Problem Chosen	2021	Team Control Number
D	**MCM/ICM Summary Sheet**	**2101569**

How music evolutes: Analysis On Index Evaluation of Music Influence

— Based on node centrality

Summary

Nowadays, music has become an incomprehensible part of cultural heritage and modern people's daily life. Music has many diffrerent components, these factors influence the composer deeply. The aim of this report is to build a model to measure the influence of different artists and songs in all kinds of genres, and identify the prominent music character in determining the direction of music development. Five models were built.

Model 1 is established for measuring the similarity between two pieces of music. First, we clean up the dataset. Then, PCA method is used to reduce the dimension, and eight independent variables and their respective data are obtained. According to the mutual independence of data, we can measure the degree of similarity by calculating Mahalanobis distance, and finally form a formula to measure music similarity by constructing modification function.

Model 2, is established for measuring the music influence of influencers. Inspired by the centrality of feature vector points in social network analysis (SNA) and the principle of three degrees of influence in information communication, we construct an index MI to measure music influence. At the same time, due to the basic characteristics of influence weakening with time, it is similar to the half-life in physics, so the attenuation correction coefficient is added.

Model 3 is established for comparing the influence and similarity between genres and within genre. First of all, through the analysis of variance in significance, we analyze the overall data, and come to the conclusion that there is a big gap between the similarity within genre and between genres. The hierarchical clustering is used to get the closely related genres and the different genres, and then the specific similar music features are measured by geometric figures.

Model 4, this model measures the extent of follower being affected. Through the method of gray correlation analysis and calculation of correlation coefficient, using Nightingale rose chart for visualization, we can draw the conclusion that the similarity within the genre is high, and the influencer has a great influence on its follower (the music similarity between the two is high).

Model 5 is the time sequence analysis of the whole and different genres of music. On the basis of the previous model, we first find the key change time through time series analysis, and then construct the RR index to evaluate the importance of network nodes on the influence model.

In addition, we find a real data set to verify the results of the model. The coincidence degree of our model's high impact musicians and professional data set is more than 80%. At the same time, the social comprehensive factors of the key years are analyzed. At the same time, we conclude that the change of music influence is closely related to social and historical changes and economy.

Keywords: PCA ANOVA SNA Gray analysis

Contents

1 Introduction

1.1 Problem Background

Nowadays, music has become an incomprehensible part of cultural heritage and modern people's daily life. Music has many diffrerent components, such as componers' innate ingenuity, current social or political events and etc. These factors influence the composer deeply, and affect their artistic work greatly in the future. In order to understand the evolution of music, we need to understand and measure the influence of previously produced music on new music and musical artists.

1.2 Restatement of the Problem

- Use influence data to build a directed influencer network.Explore the main parameters for 'music influence'. Developing a mathematical model for measuring music similarity.
- Compare the similarities between genres and within genre. Explain the music trend thet changes over time. Use the similarity network model to describe the influencer's ability.Identify the most contagious factor in music composition.
- Identify the characters that signify revolutions in music history. And the music artists that lead revolutions. Reveal the indicators that shows dynamic influencers.
- Explain systemetically how do our mathematical model embodies cultural factors such as social, political and technological changes.

1.3 Our Efforts

To achieve this final goal, we need to quantify and clearify the artist's **influence ability**. Our work are as follows.

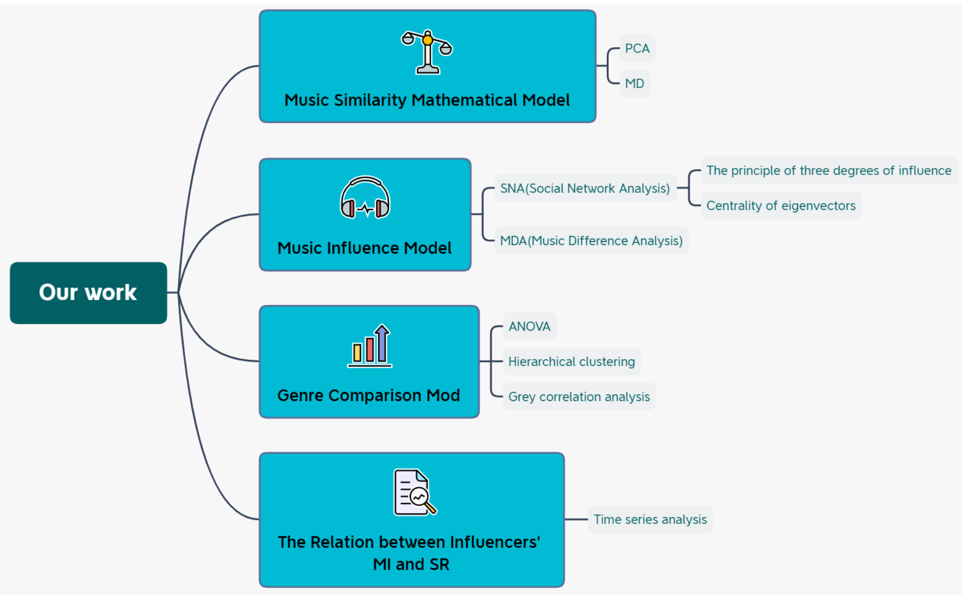


Figure 1: Visualization of process

1. Based on *influence_data* data set, we build an estimation model of *music influence*.

2. Using *full_music_data* data set to build a model for measuring music similarity explicitly.
3. Making use of the model above to analyze the music trend for concrete situations. This artical shows the critical music revolution and the most contagious characteristic of music.Identify the dynamic influencer of that era.

2 Assumptions and Justications

To simplify the given problems and modify it more appropriate for simulating real changing trend of music, we make the following basic hypotheses, each of which is properly justified.

- **Assumption 1:** We assume that the similarity between different musics can be measured by 14 representitive characters of theirs.
  $\hookrightarrow$ **Justification:** Based on the data that we have found and the professional Papers in the field of music, we found that these characters can be assumed as the distinctive indicators for describing songs.
- **Assumption 2:** We assume that one music artist can only have one main style of music over his lifetime.
  $\hookrightarrow$ **Justification:** Since we discover that most music artists often stick to their initial music genre. Though some artists may change their music genre several times, they often only have one most remarkable and reprensentitive music genre in their lifetime.
- **Assumption 3:** We assume that the music similarity decays over time.
  $\hookrightarrow$ **Justification:**Since audience are more likely to chase after the current mainstream music instead of the classical music, it's reasonable for us to think that the influence of old music genre decreases as time pass by.

3 Model preparation

3.1 Notations

We list the symbols and notations used in this paper in Table 1:

Table 1: Surface water regression coefficient

Symbols	Definiton
MI	Music influence
SR	Similarity rate
MD	Mahalanobis distance
RR	The importance of one influencer

3.2 Data Discription

The data that we use and their contents are listed as follows.

Table 2: Data cleaning for music influence analysis

Data Sets	Contents
$Data\_by\_year$	Describing the mean values of the characters of one year.
$Data\_by\_artist$	Describing the mean valuses of one music artist.
$Full\_music\_data$	The concrete data of music characters for one hundred thousand songs.
$Influence\_data$	Fifty thousand music artists ,his own music genre, and his flower's data.

3.3 Data Cleaning

The *full_music_data* that we use has about one hundread lists of the concrete data describing music charateristics. The huge amount of data made it hard to analyze and extract influential characters. In order to make it more executable, we decided to omit the data that varies greatly from the general distribution of data. To achieve this goal, we adopt the following measures:

Revised Item	**Concrete change**
artist name	The author's name is garbled. This is not convenient for us to study the relationship between artists and eliminate them
tempo	Tempo means the tempo of the track. If the tempo per minute of the track is 0, there is no rhythm in the track, and it is considered that it does not belong to the mainstream music or there may be data errors, it will be eliminated.
loudness	Data with loudness equal to - 60. The common range of loudness is - 60 to 0. Loudness = - 60 data loudness is too low, abandon this part of the data.
popularity	Data with too small population. For a piece of music, if the popularity of the track is too low, it is considered that it does not belong to the typical mainstream music, there is no need for machine study of this kind of data.
duration	Data whose duration is too large or too small. If the duration of a piece of music is too long or too small, then the audio may not be music, but audio books or mobile phone ringtones, etc., and it will be eliminated.
liveness	Data with liveness higher than 0.8. If liveness is higher than 0.8, the music is likely to be recorded live. The audience on the scene may have an impact on the attributes of music, so this part of data is removed to avoid adverse effects.
song title	There're some unusual random code in song title. Some of the song titles have "*" and other symbols, which is not conducive to our next analysis and elimination.

4 Model I: Music Similarity Mathematical Model

Since our model makes SR a integral indicator for calculating MI, we'll need to calculate SR first. In order to estimate the similarity between two songs, we decided to use *mahalanobis distance*MD to evaluate it.

However, using MD requires the indicators to be linear independent from each other. So, we decided to use Principal Component Analysis (PCA) to process on data previously.

After PCA analysis, we can achieve two linear independent integer and their data. Then we calculate MD using this data. Since the distance is in inverse proportional to SR, we'll need to construct an modification function to change the relation.

4.1 PCA

4.1.1 Process of PCA Algorithm

As we want to quantify the similarity between songs, we decide omit the characters that're clearly irrevelent to SR, including "artist name", "song title", "release date". So there's about 14 characters, and we process PCA analysis on these data.

Algorithm 1 Process of PCA for music character indicators.

Input: The matrix M_{nm}, which has n rows of m dimension data;

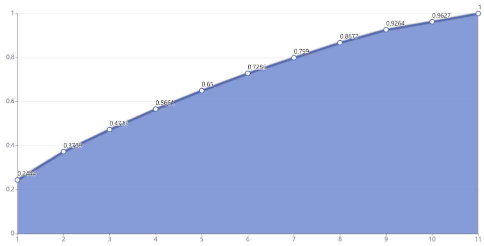
Output: linear independent indicators of the initial data ;

1: the original data is divided into n rows and m columns matrix X by columns;
2: each row of X is zero averaged, that is, the average value of this row is subtracted;
3: calculate the covariance matrix;
4: the eigenvalues and corresponding eigenvectors of the covariance matrix are obtained;
5: the eigenvectors are arranged into a matrix from top to bottom according to the corresponding eigenvalues, and the first k rows are taken to form the matrix P;
6: **return** ;

4.1.2 Results of PCA Analysis

To simplify our data, we did a dimension reduction of data, using the Principal Component Analysis (PCA). The core idea of PCA is to find several new variables to reflect the characteristics of our studying objects and reduce the dimensions needed to describe them. First, we find the covariance matrix C of the dataset, then calculate its eigenvector, and finally project the original data into the space formed by the eigenvector. When the sum of eigenvalues exceeds 0.85, the PCA method is considered reasonable. According to our calculation, the Gravel map for choosing number of eigenvalue is as follows:

Figure 2: Gravel map



It is not difficult to see that when the eigenvalue is added to the eighth item, the cumulative sum has reached 0.8677, higher than the required 0.85. Therefore, we reduce the original 12 dimensional data to 8 dimensions to pave the way for the subsequent calculation and analysis.

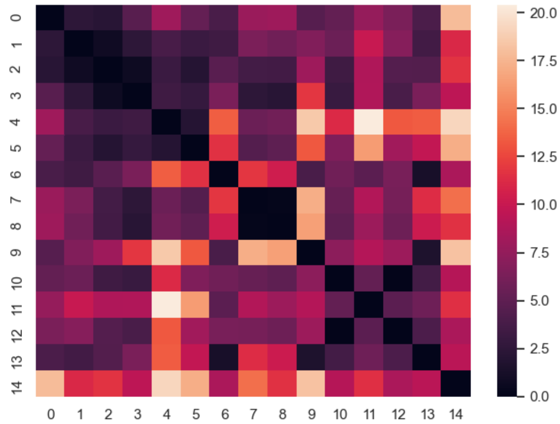
4.2 Calculation of SR

Mahalanobis distance (MD) was raised by the Indian statistician, P. C. Mahalanobis. It's used to measure the similarity between two samples. Different from Euclidean Distance, MD can overcome the problem of scale inconsistency. What's more, MD isn't affected by dimensions, as well as eliminating the correlation between variables. The formula is as follows:

$$D_M(x) = \sqrt{(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

A full-rank covariance matrix is necessary if we choose to use MD. Obviously, our data meet the requirements of calculating MD. After getting MDs, we take the reciprocal as a measurement of similarity between different artists and music.

Figure 3: Thermodynamic diagram of MD within genre



Since the MD data that we have calculated is in inverse relation to the degree of similarity between two songs, construct an modification function to change the relation. It's listed below:

$$SR = \frac{1000}{0.01 + MD}$$

4.2.1 SR of Artists between and within genre

As we have built a model to calculate SR, we used it calculate artists from different genres. We calculate the SR within genres. The SR within Blues is 0.97, while SR between blues and Spoken is 0.57. We use the same method to calculate other genres, and the final result is similar.

So, Artists between genres is very different from each other, while artists within one genre is much more similar.

5 Model II:Music Influence Model

Our model decides to use SNA (Social Network Analysis) theory. As this theory can help describe the significance and influential ability of the point in one network. So we use centrality of eigenvectors(CE) to calculate it, and we added two other factors to optimize it.

One is the similarity between followers and their followers. Another important part is that we need to consider that one's influential ability decreases over time. Combine the above two factors,we can create a influnce rate function, sot as to calculate the influence ability of one musical artist.

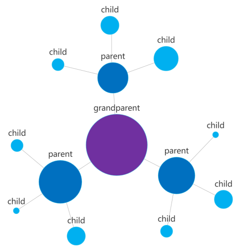
5.1 Influnce Rate Function

Our assumption is based on the assumption that the influence ability of an artist can be quantified by the number of his apprentices. However, it's too biased to consider a song simply by its access. In fact, a singer can be inspired by more than one person. Some people give him more inspiration, while others give him less. So we measure the influence of elicitors on followers through the similarity of various indicators of musicians to musicians defined by us.

Inspried by the principle of three degrees of influence, if we go beyond the three degrees of separation, our influence will gradually disappear. We only consider the SR of influencer, follower and sub follower in our formula.

$$f(SR) = \begin{cases} SR, & \text{if follower and influcer are of same genre} \\ 0.8 * SR, & \text{others} \end{cases}$$

Figure 4: Visualization of three dimensional theory



However, the influence of one artist decreases over time. So we need to add coeffient to show this kind of effect. Similar to the halflives of elements in physics,we add the regular term of γ^i.*i:The i-generation follower, j:The serial number of the follower in one generation*

After the experiment, we find that 0.5 is more suitable for the regularization term(γ); Then, for a certain artist, we sum up all the influence exerting on others in order to get a total influence of him, which can be seen as his Music Influence. The sum of such inspired following pairs for all generations is the total influence of this person. Altogether, we get the following fomula:

$$\mathrm{MI} = \sum_{i=1}^{2} \gamma^{i-1} * \sum_{j} w_{\mathrm{j}} * f(SR) \tag{1}$$

5.2 Subnetwork of Directed Influencer

We choose to analyze the influence sub network of Bee Gees. Using the above influence equation, we get the following visual network image:

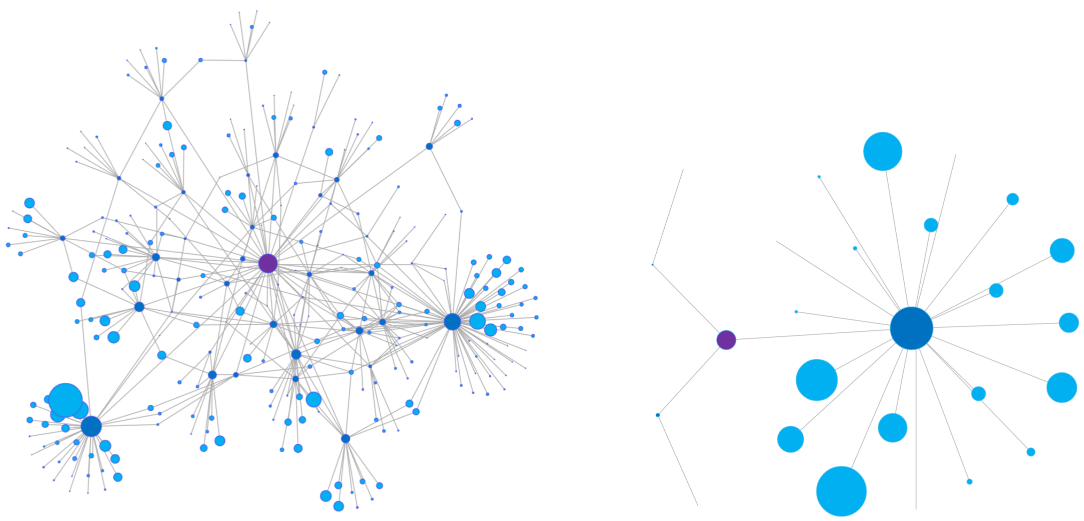


Figure 5: Subnetwork picture

For the picture on the left, we can see it has a lot of important influential follower sub nodes, and the followers also tend to have greater influence. While the number of network nodes in the picture on the right is obviously less.

Because our influence model is based on the number of followers, the influence of the nodes on the left is much larger than that on the right.

6 Model III:Genre Comparison Model

6.1 Comparision method

Since now we can calculate the SR and MI data between different genres and within one genre, we can use **Analysis of Variance** (ANOVA) method to analysis the data of this two sources. This mathmatical method was a part of the **significance test** method.

Significance test starts with making a hypothesis about the parameters of the population (random variables) or the distribution form of the population in advance, and then use the sample information to judge whether the hypothesis is reasonable, that is, to judge whether there is a significant difference between the real situation of the population and the original hypothesis.

In this way we can vividly show whether there's clear difference of SR between genre and within genre.

6.1.1 Process of ANOVA Algorithm

Algorithm 2 Process of ANOVA, proccessing on SR data

Input: SR data between different genres and within one genre;
Output:

1: calculate the sum of squares;

$$SS_T = \sum X^2 - \left(G^2/N\right) \tag{1}$$

$$SS_B = \sum \left[n_i(\bar{X} - \bar{G})^2\right] = \sum T_i^2/n_i - {}^{G^2}/N \tag{2}$$

$$SS_W = \sum SS_i \tag{3}$$

2: calculate the degree of freedom;
3: calculate the mean square;
4: calculate the F value;
5: check the f-value table for F-test and make judgment;
6: display the ANOVA table;
7: **return** ;

6.1.2 Analysis of the ANOVA results

Next we will compare the similarities of songs within and between genres. Let's take pop&rock and classical music as examples. In these two kinds of genres, ten samples are randomly selected as the research objects, and their Mahalanobis distances are plotted as follows.

Figure 7: MD within Classical

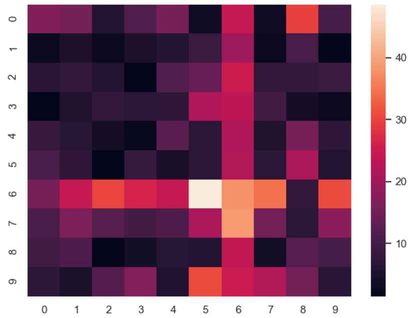


Figure 8: MD within Pop

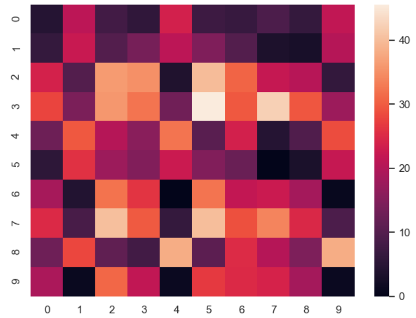
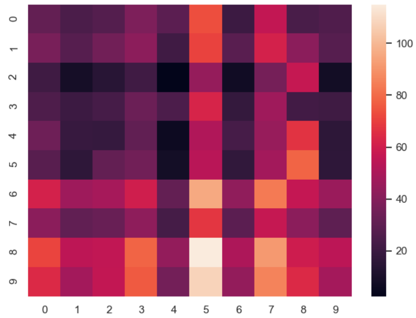


Figure 9: MD between them



The smaller the Mahalanobis distance, the greater the correlation. It is not difficult to see that SR within them is very low(average around 10), while the value of MD between them is high with the maximum value of 100.

So the correlation between the two genres is significantly less than that within the two genres themselves. We do a double sample t-test on the collected music samples.We use the MD data above to do ANOVA analysis, and the final T is 14.45 ,and the value of P is 0.So there's great difference within these two data sets.As it can be shown in the Figure 6.1.2:

Double sample t-test between pop-pop and pop-classical

	N	Mean	Standard deviation	Mean standard error
Pop	100	12.5	9.59	0.96
Pop-classical	100	39.3	15.9	1.6
Difference estimation	26.82		**T-value**	14.45
Difference 99%CI	(21.98,31.66)		**P-value**	0.000

Figure 9: ANOVA

However, it can be seen from the heat chart that the minimum MD between two different genres will not exceed 20, while the maximum Mahalanobis distance between the same genres may exceed 40. Therefore, although the similarity within genres is higher on the whole, it can not be ruled out that the music similarity of individual different genres is higher than that of some of the same genres.

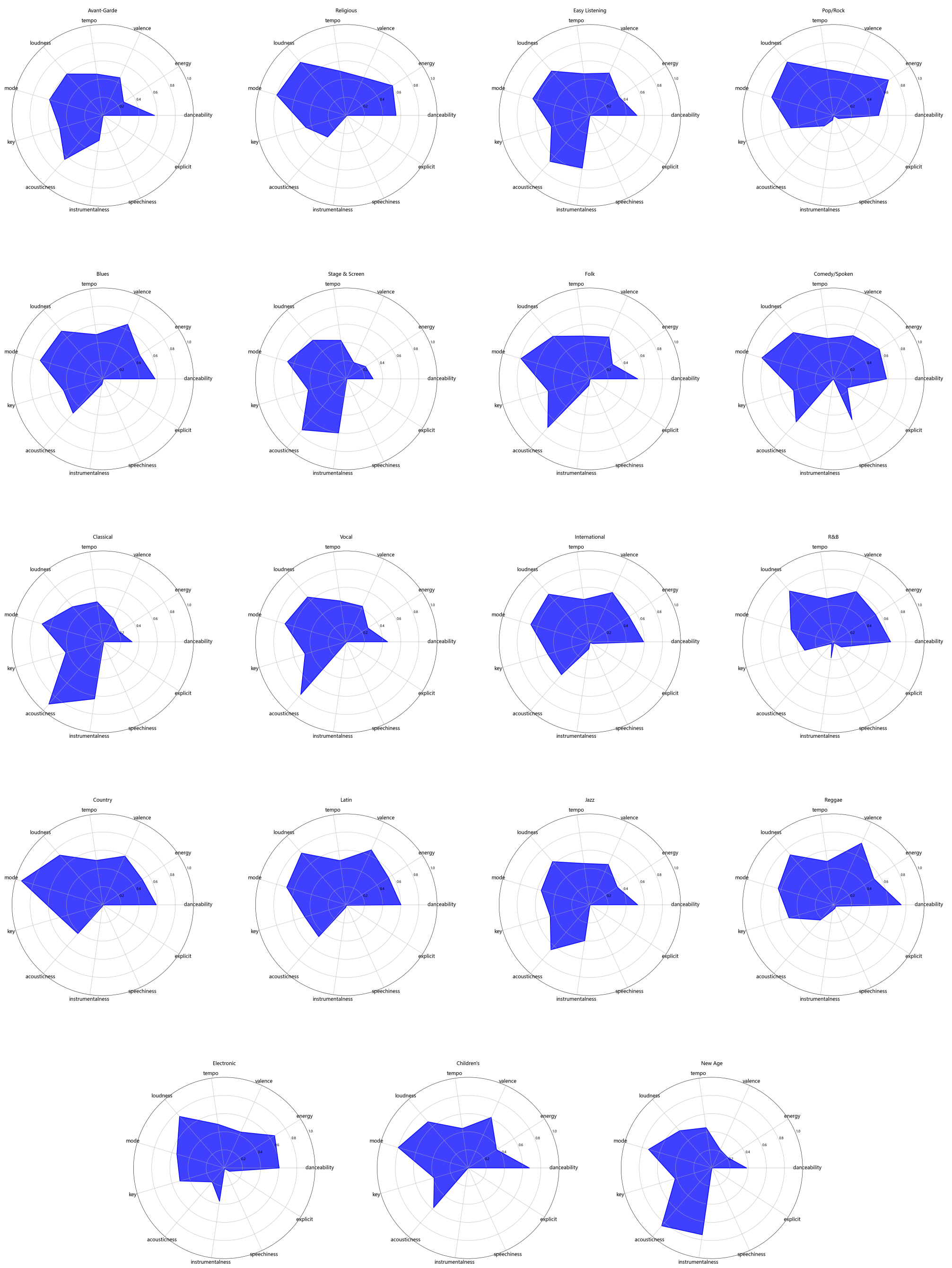
6.2 Analysis of SR and MI of Genres

For analysing the changing trend of 8 representative characters of genres, we has use the average comprehensive sorting vector method to serve the purpose of Data standardization.

$$\text{representative num} = \frac{\sum charaters}{n}$$

After data processing, we use the collected data to make the rader maps, vividly showing the difference between genres and the representative characters of one genre.

Figure 10: Radar Map of genres



6.2.1 Difference between Genre

Take Latin, New Age and Blues for example, we can see that Latin's data in loudness, tempo and valence is more outstanding while low in speachiness and instumentalness. So it sounds more rhythmic. The drums are strong.

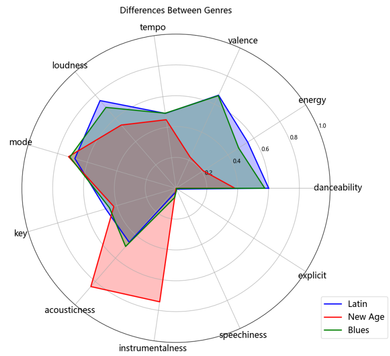
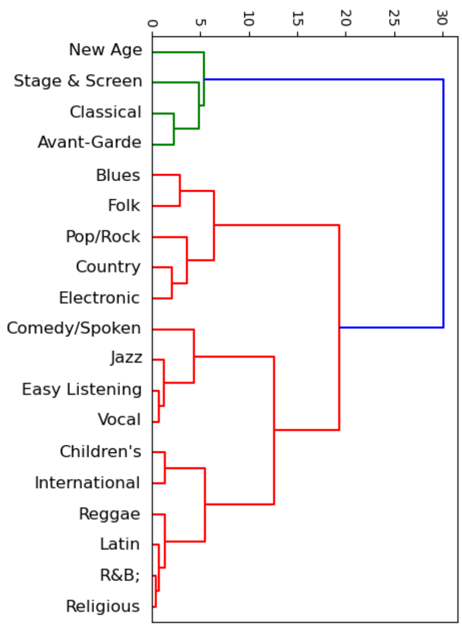


Figure 11: difference between 3 genres

While music of New Age genre focuses on acousticness and instumentalness.Consequently, many rhythms of new age exist between the natural rhythm of music and the strength of notes. It is natural and free, with a wind like streamline posture.For music of Blues time is more average in these kinds of datas. It has no obvious high point.Since Blues is a kind of vocal music and instrument music based on pentatonic scale. And another characteristic of blues is its special harmony, so it's sounds more smooth.

6.2.2 Hierarchical Clustering

Combined with genre SR data analysis and geometric estimation of feature mapping between genres, and then hierarchical clustering for these subjects' SR data, we can get the corresponding high correlation and low correlation genres:



7 Model IV:The Relation between Influencers' MI and SR

7.1 The Extent to which Follower is Influenced by Influencer

For the purpose of calculating the most infectious music characteristic, we'll need to quantify the SR values between influencer and his followers. Comapre this data with the SR data between influencer and other genres.

We chose two representative artists.And the SR data that we get is shown as follows:

artist	SR with followers	SR with others
Bob Dylan	0.183890642	0.091076
Bee Gees	49.89082423	2.250704

7.2 Gray analysis:Contagious Music Character

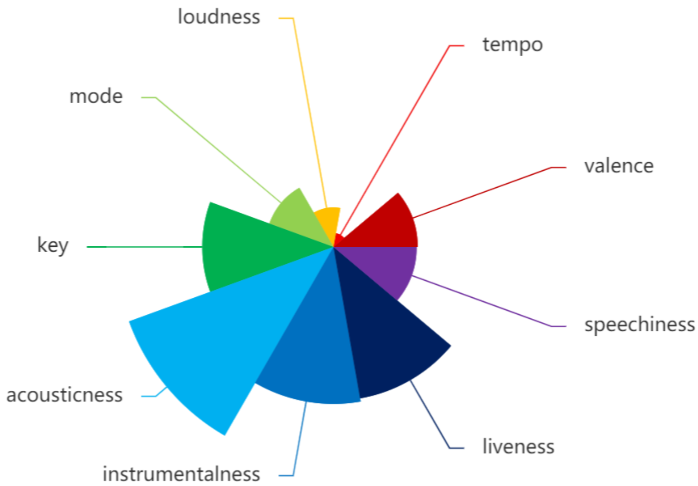
In order to compare the correlation between the music characteristics and the actual influence degree, we decided to select a representative influencer, and use his SR value and other music characteristics for Grey correlation analysis.And the value of R, which is used to express the degree of correlation,is listed as follows:

	Valance	Tempo	Loudness
R	0.079	0.013	0.037
	Mode	Key	Acousticness
R	0.064	0.123	0.204
	Instrumentalness	Liveness	Speechiness
R	0.147	0.144	0.078

The higher the value R is, more related it is with the level of SR,which means it's more contagious music character.

Furthermore,we have calculated the correlation coefficient.And the data are shown in the Figure 7.2:

Figure 12: correlation coefficient of characters

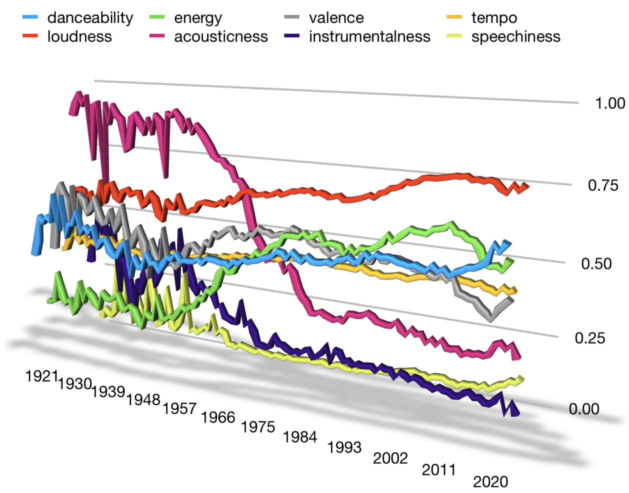


We can see from the paragraph that, acousticnecc ,instumentalness and livness are the more constagious music features. While tempo and loudness is not very infectious.

8 Model V:Time Series Model of Music Influence

8.1 The General Change of Music characteristic

Figure 13: music changes over 100 years

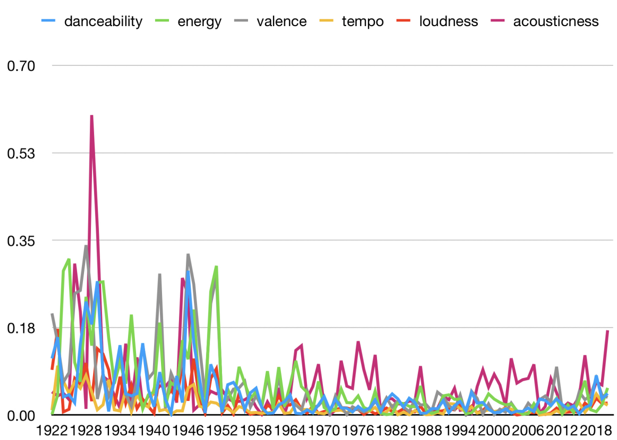


For this picture, the vertical axis of the Figure 14(a) shows the proportion of this musical feature in the total, and the horizontal axis is its time axis.

Figure 15(a) is the first-order difference of each index in the left figure. The calculation formula of longitudinal axis is as follows:

$$\frac{dy}{dx} \approx \frac{\Delta y}{\Delta x}$$

Figure 14: first-order difference of each index



These figures intuitively tell us that the proportion of pure music dropped sharply from 1957 to 1975, which shows the decline of acoustic index; And the same decline of instrumental index shows that the proportion of human voice in music is gradually increasing compared with that of musical instruments.

On the contrary, the energy index is increasing year by year. Fast and intensive music is increasing year by year. The loudness index is also rising, indicating that the music volume decibel is gradually increasing.

We can get the conclusion that major music changes are most likely around the 1960s or 1930s.

8.2 Important Node in the Music Influence Network

For the purpose of determination of key nodes in complex networks, our group has adopt the method of time series analysis.

Assumption: Because there is no genre label in the original data, we reasonably assume that an artist's life style is relatively stable. So we use the genre tag in the influence data set to assign genre tags to songs and singers in other data sets.

8.2.1 Indicators of Dynamic Influencer

To find the important music feature data of this period is locked for network key node analysis. Then determine which musicians or features are key nodes in the network without time dimension ($such\ as\ the\ 1957 music\ influence\ network$). We do this by using RR index.

RR is the value that we used to measure whether someone is a dynamic influencer. It's calculated by the following equation:

$$\mathrm{RR} = (\text{offset value})^{0.5} * \mathrm{MI} \tag{2}$$

The offset value is to describe the change of this factor to the overall data, while the influence (MI) is to describe the change ability of this person to the overall data. It equals the MI between the centre of the web and this node. The multiplication is the change ability and importance of this network node to the overall topology.

8.2.2 Revolutionary Music Artists

Using the above equation, we find the revolutionary musicians in the era of Music Revolution (1960s). The RR values of head musicians in the 1960s are listed as follows:

	The Beatles	Bob Dylan	The Rolling Stones	Jerry Lee Lewis	The Who
RR	1678.33	1543.10	1396.81	1274.01	939.47

Compared with the actual achievements of these musicians, we find that there is a high degree of data relevance and consistency. Here are the three most influential musicians we identified in the 1960s and their actual achievements:

Figure 15: The most influential musicians of the 1960s

8.3 The Changes of Genres

We choose the mainstream genres to analyze the change of music genre.It's shown as follows:

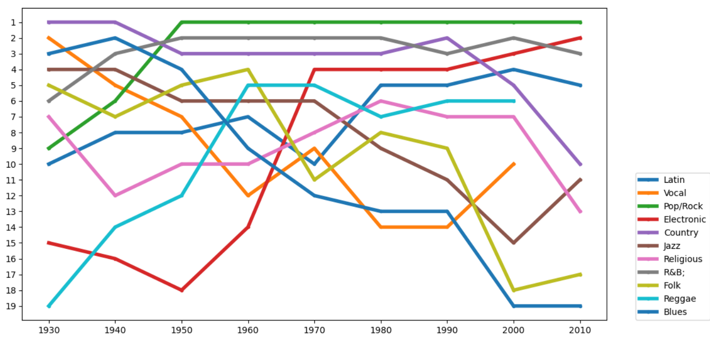


Figure 16: Genres' change

From the Figure 8.3, we can see that pop music has been on the rise over time. From 1940s to 1970s, the proportion of influence has a huge leap, which is consistent with the revolutionary development of pop music in history around the time node of the 1960s. While the proportion of R&B music influence tends to be stable, both of which are consistent with the actual mainstream music.

9 Analysis of other Factors in the Model

After detailed data analysis, it is found that the most important time of English Revolution is ages from 1950's to 1970's. The most critical year is around the 1960s.

Through the above RR index, this paper analyzes the musical characters in these two years. It is concluded that the influential musical figures in 1960 are the Beatles, the Rolling Stones and Bob Dylan. For the emergence of these revolutionary figures, we mainly take the Beatles band as an example to analyze from three aspects and find out the economic, political, cultural and technological factors in our model:

- **Political factors:**
  1. **War factors:** From the end of World War II in 1945 to the 1970s, Vietnam War, racism and other political factors led this culture to the Beatles, who was also the most influential revolutionary musician in our evaluation in the 1960s.
  2. **Globalization:** In the 1960s, the transportation between Europe and the United States was more convenient, and the emergence of airplanes made it more convenient for musicians to perform.
- **Cultural factors:**
  1. **The decline of religion:** The cultural factor that religious culture no longer occupies the main political position. This shows that the overall influence of music of religious is decreasing year by year in the model.
  2. **Hippie culture:** Hippie culture is in a prosperous period. The youth in this period are fond of fantasy, but they are dissatisfied with the disappointing reality of succumbing to the dark, so they begin to rebel and absurdity, which has become the characteristics of hippie culture.
- **Technological factors:**
  1. **invention of cassetes:** In 1963, Philips of the Netherlands developed the world's first cassette tape. Technological innovation helped the prosperity of music.
- **Economic factors:**
  1. **Golden Age:** After World War II, there was a golden age of American economic growth. From $523.3 billion in 1961 to $1063.4 billion in 1971. The prosperity of economy gives birth to the prosperity of culture.

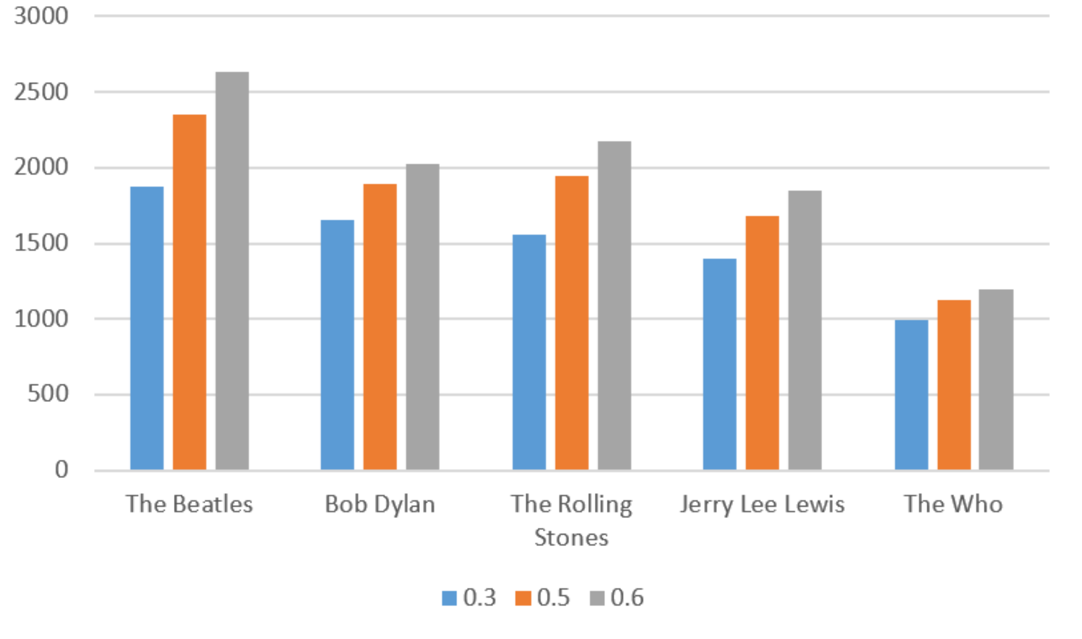
All the above reasons show that our model is reasonable to predict the trend of music changes.

10 Evaluation of the Model

10.1 Sensativity Analysis

For the constant of influence in our model, the RR value is affected by the coefficient. The confidence interval of the coefficient is between 0.1 and 0.8, and the relative rank of the influence obtained by changing the value of γ remains unchanged. This shows the robustness and stability of our model to γ.

Figure 17: Sensativity Analysis



10.2 Validating the Model

We looked for a relatively professional official music influence ranking, (100 greatest artists on Rolling Stone magazine). By comparing the influence ranking of our model with the actual data set, we find that the data are highly coincident, and the influential figures in the head are almost the same (including the Beatles, Bob Dylan, rolling stones, etc.)

11 Conclusions

11.1 Summary of Results

11.1.1 Results of Problem 1

According to our ModelI , We get this formula to calculate the similarity between the two pieces of music:

$$SR = \frac{1000}{0.01 + MD}$$

According to the results of ModelI, we find that the music of artists in the same genre is more similar.

11.1.2 Results of Problem 2

According to our ModelII, We get this formula to calculate musical influence of one artist:

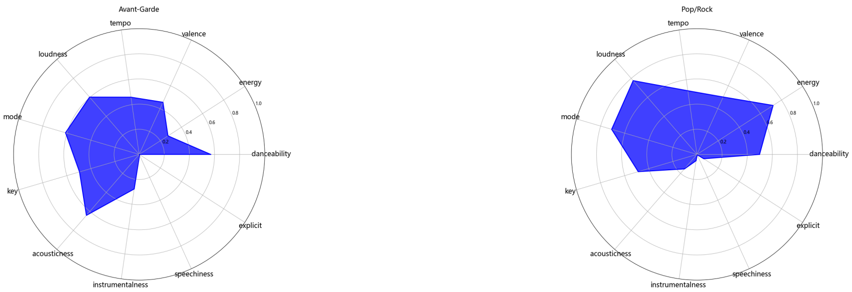
$$\mathrm{MI} = \sum_{i=1}^{2} \gamma^{i-1} * \sum_{j} w_{\mathrm{j}} * f(SR)$$

And we find that the probability that the influencer with more followers has more influence is far greater than that with less followers.

11.1.3 Results of problem 3

According to Our ModelIII, we identify there was a big gap between the similarity within genres and between genres. Figure 19(a) show the prominent feautres of 3 representative genres:

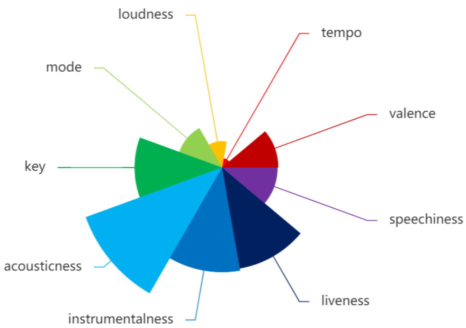
Figure 18: Radar Map of genres



11.1.4 Results of problem 4

Through the gray correlation analysis, we found that influencer does have a great influence on follower's music, and the correlation degree is usually more than 80At the same time, some musical features are more continuous,they are listed as follows:

Figure 19: correlation coefficient of characters

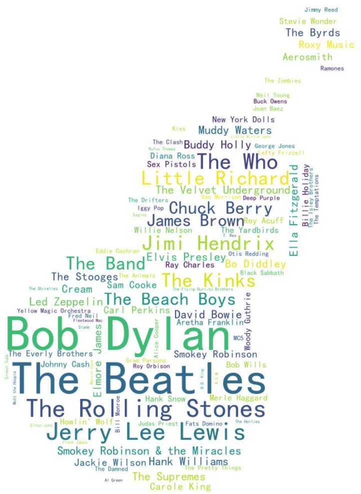


11.1.5 Results of problem 5,6

For these two questions, we adopt the RR index as the indicator of the impotance of one artist. The calculation formula is listed below: $\mathrm{RR} = (\text{offset value})^{0.5} * \mathrm{MI}$.

At the same time, we find that the time of music revolution is most likely in 1950-1970, especially in 1960. At the same time, the most infulential artists found by our model is shown as follows:

Figure 20: The influential musicians



11.1.6 Results of problem 7

Our model is in good agreement with the historical data of the actual music revolution. Taking the 1960s as an example, it coincides with many other factors int that society:

- **Economy:** the golden age
- **Politics:** war factor,globalization
- **Culture:** the decline of religious culture,Hippie culture
- **Technology:** invention of cassetes

11.2 Strengths

- **Data utilization**
  We make full use of the data given by the topic, integrate and clean the four data sets, and make effective use of most of the data given by the topic.
- **Multidimensional method**
  In the process of building the model, PCA, gray prediction, Mahalanobis distance and other methods are used to make our model more convincing.
- **Innovation**
  We have many innovative ideas in the model. For example, when looking for dynamic influencers, we use Mahalanobis distance and the square root of outliers to measure the change ability of artists.
- **Accuracy**
  We compare the influence ranking of the model with Rolling Stone magazine, and the results are in good agreement.

11.3 Possible Improvement

- If we have data of more dimensions, such as the quantitative influence of historical and social factors, we can get more accurate conclusions by improving the model.
- Due to our assumption of fuzziness, the model may have errors under extreme data.

References

[1] Zhao Xiaohui. Research on measurement and maximization of social network node influence [D]. Shandong Normal University, 2019

[2] Hao Sun, Xiaoxia Cai, Zufeng Zhang. Research on key network node identification based on depth - first search algorithm. 2017

[3] https://www.rollingstone.com/music/music-lists/100-greatest-artists-147446/public-enemy-6-89379/

[4] Qiao Zhi, Zhao Lijun, Jiang Xinkai, et al. A Navigation Probability Map in Pedestrian Dynamic Environment Based on Influencer Recognition Model. 2020, 21(1):19-19.

[5] Harshavardhana Natarajan, Sundar Singh Sivam Sundarlingam Paramasivam, Durai Kumaran, et al. Application of TOPSIS and ANOVA Methods to Evaluate the Electrical Resistance Spot Welding Process Parameters of AISI 1020 Steel. 2020, 12(4):388-393.

[6] Erdős, P.; Rényi, A. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci. 1960, 5, 17–60.

[7] R. Albert and A.-L. Barabási, Rev. Mod. Phys. 74, 47 (2002), DOI: 10.1103/RevModPhys.74.47. Crossref, ISI, ADS, Google Scholar

Appendices

Appendix A Document

Dear ICM Society:

As we all know, music has been an important part of human culture. Music has a variety of attributes. Due to the subtle difference and combination between such attributes, many genres have been derived. These genres and music have distinctive historical characteristics. The artists of a certain era will be more or less inspired and influenced by their predecessors or contemporaries when they create. How do music and genres evolve? How do musicians influence each other? This is of great interest to us. Because of our interest and social responsibility, in order to explore the evolution of music and the interaction between music better, we established a mathematical model based on the node centrality in the network and measurement similarity.

According to our model, based on the data set of music and artists given by ICM, we find that in the past 100 years, the evolution of the main genres of contemporary music is shown in the Figure 22.

Figure 22: Music trend in past 100 years

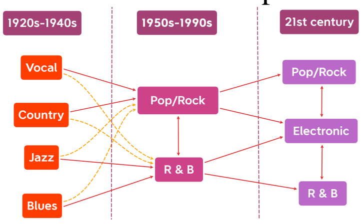


Figure 23: Top 3 most influential artists

In the 50s-70s, great changes have taken place within music genres, especially in the 60s. We believe that this is the result of the comprehensive effect of politics, culture, technology, economy and other factors. Around the 1960s, the process of globalization developed. Regional ties have been strengthened, and regional, ethnic, religious and cultural barriers have been gradually broken. With the rise of hippie culture and the invention of tape, American economy entered a prosperous period. Among all the musicians, the Beatles, Bob Dylan and the Rolling Stone are the most influential and innovative artists among all artists.As is shown in Figure 23.

Due to the limited data set given by ICM, it is difficult to analyze and compare more subdivided genres in our model. If we were given a more complete impact network data set, the impact assessment of artists will be more accurate; if we were given a data set with more subdivided tags of songs and artists, we can analyze the characteristics of songs in the subdivided genres and their changes in history. If we were given the historical, political events and economic indicators of the country and region, we may be able to keenly capture the subtle influence of these factors on the evolution of music in this period and the follow-up.

We recommend that researchers can start from social history and economy, from the distribution of nationality, religion, race and sexual orientation of singers, and from the theme of songs, such as anti war and environmental protection, to analyze the changes of music genre and artist influence.

Appendix B Codes

B.1 Calculation of RR

```
import numpy as np
import pandas as pd
import json

cov_matrix = pd.read_csv("2021_ICM_Problem_D_Data/8cov.csv")
def get_avg_by_yr_and_style(year, style):   #return type: numpy
        df_style = pd.read_csv("style_year_avg/" + style + "_year_avg.csv")
        df_style = df_style[df_style['year'] == int(year)]
        return np.array(df_style.iloc[0][1:9])
def get_similarity(x, y):
        if (x.shape != y.shape):
                return 999
        t = np.array((x - y).T)
        return  t.dot(cov_matrix).dot(x - y)
ft_df = pd.read_csv("filtered_artist.csv")
mar_dis = []
rr = []
rr_02 = []
rr_03 = []
for i in range(len(ft_df)):
        row = ft_df.iloc[i]
        year = row['year']
        style = row['style']
        avg = get_avg_by_yr_and_style(year, style)
        this_atrb = np.array(row[1:9])
        md = get_similarity(this_atrb, avg)
        mar_dis.append(md)
        rr_val = row['sim_rate'] * np.power(md, 0.5)
        rr_02val = row['sim_rate'] * np.power(md, 0.6)
        rr_03val = row['sim_rate'] * np.power(md, 0.3)
        # rr_02val = row['sim_rate'] * np.power(md, 0.6)
rr.append(rr_val)
rr_02.append(rr_02val)
rr_03.append(rr_03val)
ft_df['mar_dis'] = mar_dis
ft_df['RR_0.5'] = rr
ft_df['RR_0.6'] = rr_02
ft_df['RR_0.3'] = rr_03
ft_df.to_csv("filtered_data_with_rr.csv")
```

B.2 Extraction of network matrix

```
import numpy as np
import pandas as pd
csv_data = pd.read_csv("2021_ICM_Problem_D_Data/influence_data_1.csv")
list1 = csv_data['influencer_main_genre'].values.tolist()
list2 = csv_data['follower_main_genre'].values.tolist()
res = set(list1).union(set(list2))
resList = list(res)
csv_result = pd.DataFrame(0, index=resList, columns=resList)
for i in range(len(csv_data)):
        start = csv_data.iat[i, 0]
        end = csv_data.iat[i, 1]
csv_result[start][end] += 1

csv_result.to_csv("2021_ICM_Problem_D_Data/influence_data_2.csv")
```