| Problem Chosen | 2021 MCM/ICM Summary Sheet | Team Control Number |
|---|---|---|
| D | | 2111259 |

# The study of the similarity of artists and their genres


**Summary**

Music evolves with social changes. To understand the role of music in the human collective experience, the evolutionary and revolutionary trends of artists and genres following social changes are examined.

For questions one and two, an **adjacency list** is constructed to describe the artist's music influence network. The relationship between influencers and followers is established **through the weighted method of directed paths**. Through the **recursive algorithm**, artist's "music influence" , that is, the "new generation talent influence" is obtained. This parameter reveals the artist's attraction to new talents and contribution to the development of music. To determine the specific value of the path weight, we use the **modified cosine similarity algorithm** to obtain the similarity relationship between every two artists, and **use this as the weighting parameter of the graph**. On this basis, it is concluded that artists of the same type are more similar than artists of different types.

For question three, the inter-genre and the inner-genre are analyzed separately. We found the average music characteristics of each genre and calculated the **similarity of the average characteristics of different genres** to get the difference of music characteristics between genres. At the same time, we calculate the **similarity of every two artists in the genre** to get the similarity ranking of the artists in the genre. Besides, we summed up the influence of music within the genre and found that the pop/rock genre has the greatest influence, and through the mapping of the music characteristics of each genre to the chronological analysis, we obtained the popular genre change table.

For question four, we compare the average similarity between every two artists with the average similarity of direct influence and conclude the influence of those affected by fan creation. Furthermore, we calculated the **Pearson correlation coefficients** of various music characteristics, and found that the energy and acoustics music characteristics of the influencer were the most "infectious".

For questions five and six, similarity calculations are performed year by year, and music characteristics are analyzed for years with low similarity. Then, look for musicians with great musical influence in that year, and analyze the similarity between the musicians and the musical characteristics of that year. Take 1967 as an example, the Beatles has the most influence and has a high similarity with the musical characteristics of the changing year. Music revolutionaries. In addition, we use the **entropy method** to calculate the change weights of music characteristics in a ten-year interval. The larger the weight, the greater the degree of dispersion of the characteristics. Finally, we found that the energy and loudness characteristics of the pop/rock genre continued to rise after the 1960s, indicating that the pop/rock genre has become more active.

For question seven, we combined the cited relevant literature to analyze the mutual influence between music and culture. For music influence culture, the method of searching high-impact artists has been used. As for music embodies cultural changes, the data obtained in question six is the key to the Pop/Rock example.



**Keywords**: cosine similarity, entropy, Pearson correlation coefficient

# Contents

# 1 Introduction

## 1.1 Problem Background

Music is an art form whose medium is sound and silence.[1] In the thousand-year history of mankind, music is an indispensable part of the development of civilization. When artists are creating music, many factors affect them, including their musical creativity, personal experience, current hot news, and social status.

Sometimes, music also undergoes revolutionary changes, creating new sounds or rhythms, introducing a new genre and trend, or changing the pattern of existing genres. By analyzing song networks and their musical characteristics in different periods, we can begin to capture the influence of music artists on each other, understand the similarities and differences between and within genres, so as to better understand how music has evolved with social and cultural changes of.

## 1.2 Problem Restatement

To examine artist genres and evolutionary trends, and build a model of musical influence. We have the following problems waiting to be solved:

- Establish a music influence network to capture the parameters of "music impression".
- Develop a music similarity measure.
- Compare the relationship between artists within genres and between genres, and get the relationship between genres.
- The role of music influencers on followers.
- Find a revolution on the network that may mark the evolution of music.
- Analyze the evolution of a genre over time and give an explanation.
- Use models to express the cultural influence of music in time or environment, and identify the influence of socio-political or technological changes.

# 2 Basic Assumptions

- It is assumed that artists in the network can indirectly influence other artists through the transmission of nodes.
- The similarity between artists has nothing to do with the number of their respective songs.
- It is assumed that the musical characteristics of each artist only contribute to the time when they became active.
- Suppose that if a certain genre has the most similar musical characteristics with that of the decade, the genre is the most popular in that decade.

# 3 Notations

The primary notations used in this paper are listed in Table 1.

Table 1: Notations

| Symbol | Definition |
|---|---|
| $\alpha$ | Attenuation factor |
| $\vec{a}, \vec{b}$ | A multidimensional vector |
| $\theta$ | The angle between two vectors |
| $R_{ui}$ | Represents the $u$-th dimension data of $i$ of the vector |
| $x_i$ (i=1,2...n) | The $i$-th component of the $x$ data set |
| $y_i$ (i=1,2...n) | Normalized result of $a$ |
| $P_i$ | Music influence parameter of the artist corresponding to the node $i$ |
| $P_{i_j}$ | Music influence parameter of the artist corresponding to the $j$-th directly connected child node of the node $i$ |
| $O_{ij}$ | Similarity of music characteristics between the artists corresponding to the node $i$ and the node $j$ |
| $k$ | The number of child nodes directly connected to the $i$-th node |
| $U$, $V$ | Data sets, and the elements between them have a one-to-one correspondence |
| $N$ | Total number of corresponding element pairs in $U$ and $V$ |
| $X_i$ | The $j$-th element in the $i$-th data set |
| $Y_i$ | The normalized data set |
| $Y_{ij}$ | The $j$-th element in the $i$-th data set |
| $p_{ij}$ | The normalized result of $Y_{ij}$ |
| $E$ | The information entropy |
| $W$ | The weight of each indicator |

# 4 Model developments

## 4.1 Data Processing

Since the evaluation indicators in the *data_by_artist* data set, *data_by_year* data set and *full_music_data* data set are different in nature, dimension and magnitude, we have standardized the data. After comparing the tables, we found that the musician with ID 477787 did not have detailed data, so it was deleted.

## 4.2 Music Influence Network and Similarity

### 4.2.1 Music influences the network

Influencers and followers are regarded as nodes in the network, and the directed links between nodes represent the influence relationship between nodes. The relationship is quantified as the link weight to form a weighted directed graph as a whole. Figure 1 shows the network structure diagram.

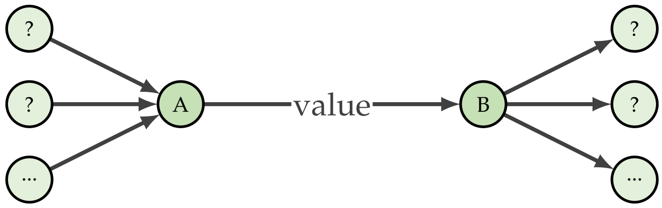
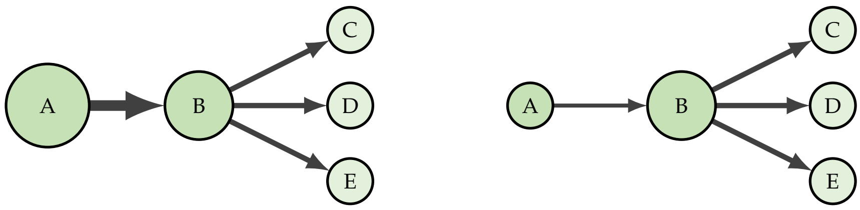



Figure 1: Weighted directed graph Model

On this basis, in order to explain the connection weight more specifically, let's take the relationship in the following Figure 2 as an example. Node A is the direct influencer of B, B is the direct influencer of CDE, and A is the indirect influencer of CDE. Next, we will analyze the musical characteristics of A and B. If the musical characteristics of A and B are similar, indicating that B is greatly affected by A, the indirect influence of A on CDE is greater as shown in Figure 2a. If the result is not similar, indicating that B is less affected by A, then A has a less indirect impact on CDE, as shown in Figure 2b.




(a) Strong influence (b) Weak influence

Figure 2: Example of different influence

In addition, the indirect impact should be smaller than the actual impact, so an attenuation factor $\alpha$ is introduced, and the final value is 0.2 through data comparison. Finally, the similarity between nodes are multiplied by the attenuation factor as the link weight, as shown in the Page 5 Figure 3.

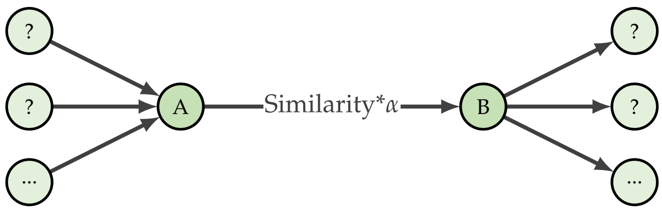



Figure 3: Weighted analysis diagram

#### 4.2.2 Artists similarity measure

From the *data_by_artist* data set, the musical characteristics of each artist can be regarded as a multi-dimensional vector. The cosine similarity algorithm can evaluate the similarity of two vectors by calculating the cosine of the angle between them. Therefore, the cosine similarity algorithm can be used to calculate the similarity of the musical characteristics of two artists. The calculation formula is as Formula 1.

$$cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \cdot \left\|\vec{b}\right\|} \tag{1}$$

$\vec{a}$ and $\vec{b}$ are two multidimensional vectors. $\theta$ is the angle between the two vectors.

Further, to solve the problem that the vector angle is too small and the vector length is very different, but the difference in the cosine value is small and the two vectors are judged to be similar, the modified cosine similarity algorithm is introduced, and the formula is as Formula 2.

$$sim\,(i,j) = \frac{\sum\left(R_{ui}-\bar{R_u}\right)\left(R_{uj}-\bar{R_u}\right)}{\sqrt{\sum\left(R_{ui}-\bar{R_u}\right)^2}\sqrt{\sum\left(R_{uj}-\bar{R_u}\right)^2}} \tag{2}$$

$i$ and $j$ are vector numbers. $R$ is a vector collection of multiple vectors. $R_{ui}$ represents the $u$-th dimension data of $i$ of the vector. $\bar{R_u}$ indicates the average value of the $u$-th dimension of all data.

Finally, since the value range of the cosine value is [-1, 1], we normalize the data obtained by the modified cosine similarity algorithm to obtain the final similarity measure between artists. The standardized formula is as Formula 3.

$$y_i = \frac{x_i - min\{x_j\}}{max\{x_j\} - min\{x_j\}} \tag{3}$$

$y_i$ is the normalized result of $x_i$. $x_i$ is the $i$-th component of the $x$ data set.

#### 4.2.3 Network model supplement

Based on the model above, there is no doubt that artists can influence their predecessors, but this part of the influence has a limited impetus to the development of music. To make the "music influence" parameter more realistic, that is, from the perspective of development, music development needs to be injected with fresh blood. Therefore, we define "music influence" as the "new generation talent influence", that is, the calculation of artist "music influence" includes only peer artists and younger artists.

In addition, due to the influence of music, the network contains loops. Based on our assumption, in the calculation of the influence of a single artist, there will be no situation that affects itself, so the loop is broken in the calculation of the single influence, which is shown in Figure 4.

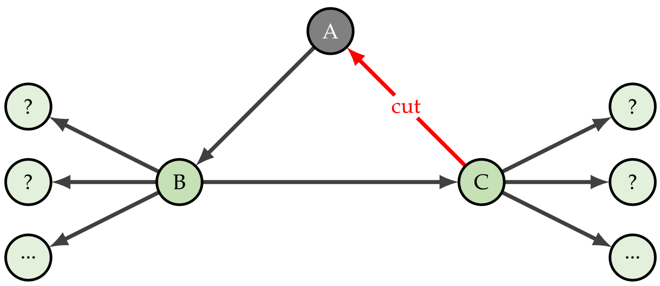



Figure 4: Weighted directed graph Model

#### 4.2.4 Music influence

Combining the above content, we can finally get the recursive relationship of influence as Formula 4.

$$P_i = k + \alpha \sum_{j=1}^{k} P_{i_j} O_{ii_j} \tag{4}$$

$P_i$ is the music influence parameter of the artist corresponding to the $i$-th node. $k$ is the number of child nodes directly connected to the $i$-th node. $\alpha$ is the attenuation factor. $P_{i_j}$ is the music influence parameter of the corresponding artist of the $j$-th directly connected child node of the $i$-th node. $O_{i_j}$ is the similarity of music characteristics between the artist corresponding to the $i$-th node and the artist corresponding to the $j$-th directly connected child node.

In the network we have established, "music influence" which is the same thing as "new generation talent influence" can be used to reveal which music masters in the genre and the proportion of music masters in each genre. Thus, we select the top five "music influence" artists of the genre as masters in the following Page 7 Table 2, and analyze the top 500 music influence artists to draw the image as shown in Page 7 Figure 5.

Table 2: Top 5 Influencers

| name | influence | genre | year |
|---|---|---|---|
| The Beatles | 790.9253 | Pop/Rock | 1960 |
| Bob Dylan | 523.7476 | Pop/Rock | 1960 |
| Hank Williams | 451.7476 | Country | 1930 |
| The Rolling Stones | 434.6987 | Pop/Rock | 1960 |
| David Bowie | 416.7795 | Pop/Rock | 1960 |

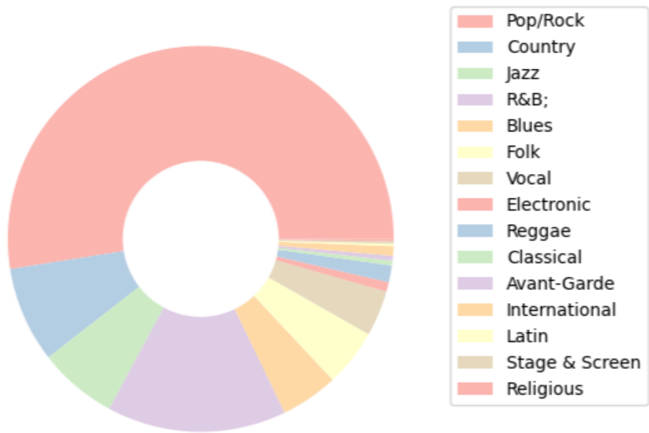



Figure 5: Top 500 Influencers

#### 4.2.5 Artists similarity comparison

Let us take the pop/rock genre as an example. We first calculate the modified cosine similarity for every two artists in the genre, and get the average similarity in the pop/rock genre by summing and averaging. Then we calculate the similarity of any two artists in the network, and average the sum.

$$\bar{O} = \frac{\sum_i^n \sum_j^n O_{ij}}{\frac{1}{2}n(n-1)} \, (1 \le i < j \le n) \tag{5}$$

$\bar{O}$ is the average similarity of artists. $O_{ij}$ is the similarity of the music characteristics of the artist corresponding to node $i$ and the artist corresponding to node $j$. $n$ is the number of all nodes.

Finally, we get the average similarity of artists in the pop/rock genre and the overall average similarity of artists as Page 8 Figure 6.

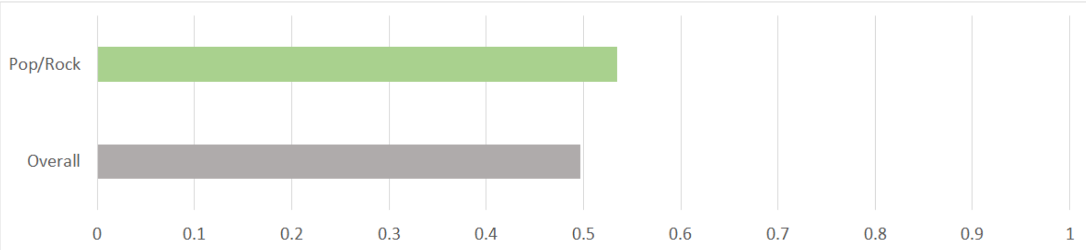



Figure 6: Similarity Comparision (Pop/Rock)

By analyzing the figure, we can see that the average artist similarity of the pop/rock genre is higher than the overall average similarity, that is, the artists within the pop/rock genre are more similar.

## 4.3 Genre Analysis

In the analysis of different genres, artists who do not know the genre will not participate in the genre analysis, that is, the genre "*unknown*" will not participate in the analysis.

### 4.3.1 Inter-genre analysis

We separately sum the music characteristics of all artists in each genre that have been preprocessed and get the average music characteristics of each genre. Perform modified cosine similarity calculation for different genres, and finally perform normalization. The following Table 3 shows part of our result.

Table 3: Table of Similarity between genres

| | Pop/Rock | Country | Jazz | R&B; | Blues | Folk | ... |
|---|---|---|---|---|---|---|---|
| Pop/Rock | NULL | 0.383519 | 0.066974 | 0.390405 | 0.074175 | 0.081382 | ... |
| Country | NULL | NULL | 0.407583 | 0.618256 | 0.81085 | 0.757632 | ... |
| Jazz | NULL | NULL | NULL | 0.359063 | 0.793443 | 0.831264 | ... |
| R&B; | NULL | NULL | NULL | NULL | 0.577393 | 0.395399 | ... |
| Blues | NULL | NULL | NULL | NULL | NULL | 0.926958 | ... |
| Folk | NULL | NULL | NULL | NULL | NULL | NULL | ... |
| ... | ... | ... | ... | ... | ... | ... | ... |

We analyze the music characteristics of the two genres with the lowest similarity, namely Pop/Rock-International, and it is shown that almost all the characteristics of these two genres in Table 4 are distributed on the upper and lower sides of the average line respectively, which can be considered as opposite. It is also shown that the similarity between Avant-Garde and New Age is "1"(relative value), so it can be considered that these two genres are related to a certain extent.

Table 4: The average characteristics of Pop/Rock and International

| | danceability | energy | valence | tempo | loudness |
|---|---|---|---|---|---|
| Pop/Rock | -0.22646 | 0.433394 | -0.0851 | 0.203152 | 0.31915 |
| International | 0.092088 | -0.55538 | 0.274276 | -0.27295 | -0.52357 |
| | mode | key | acousticness | instrumentalness | liveness |
| Pop/Rock | 0.105595 | 0.015097 | -0.44957 | -0.11148 | 0.04429 |
| International | -0.0492 | -0.04414 | 0.843321 | 0.04737 | -0.09579 |
| | speechiness | duration_ms | popularity | | |
| Pop/Rock | -0.1031 | -0.08031 | 0.252042 | | |
| International | 0.063525 | 0.364976 | -0.38035 | | |

Next, we use the music influence network to calculate the "music influence" of each artist and sum up the musical influence of artists in each genre to obtain the musical influence of the genre.

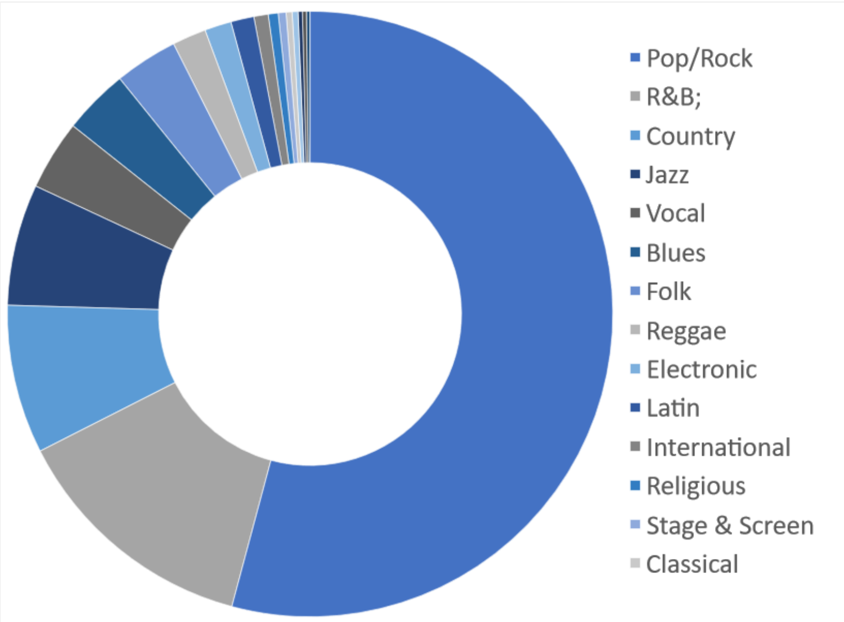



Figure 7: Genre Music Influence

From the Figure 7, it can be concluded that pop/rock, R&B and country are the three genres with the highest "music influence", among which the "music influence" of the pop/rock genre far exceeds other genres, indicating that the pop/rock genre attracts the newest talents and the most popular and developing best.

### 4.3.2 Genre internal analysis

We calculate the similarity of every two artists within each genre and take the average to get the average internal similarity of each genre, as shown in the following Figure 8 in Page 10.

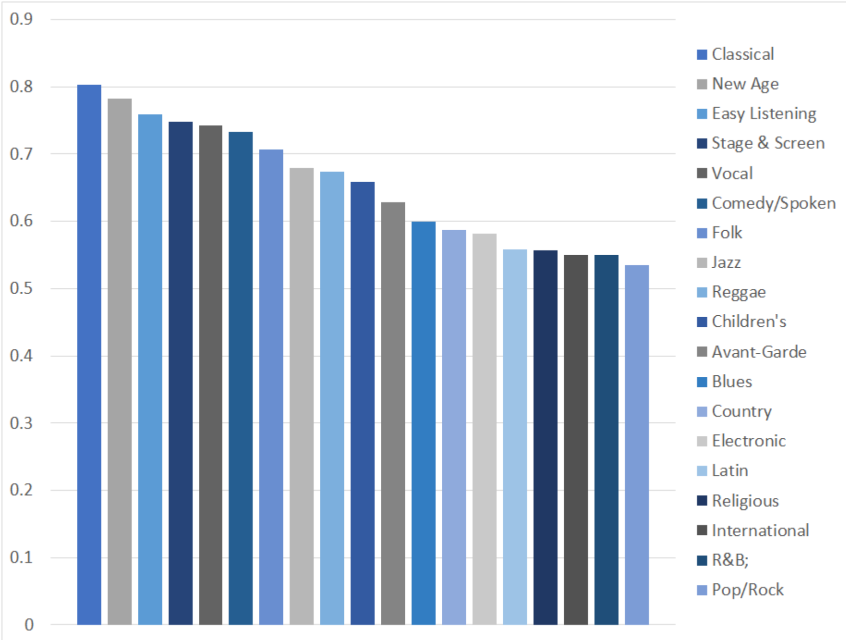



Figure 8: Internal similarity of genres

From the Figure 8, it can be known that Classical, New Age and Easy Listening are the three most similar genres, indicating that the music characteristics of different artists within these genres are similar. Similarly, pop/rock, R&B and international are the three genres with the lowest similarity, which indicates that different artists within these genres have diverse styles and a high degree of discreetness in music characteristics.

#### 4.3.3 Popular genres change over time

According to the hypothesis, through the data by artist data set, we can get the time when every artist started to be active, and use this to determine that the musical characteristics of every artist only contributed to their respective decade. Average the musical characteristics of artists in each genre that have been active in each decade (10 years), and obtain the average musical characteristics of each genre per decade. Next, we obtain the overall average music characteristics of all genres per decade in the data by year data set. By analyzing the similarity between the average music characteristics of each genre and the average music characteristics of the decade, we can find the genre with the highest similarity in every decade. According to the hypothesis, it can be considered that this genre is the most popular in that decade. Popular genres change over time as shown in the Table 5.

Table 5: Popular genres by year

| Rank | 1930 | 1940 | 1950 | 1960 | 1970 |
|---|---|---|---|---|---|
| 1 | Jazz | International | Blues | Folk | R&B; |
| 2 | Vocal | Vocal | Folk | Country | International |
| 3 | Folk | Jazz | Country | International | Blues |
| Rank | 1980 | 1990 | 2000 | 2010 | |
| 1 | Pop/Rock | Pop/Rock | Pop/Rock | Pop/Rock | |
| 2 | Electronic | Religious | Country | Electronic | |
| 3 | Unknown | Country | R&B; | R&B; | |

It can be seen form the table that from 1930 to 1979, Jazz, International, Blues, Folk, Vocal and Country all took turns occupying the top three influential lists. But after 1980, Pop/Rock ranked first in the influence list by absolute advantage.

## 4.4 Influencer's Music Influence Analysis

In Figure 6 we have obtained the overall average similarity. Based on this, we found that the average similarity between artists with mutual influence in the network, as shown in the following Figure 9.

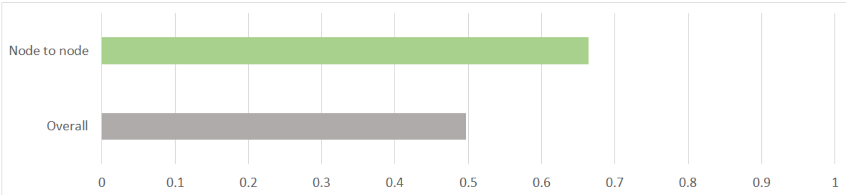



Figure 9: Similarity between artists with mutual influence

From the greater average similarity of directly influencing artists, it can be known that "influencers" will affect the music created by followers, and followers will learn the music styles of influencers, and they have more similarities.

To further determine the music characteristics affected by followers, the Pearson correlation coefficient algorithm is used to determine the relationship between the influencers' and the followers' music characteristics, which is determined by the following Formula 6.

$$\rho_{U,V} = \frac{\sum UV - \frac{\sum U \sum V}{N}}{\sqrt{\left(\sum U^2 - \frac{(\sum U)^2}{N}\right)\left(\sum V^2 - \frac{(\sum V)^2}{N}\right)}} \tag{6}$$

$U$ and $V$ represent data sets, and the elements between them have a one-to-one correspondence. $N$ is the total number of corresponding element pairs in $U$ and $V$. If $U$ and $V$ are highly correlated, the Pearson correlation coefficient tends to 1, otherwise, it tends to 0.

In the network, we take the 13 music characteristics of the influencer as $U$ and the corresponding music characteristic of the affected person as $V$, and obtain the two-dimensional data $(U, V)$ of 13 sets of music characteristics, and then calculate the data of each group of music characteristics. Pearson correlation coefficient, the Pearson correlation coefficient of 13 music characteristics is obtained, and the histogram is drawn as Figure 10.

Analyzing the histogram in Figure 10, we can find that the Pearson coefficients of these two musical characteristics are very prominent, which means that they are highly transmissible in the influence network.

## 4.5 Music Revolution Analysis

First, we perform year-by-year similarity calculation on the music characteristics of the data by year data set, and calculate the music similarity between each adjacent two

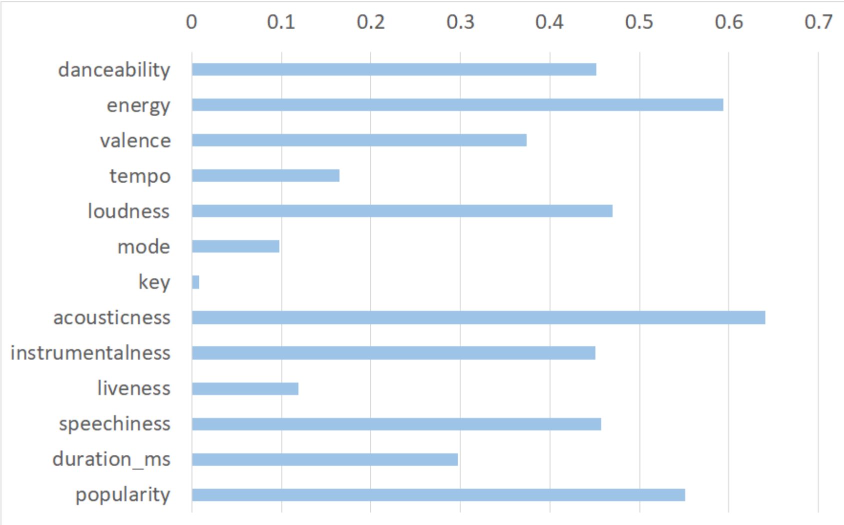



Figure 10: Pearson's correlation coefficient histogram

years, as shown in the following Figure 11.

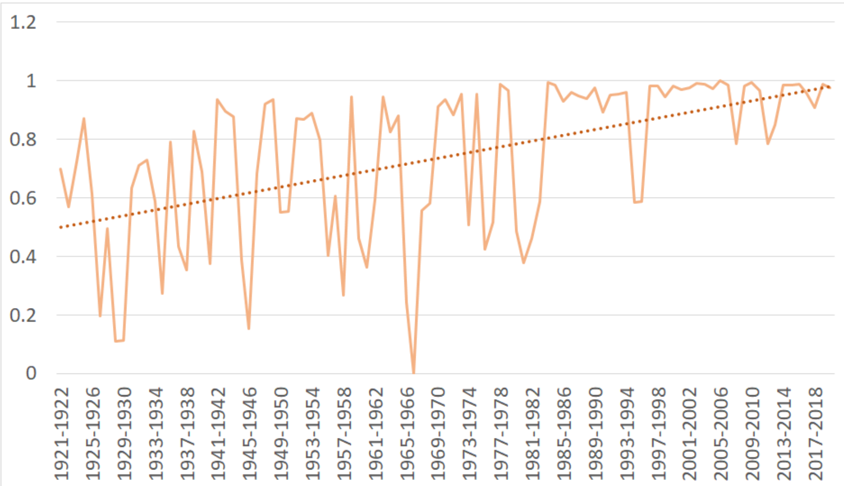



Figure 11: Yearly Music Similarity

As can be seen from the figure, the similarity of music characteristics of 1929, 1945, 1967 and 1981 with those of the previous year is the lowest, indicating that some of these years have great changes in music characteristics and may have undergone musical changes. Next, we look for artists whose music has a strong influence on the network and whose music features are highly similar to those of the years of change. Then the artists that are eventually found can be considered as musical revolutionaries.

Taking 1967 as an example, we found that The Beatles, The Rolling Stones and David Bowie had a great influence, and their similarity with that year's music features reached 0.9. Therefore, it is believed that they are revolutionaries of the 1967 music revolution.

## 4.6 Pop/Rock Genre Analysis

The entropy weight method(EWM) can use the entropy value to judge the degree of dispersion of an index. The smaller the information entropy value, the greater the degree of dispersion of the index. By calculating the entropy value of each music feature, the dynamic influence index can be obtained.

If there are $l$ $X_i$ data sets, $X_{ij}$ ($i < l$, $j < n$)represents the $j$-th element in the $i$-th data set. We use formula 3 to normalize $X_i$ to get the $l$ normalized data set $Y_i$.

$$X_i = \{x_1, x_2, ..., x_n\} \tag{7}$$

We define standardized treatment for $Y_{ij}$ as in formula 8. $p_{ij}$ is the normalized result of $Y_{ij}$.

$$p_{ij} = \frac{Yij}{\sum_{i=1}^{l} Y_{ij}} \tag{8}$$

If $p_{ij}$=0, then $\lim_{p_{ij}\to 0} p_{ij} ln p_{ij} = 0$.

For the information entropy E of a set of data, we can obtain the following Formula 9.

$$E_j = -\frac{1}{lnl}\sum_{i=1}^{l} p_{ij} ln p_{ij} \tag{9}$$

Finally, we can calculate the weight of each indicator through information entropy, as shown in Formula 10. If the weight of the index is large, it means that the index has a high entropy weight, a high degree of dispersion, and is unstable. If the weight is small, the entropy weight is small and the index is relatively stable.

$$W_i = \frac{1 - E_i}{l - \sum E_i} \, (i = 1, 2, ..., l) \tag{10}$$

We calculated the average music characteristics of pop/rock each year in the full music data data set, using 10 years as an interval, and using the entropy method to calculate the change weights of music characteristics within ten years. The greater the weight of the music characteristics, the greater the value of the music characteristics. The greater the degree of dispersion within ten years, the greater the dynamic impact. Keep the first three entropy values of every ten years, the output table is as follows.

Table 6: Top three characteristics of entropy method

| | No.1 | No.2 | No.3 |
|---|---|---|---|
| 1940 | energy | valence | tempo |
| 1950 | speechiness | instrumentalness | energy |
| 1960 | duration_ms | danceability | valence |
| 1970 | valence | loudness | liveness |
| 1980 | liveness | loudness | tempo |
| 1990 | instrumentalness | speechiness | loudness |
| 2000 | valence | loudness | liveness |
| 2010 | liveness | duration_ms | energy |

It can be seen from Table 6 in Page 13 that the valence, liveness and loudness music characteristics of the pop/rock genre have been in the region of high entropy weight for the past 70 years, that is, dynamic music characteristics with greater volatility. These musical characteristics are indicators that reflect the dynamic changes of music. Standardize the four music characteristics and draw the image as shown in Page

14 Figure 12. It can be seen that in the 1940s and 1950s, the pop/rock genre was still in the groping stage, and the genre music characteristics changed greatly. After the 1960s, the energy and loudness music characteristics were taken as examples. They showed an overall upward trend over time. The loudness and activity of pop/rock is getting bigger and bigger.

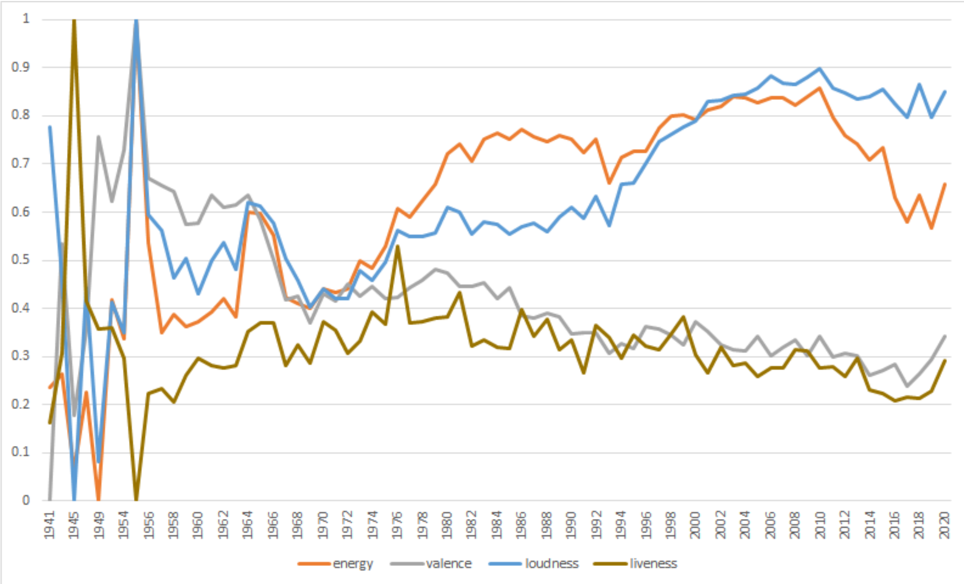



Figure 12: Pop/Rock Music characteristics change over time

## 4.7 The relationship between music and culture

To explore the relationship between music and culture, it is natural to discuss the influence of music on culture and the reflection of music on social changes. The following details the main ideas of using the influence network model constructed by our group to analyze the above two relationships.

### 4.7.1 The impact of music on culture

To get the main influence of music on culture in a particular era, we need to explore the corresponding era subnet of the influence network. First, search for several artists with high influence (five artists for example) and list them in the form of a table. Analyzing their genres and musical styles, and combining historical events, we can know where the main influence of music on culture in this era is reflected.

### 4.7.2 Music reflects cultural changes

The music itself is constantly changing and evolving. This process is slow and stable. But in some cases, major external events can make music a great leap and improve. To illustrate the idea of identifying external influences, we might as well take the Pop/Rock genre as an example for analysis, and the same is true for the analysis of external influences that cause changes in all music genres.

In order to get the music changes that occurred in a certain period, we use the key music characteristics obtained through the entropy method in 4.6 to analyze. The size and change trend of these selected music characteristics are comprehensively analyzed to finally identify external influences. Take 2000s for example, the top three dramatic musical characteristics of the changes are valence, loudness and liveness. The trends of these three musical characteristics of Pop/Rock in the 2000s can be found in the Figure 13 in Page 16.

Among them, the trend of the loudness curve can be clearly found. After the 1980s, the popularization of digital technology turned into a wave of economic growth in the 21st century. The spread of commercial music has continued to increase, and the people's economic purchasing power in some countries has also grown steadily.[2] In addition, studies have shown that under the multiple influence of listener preferences and human instincts, louder music can achieve better sales in the world music market. This phenomenon is also called "loudness war".[3] Combining academic analysis articles and historical facts of other music, the sudden increase in loudness reflects the dual impact of the advancement of digital technology and economic globalization.

# 5 Model Evaluation

## 5.1 Model stability analysis

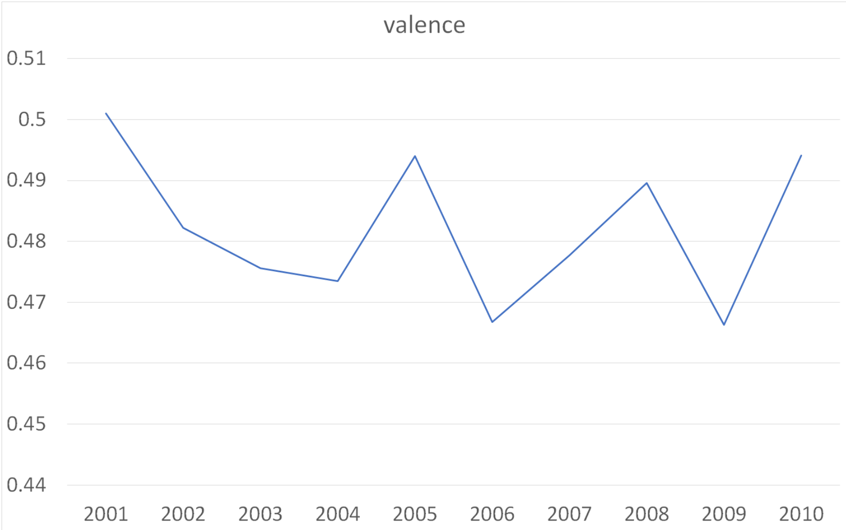
We conduct stability analysis on the main model established, namely the music influence model. It can be seen that we initially set the attenuation factor $\alpha$ to 0.2. In order to explore the impact of changing the attenuation factor on the model results, output the top 20 "music influence" artists when the attenuation factor is 0.2, and then adjust the attenuation factor to 0.1 and output the first two Ten "music influence" artists. The data are shown in Tables 7 and 8 below.

Table 7: Top 20 Influencers by $\alpha = 0.1$

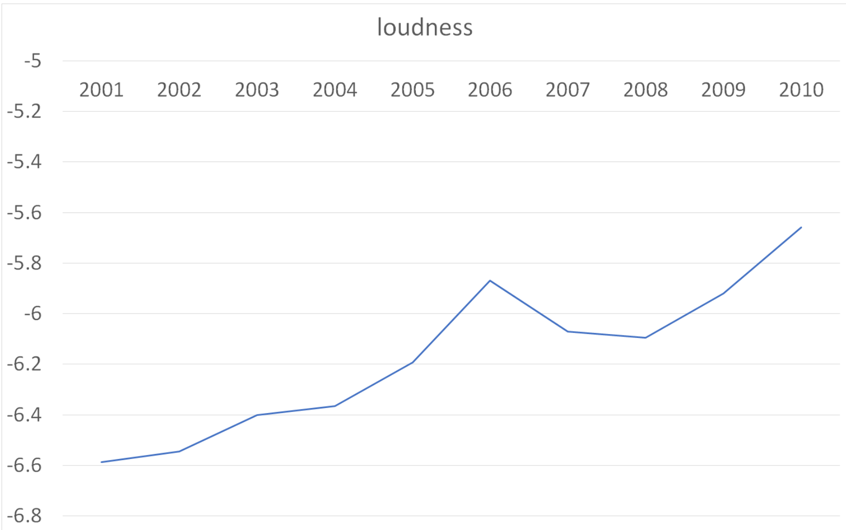
| name | influence |
|---|---|
| The Beatles | 693.22 |
| Bob Dylan | 445.13 |
| The Rolling Stones | 371.14 |
| David Bowie | 316.27 |
| Hank Williams | 297.41 |
| Led Zeppelin | 279.41 |
| The Kinks | 256.32 |
| Chuck Berry | 240.96 |
| Jimi Hendrix | 237.81 |
| The Velvet Underground | 236.58 |
| Elvis Presley | 233.57 |
| Sex Pistols | 224.98 |
| Black Sabbath | 215.77 |
| Pink Floyd | 210.99 |
| The Beach Boys | 209.90 |
| Miles Davis | 209.70 |
| The Clash | 204.51 |
| Marvin Gaye | 200.16 |
| James Brown | 198.91 |
| The Who | 194.02 |

Table 8: Top 20 Influencers by $\alpha = 0.2$

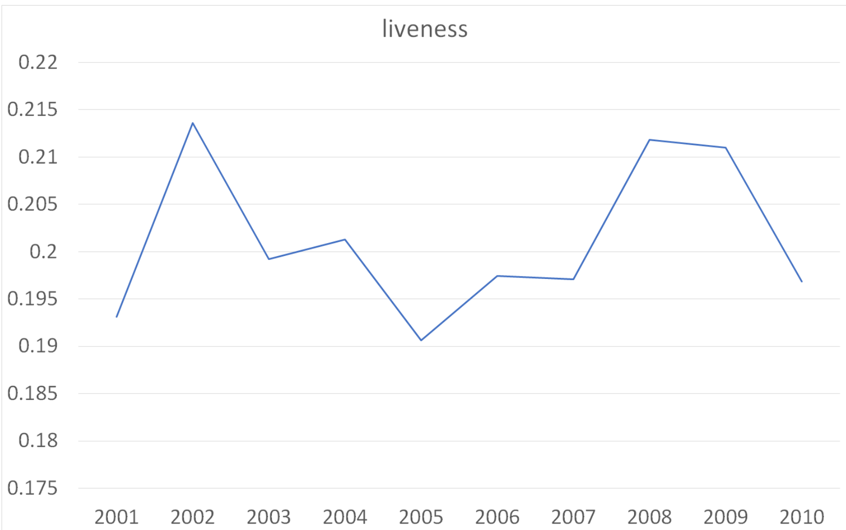
| name | influence |
|---|---|
| The Beatles | 790.92 |
| Bob Dylan | 523.74 |
| Hank Williams | 451.75 |
| The Rolling Stones | 434.70 |
| David Bowie | 416.78 |
| Chuck Berry | 359.19 |
| Led Zeppelin | 355.93 |
| The Kinks | 341.83 |
| Sex Pistols | 321.38 |
| Elvis Presley | 316.79 |
| The Velvet Underground | 308.63 |
| The Band | 302.49 |
| The Clash | 295.56 |
| Pink Floyd | 294.44 |
| Jimi Hendrix | 288.19 |
| Miles Davis | 277.25 |
| Woody Guthrie | 276.61 |
| Johnny Cash | 276.44 |
| Black Sabbath | 274.13 |
| Ray Charles | 260.22 |




(a) valence




(b) loudness




(c) liveness

Figure 13: Three main musical characteristics of Pop/Rock

As can be seen in Table 9, the top five artists whose music influences only changed their rankings. And after changing the attenuation factor, only two artists dropped out of the top ten and changed. For music impact data, let's take the Beatles as an example. The attenuation factor dropped from 0.2 to 0.1, and the music impact only dropped by 100, which is about 1/8. From this we can see that the introduction of the attenuation factor makes the indirect influence have a certain effect, but the effect is not too large, indicating that the model is very stable and correct.

### 5.2 Model correctness assessment

The Beatles were an English rock band formed in Liverpool in 1960. The group are regarded as the most influential band of all time. They were integral to the development of 1960s counterculture and popular music's recognition as an art form.[4]

Table 9: Top 5 genre of influence in 1990s

| name | Influence | similarity |
|---|---|---|
| The Beatles | 693.2233 | 0.925182 |
| Bob Dylan | 445.134 | 0.624072 |
| The Rolling Stones | 371.1467 | 0.887795 |
| David Bowie | 316.2759 | 0.915778 |
| Led Zeppelin | 279.4093 | 0.553645 |

According to the data in Table 9, the Beatles were considered by our model to be the greatest musical revolutionist in 1967, with a maximum "musical influence" of 693 and a similarity of 0.92 to the musical characteristics of the year.

## 6 Strengths and Weaknesses

### 6.1 Strengths

- The weighted directed graph is used to clearly construct the music influence network, and the indirect influence factors are considered in the "music influence" parameter, and the influence result is highly accurate compared with historical facts.
- In the algorithm, we used the cosine similarity algorithm, Pearson correlation coefficient, and entropy weight method respectively according to the specific needs to enrich the structure of the model. Based on these algorithms, we created most of the comparative evaluation models.
- In data processing, to reduce the algorithm error, we use the modified cosine similarity algorithm. In the data processing, we only use the standardization process to reduce the dimensional error while keeping the original data as much as possible and remove the number of songs in the hypothesis. Other data Both are used for the establishment of the model, and the use of data is maximized.

## 6.2 Weaknesses

- To pursue data accuracy, we did not use principal component analysis to reduce the dimensionality of multi-dimensional data and discard the data reasonably, resulting in a long program running time.
- In the music change model, we did not use the song release date in the $full\_music\_data$ data set as a sign for judging the music change artist. Instead, we find the music change artist in the active year of the artist in ten years. The data accuracy needs to be improved.

# Document for ICM

**To:** ICM
**From:** Team 2111259
**Date:** February 9th, 2021
**Subject:** A brief introduction of the musical influence model

Before starting to introduce you to our model, please allow us to express our sincere admiration for your association's efforts in the field of integrated music.

In our music influence network model, as long as we obtain the influence relationship data set between artists and the artist's music characteristic data set, we can construct the music influence with the comprehensive influence and influence relationship weight with the artist as the node.

On the music influence network, in addition to being able to directly obtain the influence parameters of each artist, it will be effortless to retrieve music influence relations. Research on its subnets can be carried out smoothly with only part of its data.

According to the obtained music influence network and the three data sets obtained, not only can our model derive the ranking of music converters of the era, but it can analyze the reflection of music on society after reducing the dimensions of the data by calculating music characteristics through the entropy method.

It is worth noting that although our model has relatively high efficiency in the analysis of 13 music features, as the number of music features increases, our model will inevitably have a sharp increase in computing time. In this case, in addition to switching to a computer with higher computing power, a feasible method is to use principal component analysis to reduce the dimensions of all these music features.

If data sets of relationships between music artists and figures in other fields can be provided to us, a special cross-domain subnet will be established. With this subnet, it will be easier to analyze the impact of music on culture, rather than just analyzing the changes in music characteristics. For example, politicians can be associated with government policies, and teachers can be associated with educational tendencies.

In addition, due to the limitations of the actual conditions in this competition, our team has simplified the calculation process of some problems to varying degrees. Once given sufficient time and computer resources, the results obtained by our model will retain more complete information of the original data set and obtain more accurate results.

All our team members are looking forward to your evaluation of our model, which will be an important piece of information for us to improve this model.

# References

[1] "Music - Wikipedia, the free encyclopedia." [Online]. Available: https://en.wikipedia.beta.wmflabs.org/wiki/Music

[2] D. Held, *A Globalizing World?: Culture, Economics, Politics*. Routledge, Aug. 2004, google-Books-ID: i7u0lbvJoaIC.

[3] E. Vickers, "The Loudness War: Do Louder, Hypercompressed Recordings Sell Better?" *Journal of the Audio Engineering Society*, vol. 59, no. 5, pp. 346–351, Jun. 2011, publisher: Audio Engineering Society. [Online]. Available: https://www.aes.org/e-lib/browse.cfm?elib=15934

[4] "The Beatles - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/The_Beatles

# Appendix: Other Figures

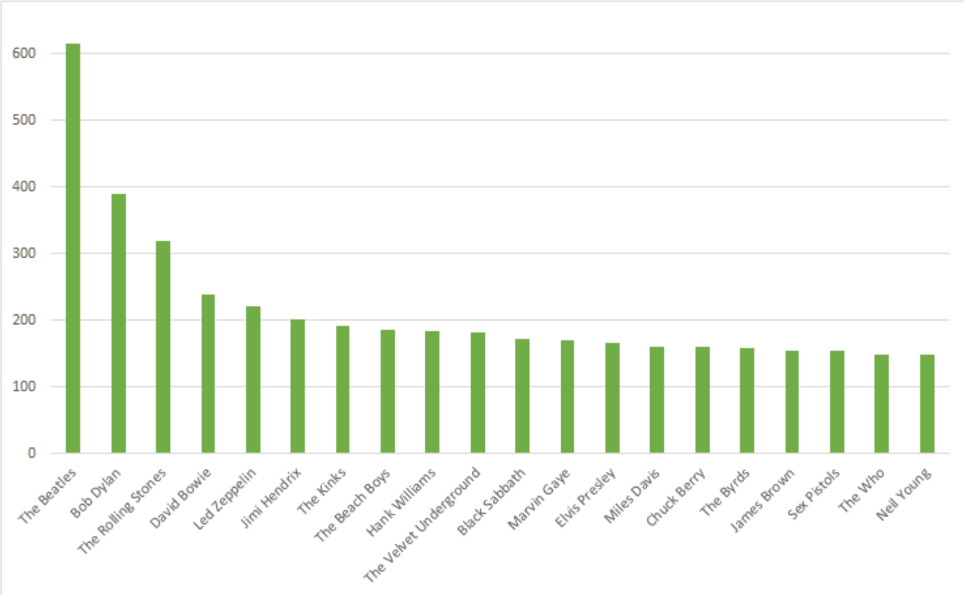



Figure 14: Number of artists who have direct influence