correlation between all influencers and their followers, and obtained a series of correlation degree average. Next, find the degree of similarity between unrelated musicians, and find the average. Since the data for uncorrelated musicians is randomly selected, we can take multiple samples to compare the relationship between fixed and uncertain values. After many comparisons, we can analyze whether it is always greater than and determine whether influencers affect followers.

## 8 Impact Analysis

### 8.1 Impact Evaluation

A musician influences other people's styles etc., and a follower usually has many influencers. In order to judge whether a follower is really influenced by its influencer, we randomly selected several sets of data, from which, we calculated the correlation between all influencers and their followers, and obtained a series of correlation degree average. Next, find the degree of similarity between unrelated musicians, and find the average. Since the data for uncorrelated musicians is randomly selected, we can take multiple samples to compare the relationship between fixed and uncertain values. After many comparisons, we can analyze whether it is always greater than and determine whether influencers affect followers.

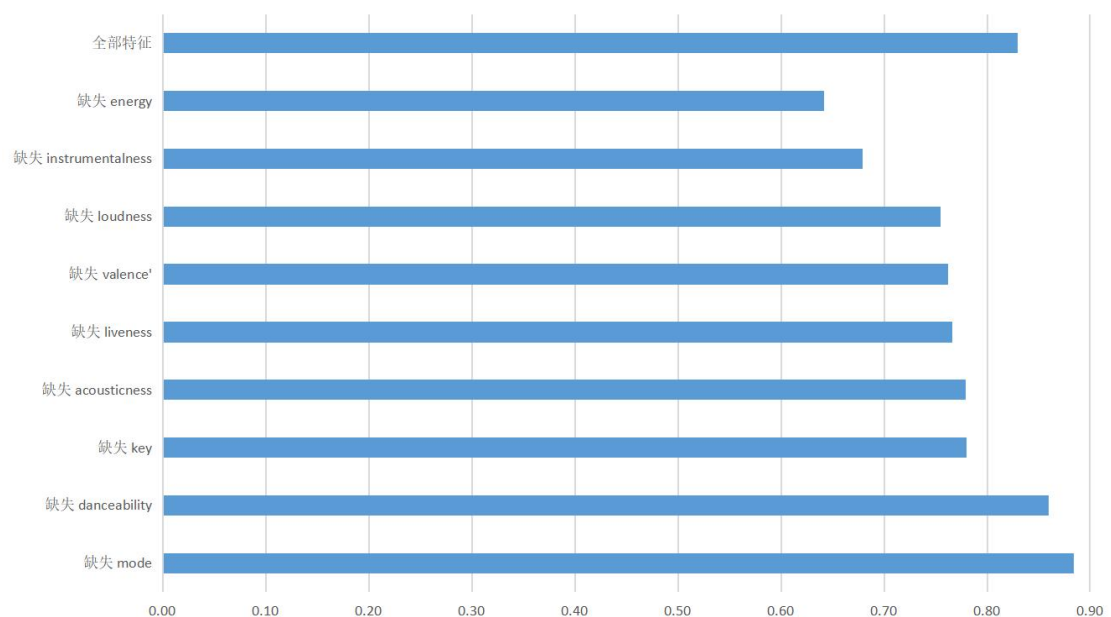


Picture 14: influencers affect

We randomly select values to compare with and, repeat the process 100 times, and get the average value, the result is shown in the figure. The results show that most, if not always, greater than, suggesting that influencers do influence their followers and the music they create to some degree.

## 8.2 Contagion Analysis of Impact

We have already analyzed that when a follower is influenced by an influencer, usually the style of his works will be close to the style of the influencer. To determine whether certain musical characteristics are more contagious, we can directly select the most influential musicians, analyze the musical characteristics of their musical compositions, and compare them with less influential musicians to compare their musical character. We still randomly selected 100 sets of follower and influencer data. After the normalized data processing is completed, follower and influencer characteristics are compared for similarity. In order to analyze which feature of the influencer has a stronger influence on followers, that is, "contagion", we choose to remove each feature value one by one, and then compare the similarity with the influencer after removing the feature value, The result is shown in the figure.



Picture 15: similarity model' result

It can be seen from the figure that after removing the energy feature, the similarity between followers and influencers has dropped significantly compared with the original, indicating that energy has the greatest impact on similarity, and energy affects the similarity between followers and influencers. Indicating that the Energy feature is more contagious.

## 9 The music revolution

### 9.1 An analysis of the musical revolution

As you can see from the previous article, music changes over time. In the history of music, it will be affected by a variety of factors such as politics and economics, and each influence will be reflected in the changes in music. Therefore, we analyze full\_music\_data datasets. Because the development process of each genre is different, although different genres have the same economic and social environment and other conditions, but the development of each genre itself is not the same, such as the popularity of The Beatles, which has a huge impact on Pop/Rock, and this type of situation is the specific situation of each genre. It is first categorized and processed by genre of the song, and then each genre is discussed separately.

In order to discuss the changes that have taken place in different musical genres, we sort the different data of the genres by time, and then count the characteristic values of the music at each time point, so we get the average data of the musical characteristics of each year as the year develops.

To calculate periods of change, we model changes in the data. Revolution means that the style of music has changed significantly compared to before, so the idea is that for a certain point in time, based on the data available before that year, the data prediction for that year, when the predicted value is far from the actual value, we think that the music has changed.

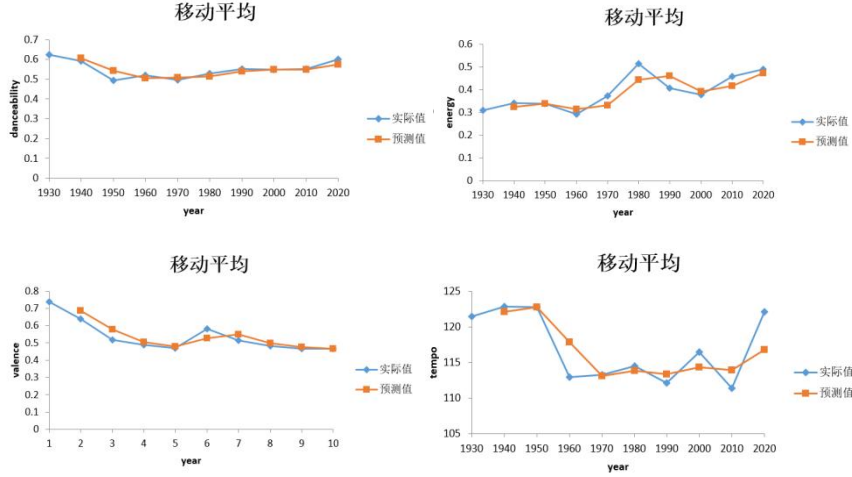
To predict the development of music, we use the moving average method for calculations. This method averages a series of changing data according to the specified number of data, and calculates the ordinal time average containing a certain number of items according to the time series one by one, so as to reflect the long-term trend. The advantage of this method is that the data given does not exist every year, some genres do not exist songs in a certain year, using this method can eliminate the influence of random fluctuations, show the direction and trend of events, and then analyze the long-term trend of the forecast sequence according to the trend line.

Definition 5: Moving average calculation formula

$$F_t = \omega_1 A_{t-1} + \omega_2 A_{t-2} + \dots + \omega_n A_{t-n}$$

In the formula,  $n$  is the period,  $\omega_i$  is the weight of each stage,  $A_{t-i}$  is the data of the time period, and  $F_t$  is the predicted value.

Let's take Jazz music as an example, take 10 years as the development cycle, and get the predicted values of each feature as shown in the figure:



Picture 16: moving average prediction results

As can be seen from the results, there are many predicted points that differ from the actual value, but not all points have produced musical changes, and in order to measure which points have had musical changes, we use prediction limits to analyze the data. Only when the difference between the predicted value and the actual value is greater than the set limit is considered to have undergone a musical change during that period.

Definition 6: Prediction Upper Prediction Lower Prediction Bound

$$\bar{\mu} = \sum_{i=1}^{10} Data_i$$

$$\sigma_p = \sqrt{\frac{1}{10} \sum_{i=1}^{10} (Data_i - \bar{\mu})^2}$$

$$P_{\max} = \bar{\mu} + 2\sigma_p$$

$$P_{\min} = \bar{\mu} - 2\sigma_p$$

The confidence interval is determined by the prediction limit as  $(P_{\min}, P_{\max})$ , the data within the interval is considered reasonable, the actual value is small from the predicted value, and the

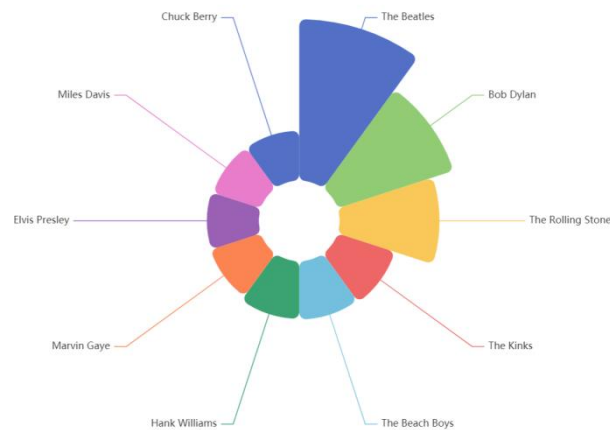
data is considered to have changed outside the confidence interval, and the time period is considered to have changed.

By comparing whether the data were within confidence intervals, we determined that there were large differences between 1940 and 1950. We think that in the '50s, jazz revolutionized. This was the time when Bebop began to develop gradually in the early 40s and quickly matured around 1945. Bebop is an important jazz style, with its faster performance of repertoire, emphasis on more complex chord changes, and the freedom of superb personal improvisation techniques, which became a revolutionary new style and new sound for jazz, laying the foundation for modern jazz, in line with the major changes in jazz in history.

## **9.2 Artists representing revolutionaries**

Although concerts are influenced by many factors, the transformation of music in one genre indicates the emergence of influential artists who gained several fans in his style and had a transformative impact on the genre. We measure a musician's influence by the number of followers they have. Moreover, when analyzing the influence of a certain musical genre by those musicians, we do not consider only the influence of outstanding musicians within the genre, but the influence of musicians of all genres, and the influential musicians within the genre, so as to comprehensively evaluate the impact on musical change. Therefore, in order to find the musicians who had a greater impact on the change of Jazz Music 2050s, we analyzed all artists at that time to measure their impact on the change in Jazz music according to the number of their followers and the number of Jazz genre musicians among their followers, as well as the musicians who had a high influence on Jazz music during this time period.

After comprehensive evaluation and ranking, we selected the top 10 musicians who influenced the transformative changes in jazz music in the 50s of the 20th century, and mapped their influence on jazz:

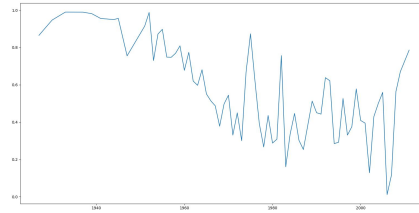


Picture 17: musicians' influence on jazz

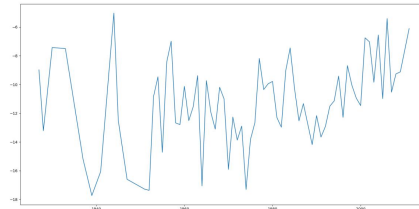
From the picture we can see that The Beatles had the greatest influence on jazz at that time, The Beatles was in a very popular stage at that time, although The Beatles' music genre was Pop/Rock rather than Jazz, but Jazz became more cheerful and faster after 1950, and we have reason to think that Jazz was indeed influenced by The Beatles in its development. Although not the same musical genre, The Beatles still had a driving force on jazz's revolution.

## 10 Dynamic indicators in the evolution of music genres

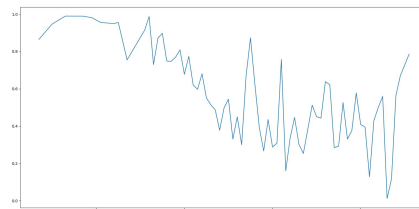
Music is the jewel in the palm of human art and the treasure of the human spiritual world. It is the product of the combined effect of people's spiritual outlook and social atmosphere in a certain era. Therefore, the characteristics of music often have the imprint of the times, which can well show the style of an era. The importance of time variables is self-evident. Very important historical events may occur within a certain period of time. As a part of social history, human beings will inevitably participate in these historical events. Similarly, the emotions and thoughts of musicians will also be affected by these important historical events, and they will intentionally or unintentionally bring these thoughts or emotions affected by historical events into their works, which leads to historical events. The time node of the event coincides with the time node of the revolution in music works, which provides a basis and support for our analysis. Based on this, our team analyzed the danceability, loudness, acousticness and popularity characteristics of blues music, and drew these characteristics in chronological order as a line chart that can read the changing trend, so as to conduct our team's analysis.



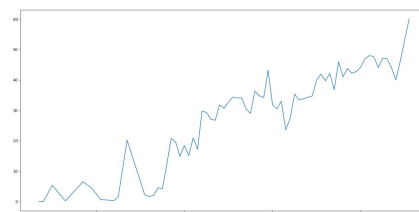
Picture 18: danceability over time



Picture 19: loudness over time



Picture 20: acousticness over time

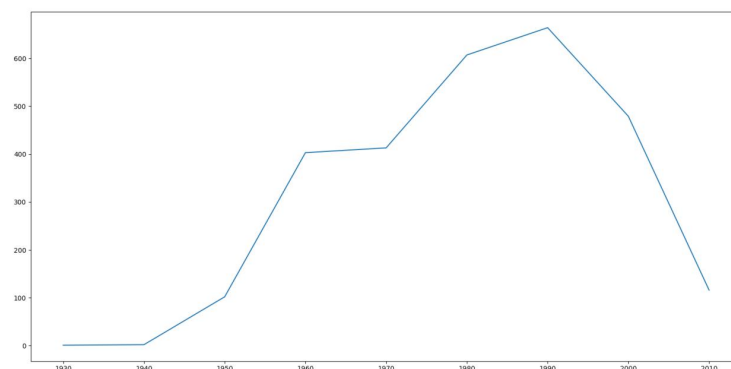


Picture 21: popularity over time

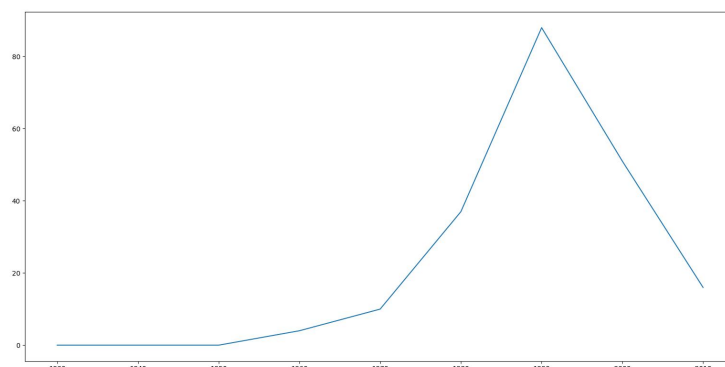
In the 1930s and 1940s, many major events took place in world history, and among them, the historical event with the most profound impact on human history is undoubtedly the Second World War, which lasted for several years. During this tragic war, the society was turbulent and human beings fell into a very heavy suffering. The hearts of countless musicians were also touched by this tragic society, which drove them to create unique music and change the world. characteristics of music. For example, the loudness characteristics of blues music declined sharply around 1940, and we have reason to infer that this was directly or indirectly caused by the Second World War.

The danceability feature of blues music has undergone obvious changes around 1970, which may be related to the famous musicians of the time, such as Stevie Wonder, James Brown, Aretha Franklin, Marvin Gaye and other famous musicians. impetus. In the subsequent time, this feature has not had a stable trend, which shows that with the gradual increase in the popularity of blues music, more musicians have joined the ranks of creating blues music, their styles are very different, and their danceability is low Among the music creators and music creators with high danceability, there are music masters who can affect the characteristics of music, which is a good evidence for the steady improvement of popularity.

Blues music is characterized by melancholy, and its popularity has steadily increased in recent decades. When we combine the other three attributes, it seems that there has not been a very obvious change. Therefore, our team speculates that this may reflect people's development in recent decades. a state of mind. We speculate that people's psychological state in these decades is fragile and emotional, and people are more vulnerable to psychological injury than before, so we find resonance in the melancholy atmosphere of blues music. We speculate that this state of people gradually intensifies with the development of society, which may have a lot to do with the rapid development of human society or the gradual increase in people's life pressure. Of course, limited by the lack of data, we do not have sufficient data to support our ideas. This is just a reasonable guess based on known data.



Picture 22:Pop/Rock debut musicians over time



Picture 23: Electronic debut musicians over time

In addition, we also found that the change in the number of debut musicians of a kind of music over time can often reflect the development of a music. For example, by observing the graph of the number of debut musicians in Pop/Rock over time, we can clearly It can be clearly seen that Pop/Rock entered a period of great development in the 1950s and 1970s, and entered its heyday around 1990. Since 1990, Pop/Rock music has shown a decline. In addition, through this image we can also analyze the origin time of a type of music. For example, pop/rock music had almost no musicians who could represent the genre debut before 1940, but after 1940, it began to grow steadily, so we have reason to infer that pop/rock originated around 1940. Similarly, we can learn that electronic music began to appear around 1950, and entered a period of great development after 1970.

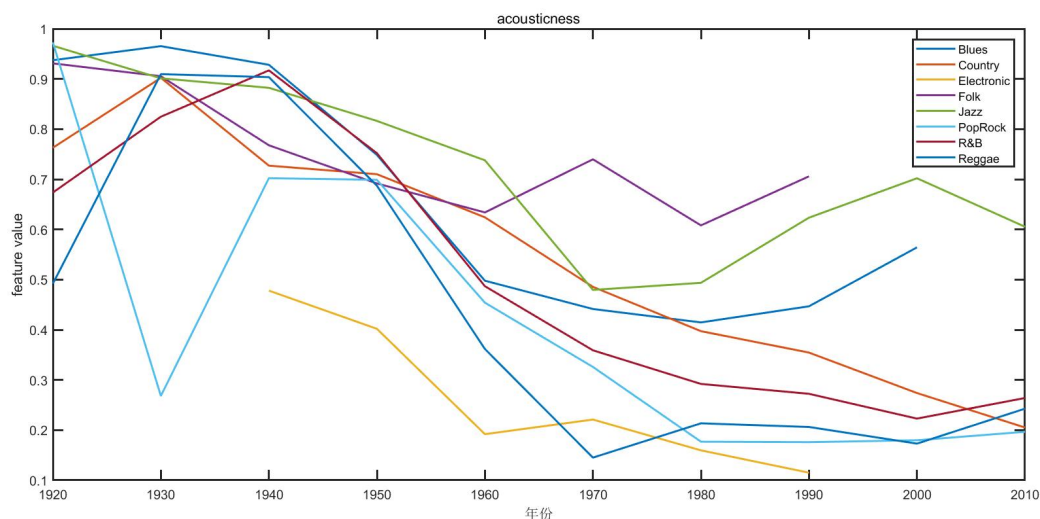


## 11 The influence of music

In order to explore the cultural impact of music on time or environment, as well as the connection between social, political or technological change, we consider the following three aspects:

### 11.1 Acousticness of music

Acousticness is a measure of confidence that an audio track is acoustic. Through it, we can know how likely it is that the song used electronic amplification or technological enhancement. When the confidence level of the acoustics is lower, the more likely the song will be enhanced with electronic amplification or technology. Therefore, we took the data of all eight types of songs in Fullmusicdata, normalized their acousticness, statistically averaged different years in units of ten years, and drew a line chart of the characteristic of several music genres over time.



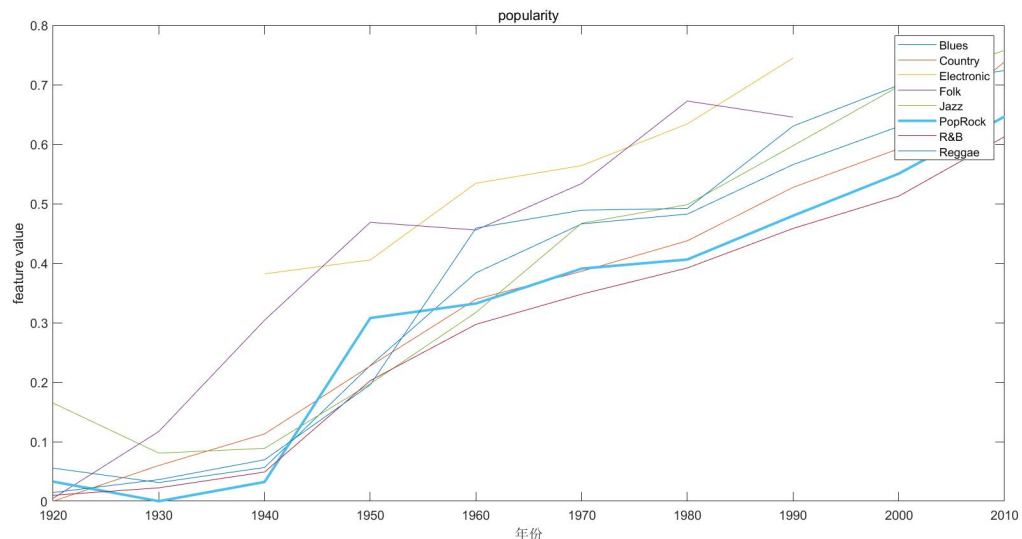
Picture 24: changes

As can be seen from the line chart, the confidence that most music genres were acoustic before 1950 was mostly high, and from the 50s to the 80s began to decline sharply, and then the decline in the value of this characteristic of each genre slowed down or recovered, that is, it stabilized.

Looking through the relevant history, we learn that the first computer Eniac was born in 1946, and in the following decades, computer and other electronic science and technology developed rapidly. The use of a large number of electronic music equipment has made the use of technology-enhanced electronic amplification in music common, and we believe that the widespread use of technology-enhanced electronic amplification has led to a successive and substantial decline in this confidence level in the decades after 1950. Until 1970, the decline in confidence slowed down in some genres and rebounded in others.

## 11.2 Changes in prevalence of each category over time

The reaction of music listeners to music often shows the development of music and the connection between social popular culture, and we plot the popularity of each category over time in the same way that we draw acousticness.



Picture 25: changes of acousticness

As can be seen from the figure:

1: After the popularity of various genres increased by leaps and bounds in the 40s and 50s of the 19th century, the popularity increased year by year in the following decades, and faster than before 1940.

Linked to historical events, technological revolutions, especially in the field of computer electronics, and the popularity of music tapes, tape recorders, etc., have made various music more popular, so it affects more listeners, so all kinds of music have become more popular year by year.

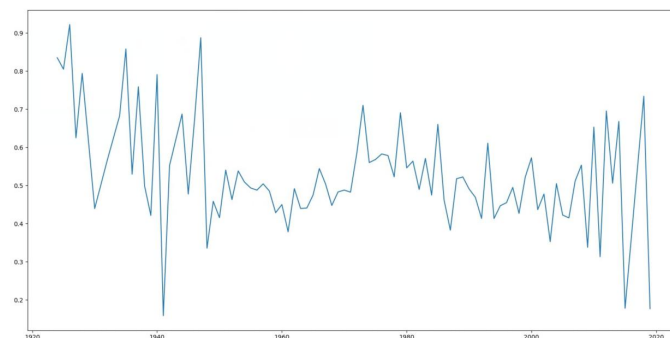
2: Pop Rock's popularity growth between 1940 and 1950 was particularly impressive. For ease of view, we have marked it with a thick line in the figure.

Through looking for information, we learned that in the 50s of the 20th century, Elvis Presley, Elvis Presley debuted, his music transcended racial and cultural boundaries, blended country music, blues music and rock music, formed a distinct personality style, shocked the pop band at that time, so that pop rock quickly swept the world orchestra, becoming an amazing miracle in the history of world popular music.

3: The statistics of electronic music were almost not popular before 1940, and the popularity explosion in 1940 was growth. Looking up Wikipedia, we learn that AEG tested recordings in the 1940s, and during this period the audio tape was invented, and then spread widely around the world, eventually leading to the development of electroacoustic tape music. The first known example of electronic-acoustic tape music was composed by Halim El-Dabba in 1944. He is widely known for his electronic music compositions that became popular in the 1950s at the Columbia-Princeton Electronic Music Center.

### 11.3 Valence of music

In order to observe whether the emotional expression of music is related to social development, we similarly draw a line chart of the valence characteristics of jazz music with the change of time.



Picture 26: changes of Valence

1: As can be seen from the figure, the emotional enthusiasm of jazz music declined significantly around 1930.

By looking at history, we can know that between 1929 and 1933, the United States had a great economic crisis, and the people lived in difficulty. People are generally negative.

2: Around 1941, the mood expressed by the sound had another obvious cliff-like decline, almost falling to the bottom.

Through our research, we learned that Pearl Harbor occurred in 1941. Japan dragged the United States into World War II. The US Pacific Fleet was badly damaged in the incident. Therefore, we infer that emotional positivity appears here, and the bottom-of-the-road fall is due to the widespread existence of negative emotions such as heavy mourning for the victims of the event, and the enthusiasm for musical expression is very low. After that, it quickly picked up.

3: Around 1950, the emotional positivity of music declined significantly again and continued to be sluggish. It didn't get better until around 1970.

By looking up the history of the United States, we find that between 1950 and 1970, the black American movement broke out. Black people were not difficult to racial segregation

and discrimination, set off nonviolent protests, and fought for democratic rights in the civil rights movement. It can be speculated that the decline in emotional enthusiasm during this period is related to the oppression of black people and the outbreak of the civil rights movement. The emotional tendencies of the songs have become positive again, which we believe has to do with the future of African Americans, and based on these results, we believe that there is an implicit relationship between musical emotions and social change.

## 12 Model Evaluation and Further Discussion

In Task 1, we use a combination of point centrality and eigenvector centrality to calculate the magnitude of influence for each person in a given network. The calculation formula is as follows:

$$E_{C_i} = c \sum_{j=1}^n w_{ij} d_{ij} + (1 - c) \sum_{j=1}^n a_{ij} x_{ij}$$

$$MI_i = \frac{E_{C_i}}{\max\{E_{C_i}, i = 1, 2, \dots, n\}}$$

In the first formula, in order to better combine the two formulas, we define the parameter  $c$ , the range of parameter  $c$  is  $[0, 1]$ , when  $c = 1$  the calculation formula degenerates to a complete point centrality calculation formula, in Task 1, we take 0.9 for  $c$ . To test its sensitivity, let's still take the network in Task1 as an example to test the magnitude of the change in the result calculated when  $c=0.9$  by changing the  $c$  values to 0.1, 0.3 and 0.6, respectively. In the test results we choose 10% as the stability threshold, i.e. When the following formula holds, the model is considered stable.

$$\left| \frac{MI_{fact} - MI_{test}}{MI_{fact}} \right| \times 100\% < 10\%$$

It is not difficult to find that the change of the  $C$  value will not cause more than 10% of the change in the final result by comparing the change range of the data, so we can consider that the model has a certain stability.

In Task 4, we combine Euler distance similarity with cosine similarity, comprehensively consider the advantages and disadvantages of the two calculation methods, and the final calculation formula is as follows:

$$S_{global} = \lambda S_E + (1 - \lambda) S_{cos}$$

In order to realize the combination of the two formulas, we define a parameter  $\lambda$ , which has a value range of  $[0, 1]$ , and when the value is 0, the global similarity degenerates to cosine similarity; When the value is 1, the global similarity degenerates to Euclidean distance similarity. In Task2, we take the  $\lambda$  value of 0.5. Similarly, we use the above method to test its sensitivity, i.e. When the following formula holds, the model is considered stable.

$$\left| \frac{S_{global_{fact}} - S_{global_{test}}}{S_{global_{fact}}} \right| \times 100\% < 10\%$$

We take the values of 0.3 and 0.7 respectively, and compare the final result with the result we achieved in Task, and finally come up with the following table to show the difference before

and after changing the  $\lambda$  value.

It is not difficult to find by comparing the data that changing the value of the parameter will not cause more than 10% change in the final data result, so we can consider the model to be stable.

## 13 Document

Dear Sir/Madam,

We create a spatial node for each influencer and follower to get an initial network that describes the most basic interrelationships, which are prototypes for subsequent changes. In the case of my prototype graph, we carefully analyzed more factors that might contribute to the influence of music, and from the known data table important indicators, weighted three impact prototype graphs, we got the final "musical impact" score, weighted network with 5000 randomly selected data, and the size of the nodes that were used to obtain the corresponding musician's "musical impact". Therefore, the value of using our approach to understand the impact of music through the web is intuitive and precise.

We can combine small data sets to form a large data set. We find the same musician in two small datasets and combine all the information about that musician into one big data set. Even if the dataset is a bit confusing, we can find all the information about the musician by indexing the musician's name, and then finish all the information about the musician in the larger dataset. This gives us a large dataset with comprehensive information. In addition to taking the names of musicians as a small unit, large data sets can also be used as a small unit to analyze and average all musical characteristics of all musicians in the genre. It is easier to compare the relationship between musicians and genre units in musician units. The premise is to integrate and improve the four small data sets given to obtain a large data set with comprehensive information coverage.

In the years when data is intermittently missing and the data interval is small, we can interpolate the data to complete the data, making it easier to process the data and get more intuitive visualizations. Music is an art, and art is a mapping of culture. Music has been a language of information for a long time. In this case, there is an evolution of music. When society is full of pressure, people are more eager to have positive and stimulating rock music; Similarly, when society is calm and quiet, people want songs with soft and deep thoughts. When a certain type of music is in decline, we don't have to be pessimistic. There are two reasons for this: first, music is firmly integrated with other musical genres to form a new form of music; Second, in the current social context, this genre of music is no longer needed. From this, we can see the reflection of an art on the theme of the times in the progress of the times. Culture is a delicate brush in the great river of the times, music is a brilliant color in the cultural brush, and it is an emotional bridge connecting people and things. To complement each other, we should be able to hear the music well.

Team  
2023/1/10