Problem Chosen	2021 MCM/ICM Summary Sheet	Team Control Number
D		**2120266**

Seek evolution from directed influencer network

Summary

Music, the soul of mankind, only those who understand music can understand life. In order to examine the revolution and change trends of artists and genres, we construct the analysis framework of **"Building a directed influencer network–Exploring musical influence indicators–Seeking the similarities and diffreences within and between genres–Capturing the influence of musicians on each other"** , which mainly solves the following problems:

For question 1, we constructed a directed complex network model where influencers point to followers. And by creating a sub-network based only on the following relationship between musicians,and three sub-networks of **the same year-artist**, **the same genre-artist**, and **the same year-same genre-artist**, we dig out eight indicators, which are the importance, the competitiveness of the genre and so on. A **topsis model** modified based on the **entropy method** was established to obtain the artist's music influence ranking. We found that the Beatles are the most influential musician in the social network.

For question two to four, we first perform Kmeans clustering on musicians of the same genre and divide them into several sub-genres. Then we divide all the features into three categories: **music characteristic, sound characteristic and music description**, perform PCA respectively, and extract seven features as indicators to measure music similarity, establish a music similarity measurement model based on Cosine Similarity, and discover the artists within genres are more similar than the artists between genres. Then through the **OpenOrd algorithm**, it is found that musicians of the same sub-genre have more similar musical styles. We put forward the concept of the ancestor in the music field based on the tree structure. With the help of the union-find algorithm, we calculate the similarity between the ancestor node and the child node in the 6 musical features, find that influencers did influence their respective followers and **artists with higher influence have greater influence on followers**, which is mainly reflected in the three musical characteristics of emotion, rhythm, and purity, and these three characteristics are more related to popularity.

For question five to seven, we discovered three revolutionary characteristics of popularity, acousticness, and energy by observing the inflection points of the various characteristics process and the nodes with high or low growth rates in the music evolution. Combining directed influencer network, it is found that **musicians headed by JohnnyCash and Miles Davis have caused great changes in the 1950s**. We studied R&B in depth and then found that before 1950, all six indicators fluctuated violently. After the R&B was formally formed in 1950, the music style of the genre began to unify. At the same time, we evaluate each influencer and find out the dynamic influencers among them. And we analyzed the influencer value of R&B genre over time. According to the network structure of musicians and related research conclusions, we find that the development and characteristic changes of music in the 20th century are closely related to major events, important breakthroughs in music history, and new ways of creating music.

Finally, we analyzed the Strengths and weaknesses of the model and made a conclusion.

Keywords: OpenOrd; Complex network; Union-find;

Our Letter

Dear President of the Association:

With the strengthness of the composer's awareness of autonomy, the broadening of the scope of music equipment, and the diversification of the audience's aesthetic tastes, it is necessary to start from a new perspective to measure musical influence. According to the requirements of your association, we are very happy to have the opportunity to introduce our research and findings to you, and hope that it can provide you with some valuable insights.

Our work is based on networks of artists. They have their own main genres, musical styles and other characteristics, and these data are buried in various sub-networks divided by time, genre, and association. In order to explore this potential treasure house, we first comprehensively assessed the influence of each musician in the network through a set of more universal models, and had the most basic grasp of the nodes of the network. After that, we further mined the information contained in each artist, each node in the network and proposed a set of models to measure the characteristics of musicians, and further explored the information in the network combined with the knowledge of mathematics and music science. Through this information, we used tools such as OpenOrd algorithm and dimensionality reduction algorithm to measure the connection and similarity between and within each genre, and found deeper "sub-genres" and crossovers multiple genres of "super genres". We have also discovered a group of musicians who have had a profound impact on the world, helping to clarify the music inheritance relationship in the complicated modern music history. Finally, we combined the explicit key historical nodes to find the reflection of the music industry on these worlds, and found out the close relationship between music and the whole world.

The music of the 20th century is completely different from the music of the previous thousands of years, but the music of the 21st century has made a greater leap on this. We can still find out the progress of the 20th century music from the existing data. It is still difficult to do anything in the 21st century, where the network is complex and the influencing factors are more diverse. So with more or richer data, we hope that in the same way, we can clarify the development and connection of new music in the 21st century, and even make bold predictions about the future development of music.

We are really appreciated for this opportunity to assist you in building up an online marketing strategy,and we are convinced that our proposal can be utilized in improvement of your competence for the threeproducts. Please feel free to contact us for further information on the project.

Yours Sincerely,

Team #2120266 of 2021 MCM

Contents

1 Introduction

1.1 Problem Background

As one of our highest expressions of thoughts and creativity, music has always been a difficult realm to capture, model, and understand [4]. However, music has always been a part of human society and an important part of cultural heritage.

At the same time, when an artist is creating music, the performance, style, and meaning of the music vary according to the cultural and social background.An artist learns from predecessors of different genres at different times. Artist can list many musicians who they think have influenced his music. Although the artistic genres of some musicians are not the same as theirs.

The influence of some artists can even make a musician who follows his studies change his genre and musical style. These influential musicians are usually likely to have made great contributions to the transformation of music genres in the era. Of course, the change of music genre does not only depend on these musicians. Sometimes it is the changes of the times and social background that make the musicians' thinking have some changes. These changes are often a revolutionary changes.

According to the understanding of the above, we can understand about the development of music in the whole society over time.

1.2 Assumptions

Assumption 1. It is assumed that the genre of the musician will not change.

Assumption 2. Do not consider the artist's innate innate ingenuity or other personal experience on the music they create.

Assumption 3. Assuming all the artists during 1920-2020 were recorded without omission which means the influence of musicians not in the data set is not considered.

1.3 Nomenclature

Table 1: Symbol Table

Variable	Description
$C_d(x)$	outside degree centrality
$C_b(x)$	betweenness centrality
$C_c(x)$	closeness centrality
$D_C(x)$	use the distance between Ax and the ideal vector Ac to define importance
x_{ij}	the element in theInfluence_data
nd	negative_degree
$\overrightarrow{A_x}$	measurement
Y	year
G	genre
r	Correlation coefficient of the same indicator in the same genre

2 Explore musical influence

In order to ensure the accuracy of the subsequent data processing and to avoid double edges when establishing a directed network, we processed the data with outliers and deleted all outliers. (For example a follower's name is the same as the influencer's name.) In the following, we will finish the steps to build up and validate the model.

2.1 Create multiple directed networks of musical influence

We use the 'influence_data.csv' dataset to create multiple music influence directed networks,which were divided into 4 types: Importance is a measure of the importance of points in the network. Here we measure the importance of influencers in the network.

2.1.1 Overall following network

we define a vector $\overrightarrow{A_x}$ which contains three measures as the form below:

$$\overrightarrow{A_x} = \left(\frac{C_d(x)}{\max C_d(x)}, \frac{C_b(x)}{\max C_b(x)}, \frac{C_c(x)}{\max C_c(x)} \right) \tag{1}$$

In Freeman's research1979, **outside degree centrality** $C_d(x)$, **betweenness centrality** $C_b(x)$ and **closeness centrality** $C_c(x)$ can be used to identify masters who have significant influence or impact in a network [3]. These values are defined as below:

$$C_d(x) = \sum_i a_{xi}, \quad C_b(x) = \sum_i^n \sum_j^n g_{ij}(x), \quad C_c(x) = \sum_{i=1}^n l_{ix} \tag{2}$$

In the same time,the ideal model of an author who has the most significant influence within the network will have his/her own measure vector.The ideal vector will be defined as $A_c = (A_1^C, A_2^C, A_3^C) = (1, 1, 0)$,We use the distance between $\overrightarrow{A_x}$ and the ideal vector$\overrightarrow{A_c}$ to define importance.We define the importance as $D_C(x)$:

$$D_C(x) = \sqrt{\left(A_{x1} - A_1^C\right)^2 + \left(A_{x2} - A_2^C\right)^2 + \left(A_{x3} - A_3^C\right)^2} \tag{3}$$

$D_C(x)$ is the importance value we need to get, and we sorted it according to the importance value of each influencer, as shown in the table 2:

Table 2: Top 5 ranking of importance

influencer_id	**influencer_name**	**distance importance**
754032	The Beatles	4.06942391
66915	Bob Dylan	2.332947219
894465	The Rolling Stones	1.804525432
548397	Jonas Brothers	1.726553279
2811293	Meghan Trainor	1.720445442

It can be seen from table 2 that The Beatles is the most important influencer. His influence can reach 4.07, which is 1.74 higher than Bob Dylan, whose importance is 2.33, and the importance of Bob Dylan is also far greater than the importance of the eight artists behind, which shows that the importance of these two artists on the Internet is extremely high.

2.1.2 Subnets according to years

- Competitiveness in the genre

The number of artists in the genre. The greater the number of artists in the genre, the greater the influence.

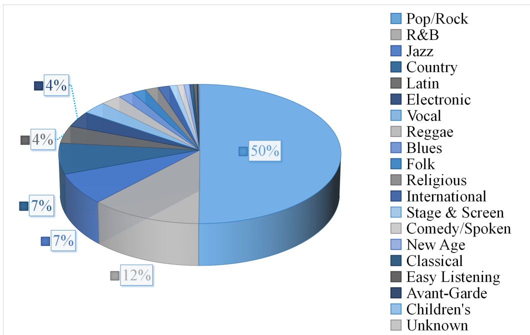


Figure 1: The number of artists in the field.

From the figure above we can see that,the number of artists in Pop/Rock is the biggest.

- Number of followers across areas

The degree of attraction of an artist's work to artists not in the field can also be said to be a purely artistic measure of the artist's work.

Influencer_id 754032 have the most follwers across areas.

2.1.3 Subnets according to genre

- Artistic career competitiveness

The total number of artists in all fields in the ten years, which means difficulty for artists to divide the influence of this cake.

The greater the competitiveness, the greater the influence.

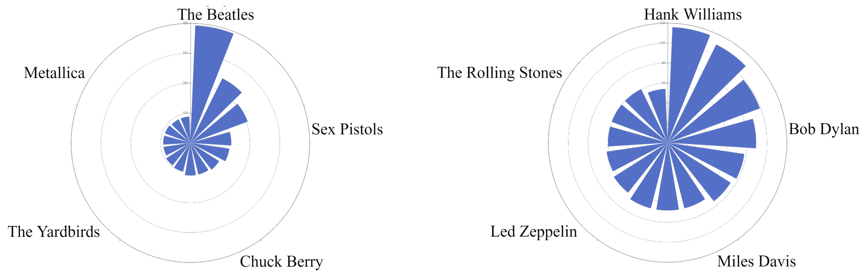
The total number of artists in all fields is 771 in 1960,which is the most of all years. While the total number of artists in all fields is 15 in 2010, which is the least of all years.

- Number of followers in art career

There is valid evidence to prove that most musicians' careers will not exceed 30 years. This can also be observed in the data. In more than 40,000 data, the start time of the genre of followers is earlier than that of influencers. There are only more than 700 pieces of data for 30 years and more.

- Number of followers outside art career

The artist's influence on the future era after his creative career is over.



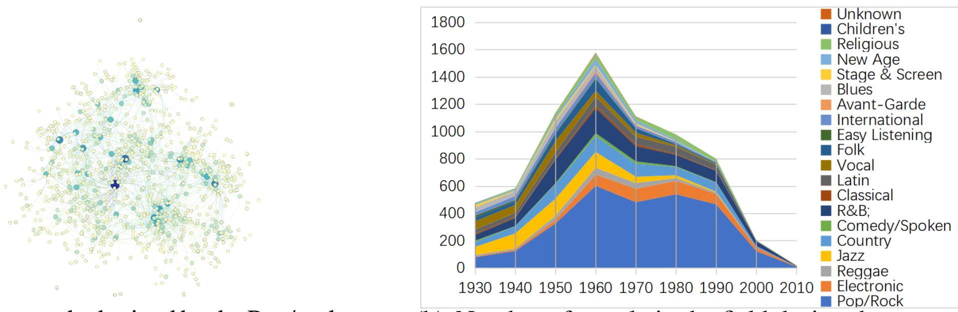
(a) Number of followers in art career (b) Number of followers outside art career

Figure 2: Number of followers in and outside the art career

2.1.4 Subnets according to years and genre

The following figure shows a directed network from influencers to followers obtained by the Pop/rock genre in 1990.

- Number of people in the field during the career



(a) Directed network obtained by the Pop/rock genre in 1990. (b) Number of people in the field during the career

Figure 3: Network obtained by the Pop/rock genre in 1990 and number of people in the field during the career

Ordinary geniuses in the era of contending hundreds of schools of thought can often shine in the era of dim starlight, which can be understood as the difficulty of art in gaining influence in the field during which the artist is active.

- influence ranking in the field every ten years

Filter out all the artists in the field of this year, and then rank them according to the number of influencers (the lower the ranking, the greater the influence).

Correlation between eight indicators

After obtaining the 8 indicators of the 4 types of directed graphs, we perform correlation analysis on these 8 indicators, and the correlation coefficients obtained are shown in the following figure:

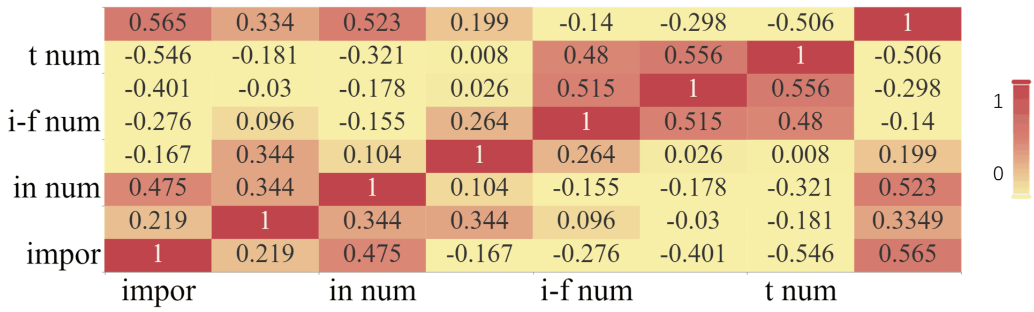


Figure 4: Heat map of each indicator

It can be seen that the correlation of various indicators is not high, so we do not do dimensionality reduction processing on them.

2.2 The EM-TOPSIS evaluation model

We have established a topsis model based on the entropy method to get a comprehensive evaluation score. The specific steps are as follows:

Step 1. Normalize the original matrix

We want to normalize the original matrix, which means to uniformly transform all indicator types into extremely large indicators.

Among the above eight indicators, only influence ranking in the field every ten years is a minimal indicator. The smaller the ranking, the greater the influence. Here we will convert it into a large indicator to directly filter out the artist's field.

Step 2. Forward matrix normalization

There are a total of n pieces of data, and 8 feature evaluation indicators form a normalized matrix X. We denote the standardized matrix as Z, and each element in Z. Our purpose of normalizing the forward matrix is to eliminate the influence of different dimensions.

Step 3. The entropy weight method

First we Calculate weight of i in the index j get p_{ij}. Then we calculate the information entropy of each index and calculate the difference coefficient in the index of j to get e_j and d_j. At last We use this formula to calculate the entropy weight in the index of j:

$$W_j = d_j / \sum_{j=1}^{8} d_j \tag{4}$$

Step 4. Calculate and normalize the score

- Define the maximum and minimum

$$\begin{aligned} \mathbf{Z}^+ &= \begin{pmatrix} z_1^+ & z_2^+ & z_3^+ & z_4^+ & z_5^+ & z_6^+ & z_7^+ & z_8^+ \end{pmatrix} \\ \mathbf{Z}^- &= \begin{pmatrix} z_1^- & z_2^- & z_3^- & z_4^- & z_5^- & z_6^- & z_7^- & z_8^- \end{pmatrix} \end{aligned} \tag{5}$$

where $\mathbf{Z}_j^+ = \max_i \left(x'_{ij}\right), \mathbf{Z}_j^- = \min_i \left(x'_{ij}\right)$

- Define the distance between the i-th evaluation object and the maximum value and the minimum value

$$D_i^+ = \sqrt{\sum_{j=1}^{m} \omega_j \left(Z_j^+ - z_{ij}\right)^2} \quad D_i^- = \sqrt{\sum_{j=1}^{m} \omega_j \left(Z_j^- - z_{ij}\right)^2} \tag{6}$$

- Calculate the unnormalized score of the i-th evaluation object and normalize to get the final score

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{7}$$

2.3 Results and analysis

We evaluate an influencer's influence by the method, and get the music influence evaluation score of this influencer, and sort the order in the following table.

Table 3: Top 10 ranking of scores

influencer_id	influencer_name	score
754032	The Beatles	0.004985323
66915	Bob Dylan	0.003425467
894465	The Rolling Stones	0.002681564
549797	Hank Williams	0.002284731
316834	Marvin Gaye	0.00219555
531986	David Bowie	0.001912784
379125	Neil Young	0.001788269
139026	Led Zeppelin	0.001772536
100160	The Kinks	0.001746583
423829	Miles Davis	0.001709217

In the established network, we have obtained the influence of individual influencers and made a ranking of them.

3 Calculate similarity and correlation

3.1 Data preprocessing

Step 1. Standardlization:

We standardlize tempo, key, popularity and year with minmax standardlization.On the other hand, we changes the loudness value less than -40 to -40, delete a few huge duration_ms, and performs minmax standardlizationon these two features.

Step 2. Dimensionality reduction

First, we conducted a general correlation analysis on the data.Then We divide data into three categories according to music characteristics, sound characteristics, and music description, and perform PCA dimensionality reduction respectively.

Table 4: The results of PCA

Features	First	Second	Third	sum
music characteristics	0.4116	0.2153	0.1985	0.8254
sound characteristics	0.4862	0.2423	0.12941	0.8579
music description	0.8497			0.8497

As a result, 7 new features corresponding to each song are obtained, that is, each song corresponds to a 7-dimensional vector.

K-means clustering

Step 1. Get each artist's genre from influence_data.csv in one-to-one correspondence, and expand the full_music_data.csv data set.

Step 2. According to the full_music_data.csv data set, we get the correspondence between the song-artist genre.

Step 3. Perform Kmeans clustering on the data, and set the cluster to 3.

The clustering **results** are as follows in Figure 5:

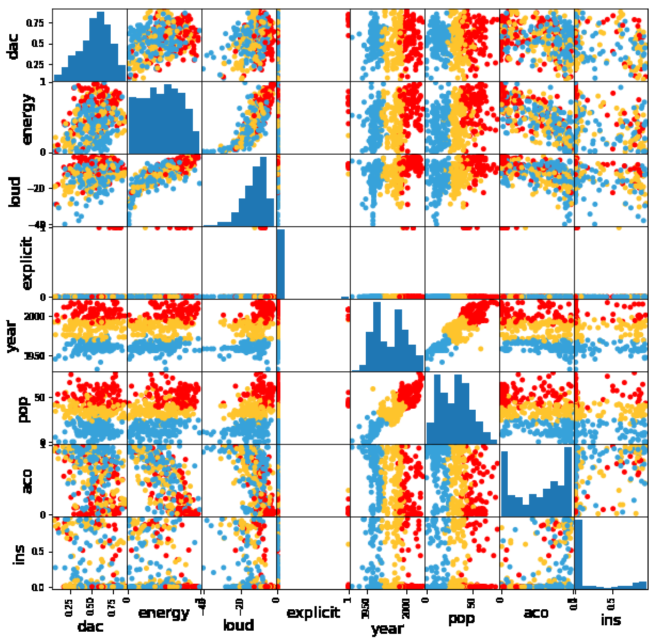


Figure 5: Clustering results

From the figure above, We can see that genres can be clearly classified in above 8 features.

3.2 Calculate difference with Cosine Similarity

The **reason** why we measure the similarity of music using the cosine similarity of the music vector instead of the Euclidean distance is because the same style of music may have a numerical difference in its corresponding "music vector" because of its radical or soft style. However, this situation can be more focused on distinguishing the difference of the direction of different "music vectors" rather than the cosine similarity avoidance of numerical differences, or in other words, the direction of each "music vector" is the style of the song.

Cosine similarity has a special property that makes it suitable for metric learning:the resulting similarity measure is always within the range of 0 and +1, which allows the objective function to be simple and effective. [6]

Cosine similarity (CS) between two vector is defined as:

$$\cos\theta = \frac{xy}{||x|| \times ||y||} \tag{8}$$

We use the Cosine Similarity(CS) to calculate the similarity between all the songs of a singer and all the songs of the singers of the same genre, the similarity of all the songs of a singer and all the songs of the other two types of genres. The **result** is as follows:

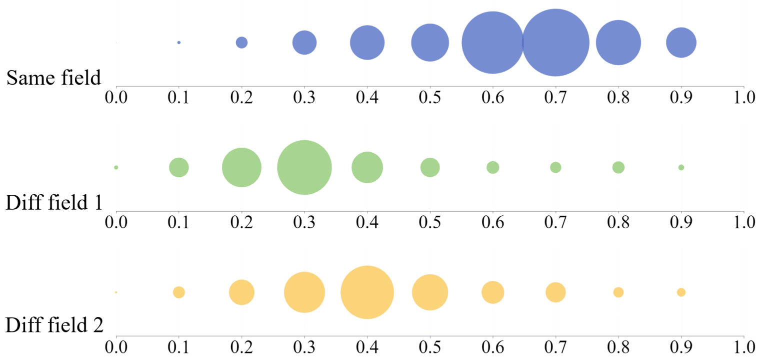


Figure 6: the similarity between all the songs of a singer and all the songs of the singers of the same genre, the similarity of all the songs of a singer and all the songs of the other two types of genres.

The closer the cosine similarity is to 1, the higher the similarity between the two vectors.

The first axis represents the similarity between all the songs of a singer and all the songs of the singers of the same genre. We can see that the song similarity of singers of the same genre is mostly distributed between 0.6 and 0.8, mostly greater than 0.8, which is the same as the expected result, and the song similarity of singers of the same genre is higher.

The second and the third axis represents the similarity of all the songs of a singer and all the songs of the other two types of genres. We can see that the similarities between the songs of the first genre singer and the songs of the second genre are mostly between 0.2 and 0.4. The similarities between the songs of the first genre singer and the songs of the third genre are mostly between 0.3 and 0.5. The similarity is not very high.

So we came to the conclusion that the similarity within genres is higher than the similarity between genres.

3.3 Calculate correlation coefficient

For the similarity within a genre, we judge the distribution of a certain characteristic value of a certain genre and compare it to find that a certain characteristic of a certain genre is more decisive, but the other indicators are less decisive. It can be concluded whether the characteristics of this genre are similar.

3.3.1 Index selection

After consulting information and consulting professionals, we extracted 6 secondary indicators from 11 primary indicators. There are 11 first-level indicators, as well as the second-level indicators extracted from them, as well as their description, range and nature of change.

Here we will slightly improve the mode indicator according to the meaning of mode.

mode: An indication of modality (major or minor), the type of scale from which its melodic content is derived, of a track. Major is represented by 1 and minor is 0.The general characteristics of major music are positive emotionally. The general mood of minor music is negative. In order to balance the value of the three components of the emotion indicator, we changed mode=0 or 1 to mode=0.35 or 0.65 because the numerical difference brought by the previous mode accounted for too much in the numerical difference of the emotion indicators of each song.

3.3.2 Pearson correlation

Correlation is a measure of a monotonic association between 2 variables. A monotonic relationship between 2 variables is a one in which [7]:

(1) as the value of 1 variable increases, so does the value of the other variable

(2) as the value of 1 variable increases, the other variable value decreases.

What we use here is the Pearson correlation coefficient to calculate the similarity within and between genres. The Pearson correlation coefficient is used to measure the correlation between two variables X and Y, and its value is between -1 and 1:

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

r represents the correlation coefficient of the same indicator in the same genre.

We draw box plots based on the Pearson correlation coefficient of each feature of each genre:

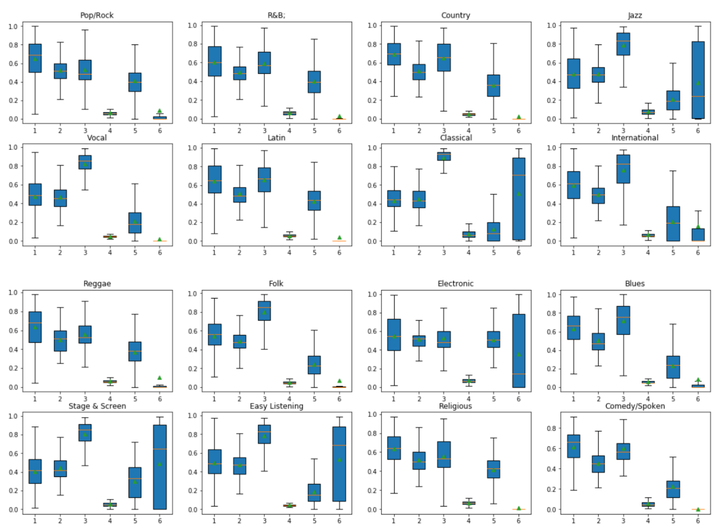


Figure 7: Box plots of similarities and differences within genres

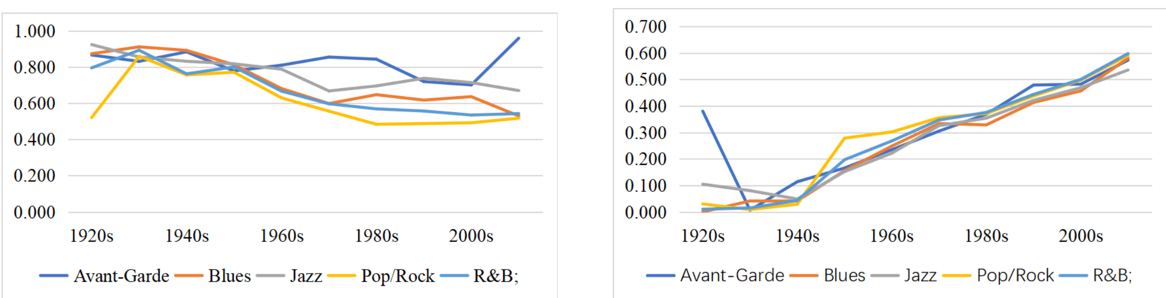
If the rectangular block is relatively short, it means that the index is relatively concentrated and reflects the similarity within the genre.

If the rectangular block is relatively long, it means that the index is relatively scattered and reflects the differences within the genre.

From the Figure 7 above, we can see that:

- Similarity within genre
  - The overall duration of Pop/Rock is relatively short, the general sense of rhythm is relatively strong, the mood is relatively positive, and the songs are generally sung by people.
  - The purity of classical music is higher.
  - The rhythm of most genres is more concentrated in the range of 0.3 to 0.6.
  - Comedy/Spoken, Latin generally have a more positive mood, which is more in line with their musical characteristics.
- Differences within genre
  - Jazz, easy listening, whether the track contains different vocals, the distribution is more scattered.
  - The popularity of stage&screen is also scattered.
  - In Reggae, the emotional tendency of music is also quite different, which shows that some songs are more active and some are more negative.
- Influence between genres
  - With the rapid increase in the influence of popular music, some of the characteristics of other music (music genres similar to popular music) have become more similar to popular music, and emotions have become more positive.
  - With the development of Country music, other similar music genres have become more emotional and more pure.

Through the above calculation method, we have calculated the comparison chart of the six indicators of each genre over time. The following shows only five genres:



(a) Purity (b) popularity

Figure 8: The changing trend of different genres with the same characteristics over time

Over time, the rhythm of genres has not changed significantly.Over time, the popularity of each genre is on the rise.Over time, the overall purity of each school has shown a slight downward trend.

First, the characteristics corresponding to various songs are divided into 4 categories through the hexagram chart.

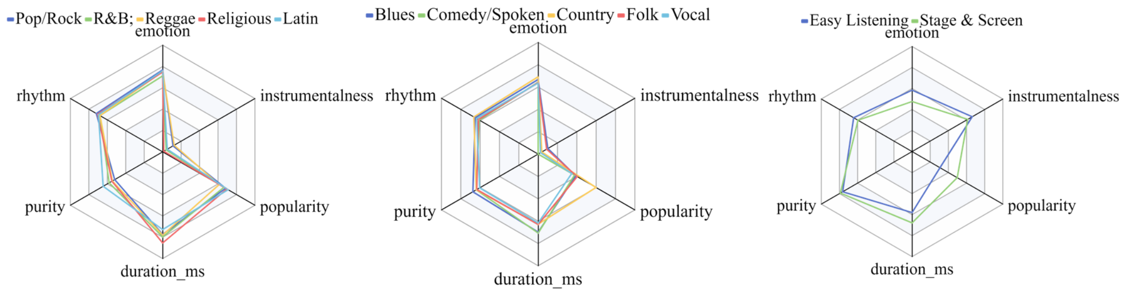


Figure 9: Three types of feature distribution

We found that although the shapes of the first and second types are similar, they are partially different in value. This is specifically reflected in the fact that the rhythm, emotionality and duration of the first type are higher than those of the second type, and the instrumentalness value of the two Values are very low

Compared the third categorywith the first two categories, the third category has higher musical purity. The rhythm, emotion, and duration are not too different, but the instrumentalness value is much higher, which represents the instrument of this type of music. The proportion is very high.

3.4 OpenOrd Model

Force-directed layout algorithm is a kind of drawing algorithm, which draws based only on the structure of the graph itself.

We regard each cluster of points in the figure as a unit, and there are several lines between each cluster of points and other clusters. The gravitational force generated by these lines is related to the number of lines between the two clusters (The connection between the dot groups) is related, so the distance between the two dot groups in the force-directed layout diagram also represents the closeness of the relationship between the two dot groups.

We use the OpenOrd algorithm to realize the concept of gravitation-repulsion, and at the same time, through edge cutting, ignoring part of the long edges, covering part of the blank in the network diagram, etc., making the image expressiveness better.

Suppose we have an undirected weighted graph $G = (V, E)$, where the vertices are given by $V = v_1, ..., v_n$ and the edges are given by $E = e_{ij}$. Let $W = w_{ij}$ be the adjacency matrix corresponding to the graph G so that edge e_{ij} has weight w_{ij}. Since the graph is undirected, we know that $w_{ij} = w_{ji}$ so that W is symmetric.

The goal of OpenOrd is to draw G in two dimensions. Let $x_i = (x_{i,1}, x_{i,2})$ denote the position of v_i in the plane. OpenOrd draws G by attempting to solve:

$$\min_{\mathbf{x}_1,\ldots,\mathbf{x}_n} \sum_i \left(\sum_j \left(w_{ij} d\left(\mathbf{x}_i, \mathbf{x}_j\right)^2 \right) + D_{\mathbf{x}_i} \right) \tag{10}$$

where D_{x_i} denotes the density of the points $x_1, ..., x_n$ near x_i. The sum contains both an attractive and a repulsive term. The attractive term $\sum_i \left(\sum_j \left(w_{ij} d\left(\mathbf{x}_i, \mathbf{x}_j\right)^2 \right) \right.$ attempts to draw together vertices which have strong relations via w_{ij}. The repulsive term D_{x_i} attempts to push vertices into areas of the plane that are sparsely populated. [5]

Simulated annealing, which involves the probabilistic decision to take moves that actually increase the energy associated with the node can solve the problem. [2]

Control node clusters and the number of blank areas in the layout through Edge-Cutting. In order to control node clustering, our heuristic method affects the relative importance of the repulsive and gravitational components in equation. In order to control the gap, we allow to ignore some long edges and cover in the optimization process of the objective function. Part of the blank in the network diagram. Make the image rendering process look better through the above edge cutting method.

Through the above method, we draw the influence_data including the entire network as shown below:

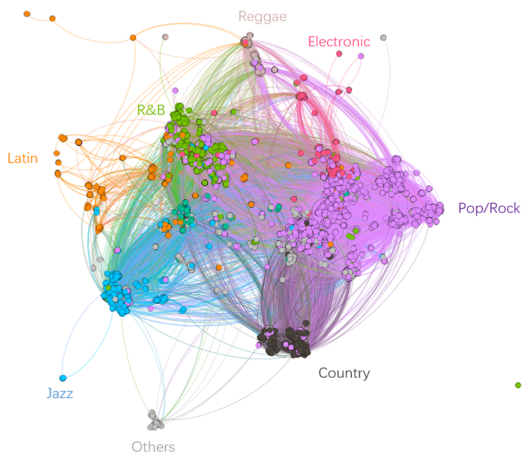


Figure 10: Entire network according to genres

Through the above images, combined with the "closer the connection, the closer the position" in the OpenOrd algorithm, we can get the following properties of various music genres:

- The number of popular musicians is the largest, and they have the closest connection with electronic and country. At the same time, some popular musicians are more closely connected with Jazz, country, and R&B musicians than with other popular musicians. They are "maverick people"; the genres that have less connection with pop music are Latin, Reggae, Jazz.
- R&B genre is relatively closely related to most genres (except Country)
- Electronic and Reggae-Jazz and Latin-Country are relatively sparsely connected

3.5 Union-find algorithm

ancestor node: A small number of artists can be found in each field. These artists affect all artists, and all artists can find their only "ancestor node".

We use the union-find algorithm to find the ancestor node of all nodes.

Union-find algorithm is mainly used to solve the problem of grouping some elements. It manages a series of disjoint sets.

Our steps to calculate the ancestor node are as follows:

Step 1. initialization : Initialize the set of each point to itself.

Generally speaking, this step only needs to be executed once each time the data structure is used, and the time complexity is O(N) regardless of the implementation method.

All elements in a set are organized into a tree structure with a representative element as the root

Step 2. Find: Find the collection where the element is located(the root node).

To determine the set of x is to determine the representative of the set. You can move up the tree structure continuously along parent[x] until you reach the root node.

Step 3. Merge: Combine the collections of the two elements into one collection.

Generally speaking, before merging, you should first determine whether two elements belong to the same set. This can be achieved by the above "find" operation.

For each element parent[x] points to the parent node of x in the tree structure. If x is the root node, then let parent[x] = x.

3.6 Results analysis

1. Influence of influencers

We get the number of ancestor nodes in each field .There are 1171 ancestor nodes in the Pop music field, and there are 339, 98, 263, and 297 in electronic music field, Reggae music field, Jazz music field and Country music field.

In order to numerically show the similarity between the 6 index characteristics of the ancestor node and the 6 index characteristics of its child nodes, we define that the difference between a certain characteristic value of the child node and the corresponding characteristic value of the parent node is within the following range. The node and its ancestor node are similar in this indicator. The difference between the index and the influencer is considered similar within the following range: $'emotion' + -0.5$, $'rhythm' + -5$, $'purity' + -0.25$, $'duration\_ms' + -20000$, $'popularity' + -10$, $'instrumentalness' + -0.01$.

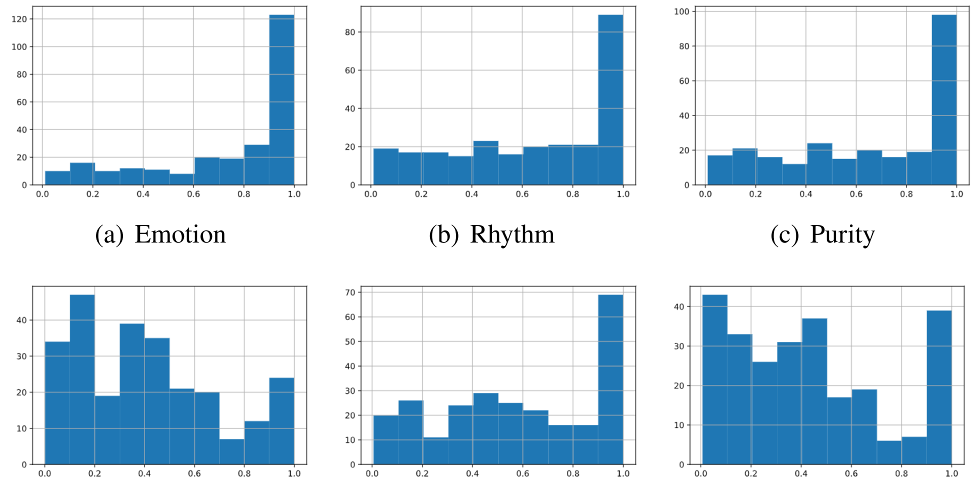
Then we separately count the proportions of all the child nodes of each ancestor node, the number of child nodes with similar characteristics to its ancestor node, and obtain the influence of each ancestor node's 6 indicators on all its child nodes (reflecting A certain index of its child node and the index of the ancestor node are similar in proportion, then the index of the ancestor node influence the respective artists).

In the statistical results, we remove the ancestor node with the sum of 6 ratios of 6, because the reason for this is that the ancestor node has only one follower of itself.

The statistical results are the numerical distribution of'emotion','rhythm','purity','duration_m -s','popularity', and 'instrumentalness':

As can be seen from the figure below, the identified influencers in fact influence the respective artists.

It is specifically reflected in the three indicators of'emotion','rhythm', and'purity'. The ancestor node has a high degree of similarity with the child nodes. In these four characteristics, the ancestor node will affect the child nodes.



(a) Emotion (b) Rhythm (c) Purity

(d) Duration_ms (e) Popularity (f) Instrumentalness

Figure 11: numerical distribution of sin genres

According to the degree of importance of the musicians we obtained, we selected 5 influential musicians from the ancestor nodes. They are The Beatles, The Rolling Stones, Miles Davis, Jimi Hendrix and Stevie Wonder we are all familiar with.

According to the previous calculation method, we calculated the similarity of 6 indicators between followers of influential artists and artists:

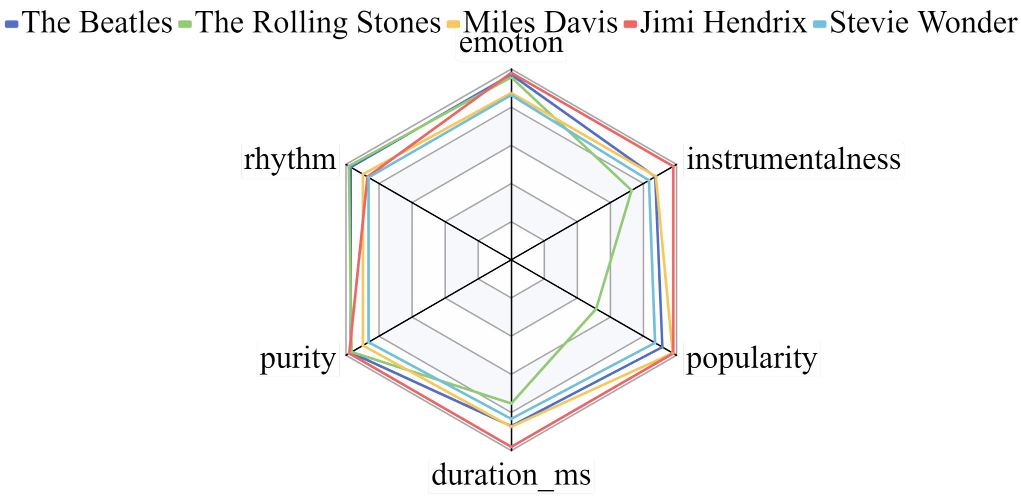


Figure 12: the similarity of 6 indicators between followers of influential artists and artists

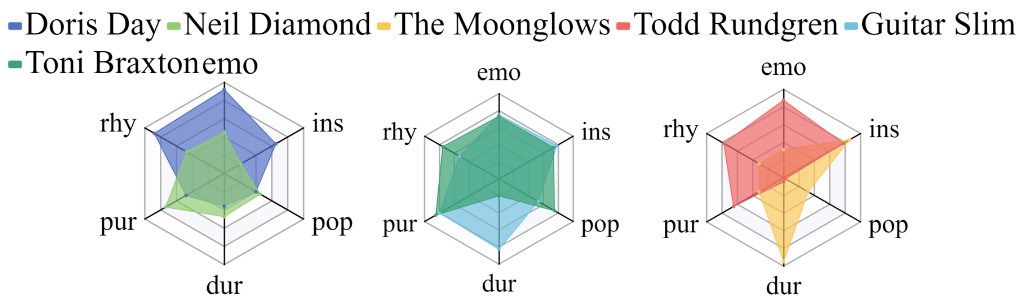


Figure 13: The similarity between the followers of ordinary artists and the 6 indicators of artists

Then randomly select 5 ordinary musicians (musicians with no strong influence) for comparison. Obviously, for artists with higher influence, their followers are more similar to the artists they follow in various musical characteristics, especially in the emotion, rhythm, and purity that have proved to be more influential; and for ordinary the similarities between artists and their followers have their own merits.

This proves that the ‘influencers’ actually affect the music created by the followers.

2. Correlation

The Full_music_data dataset uses correlation analysis to explore the correlation between various indicators and popularity. The correlation is higher, which means it is more infectious.

According to the following data, music features with a correlation coefficient greater than 0.2 are considered to be more contagious. We get **results**:There are five indicators of energy, loudness, year, acousticness and instrumentalness are considered more infectious, especially loudness and acousticness.

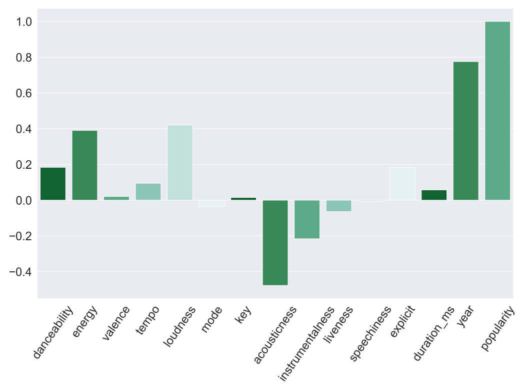


Figure 14: Correlation between various indicators and popularity

4 Analyze the evolution of music

4.1 Data Preprocessing

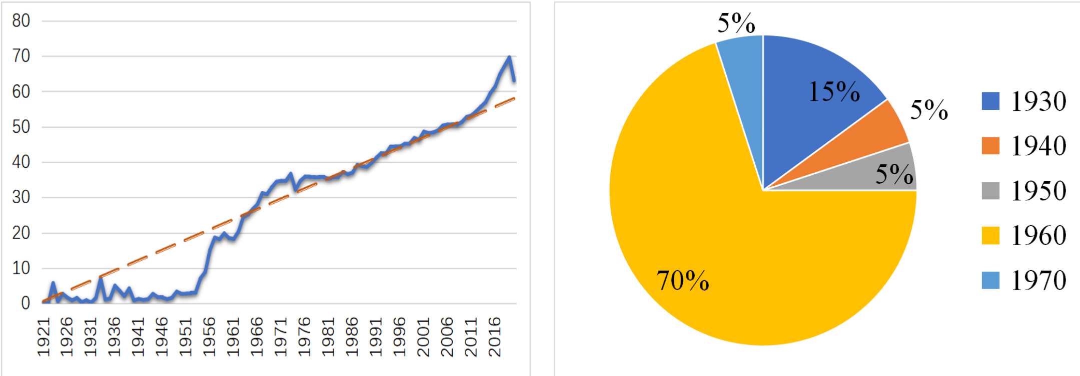
The three indicators all have a rapid change trend between 1955 and 1975, so we guess that there may be changers in this time period. The following combined data to verify our guess.

Step 1. From the full_masic_data data set, extract songs whose creation time is in the interval of 1955-1975. There are more than 90,000 songs in total. In the past 20 years, there have been 34,000 songs, accounting for one-third, which also shows that this stage is an exceptionally brilliant time for music creation.

Step 2. The change makers understand that at the beginning of this stage from 1955-1960, the musical characteristics created by other artists are significantly different from those created by other artists and are in line with the overall trend of change.

4.2 Impact on characteristics

4.2.1 Popularity



(a) The trend of popularity (b) Proportion of influencer every ten years

Figure 15: popularity

From the figure (a) we can see:

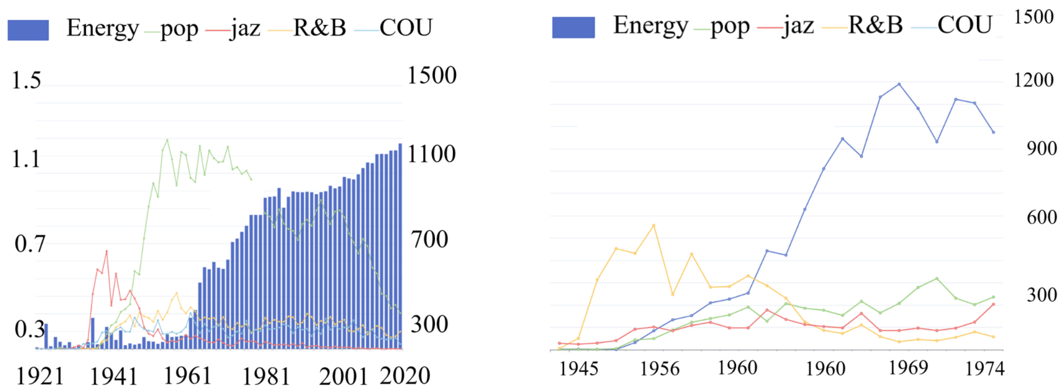
1. The overall indicator is on the rise
2. We divided into three time periods for specific analysis
    - 1921-1954 Popularity showed a trend of steady fluctuations overall.
    - 1955-1975 Popularity has the rapid growth stage.
    - 1976-2020 The growth of the popularity slows down.
3. From 1955 to 1960, there were a total of 8,820 songs, 3806 songs with a popularity index greater than the average value of 17, and among these songs, there were six musicians who wrote songs greater than 100.

We can explain why popularity has the rapid growth stage from Figure (b) :

Table 5: six musicians who wrote songs greater than 100

name	active_start	count
Glenn Gould	1940	128
Miles Davis	1940	148
Ella Fitzgerald	1930	197
Frank Sinatra	1930	177
Elvis Presley	1950	145
Billie Holiday	1930	130

According to the ranking of artist influence calculated by the first question, it can be seen that among the top 20 most influential artists, 80% were active in 1955-1975. At the same time, we also learned that a group of extremely influential musicians suddenly emerged in the 1960s and 1970s, such as The Beatles, Bob Dylan, The Rolling Stones, David Bowie, Neil Young, Led Zeppelin and so on, created a large number of popular songs, so the music influence has increased greatly at this stage.

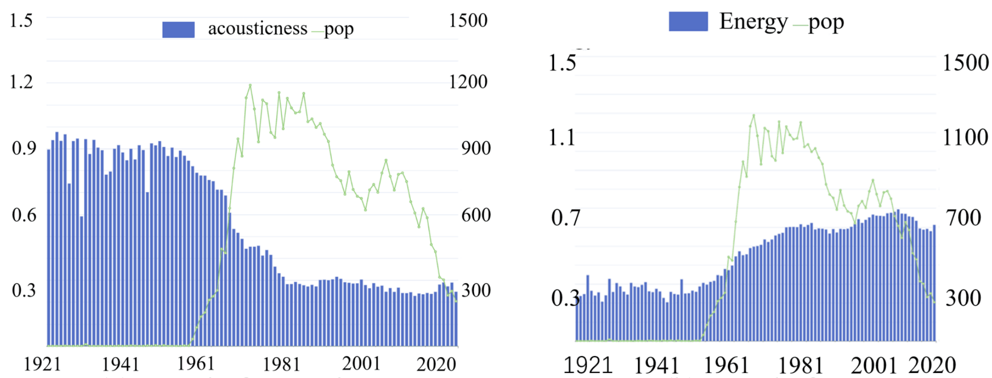


(a) A comprehensive analysis of the annual statisticsof the four types of music and the populairit of themusic industry each year

(b) 1945 after the end of World War II

Figure 16: popularity changes in music evolution

It can be seen from the figure that with the decline of jazz, the overall popularity of music has not decreased, but has risen sharply in response to the rise of popular music. Specifically, around 1960, the trend changed from jazz music to pop music, and about 10 years later opened the era of popularization of music that most people prefer and are more willing to accept.



(a) Trend of acousticness

(b) Trend of energy

Figure 17: popularity

In the two figures above, from the 1950s, with the rapid increase of popular music, two major leaps appeared in the evolution of music: during the 20 years of rapid development of popular music into mainstream music, energy gradually increased Until it doubles, the acousticness index gradually decreases to one-third of the original. This is inseparable from the vibrant pop music and the various processing techniques used in the production process.

At the same time, the impact of popular music is far-reaching. After 1985, when the number of popular music was declining, the two characteristics it brought are still maintained, that is, the changes in these two characteristics driven by popular music. After 1985, it has deeply influenced other types of music.

It can be seen from the Figure (a):

From 1955-1960, there were a total of 8,820 songs, and 3381 songs had an acousticness index greater than the average value of 0.79. Among these songs, there were three musicians, JohnnyCash, Miles Davis, and Ella Fitzgerald.

Figure (b) shows that:

From 1955-1960, there were a total of 8,820 songs, and 3,785 songs had an energy index greater than the average value of 0.31. Among these songs, there were two musicians who created songs greater than 100: JohnnyCash,Miles Davis.

4.2.2 Conclusion

The game changer is a group of musicians headed by JohnnyCash and Miles Davis. In the 1850s, the two musical characteristics of energy and Acousticness set off a wave, and the music has become more energetic since then, with more technology enhancements.

4.3 Influence process

We chose R&B for analysis:

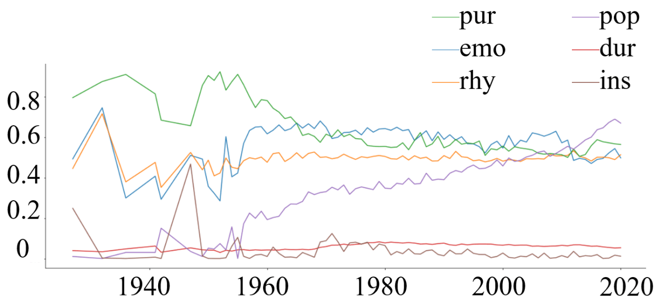


Figure 18: analysis for R&B

It is not difficult to find that before 1950, the six characteristics of R&B music were all very disorderly. On the one hand, the early R&B music before 1950 was mainly produced by the black population at the bottom; the melody is simple, the lyrics are rough, and the singing and orchestration are random. , Oral singing, it is difficult to record scores, resulting in less collection and little similarity between them. After that, the American Billboard magazine put forward the concept of R&B in the late 1940s, which allowed R&B to be divided systematically, and the list of tracks continued to increase. That is to say, the changes in the 6 characteristic values after 1950 are relatively stable. There are several important nodes in the development history of R&B music:

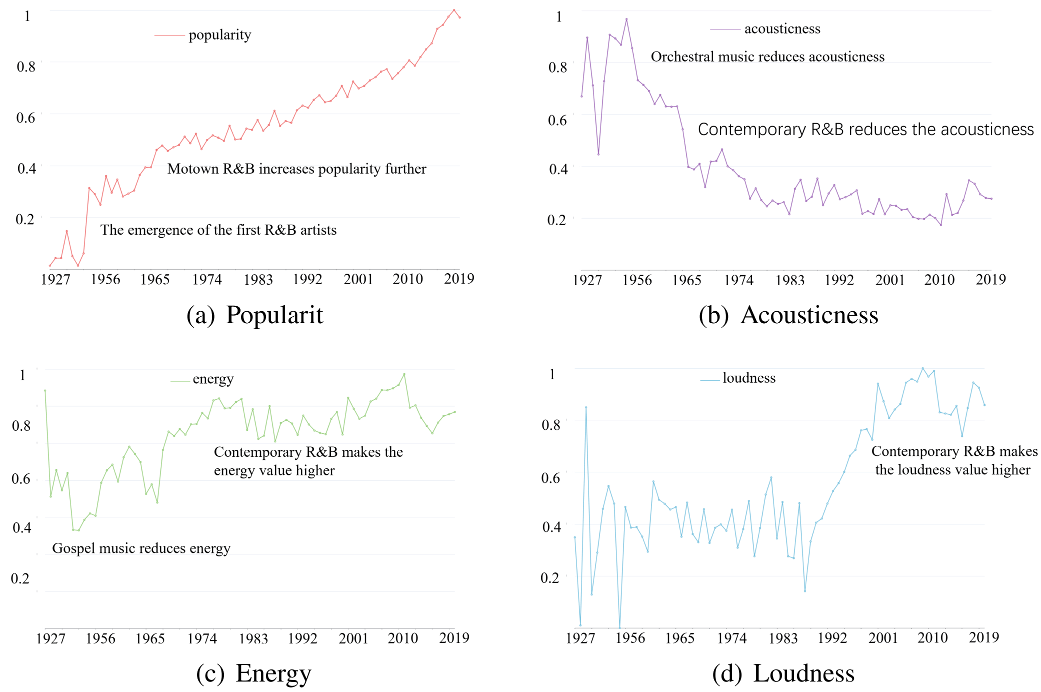
From the 1950s to the 1960s, R&B was recognized as a trend in the music market around the world. The first batch of R&B musicians such as Ray Charles, James Brown, Jackie Wilson, etc. appeared, and the popularity value increased.

1. The emergence of Motown R&B in the 1960s.

According to a 1971 article in Rolling Stone [1].Because black artists of the R&B genre of this era have been influenced by church music since childhood, this style of R&B frequently uses gospel music as a background, which reduces the energy value of this period; repetitive and complicated use of wind and string music , The synthetic characteristic purity value suddenly decreases while the acousticness value increases; and the appearance of mature and unique R&B such as Motown R&B further increases the popularity of R&B.

2. The development of Contemporary R&B

After 1970's, R&B evolved into Contemporary R&B style. Its commonly used modern typical musical instruments include electronic recording and production of Loop drum rhythms, and the introduction of synthesizers and electronic music, which further reduce the R&B acousticness value of this period. And because these modern instruments are compared with previous traditional instruments , The loudness is generally greater, so the loudness increases Their impact on the popularity, acousticness, energy, loudness and other indicators of R&B music are as follows:



(a) Popularit (b) Acousticness

(c) Energy (d) Loudness

Figure 19: Impact on the popularity, acousticness, energy, loudness

4.4 Dynamic influencer

We comprehensively measure dynamic influencers through two aspects: the range of influence of each influencer and the degree of influence of each influencer on followers

In terms of the scope of influence, we mainly focus on number of child nodes of each influencer and number of grandchildren nodes of each influencer. At the same time, different weights are given to such child nodes and grandchild nodes, and the sum is obtained.In terms of the degree of influence, we use the fourth question method to find the similarity between the artist and his ancestor node in each feature, and find the proportions of each influencer's followers, 6 characteristics are similar to the influencer, and the 6 proportions And as a measure of influencer's influence on its followers.

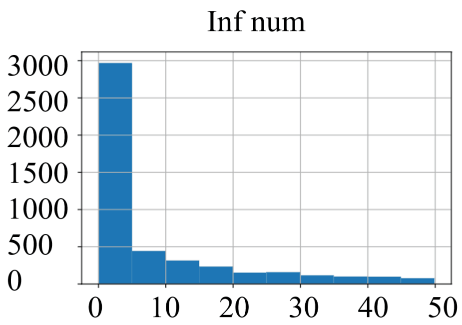
A total of more than 5000 influencers have most of the range of influence values between 0 and 50 (as shown below).

In order to balance the numerical relationship between the influence degree and the influence range, we multiply the influence degree whose value range is [0,6] by 8. And because there is a certain mutual restriction between the two indicators, that is, when the scope of influence is wide, the proportion of followers similar to influencer will be more difficult to increase, and vice versa. So we can add these two indicators to get the indicators that ultimately measure dynamic influencers.

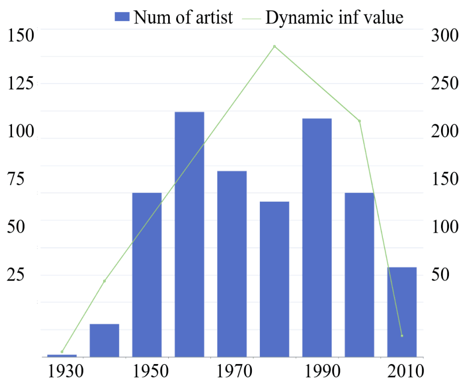
Calculate the dynamic influencers index for each influencer, and get the top 15 artists of this index:The Beatles,Bob Dylan,The Rolling Stones,Chuck Berry,Elvis Presley,Jimi Hendrix,The Velvet Underground,Hank Williams,The Kinks,Little Richard,The Who,David Bowie,James Brown,Led Zeppelin,Sex Pistols.Most of them are pop musicians.

We also analyze the dynamic influencers of R&B music over time. We draw the number of artists and dynamic influencers in each era of RB music as shown below also analyze the dynamic influencers of R&B music over time.

It is easy to find that the Figure shows that although the number of musicians in the RB music industry in 1970 and 1980 was small, their dynamic influencers were more extensive. The reason is that R&B musicians such as Chic, Prince, etc. have a large number of followers.



(a) Influence values of influencers



(b) The number of artists and dynamic influencers in each era of R&B music

Figure 20: Dynamic influencer

4.5 The impact of change on music

Regarding social, political or technological changes, we will mainly divide the 100-year data into two sections for exploration.

The first stage is the budding period of "experimental music" from the 1920s to the 1960s; but until the end of World War II, the wave of modernization began to sweep all aspects of the world. This kind of experimental music began to explode around 1945. Before 1945, only a few composers were conducting musical experiments; after 1945, the number of composers who continued to write in traditional styles became a minority, which is also in line with our analysis. After the year, a large number of pop, jazz, rb, country and other types of music emerged correspondingly.

The music of this era is not only reflected in the change of style, but also the technology of mass music generation and processing has begun to appear. Using modern techniques, Pierre Schaeffer (1910-1995) composed the first song in 1948 The specific music work "Train Etude" is composed of sound recordings such as wheel rolling, air jets, and siren. The advent of computers made music a step further than synthesizer music. Composers either use new musical instruments and new sounding methods, use sound-producing instruments, objects such as car parts, chains, iron sheets, etc.; or discover the new expressive power of traditional musical instruments. There are not only new music methods, but the original music methods have also been further explored. New performance and singing techniques, whether in string, wind, piano, percussion or vocals, have amazing discoveries like human voices. , Can imitate the sound of various musical instruments. These changes greatly increasing the emotional upper limit that music can directly transmit, and the composition of music is greatly enriched; leading to the energy value in the music trend we observe Constantly increasing, the value of acousticness and our combined comprehensive indicator—purity keeps declining. The neo-romanticism after this period is the product of the combination of modernism and romanticism. It is a new school that emerged in the 70s and 80s. It requires music to be tonality, based on traditional functional harmony, more emphasis on emotional expression, and often quotes musical materials from 19th century romantic composers. It is particularly noteworthy that a genre of "totalism" that has just emerged in the United States in the 1990s.

The composer's work must make the general audience feel good and impress their hearts, but still retain enough complex and difficult music skills to attract more interesting music listeners. The rise of this kind of genre that partly abandons intense emotional expression has led to a comprehensive monitoring of musical emotions, 'emotion', which has rapidly climbed to a peak around 1945, and has fallen by about 10% in the last 20 years.

5 Strengths and Weaknesses

5.1 Strengths

1. When dealing with the problem of artist influence evaluation, it is scientific to select evaluation indicators by creating a sub-network of the targeted influencer network. The topsis evaluation model based on the entropy method is more objective and applicability.

2. Based on the actual situation of the artist's relationship network and music characteristics, we selected some indicators that can describe the music similarity, and the indicators are forwarded to develop a set of descriptive music similarity measurement models.

3. Introduced the OpenOrd algorithm, which is a force-oriented layout algorithm, specially used to deal with very large graph structures. he model we built is simple and easy to understand, has good portability, and is suitable for most network analysis.

5.2 Weaknesses

Limited to the availability of the data (due to the limitation of limited data), we can only analyze the evolution of music, the evolution and revolutionary trends of artists and genres from a limited perspective, without considering the inherent originality or originality of artists. His personal experience has certain limitations.

6 Conclusion

Changes in music characteristics can examine evolutionary and revolutionary trends of music. This study is based on the influencers and followers of 5,854 artists , and the musical characteristics data set of 98,340 songs, build a directed influencer network, comprehensively analyze various characteristics of music, and explore the paradigm of the relationship and interaction between genres and musicians, Quantitatively analyze the evolution of music from the macro and micro perspectives by monitoring changes in various characteristics.

The advantages of this study lie in choosing models and indicators from a new perspective and in combination with the characteristics of musical evolution . The analysis level is clear, the results are stable, and the conclusions are practical. At the same time, there are still some unsolved problems in this paper, such as limitations in the process of selecting indicators, it requires in-depth research on the basis of expanding the number of indicators and sample size.

References

[1] https://www.rollingstone.com/music/music-features/the-motown-story-how-berry-gordy-jr-created-the-legendary-label-178066/.

[2] George S Davidson, Brian N Wylie, and Kevin W Boyack. Cluster stability and the use of noise in interpretation of clustering. In *infovis*, pages 23–30, 2001.

[3] Linton C Freeman, Douglas Roeder, and Robert R Mulholland. Centrality in social networks: ii. experimental results. *Social Networks*, 2(2):119–141, 1979.

[4] Peter M Todd D Gareth Loy. *Music and connectionism*. Mit Press, 1991.

[5] Shawn Martin, W Michael Brown, Richard Klavans, and Kevin W Boyack. Openord: an open-source toolbox for large graph layout. In *Visualization and Data Analysis 2011*, volume 7868, page 786806. International Society for Optics and Photonics, 2011.

[6] Hieu V Nguyen and Li Bai. Cosine similarity metric learning for face verification. In *Asian conference on computer vision*, pages 709–720. Springer, 2010.

[7] Patrick Schober, Christa Boer, and Lothar A Schwarte. Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5):1763–1768, 2018.