Problem Chosen	2021	Team Control Number
D	MCM/ICM Summary Sheet	**2105330**

Omnipotent Music
Analysis on Network and Changes in Music Characteristics

Summary

Music, is a great achievement of human civilization. By considering songs and their musical characteristics, as well as the mutual influence of musicians, we can measure the musical influence of musicians. Furthermore, we can also better understand how music affects society and culture, and how music evolves over time.

First, we proposed the Musical Influence Network (MIN) Model. We established a network based on the mutual influence of musicians, and comprehensively considered the out-degree and agglomeration clustering coefficient of nodes to analyze the musical influence of nodes.

Secondly, we proposed a Music Similarity Model based on Mahalanobis distance. We use Mahalanobis distance to measure the similarity between music nodes. Based on Mahalanobis distance, we define the degree of separation between two nodes and draw a conclusion that artists of the same genre are more similar.

Integrating the influence network and the music similarity model, we think the influencer can really influence the followers.

In order to further analyze which musical characteristics are more contagious, we use the degree of separation of the musicians as the dependent variable, and the difference of each musical characteristic of the musicians as the independent variable, then perform multiple linear regression analysis.

We analyze the variation trend of the characteristics of a certain genre, and we can get the evolution of the genre over time.

Next, we established the Genres Flourishing Degree Model. We define the market share of genres to measure the prosperity of a genre at the time. Through the model, we can get the changes in the degree of prosperity of each genre.

Finally, we proposed a Difference Equation Social Change Model.

We selected some features of music, evaluated the social changes reflected by music with their differences, defined the indicator of social changes, and used this model to identify some major events in history, such as the economic crisis, World War II, etc.

After completing the above work and conducting a sensitivity analysis, we wrote a letter to the ICM Society to introduce our results.

Keywords: R-type clustering, Min Model, Mahalanobis distance, difference equation, Multiple linear regression

Omnipotent Music
Analysis on Network and Changes in Music Characteristics

February 9, 2021

Summary

Music, is a great achievement of human civilization. By considering songs and their musical characteristics, as well as the mutual influence of musicians, we can measure the musical influence of musicians. Furthermore, we can also better understand how music affects society and culture, and how music evolves over time.

First, we proposed the Musical Influence Network (MIN) Model. We established a network based on the mutual influence of musicians, and comprehensively considered the out-degree and agglomeration clustering coefficient of nodes to analyze the musical influence of nodes.

Secondly, we proposed a Music Similarity Model based on Mahalanobis distance. We use Mahalanobis distance to measure the similarity between music nodes. Based on Mahalanobis distance, we define the degree of separation between two nodes and draw a conclusion that artists of the same genre are more similar.

Integrating the influence network and the music similarity model, we think the influencer can really influence the followers.

In order to further analyze which musical characteristics are more contagious, we use the degree of separation of the musicians as the dependent variable, and the difference of each musical characteristic of the musicians as the independent variable, then perform multiple linear regression analysis.

We analyze the variation trend of the characteristics of a certain genre, and we can get the evolution of the genre over time.

Next, we established the Genres Flourishing Degree Model. We define the market share of genres to measure the prosperity of a genre at the time. Through the model, we can get the changes in the degree of prosperity of each genre.

Finally, we proposed a Difference Equation Social Change Model.

We selected some features of music, evaluated the social changes reflected by music with their differences, defined the indicator of social changes, and used this model to identify some major events in history, such as the economic crisis, World War II, etc.

After completing the above work and conducting a sensitivity analysis, we wrote a letter to the ICM Society to introduce our results.

Keywords: R-type clustering, Min Model, Mahalanobis distance, difference equation, Multiple linear regression

Contents

1 Introduction

1.1 Background

Music is an important part of our cultural life. It can record the current mental state of mankind as well as reflect social changes. In the past 90 years, human society has developed rapidly. At the same time, music has also developed tremendously. As time goes by, a number of influential artists have emerged, resulting in a variety of music genres. Many artists have contributed to the changes in music genres. Some have influenced the musical creation of later generations, and some have made creative changes to music based on social events at the time. Studying the mutual influence between artists, the associations and differences between music genres, the relationship between genre evolution and social factors, and constructing a network of mutual influences of various factors are of great benefit to studying the evolution of music and the development of social history. If we can quantify the evolution of music, it will be of great significance to our understanding of the role music plays in human collective experience.

1.2 Our Work

Firstly, we preprocess the data:

- Using R-type clustering to reduce the number of variables.

secondly, we construct the following four models:

- The music influence network
- The similarity model
- The genre rise and decline model
- The social change model

Then we apply our models on the following problems:

- Study the influence of musicians and genres
- Compare the similarity of musicians and genres
- Study the musical evolution and revolution
- Predict changes in the external environment by analyze musical changes

2 Assumptions and Notations

2.1 Assumptions

- Influencers who have more followers have more influence.
- If followers influenced by a certain influencer influence each other more, we consider the influence of the influencer is greater.

- Once a follower is influenced by a certain influencer, the musical characteristics of his works will be more similar to those of the influencer.
- We consider positive music more danceful, energetic, and musically more cheerful. On the contrary, negative music is less danceful and weaker. It has little energy, and conveys more sadness.
- The characteristics of music reflect the overall atmosphere of society at that time to a certain extent .

2.2 Notations

Symbol	Meaning
N	Total musician number
i	The index of musician node
$\vec{X}_i$	The vector composed of each musical characteristic of node i
$C_d(i)$	The out-degree number of node i
$C_c(i)$	The clustering coefficient of node i
$F(i)$	The influence degree of node i
$D_M(i,j)$	Mahalanobis distance between nodes i and j
$Sep(i,j)$	The degree of seperation between node i and j t
G_k	The collection of musicians from the kth genre,$k = 1, ..., 20$
T_i	The social change index of year i
M_{kt}	The collection of musicians from the kth genre in year t,$k = 1, ..., 20$,$t = 1921, ..., 2020$
$pop(M_{kt})$	The sum of the popularity of all songs in the kth collection
$Scale(k,t)$	Music market share of the kth genre in year t

3 Data Processing

- All the data we use comes from the given four files.

In order to reduce the error caused by the high correlation between various variables describing music, we use R-type clustering to deal with variables. According to the order in the data set, the 11 variables named 'danceability',..., 'speechiness' are numbered from1 to11. We use the correlation coefficient as a measure of similarity between variables, and perform R-type clustering. Figure 1 is the resulting cluster map.

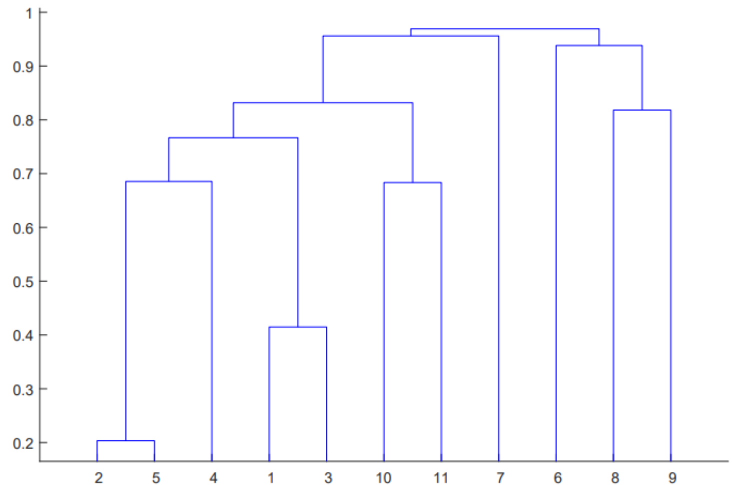

Figure 1: The R-type clustering result

According to the figure that variable 2 and variable 5, namely energy and loudness, have a high correlation, so we delete the variable loudness in the subsequent processing.

4 The Influence Model

We define an influence network as a directed graph, and each musician is a node in this graph. If musician i influences musician j, there is a directed edge from node i to node j. The whole graph is as Figure 2.

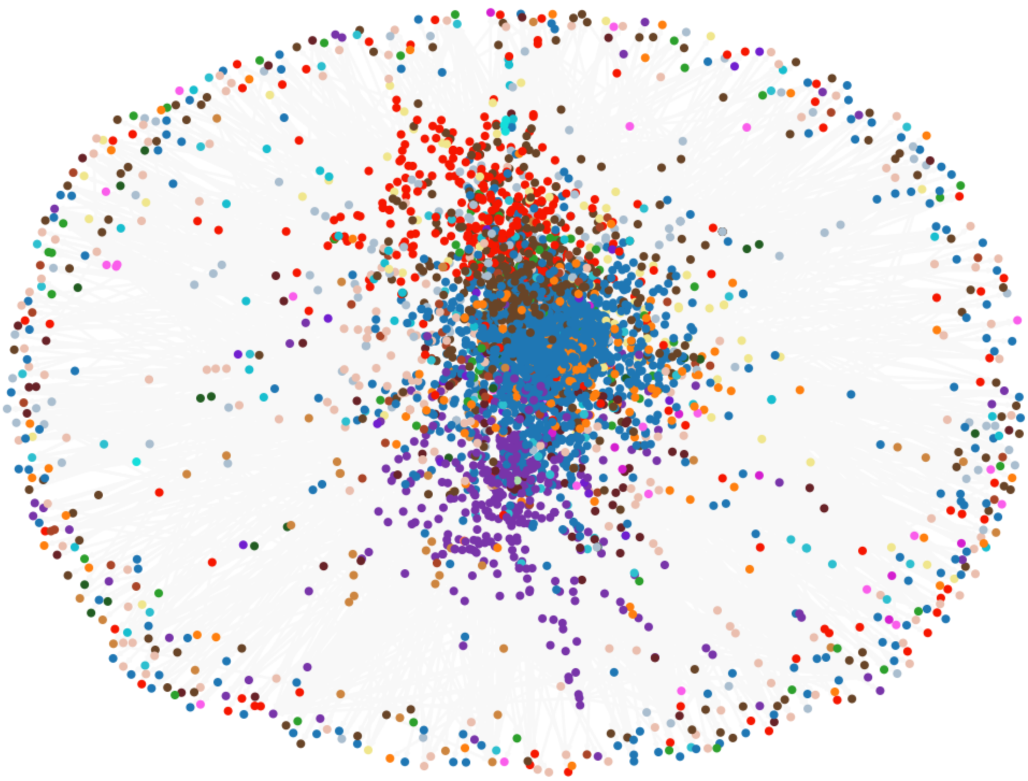

Figure 2: The influence network

- **Out-degree of a node**
  Define a 0-1 variable e_{ij}, if there is a directed edge from node i to node j, $e_{ij} = 1$, otherwise $e_{ij} = 0$.

$$C_d(i) = \sum_{j=1}^{N} e_{ij} \tag{1}$$

4.1 Musical Influence Based on out-degree-clustering coefficient Model

In our case, we use Hypothesis 1 and Hypothesis 2. We measure a certain influencer's influence by combining the number of his followers and the relationship between his followers.

- **Clustering coefficient of a node**
  N_i: The set of subscripts of i's neighboring nodes.
  K_i: The number of neighboring nodes of i.

$$C_c(i) = \frac{\sum_m \sum_n e_{mn}}{k_i(k_i - 1)}, m \in N_x, n \in N_x \tag{2}$$

- **The importance of a node**

  We believe that it is not comprehensive to consider only the out-degree of nodes or only the clustering coefficient of nodes. Therefore, we will combine them to define a new indicator: the influence of the node $F(i)$

$$F(i) = aC_d(i) + (1 - a)C_c(i) \tag{3}$$

In order to combine the two, we introduce a weight coefficient a. According to the hypothesis, we consider the number of followers more important, so we set $a = 0.6$.

Results and Analysis

We will use $F(i)$ to measure a musician i's musical influence. After sorting the $F(i)$ of n musicians, we got the ten most influential musicians. Their basic information is as shown in Figure 3.

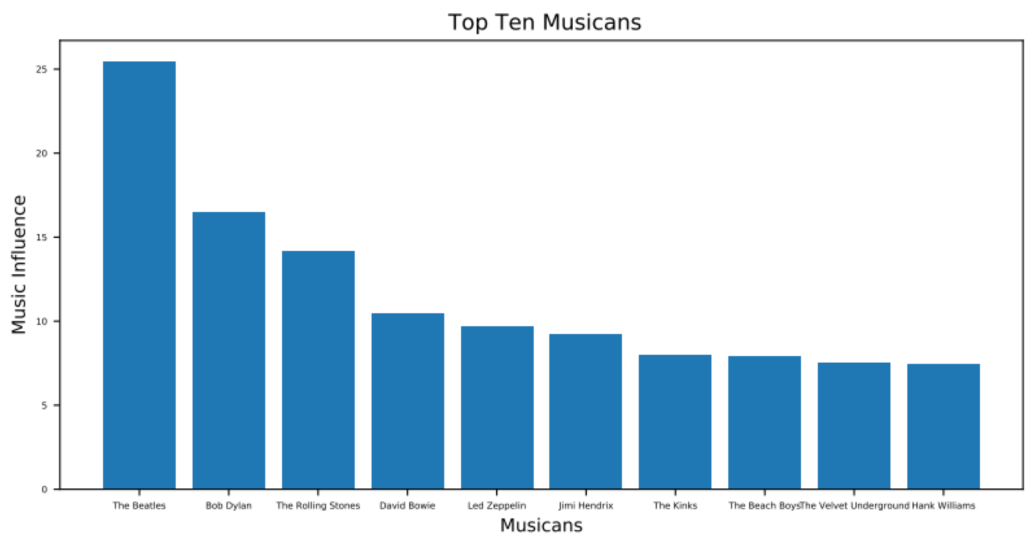

Figure 3: The top 10 musicians

In order to make a more reasonable, we will change the value of a in the sensitivity analysis to see if the result will change.

4.2 Description of a Subnetwork

Based on our influence network, we select some singers' influence subnetwork to describe. As shown in Figure 4, the greater the influence of a singer, the greater the size of the singer's correspondence and node, as well as the greater the degree of integration(the greater the number of directed edges pointing to other nodes).

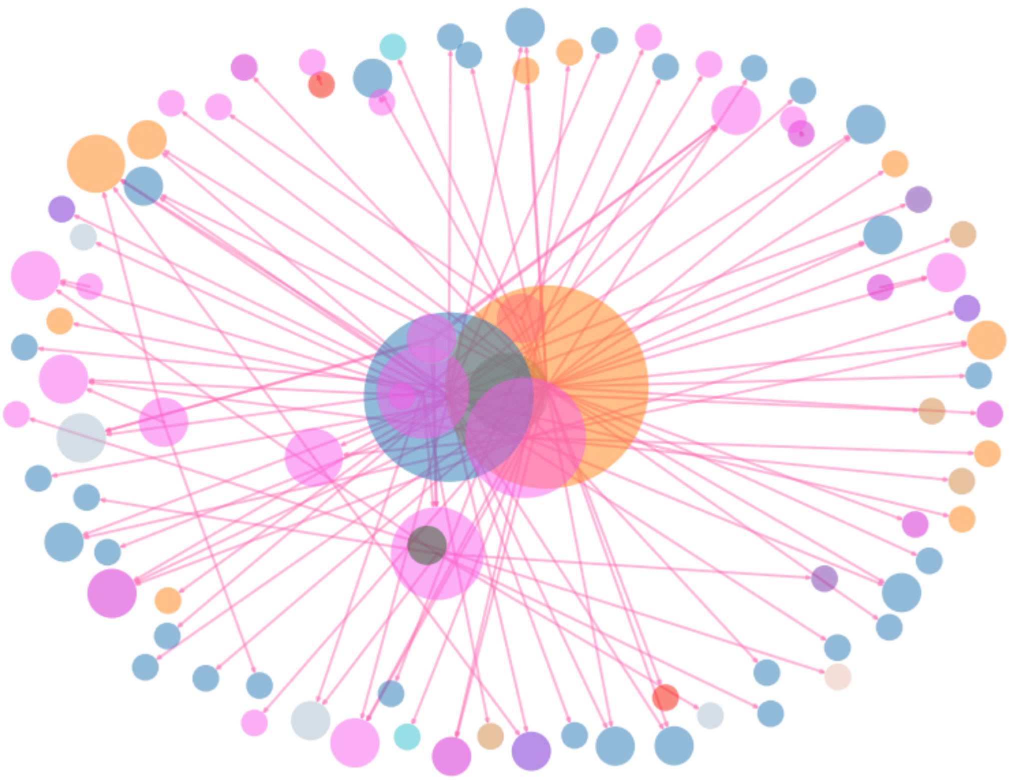

Figure 4: A subnetwork describing Classical

In this subnetwork, nodes of different colors correspond to different genres. For example, the light pink node corresponds to musicians of Classical.

5 The Similarity Model

We use the musical characteristics of a musician to represent each musician with a vector, and use the distance between the vectors to represent the similarity of them. People often use Euclidean distance to measure the similarity between two vectors. However, because different musical characteristic variables have different dimensions, making their measured values vary widely, we use Mahalanobis distance as the similarity measure of songs.

5.1 Define Similarity Using Mahalanobis Distance

Before defining the Mahalanobis distance, we first define the covariance of two vectors and the covariance matrix of n vectors.

- Covariance of two p-dimensional vectors:

$$Cov(\vec{X},\vec{Y}) = \frac{\sum_{i=1}^{p}(x_i - \overline{x})(y_i - \overline{y})}{p-1} \tag{4}$$

- Covariance matrix of n p-dimensional vectors:
  For a set of vectors$\left\{\vec{X_1},\vec{X_2},...,\vec{X_n}\right\}$, among them $\vec{X} = \{x_1, x_2, ..., x_p\}$.
  Its mean is $\vec{\mu} = \{\mu_1, \mu_2, ..., \mu_p\}$

- **The Mahalanobis distance is defined as follows:**
  For two sample points X_i, X_j from p-dimensional space:

$$D_M(i,j) = \sqrt{(\vec{X_i} - \vec{X_j})\sum^{-1}(\vec{X_i} - \vec{X_j}))} \tag{5}$$

  $\sum$ is the overall covariance matrix, estimated with sample covariance.

We believe that the smaller the Mahalanobis distance between two sample points, the more similar they are. In order to facilitate the subsequent discussion, we define an index $Sep\,(i,j)$to measure similarity: the degree of separation between nodes i and j

$$Sep(i,j) = [D_M(i,j)]^2 \tag{6}$$

6 Applications of the Influence Model and the Similarity Model

6.1 Compare the Similarity of Musicians within Genre and between Genres

- For the subsequent formula, we first give the concept of the median point of the genre. For the genre G_i, its median point is defined as follows:

$$\vec{\bar{X}}_i = \{\bar{x_1}, \bar{x_2}, ..., \bar{x_p}\} \tag{7}$$

  Among it,$\overline{x_k} = \frac{1}{|G_i|}\sum_{\vec{X}\in G_i} x_k$

- In order to measure the similarity of musicians in the same genre, we were inspired by the definition of variance. Based on the Mahalanobis distance, we proposed the value D_{in} to represent the difference within the genre, such as the difference value D_{iin} of the genre i as follows:

$$D_{iin} = \frac{1}{|G_i|}\sum_{\vec{x_k}\in G_i} Sep(\vec{X_k}, \vec{\bar{X}}_i) \tag{8}$$

- Similarly, we define value D_{out} to represent difference between different genres, such as the value D_{ijout} between genre i and genre j as follows:

$$D_{ijout} = \frac{1}{|G_i|}\sum_{\vec{x_k}\in G_i} Sep(\vec{X_k}, \vec{\bar{X}}_j) \tag{9}$$

Comparative criterion

By comparing D_{in} and D_{out}, we can draw the conclusion: if $D_{in} < D_{out}$, it means that musicians in the same genre are more similar than musicians in different genres; otherwise, musicians in different genres are more similar.

Results and analysis

We made a comparison between every two genres. In order to display the results more intuitively, we drew a heat map as Figure 5.

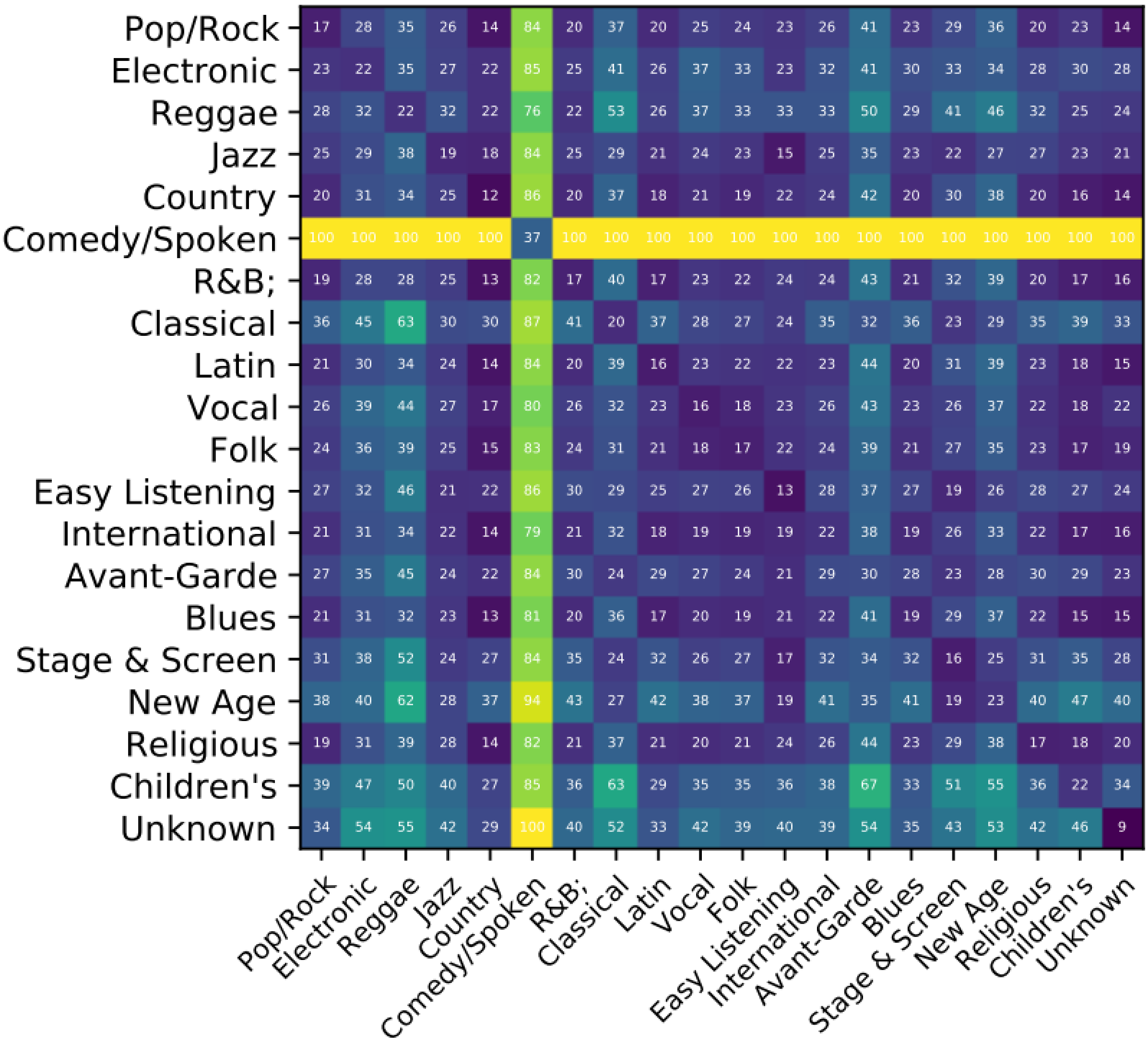

Figure 5: The heat map reflecting similarity

According to the heat map, most genres's musical characteristics are more similar within the genre, except from Children's. However, the Children's genre is still more similar within itself than between genres in most cases. So we can draw the conclusion that in most cases, musicians within genre are more similar than musicians between genres.

6.2 Compare Similarities and Influences between and within Genres

Factors distinguish a genre

We have previously studied the similarities within and between genres using various musical characteristics of music, and the results presented in Figure5 show that some genres are relatively close while others are quite different. Therefore, we believe that some musical characteristics can be used to distinguish genres. We use the various musical characteristics of the representative musicians of each genre to make Figure6 as follows.

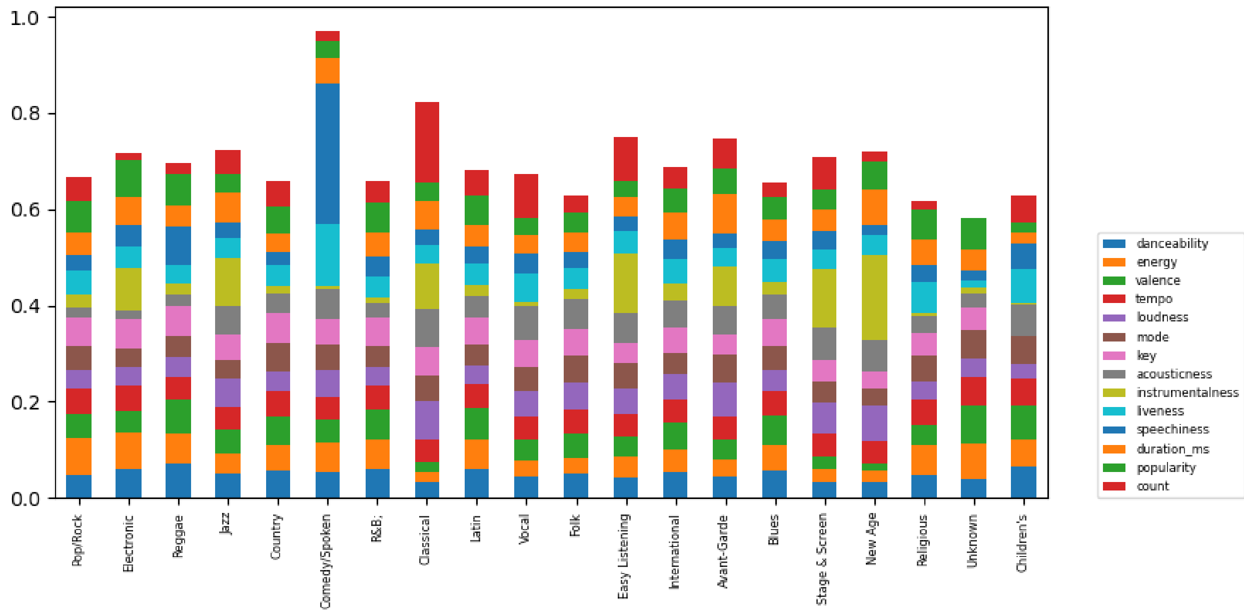

Figure 6: The musical characteristics of each genre

We can draw some conclusions that the speechiness of Comedy/Spoken is significantly high, the instrumentalness of New Age is significantly high, and the key of Children's is significantly low. These three genres can usually be distinguished by one characteristic. For other genres, we found that they have big differences in instrumentalness, which can divide genres into two groups. In addition, genres can be further divided mainly through energy, valence, and speechness.

Genres' change over time

In order to measure the rise and decline of genres over time, we defined the market share of the genre in a specific year to reflect the prosperity of the genre in a certain period of time. For the genre G_i of a certain year t

$$Scale(i,t) = \frac{pop(M_{it})}{\sum_{j=1}^{20} pop(M_{jt})} \tag{10}$$

We already know that the popularity of a song is determined by its play volume and recent play time. Therefore, in the case of the same amount of play, recent songs will be seemingly more popular than previous songs. Here we use the ratio to define the proportion of a certain genre's popularity at the time, which can well eliminate the popularity error caused by time. For each genre, we fit its market share with the year to get Figure7 below.

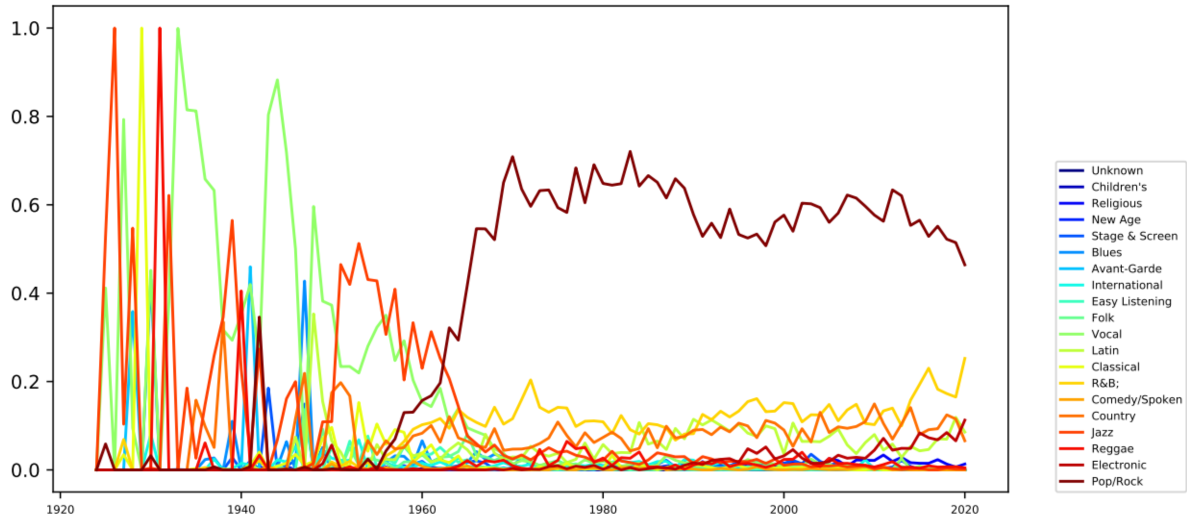

Figure 7: The genres change over time

From Figure 7, we can get the following information:

- **Pop/Rock**
  From 1920 to 1950, the market share of it was very small, almost 0. From 1950 to 1970, its market share grew rapidly, and reached a maximum of about 0.7 in the 1970s. Since 1970, its market share fluctuated around 0.6.

- **Vocal**
  Its market share was the highest in 1930, almost 1. Around 1940, the market share was as high as 0.9. Since 1940, the market share has generally shown a downward trend. Since 1970, the market share has been at a low value.

- **Jazz**
  In the 1920s and 1930s, the market share reached its maximum value, which was almost 1, and then declined. In the 1950s, the market share rose again and reached a maximum value of 0.6. Since 1950, the market share has shown a downward trend , and dropped to a low value in 1970, then remained stable.

- **R&B**
  Before the 1950s, its market share was small. From the 1950s, its market share began to rise. After the 1970s, its market share was relatively stable, fluctuating around 0.1.

We can divide genres into 3 categories

- Some genres had a relatively high market share before the 1950s and a declining market share after the 1950s, such as Vocal, Jazz, Avant-Garde, etc.

- Some genres had a relatively low market share before the 1950s and an increasing market share after the 1950s, such as Pop/Rock, R&B, Country, etc.

- Some genres' market share have not changed significantly, such as International.

We found that in the history of musical evolution, there are several important points in time between 1921 and 2020, such as around 1950 and around 1970.

Genres related to each other

We believe that there is an interactive relationship between genres, which is the result of the mutual influence of musicians between the two genres. We will study on a certain genre's influencers and calculate thier genres and proportions. The larger the proportion, the more related these two genres are. We generated the influence of 18 genres in the form of a pie chart, and we display two of them here. They are Figure8 and Figure9.

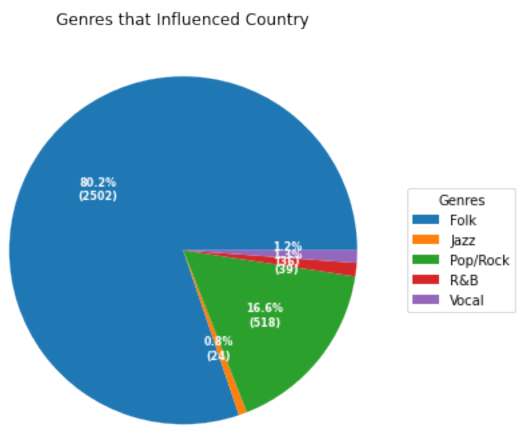

Figure 8: Genres that influenced Country

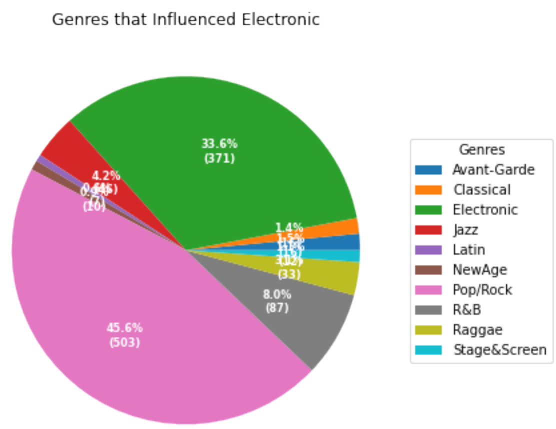

Figure 9: Genres that influenced Electronic

According to the figure, there is a certain relationship between some genres. We only list out the genres that have an impact on this genre accounted for more than 10%, and the order is in descending order of influence.The list is as follows:

Genre	Closely related genres
Classical	Pop/Rock
Comedy/Spoken	Pop/Rock
Raggae	R&B
Stage&screen	Jazz
Jazz	Pop/Rock
Vocal	Jazz
Pop/Rock	Jazz,Electronic
Religious	Pop/Rock,Jazz
Classical	Pop/Rock,Vocal
Folk	Pop/Rock,Country
Latin	Pop/Rock,Folk
Country	Pop/Rock
Electronic	Pop/Rock

6.3 Influencers's Influence

Identify Whether the Influence Actually Exists

Based on Hypothesis 3, we believe that, given a set of all musicians A. For any $i \in A$, if there is a corresponding influencer $j \in A$. Then if $Sep\,(i,j)$ is less than $\overline{Sep\,(i,b)}$ between i and all nodes b. It can be considered that the creation of i is indeed affected by j. Among them, $b \in B_i$, B_i is a set that $\forall b \in B_i$,there is no edge between node b and both i and j in the influence network. We finally use the proportion of musicians in A who can meet the above conditions to measure whether it is common for influencers to really influence the music created by their followers.

Simplified Model

- Since some followers are influenced by multiple influencers, the influence of a particular influencer on him may be weakened, and it is difficult to judge whether he is more similar to a particular influencer numerically. In order to simplify the model, we study a musician with only one influencer. We redefine set A as a set of musicians with only one corresponding influencer (that is, musician with only one in-degree in the music influence network).
- Since we have previously concluded that music in the same genre is more similar. It is of little significance to compare i with musicians in other genres. Thus, we further limit the genre all elements in the set B_i to be the same as the genre of y_i.
- Since the musical characteristics of the genre will also evolve over time, we further limit the initial active year of all elements in the set B_i to be the same as j, in order to reduce the error caused by time.

Specific Method

- Obtain the musician set A from the influence network, the size of A is n.
- For $\forall i \in A$, find its only corresponding influencer j.
- Obtain the set of musicians B_i meeting the following conditions from the influence network:

  $\forall b \in B_i$, the genre and initial active year of b and j are the same, and there is no edge between node b and both i and j in the influence network.
- If the set $B_i = \emptyset$, remove the i from A, $n = n - 1$.
- Calculate

$$S_i = Sep(i,j) \tag{11}$$

$$S_i' = \frac{1}{|B_i|}\sum_{b \in B} Sep(i,b) \tag{12}$$

Define 0-1 variable $flag(i)$

$$flag(i) = \begin{cases} 0, S_i < S_i' \\ 1, else \end{cases} \tag{13}$$

- Count, define the ratio P.

$$P = \frac{\sum_{i=1}^{n} flag(i)}{n} \tag{14}$$

- Observe the value of P

Results and analysis

We first get the set A with size 631. When generating sets B_i, it is found that 61 corresponding sets B_i are empty sets. So these 61 elements are removed from set A, and set A is reduced to 570. After calculation, among the 570 musicians in set A, there are 472 ones satisfying $S_i < S_i'$, so $P = 82.8\%$.
Based on the result, we can think that the phenomenon that influencers really affect the music created by their fans is widespread.

Musical Characteristics with More Influence

Based on the belief that influencers will really influence their fan's creations, the similarity between influencers and followers can reflect the influence of influencers. Therefore, it is not difficult to infer that the degree of separation between influencers and followers can also reflect the influence of influencers.

Comparison method

Analyze the musician i who satisfies $S_i < S_i'$ in A, and the corresponding influencer j. We take the influence of j on each musical feature as the independent variables x_1 to x_{10}, and $Sep(i, j)$ as the dependent variable Y, and then perform multiple linear regression. To analyze whether there are certain musical characteristics that have been more affected.
Among them, the influence of j on each music feature is represented by the difference between the corresponding music features of i and j.
We will compare the standardized regression coefficients and significance of each independent variable to determine which musical feature is more infectious.

$$Y = b_0 + b_1 x_1 + ... + b_{10} x_{10} \tag{15}$$

The standardized regression coefficient is

$$b_i^* = bi \cdot \frac{Cov(X_i)}{Cov(Y)} \tag{16}$$

Results and analysis

We use SPSS to perform multiple linear regression.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.465[a]	.217	.200	454.1410771	2.056

a. Predictors: (Constant), valence, mode, instrumentalness, tempo, liveness, speechiness, key, acousticness, danceability, energy

b. Dependent Variable: influence

Figure 10: The Model Summary

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	26294103.46	10	2629410.346	12.749	<.001[b]
	Residual	95078538.36	461	206244.118		
	Total	121372641.8	471			

a. Dependent Variable: influence

b. Predictors: (Constant), valence, mode, instrumentalness, tempo, liveness, speechiness, key, acousticness, danceability, energy

Figure 11: The ANOVA table

In Figure10 and Figure11, the .sig in the ANOVA table is < 0.001, so the fitting is considered to be successful. The Durbin-Watson value in the Model Summary table is 2.056. Unfortunately, the Adjusted R Squre value is 0.200, which is a little small. However, we still believe that this regression is effective.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	255.586	22.097		11.566	<.001		
	danceability	-288.365	164.175	-.096	-1.756	.080	.568	1.760
	energy	17.346	118.115	.010	.147	.883	.370	2.705
	speechiness	800.074	255.578	.150	3.130	.002	.739	1.353
	liveness	928.585	107.267	.395	8.657	<.001	.816	1.225
	instrumentalness	191.199	82.411	.103	2.320	.021	.855	1.169
	key	.904	4.428	.009	.204	.838	.801	1.248
	acousticness	-107.003	83.825	-.080	-1.276	.202	.437	2.289
	mode	-76.942	36.752	-.091	-2.094	.037	.897	1.115
	tempo	2.341	.836	.130	2.800	.005	.788	1.268
	valence	98.965	98.708	.056	1.003	.317	.537	1.862

a. Dependent Variable: influence

Figure 12: The Coefficients table

In Figure12 Based on the normalized coefficient Beta and significance in the above figure, we can draw a conclusion that the more contagious music characteristics are liveness, speechiness and tempo.

7 The Musical Evolution and Revolution

7.1 On the Whole

Significant changes

According to Figure7, we can conclude that the proportion of each genre changed greatly between 1950 and 1970. After 1950, the popularity of Vocal, Jazz, Avant-Garde and other genres showed a downward trend, while the popularity of Pop/Rock, RB, Country and other genres showed an upward trend. We think that around this period, music has undergone a change of genre. Pop/Rock developed rapidly in 1970. Since then, it has been dominating the music market. We think this is a revolutionary moment for Pop/Rock and the entire music market.

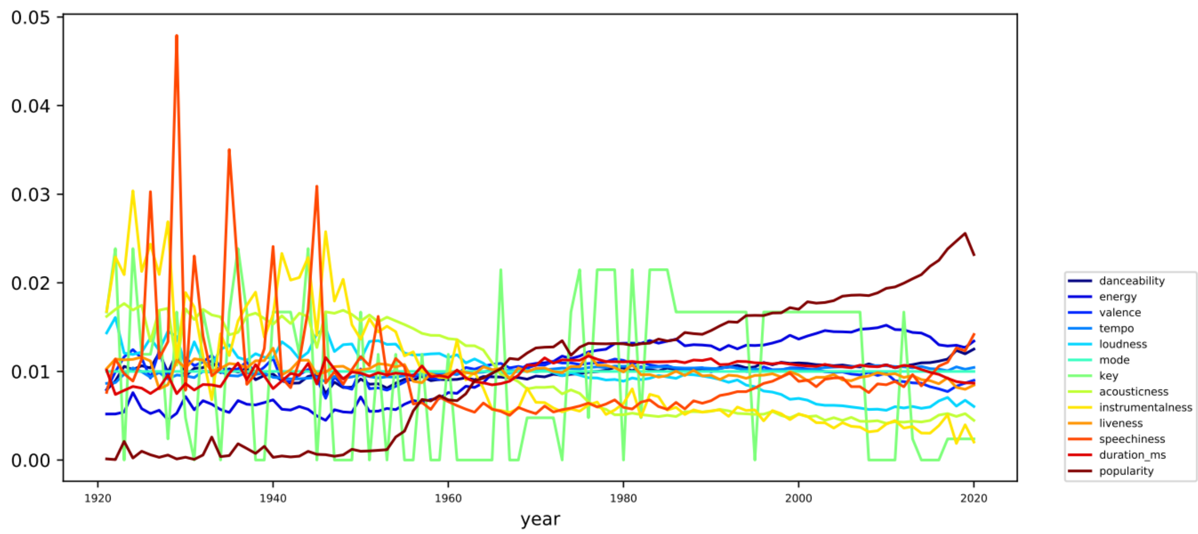

Figure 13: The musical characteristics change by time

Based on the musical characteristics from 1921 to 2020, we have made Figure13 the changes in each musical characteristic over time. This image also supports our view of the great changes in the music of the 1950s. Around 1950, the overall danceability, valence, and tempo of music showed an inflection point, all from falling to rising. Energy began to rise. Acousticness, instrumentalness and speechiness began to fall. In the 1980s valence and tempo began to decline.

Combining the analysis of the two images, we believe that 1950, 1970, and 1980 are the key time periods for music revolution. In order to further prove our point of view, we fit the value of the above music characteristics over time and find the difference. The results can prove our point.

Possible revolutionaries

Since we believe that the 1950s, 1970s, and 1980s are critical periods of musical transformation, we can pick out possible revolutionaries through the following methods.

- The active year of the revolutionaries should be around the above-mentioned critical periods. we choose 10 years before and after the critical moment here.
- Revolutionaries should have great influence. Here we limit it to the top 100 musicians in the influence network.
- The genre of the revolutionaries has undergone tremendous changes during his active years.
- The changes in the musical characteristics of the revolutionaries should be similar to the overall change trend of the musical characteristics, and the changes are significant.

Based on the conditions, we screen possible revolutionaries in the influence network.

- **Elvis Presley 1950 Pop/Rock**
  His energy and valence were very high, acousticness and instrumentalness was rather

low, and the trend is consistent with the overall trend of music. He can be considered a pioneer of Pop/Rock.

- **Marvin Gaye 1950 R&B**
  His energy and valence were high, acousticness,instrumentalness and speechiness were low, and the trend is consistent with the overall trend of music. It can be considered that the rise in R&B during the same period is closely related to his contribution.

- **The Beatles, David Bowie, The Rolling Stones, 1960 Pop/Rock**
  Their music characteristics are consistent with the overall trend of music, and all have a very large influence. Combined with the rapid expansion of Pop/Rock in 1970s, it can be considered that they greatly promoted the development of Pop/Rock and laid the foundation for Pop/Rock to occupy the mainstream.

- **Nirvana 1980 Pop/Rock**
  Its valence is significantly lower than those in the same period, and it's consistent with the overall trend of music.

7.2 In a genre

We will take Jazz as an example to analyze the influence and evolution of one genre. The analysis of other genres is similar.

Influence process of musical evolution of one genre over time

In order to show the influence of Jazz in a certain era, we separately counted the years in which Jazz influencers and their corresponding followers started their music careers. Then counted the number of corresponding musicians according to the year. The results are shown in Figure 14.

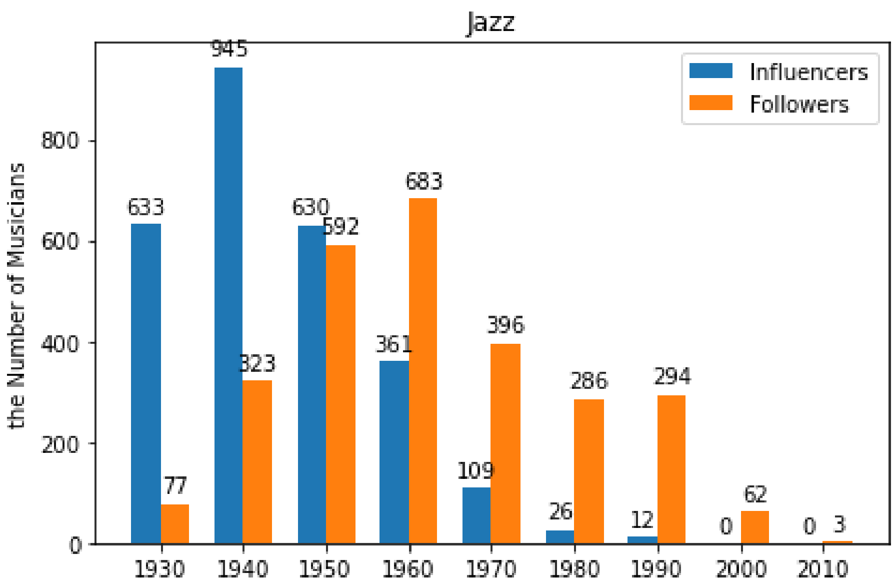

Figure 14: Then the number of corresponding musicians

According to Figure14, the number of influencers and followers changes in roughly the same trend, but there is a certain time lag in the changes between the two.

Combining the previous conclusions, the influence of influencer i on follower j can be indirectly reflected by $Sep(i,j)$. Therefore, we define the influence of genre G_i over time $Inf_i(t)$: $t = 1920 + 10k, k = 0, 1, ..., 10$ $A_i(t)$: The set of followers of genre i whose starting career time is equal to t
$B_i(t)$: The set of followers of genre i who is in $A_i(t)$ and its starting career time is earlier than t

$$Inf_i = \frac{1}{\sum_{j \in B_i(t)} \sum_{k \in A_i(t)} e_{jk} Sep(j,k)} \tag{17}$$

From the formula, we can see that the dynamic index is the number of followers of G_i in that era and the degree of separation between influencers and followers.

The evolution of a genre over time

The changes of various characteristics of Jazz over time can be plotted as shown in Figure 15

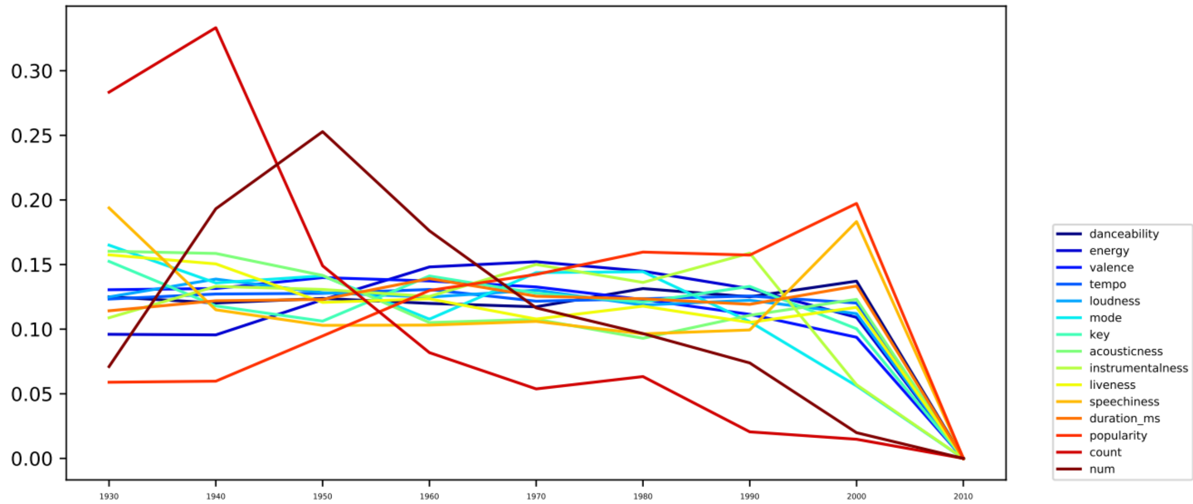

Figure 15: The changes of various characteristics of Jazz over time

Figure 15 shows the changing trend of various characteristics of Jazz.

As time goes, the musical characteristics of Jazz dose not change much, but remains stable. However, after reaching its maximum value in the $1940s$, the amount of songs has gradually declined since the $1940s$. The popularity of Jazz has slowly risen over time.

Therefore, we can infer a more general conclusion: the music characteristics of each genre will not change greatly over time, while the number of songs, popularity and other characteristics that reflect the prosperity of the genre will vary over time.

7.3 The Social Change Model

The cultural influence of music

The influence of music on culture is beyond doubt. Based on the models we build before, we can find that the more popular music at the time promoted the development of culture.

Combining figure xx (that market share figure) and our cognition of various genres of music, we can draw conclusions, such as:

- In the 1920s and 1930s, the mainstream music was Jazz and Country, with beautiful melody and strong rhythm, which had an impact on the inclusive and open social culture at that time.
- Since the 1970s, the mainstream music is Pop/Rock with the characteristics of passion and boldness, which promoted the social atmosphere of pursuing individuality and daring to resist.

Social Change Recognition Model Based on Cumulative Difference Formula

Based on Hypothesis 4 and 5, it can be reasonably inferred that when the characteristics of music change greatly, social or political changes may have taken place. Therefore, we use danceability, energy and valence of a certain year i as variables. Based on the difference equation, we define a social change index T_i, the expression is as follows:

For year i, let the overall musical characteristics danceability, energy, valence be x_i, y_i, z_i

$$T_i = |x_{i+1} - x_i| + |y_{i+1} - y_i| + |z_{i+1} - z_i| \tag{18}$$

When T_i is large, it can be inferred that the social changes at that time are large.

By calculating T(i) from 1921 to 2020, we can get an image as Figure16.

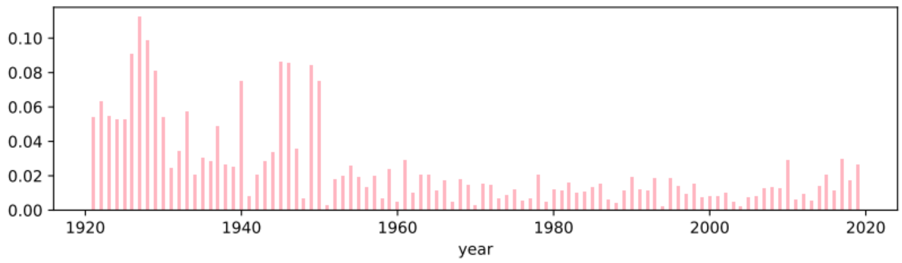

Figure 16: The social change index T_i

Combining the history of the United States to analyze the image, it is not difficult to verify the rationality of our model. The time node with larger T in the figure:

- In the 1930s, there was an economic crisis in the United States
- In the 1950s, the United States had just won World War II

Technological developments found in musical features

By analyzing some of the characteristics of music, we can identify the impact of technological changes on music. For example, the speechiness feature of music. When speechiness is greater than 0.66, the song may be a soundtrack composed entirely of spoken language, such as talk shows, audio books, etc. Analyzing the time distribution of the number of music with speechiness greater than 0.66, we can get Figure17.

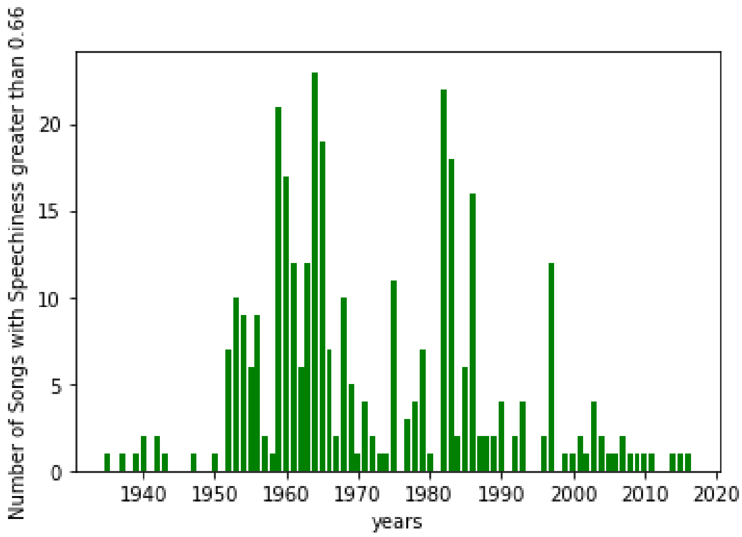

Figure 17: Number of songs with speechiness greater than 0.66

According to Figure 17, in the 1950s and 1960s, there were more types of talk show or audiobook music. We know that, driven by the third technological revolution, American audio books were born in the 1950s and became popular in the 1960s. Therefore, we can see the impact of technological changes on music.

8 Sensitivity Analysis

When measuring the influence of a musician, the out-degree-clustering coefficient model we proposed includes a constant parameter a. We selected $a = 0.6$ as the weight of the out-degree, and $(1 - a) = 0.4$ as the weight of the clustering coefficient , and summed to get our influence index.

In this section, we test the sensitivity of the out-degree-clustering coefficient model by changing a to show its reliability.

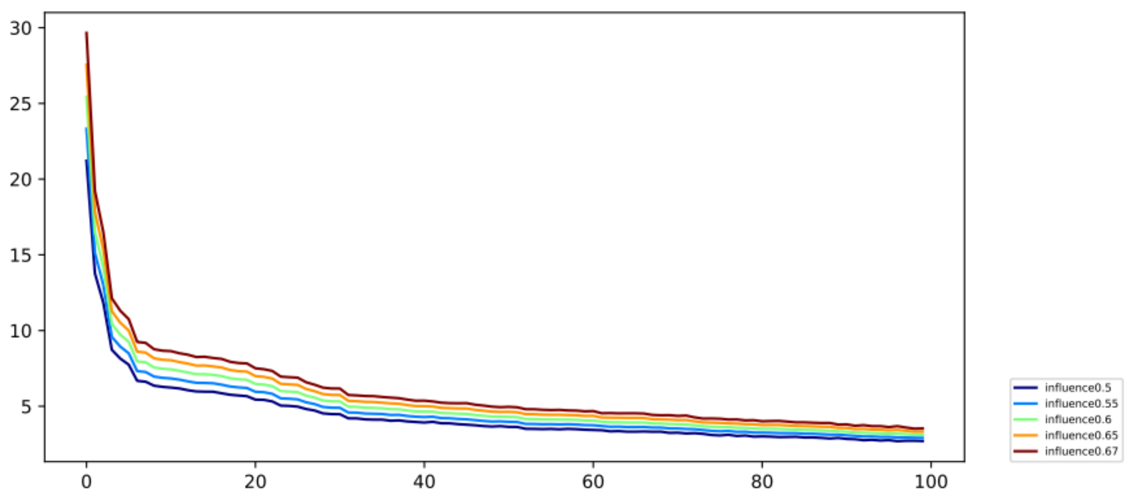

Figure 18: The partial sensitivity analysis

In Figure 18, we get the influence index of the top 100 artists with different a. Each curve represents a different a, and all show the same trend. Therefore, the out-degree-clustering coefficient model we proposed is not sensitive to the value of a.

9 Strengths and Weaknesses

Strengths

- We innovatively applied the model of measuring the importance of the network to our model of musician's influence, and verified the model's effectiveness through reasonable results and sensitivity analysis.
- We cluster the variable set in advance, which reduces the complexity of the calculation and errors from the perspective of reducing the correlation of variables.
- We innovatively apply Mahalanobis distance to represent similarity, avoiding the inconsistency of measurement and the dependency of variables.

Weakness

- Insufficient data mining. We did not make full use of the data provided, such as data in full_music_data.csv.
- Our discussion on the network model is relatively brief.
- We use the method of data integration and human judgment to solve some problems, which is relatively time-consuming and subjective.

References

[1] Ren Zhuo-Ming, Shao Feng, Liu Jian-Guo, Guo Qiang, Wang Bing-Hong,Node importance measurement based on the degree and clustering coefficient information? ,Acta Physica Sinica, 2013, (12)

[2] R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart, The Mahalanobis distance, Chemometrics and Intelligent Laboratory Systems, Volume 50, Issue 1, 2000,Pages 1-18,ISSN 0169-7439

[3] LIU Jian, CHENG Genyin,Analysis of coal consumption in China based on multiple linear regression, Journal of North China Institute of Science and Technology, 2018, 15(3).

[4] Yang Ke, Research on microblog user influence based on improved PageRank algorithm, Xi ' an University of Architecture and Technology, 2014.

[5] Aiken, L.S., West, S.G., Pitts, S.C., Baraldi, A.N. and Wurpts, I.C. (2012). Multiple Linear Regression. In Handbook of Psychology, Second Edition (eds I. Weiner, J.A. Schinka and W.F. Velicer).

[6] [1] Huang Decai, Qi Huachun. PageRank Algorithm Research [J]. Computer Engineering, 2006 (04) : 145-146 + 162.

[7] https://blog.csdn.net/guomutian911/article/details/81604545

[8] Hu Chao. The emergence of jazz and the social culture of the United States. [J] Voice of the Yellow River, 2017 (09) : 70.

10 A Letter for ICM Society

To:	The Integrative Collective Music Society
From:	Team # 2105330 of 2021 ICM
Date:	Feb 9,2021
Subject:	Omnipotent Music Analysis on Network and Changes in Music Characteristics

Dear Sir/Madam:

It is a great honor for us to build a model for you to measure the influence of music and examine the evolution and revolutionary trends of artists.

We did the following work and drew the relevant conclusions

- First, according to the data set provided, we propose a music influence network model to determine musicians' influence. The most influential artist is the Beatles. Besides, we can pick out artists who may represent the musical revolution, such as Elvis Presley, Marvin Gaye, etc based on the network.
- Second, we propose a music similarity model, which uses musical characteristics to develop a music similarity measure. Based on our model, we are 82.8% sure to draw a conclusion that artists of the same genre are more similar than artists of different genres. Besides, we can identify the differences and influences between genre.
- Then, we put forward the rise and decline model of genres to explore the evolution of genres over time. We found that between 1950 and 1970, genres changed significantly. We consider it a revolutionary moment for Pop/Rock and the entire music market in 1970.
- Finally, we introduce a difference-based social change model, which can identify the occurrence of social changes by mining music characteristics at different times. Based on our model, we have obtained two time points of dramatic social changes: during the economic crisis in the 1930s and the victory of World War II in the 1950s.

We define the value of music influence as the value of promoting the development of music, the value of finding potential musicians, the value of finding music trends, and the value of reflecting social changes. It is of great significance to fully explore the influence of music.

Considering that the data set given is only a subset of the music industry. When faced with larger data sets, we may encounter challenges such as too many genres, vague genre definition, and incomplete music data, but we still have confidence in our model.

Our work and solutions will be adjusted with more data. For a large number of genres, we plan to use cluster analysis; for data gaps, we expect to use some reasonable prediction algorithms; For more complex influence relationships, we will introduce corresponding complex network models.

Thank you again for inviting us to analyze the evolution and revolution of music.

Sincerely,
Team # 2105330 of 2021 ICM