Problem Chosen	2021 MCM/ICM Summary Sheet	Team Control Number
D		2121740

Abstract

In this paper, mathematical models are built by our team to evaluate the influence of previously produced music on new music and musical artists. Our model mainly contains two parts: a network to show the connection between influencers and followers and a quantitative model to evaluate the similarities between two artists' work. After quantitative analyses, we determined the consistency of these two models and use them to further explore different music genres' development and process of musical evolution/revolutions. In details, our model contains seven main parts:

With the data of connections between influencers and followers, we build a network to show the general process of influence. Based on the network, we raised the concept of "Influence Indicator", which is determined by the number of influencers, whether influencer chose the same genre, year difference with followers, and influence of the follower, to show the power of the influencers' influence. Details of subnetwork containing The Beatles are given as an example.

With the quantified data of characteristics of the music and type of vocals of a song or artist, we build a model to evaluate their similarities. Due to the high dimensions of the data, we applied Principal Component Analysis (PCA) to reduce dimension for simplification. Based on the features of reduced data, cosine similarity was chosen to represent the similarities.

In the process of building the above two basic models, we observe some interesting phenomenon about genres: the influence network is more complicated, and similarities are higher among artists in the same genre. Evaluating the importance of features in different genres with the Random Forest algorithm, we found that 'Acousticness', 'Energy', 'Danceability', 'Instrumentalness', and 'Speechiness' are the five main factors that distinguish a genre. Different genres sometimes also showed a similar developing trend, like during the wartime or technology rapidly developing time, and some of them share an inherent relationship in a social context.

We also find some correlation between the two basic models: the pair of influencers and followers do show higher similarities than the others. By comparing between influencers with more than 20 followers and their followers with relatively high general similarities, we find the feature of 'Acousticness' is easiest to be passed down to the followers, which corresponds to the result of what distinguishes a genre in part 3.

In the evolution process of genres shown in part 3, there exist major leaps of nearly all genres, which can be viewed as revolutions of the whole musical field. Evaluating the importance of features with the Random Forest algorithm, we find that 'Acousticness', and 'Duration', best distinguish the revolution in the 1960s-1970s; and 'Loudness' best distinguish the revolution in 1990s-2000s. Comparing the number of newly emergent followers during the revolution, The Beatles and Nirvana respectively show a leading role in the revolution in the 1960s-1970s and 1990s-2000s.

To more detailly analyses the revolutions, the Pop/Rock genre is taken as an example. By comparing the features of music by three main influencers before and after the revolution, conclusions are made that Pop/Rock music tend to be more acoustic and energetic after 1968 and more cheerful after 2000, which correspond to the result in part 5.

In the Social context, music represents characteristics of an era and also in turn presents significant involvement in that time. Our work shows how the influence indicator model and the data network help investigate such topics. On the other hand, the impact of society, technology, and politics on music can also be drawn from the data network.

keywords: musical evolution; influence network; similarity; PCA; Random Forest algorithm

Contents

1 Introduction

1.1 Background

The art of music is a cultural form of the glorious progress of human being's spiritual wealth. The study has shown that people listen to music to regulate arousal and mood, to achieve self-awareness, and as an expression of social relatedness [1]. Over the decades, music has generated multiple genres, and huge waves of musical popularity came with people developing various tastes. As a result, to better understand the influence of music on human society, the developmental process of music deserves more investigation. The study has shown that the development of music comes with a large part of imitation as well as some significant revolutions [2]. Many artists tend to believe that their work has been influenced by several previous artists, and they also agree that some huge social changes, like the political revolution, can be a major reason for the significant revolution of music.

1.2 Restatement of Problems

Our team is asked by the Integrative Collective Music (ICM) Society to develop a model that measures musical influence, which contains two main parts: evolutionary and revolutionary. Data provided include the 5,854 artists reported influencers and musical features of 98,340 songs. Based on the data set offered, Our team will:

I. Build a network to show the connections between influencers and followers;
II. Develop a model to evaluate the similarities between different songs, artists, and years;
III. Compare the similarities and influences in the unit of the genre;
IV. Explore the inherent correlation of similarities and influence network;
V. Evaluate determinants of revolutions in the music field;
VI. Give an example of Rock/Pop music's development in the revolutions;
VII. Evaluate the effects of social, political, or technological changes.

1.3 Previews Research

To better understand the problems, our team explore some previews researches.

I. Development of music field from the 1930s

By the end of the first World War, the emergence of neoclassicism music represent the renaissance of music. During the next 60 years, many new genres showed up and, especially from the mid-l990s, elements of world music have now crossed over into a vast range of other kinds of music, such as avant-garde jazz, John Zorn's work comes to mind here, some recent New Age albums like Andreas Vollenweider's Book of Roses or even dancefloor and hip hop. This phenomenon shows the world music's inclusiveness of national and ethnic identities and multicultural diversity[3].

II. Development of genres

A music genre is a conventional category that identifies some pieces of music as belonging to a shared tradition or set of conventions [4]. Although it is conceivable to create a musical style with no relation to existing genres, new styles usually appear under the influence of pre-existing genres[5]. A music genre or subgenre may be defined by the musical techniques, the cultural context, and the content and spirit of the themes[6].

1.4 Other Assumptions

To simplify our problems, we made some other assumptions:

- The data set provided is out of bias;
- The releasing year is the creating year of the music, which can represent the features of music of that year;
- The followers of followers can also be influenced by the influencer.

2 Notations

Symbol	Description	Symbol	Description
I	Influence indicator	ρ	Number of influencers
m	Whether follower share same genre with influencer	γ	Year difference
w	Influence of the follower	r	Cumulative contribution rate
S	Cosine similarity	W	Feature importance
Δ	Number of new artists per year	P	Proportion of high S
δ^2	Sum of squares of deviations	C	Follower count

3 Models

3.1 Data Pre-processing

To generate more clear and reasonable data, data pre-processing as follows is did:

I. Normalize the data in columns 'tempo','duration', and 'loudness' to 0-1 in the three databases.

II. Turn the feature 'mode' to a classifier.

III. Drop the unreasonable data, including:

(i) 43 rows of the data whose artist's name and features are the same while the id is different in data grouped by artists.

(ii) Row of data whose influencer is Mozart: one artist claiming to be influenced by an artist 60 years later from himself. After looking into the data, the influencer is found to be *Wolfgang Amadeus Mozart*, a musician from the 18^{th} century. The data becomes meaningless because the active time for Mozart is incorrect. To avoid such errors, we checked all 8 pieces of data with the year difference smaller or equal to -40 and found other data is sensible.

(iii) The data whose genre is 'unknown' when doing classification.

3.2 Influence Network

As reported by the artists themselves as well as the opinions of industry experts, connections in form of influencer-follower were built among artists in the influence data frame. Based on these connections, a network can be built and artists' influence indicator (I) can be measured.

3.2.1 Visualization of influence network

The network is developed by the tool Gephi 0.9.2 (see Figure.1).

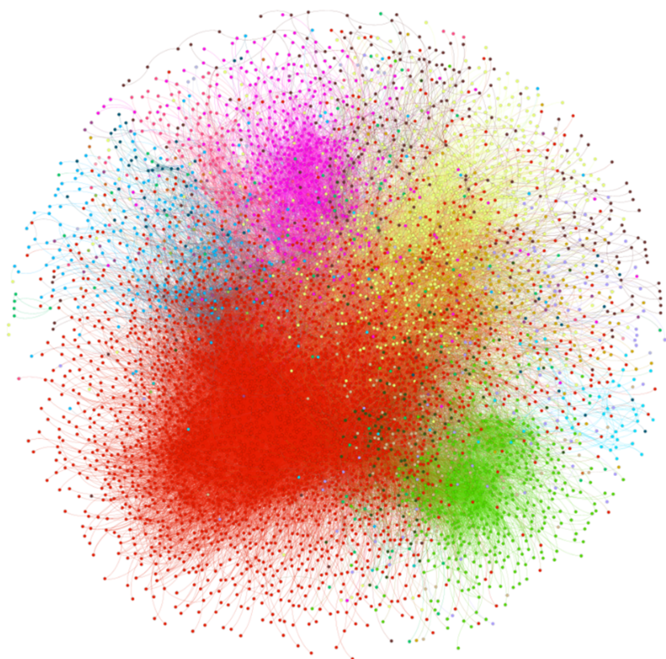
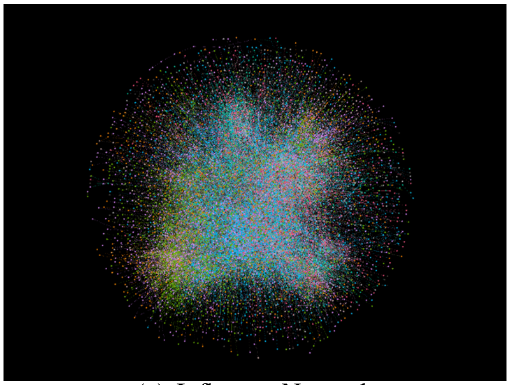


Figure 1: Influence Network Colored by Genre

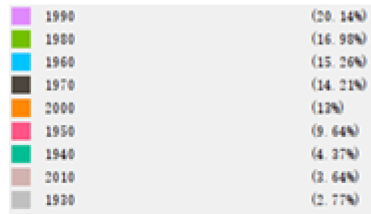
From the filtered influence data, we can attain 5,560 artists' information as the nodes. Then, we attain 4,2714 influencer-follower relationship entries as the edges connecting nodes. Considering the influence is monodirectional, the edges are directed, pointing from the influencer to the follower. The information stored in the node includes basic data of an artist, namely *the id, the name, the main genre, and the active start*. The information helps filter out unimportant nodes when making a query to investigate the influence of a specific artist.

We painted the nodes with different colors according to their genres. As shown in the figure, red nodes represent Pop/Rock musicians, which take up **50.11%** area, indicating the proportion of artists in this genre. Similarly, yellow nodes represent R&B musicians, pink nodes represent Jazz musicians, and so on. The layout of the graph subjects to force-directed graph drawing algorithms, namely the Fruchterman Reingold layout. The purpose is to position the nodes of a graph in two- or three-dimensional space so that all the edges are of more or less equal length and there are a few crossing edges as possible, by assigning forces among the set of edges and the set of nodes, based on their relative position [7]. The critical meaning of the layout is that nodes connected are mutually attracted while nodes unconnected are mutually repulsed. As intuitively shown in the figure, nodes of the same color cluster, which means influence within the genre is common and strong. Also according to the penetration of nodes with different colors, we can take a peek at how artists influence between genres.

It is worth mentioning that although the color partition of nodes is based on genre, it can be based on other attributes, and thus provides other information. Figure.2 shows nodes of different colors based on the artist's starting time to be active.



(a) Influence Network



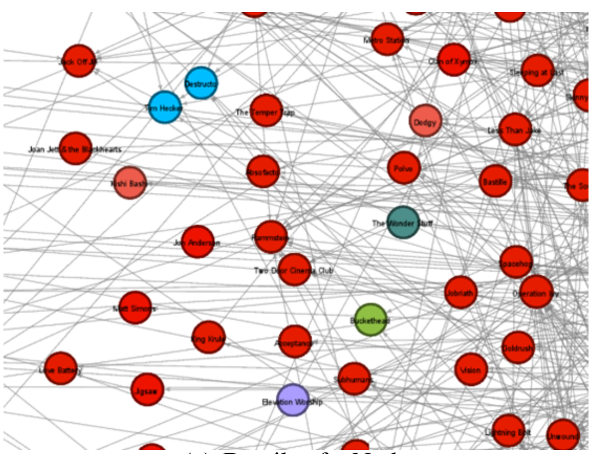
(b) partition

Figure 2: Influence Network Colored by year

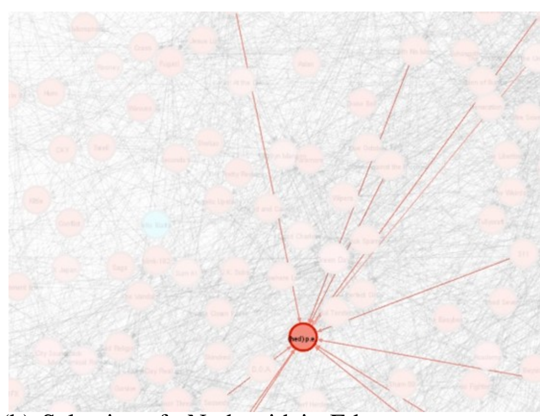
We can see that all colors are mixed together, which indicates that influence across time is common. Also, although there are only **14.21%** nodes are blue, they seem to occupy a large area in the center for the visual effect. The appearance implies that artists in the 1960s, including *The Beatles, Bob Dylan, The Rolling Stones* and so on, have great influence across time.

3.2.2 Influence indicator

Apart from the general picture, the network also shows details of in- or out-edges and neighbors of a node (see Figure.3). Using the details, we can develop the influence indicator for each artist.



(a) Details of a Node



(b) Selection of a Node with its Edges

Figure 3: Details of in- or Out-edges and Neighbors of a Node

The model for the influence indicator uses the comprehensive evaluation concept. The comprehensive evaluation method aims to turn variables in different fields and with different units into a uniform variable, namely the grade. In this problem, we can easily develop an indicator in the unit of "follower(s)". More followers indicate more influence. As a result, the problem turns into how to evaluate a follower. There are four evaluation factors for the "quality" on a follower:

I. The number of influencers the follower claims to have;
II. Whether the follower chooses the same genre with the influencer;
III. The time gap between the influencer and the follower to become active;
IV. The influence of the follower himself.

These evaluation factors are in different directions, for which we can use the comprehensive evaluation measure to turn them into a uniform and easy-to-calculate unit: "follower(s)". In other words, we give four parameters $(\rho, m, \gamma, \omega)$ to represent these four factors, and count the follower in question as

$$\rho \cdot m \cdot \gamma \cdot \omega \cdot 1 \tag{1}$$

Summing up all direct followers of the influencer in this form, we get a numerical parameter as the influence indicator. Here is mindmap for the evaluation (see Figure.4):

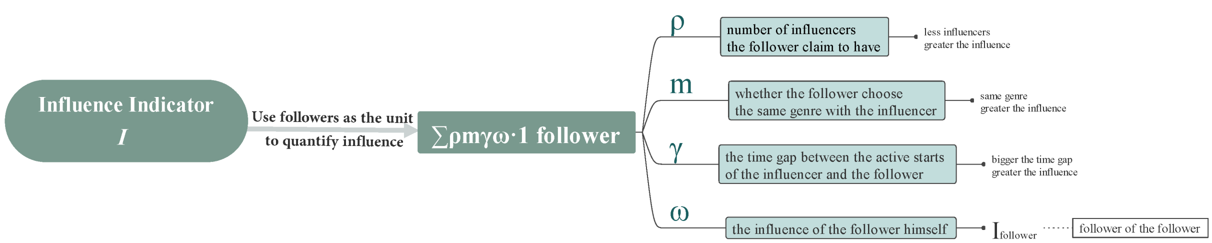


Figure 4: Mindmap for Evaluation

The first factor is the number of influencers the follower claims to have. The more influencers he has, the less importance is attached to the influencer in question. As a result, a follower shall be counted as $\rho \cdot 1$ followers. ρ is a parameter negatively related to the number of influencers the follower has. We use the distribution of the number of influencers to determine ρ (see Figure.5). 620 artists only have one influencer and one artist has 47 influencers, which is the most.

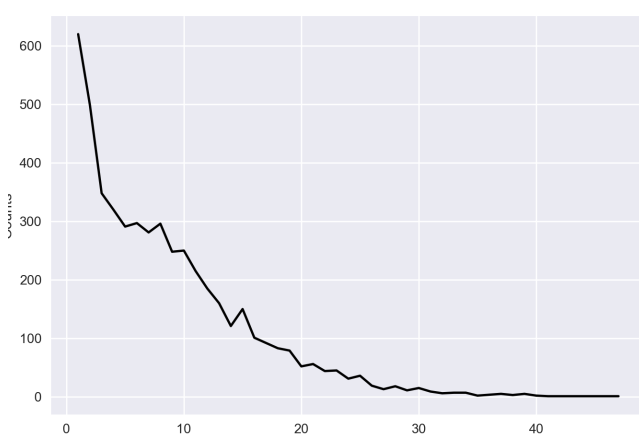


Figure 5: Distribution of Number of Influencers

However, the distribution hardly fit any functions, so we choose to make a general partition for it. From the figure,

$$\begin{cases} y \geq 300, & 1 \leq x \leq 4 \\ 200 \leq y < 300 & 5 \leq x \leq 11 \\ 50 \leq y < 200 & 12 \leq x \leq 21 \\ y < 50 & 22 \leq x \leq 47 \end{cases} \tag{2}$$

So we develop

$$\begin{cases} \rho = 1.0 & 1 \leq influencer \leq 4 \\ \rho = 0.8 & 5 \leq influencer \leq 11 \\ \rho = 0.5 & 12 \leq influencer \leq 21 \\ \rho = 0.2 & 22 \leq influencer \leq 47 \end{cases} \tag{3}$$

The second factor is whether the follower chooses the same genre as the influencer. If yes, we believe the influencer has a greater influence on the follower. Since it is a simple yes/no question, we develop m as m=1 if the follower chooses the same genre, and m=0.5 otherwise.

The third factor is the year when the follower starts to be active. The practical meaning of the factor is to evaluate how the influence spread across time. We believe that if the follower already has a career before the influencer starts to be active, the follower is more valuable. Also, if the influencer can influence an artist from many decades later, the influence proves to be greater. As a result, we want to learn more about the difference between starting active time of the influencer and the follower. Then, we attain a distribution figure of the difference between starting active time of the influencer and the follower (see Figure.6).

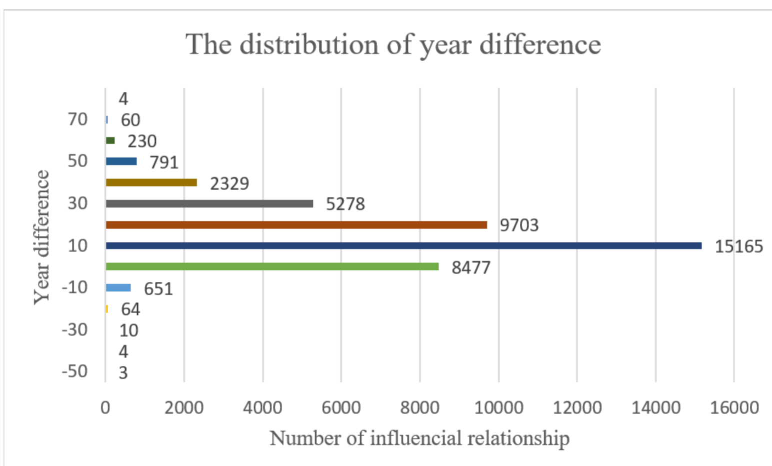


Figure 6: The Relationship Between Influencer Number and Year

Combined with practical meaning, we develop a parameter γ to count one follower as $\gamma \cdot 1$ follower based on the year difference between the influencer and the follower.

$$\begin{cases} \gamma = 1.0, & \text{year difference } \in \{0, 10, 20\} \\ \gamma = 1.2, & \text{year difference } \in \{-10, 30\} \\ \gamma = 1.5, & \text{year difference } \in \{-20, 40, 50\} \\ \gamma = 1.8, & \text{otherwise} \end{cases} \tag{4}$$

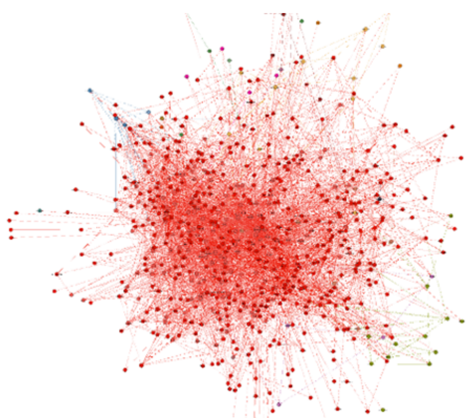
The fourth factor is the influence of the follower himself. The practical meaning of the factor is to evaluate how the influence spread indirectly. In theory, we can recursively apply the process to calculate the influence indicator for each follower and weigh them by order. However, for simplification, we can just consider the number of sub-followers and give it a weight. After calculation, the expected value of $\rho \cdot m \cdots \gamma = 0.79$. Considering indirect influence on the second-layer follower may be little, we decide the weight to be $\frac{1}{10} \cdot \rho \cdot m \cdot\ \gamma = 0.079 \approx 0.08$.

$$I = \sum \rho \cdot m \cdot \gamma \cdot \text{ follower } + 0.08 \cdot \text{ followers of followers} \tag{5}$$

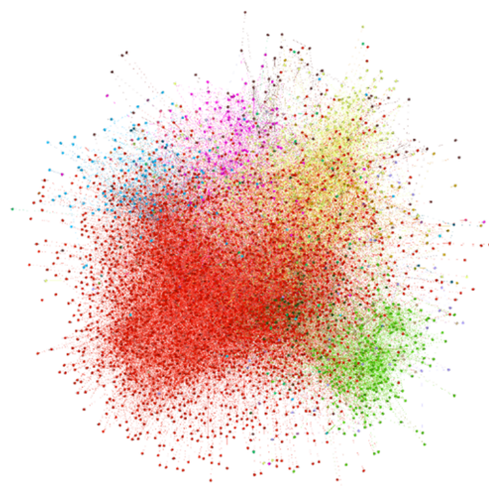
3.2.3 Example subnetwork

To draw an example, we explored the subnetwork starting from The Beatles, who has the most followers.

Figure.7 shows the followers of The Beatles (consisting of 615 nodes and 4145 edges) and Figure 8 includes the followers of followers (consisting of 3398 nodes and 32956 edges). The data directly demonstrates that The Beatles has 615 followers and 2783 sub-followers.



(a) One-layer Subnetwork starting from The Beatles



(b) Two-layer Subnetwork starting from The Beatles

Figure 7: Subnetwork Starting from The Beatles

To calculate the influence indicator of The Beatles, we need to identify ρ, m, γ for each follower. Due to the simplified partition, there are only 4 possible values for ρ, 2 for m, and 4 for γ. We only need to find out the number of followers that fit in the 32 classes separately. In practice, the measure greatly reduces the difficulty for data querying. The query result is shown in the chart below.

ρ=1.0	1	1	1	1	1	1	1	1																									Sum
ρ=0.8									1	1	1	1	1	1	1	1																	
ρ=0.5																	1	1	1	1	1	1	1	1									
ρ=0.2																									1	1	1	1	1	1	1	1	
m=1.0	1	1	1	1					1	1	1	1					1	1	1	1					1	1	1	1					
m=0.5					1	1	1	1					1	1	1	1					1	1	1	1					1	1	1	1	
γ=1.0	1				1				1				1				1				1				1				1				
γ=1.2		1				1				1				1				1				1				1				1			
γ=1.5			1				1				1				1				1				1				1				1		
γ=1.8				1				1				1				1				1				1				1				1	
followers	34	3	6	0	4	2	0	0	154	20	25	0	24	1	0	0	184	39	8	0	21	4	0	0	60	20	1	0	4	1	0	0	615
multiplier	1	1.2	1.5	1.8	0.5	0.6	0.75	0.9	0.8	0.96	1.2	1.44	0.4	0.48	0.6	0.72	0.5	0.6	0.75	0.9	0.25	0.3	0.38	0.45	0.2	0.24	0.3	0.36	0.1	0.12	0.15	0.18	377.75

Figure 8: Matrix Representation of Query Result

For example, the first column of data shows that among all 615 followers of The Beatles, there are 34 of them who have less than 5 influencers ($\rho = 1.0$), choose the same genre as The Beatles ($m = 1.0$), and start acting 0, 10, or 20 years after The Beatles ($\gamma = 1.0$). As a result, the influence indicator of The Beatles is

$$I = \sum \rho \cdot m \cdot \gamma \cdot \text{ follower } + 0.08 \cdot \text{ followers of followers } = 377.75 + 0.08 \times 2783 = 500.87 \text{ (followers)} \tag{6}$$

3.3 Model for Evaluating Similarities

With the musical features provided in the data set, we develop a model to evaluate the similarities between different songs, artists, and years.

3.3.1 Characteristic dimension reduction based principal component analysis

The data set has a high dimensional indicator. In order to establish the similarity model, principal component analysis (PCA) will be used to reduce the dimension of characteristic information.

Principal component analysis is a statistical analysis method that turns original variables into a few comprehensive indicators. That is to use fewer comprehensive indicators to replace the original more variable indicators and make these fewer comprehensive indicators can reflect the information reflected by the original more indicators as much as possible, at the same time they are independent of each other. To apply PCA, we did the following steps:

I. Standardization of raw data matrices

The raw data dimensions are as follows: 'danceability', 'energy', 'valence','tempo','loudness','mode' 'key', 'acousticness' 'instrumentalness', 'liveness','speechiness', 'duration'.

Let the original variables be X_1, X_2...X_{13}. The 5602 observations' data matrix of X is as follows:

$$X = \begin{Bmatrix} x_1 1 & x_1 2 & \cdots & x_1 p \\ x_2 1 & x_2 2 & \cdots & x_2 p \\ \vdots & \vdots & \ddots & \vdots \\ x_m 1 & x_2 & \cdots & x_m p \end{Bmatrix}$$

II. Centrally standardization of the data matrix

According to columns(Mean becomes 0, the variance turns to 1), the data matrix is centrally standardized. For convenience, the normalized data matrix is still written as X.

$$\mu = \frac{1}{m} \sum_{i=1}^{m} x_{ik} \tag{7}$$

$$\sigma_k^2 = \frac{1}{m-1} \sum_{i=1}^{m} (x_{ik} - \mu_k)^2 \tag{8}$$

$$x_{ik} = \frac{x_{ik} - \mu_k}{\sigma_k} \qquad i = 1, 2...m, k = 1, 2..p \tag{9}$$

Where i represents the artist number, k represents the musical characteristic number, and x_{ik} represents the score of the i^{th} artist's k^{th} musical characteristic. μ represents mean and σ_k represents the standard deviation of the k the musical characteristic of the total m artist.

III. Calculation of correlation coefficient

After eliminating the influence of dimension on the evaluation result, and obtaining the standardized matrix Mb. The correlation coefficient matrix R of the sample matrix X was calculated (see Figure.9).

$$R(i,j) = \frac{cov(i,j)}{\sqrt{cov(i,i)cov(j,j))}} \tag{10}$$

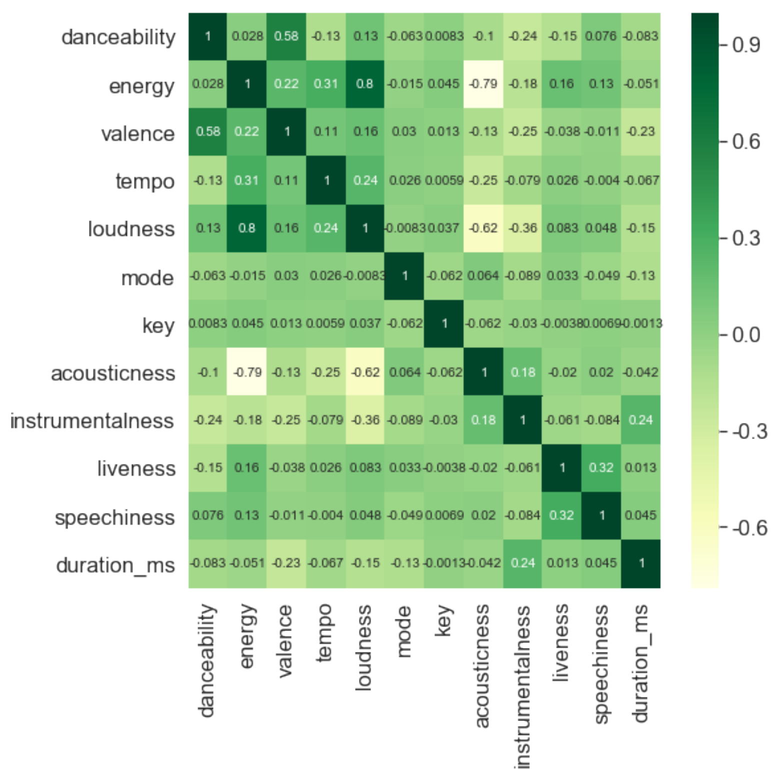


Figure 9: The Correlation Coefficient Matrix R of the Matrix X

IV. Solvent of characteristic equation

For the correlation coefficient matrix R, using the Jacobian method to solve the characteristic equation's p non-negative eigenvalues $\lambda_1, \lambda_2 ... \lambda_k$.

V. Calculation of principal component matrix

Select W principal components so that the variance of the preceding W principal components and their proportion to the total variance are

$$\eta = \frac{\sum_{i=1}^{w} \lambda_i}{\sum_{i=1}^{p} \lambda_i} \tag{11}$$

and the selected W principal components retain the information of the original P features as much as possible.The resulting principal component matrix is denoted as M.

Principle component	Variance contribution rate	Cumulative contribution rate r%
1	0.27295739	0.27295739
2	0.24371001	0.5166674
3	0.18198041	0.69864781
4	0.11493063	0.81357843
5	0.07401302	0.88759145
6	0.03631543	0.92390688
7	0.02479988	0.94870676
8	0.01869345	0.96740021
9	0.01546535	0.98286556
10	0.0091558	0.99202137
11	0.00441451	0.99643587
12	0.00356413	1

Table 1: Variance Contribution of each Principal Component

The above table is the variance contribution table of the principal component analysis of music genres. It can be seen that the cumulative contribution rate r of the first five principal components reaches 88.75%, that is, the information contained by the first five principal components accounts for 88.74% of the total information, so the first five principal components can be selected. The coefficients of each variable in the first five principal components are shown in the table below.

	0	1	2	3	4
0	0.042263	-0.075016	0.019724	-0.378058	0.321421
1	0.162628	-0.489949	0.106265	0.156550	0.062333
2	0.023868	-0.183983	0.029365	-0.558069	0.626971
3	0.018373	-0.089029	0.019834	0.031162	-0.002074
4	0.060111	-0.195528	0.040114	-0.005788	-0.063258
5	-0.924028	-0.336754	-0.136664	0.091857	0.053975
6	0.193037	-0.127860	-0.972445	0.017354	0.018190
7	-0.275053	0.685966	-0.144958	-0.285553	-0.018155
8	0.009297	0.269971	-0.008277	0.652284	0.694064
9	-0.004326	-0.028369	0.005587	0.029099	-0.095660
10	0.011723	-0.006898	0.002701	-0.020002	-0.035165
11	0.016042	0.016664	0.002677	0.055835	0.004370

Figure 10: The Coefficients Between Principal Components and Characteristics

3.3.2 Similarity evaluation based on cosine similarity

After dimensionality reduction, the first five principal components were selected to calculate the cosine similarity S of all artists.

We calculated cosine similarity S based on the first five principle component of the model to construct the similarity model between different observations. Figure.11 shows S between random 10 artists based on the first 5 principal components.

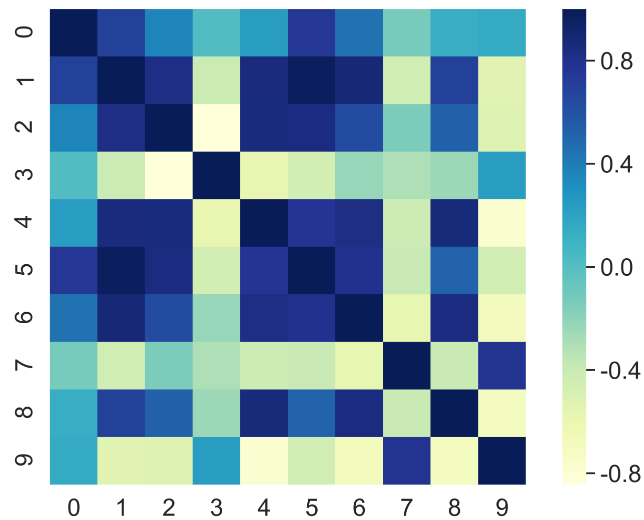


Figure 11: Cosine Similarity Between 10 Random Artists

From the Principal component analysis visualization, we can see that artists within the same genre has similar principal components score which indicates that artists within a genre are more similar than artists crossing genres, especially when the genre has some distinguishable characteristics (see Figure.12).

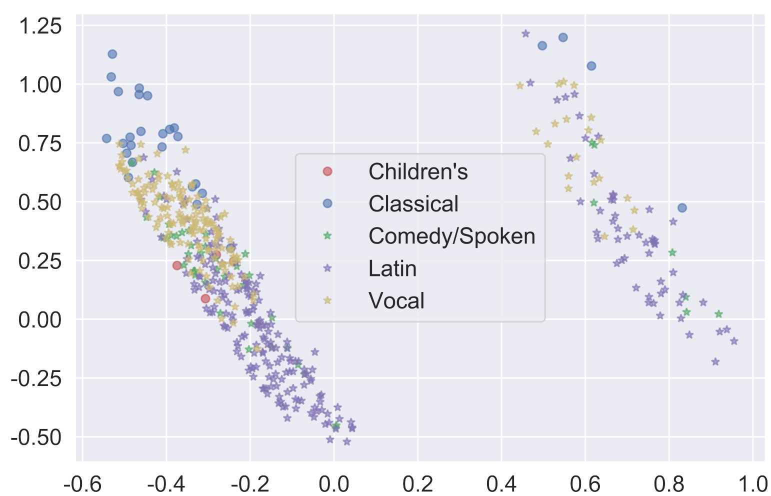


Figure 12: Principal Component Analysis Visualization

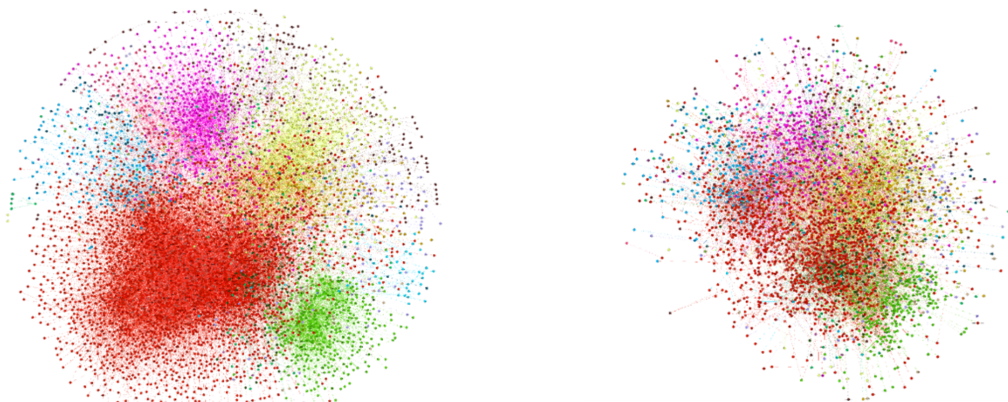
3.4 Comparison in the unit of Genre

Based on the network of influence and model of evaluating similarity S, comparisons in the unit of the genre of music are made.

3.4.1 Difference in influence/similarities within and between genres

Based on the results from section 3.3.2, a conclusion can be made that within the same genre, similarities are higher than outside the genre. Similarly, we can also show that that the influence within the genre is much stronger.

As explained in Section 3.2.1, nodes of the same color cluster, which means the influence is mainly within genres. However, penetrations of different colors also indicate influences between genres. To investigate the topic deeper, we can filter the network and only show nodes with edges within and genres (see Figure.13).



(a) Influence within genres

(b) Influence between genres

Figure 13: Comparison of Influences Between/Within Genres

In Figure.11(a), there are 5249 nodes and 32735 edges, which means that there are 5249 (93.7%) artists who have influenced or been influenced by artists within the genre, and the within-genre influence takes up 76.57%. Similarly, Figure.11(b) shows that there are 3285 (58.64%) artists who have influenced or been influenced by artists outside their own genre, and the across-genre influence takes up 23.43%.

As a result, the classification of what genre a song or an artist belongs to is important for the understanding of its musical development. Then, the problem comes out that what characteristics distinguish a genre?

3.4.2 Features distinguish a genre

To determine the main features that distinguish a genre, the relationships between different features and genre of an artist is needed. Traditional ways like Pearson product-moment correlation coefficient can hardly fit the model due to the abnormally distributed data. As a result, we adopt the Random Forest algorithm to measure the contribution of each feature for the prediction of an artist's genre to mark whether they distinguish a genre.

I. Random Forest algorithm

Random Forests an ensemble learning method for classification, regression, and other tasks that operate by constructing a multitude of decision trees at training time. During the process of prediction, the built-in function in Random Forest can rank the importance of the features W by their contributions to the prediction. As a result, the features with the highest contribution can best distinguish a genre. To apply the algorithm, quantified music features of an artist used for prediction.

II. Result

The prediction gains an accuracy of 77%. Due to the diversity of genres, the result can be viewed as reliable. The five most important features and their feature importance W are shown in Figure.14:

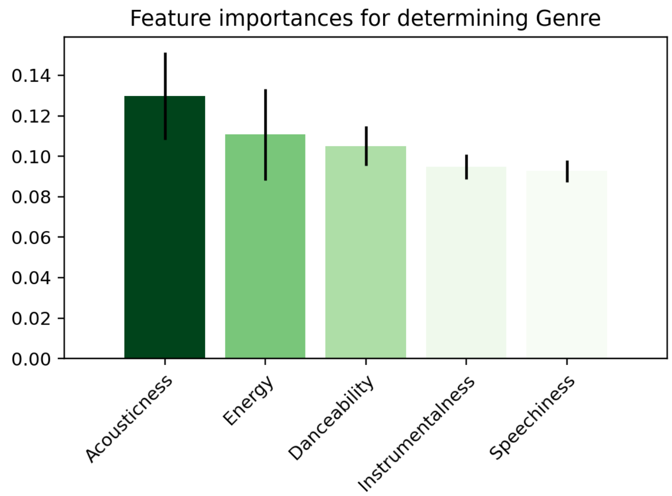


Figure 14: Five most Important Features Determining the Genre

As we can see from the graph, the features best distinguish a genre are: 'Acousticness', 'Energy','Danceability','Instrumentalness', and 'Speechiness', and these scores (highest $W = 0.129$) are apparently higher than the others (averagely $W = 0.063$).The result well corresponds to the reality:

- Acousticness describes the type of vocals. It shows whether the track is acoustic (without technology enhancements or electrical amplification). For music types like Classical or Country music, a large part of them has high Acousticness. However, for music types like Electronic, electrical amplification is a must, so it may well distinguish genres.

- Energy describes the characteristic of music. It represents a perception of intensity and activity. For example, Pop/Rock has higher energy, while Religious and Country may have small scales of this, so it may well distinguish genres.

- Danceability describes the characteristic of music. It is a measure of how suitable a track is for dancing. Music types like Jazz and Latin are suitable for dance, while music types of Religious and Classical can hardly be danced with, so it may well distinguish genres.

- Instrumentalness describes the type of vocals. It predicts whether a track contains no vocals. For music types like Folk and Comedy/Spoken hardly contain instruments, while nearly all Classical music contains, so it may well distinguish genres.

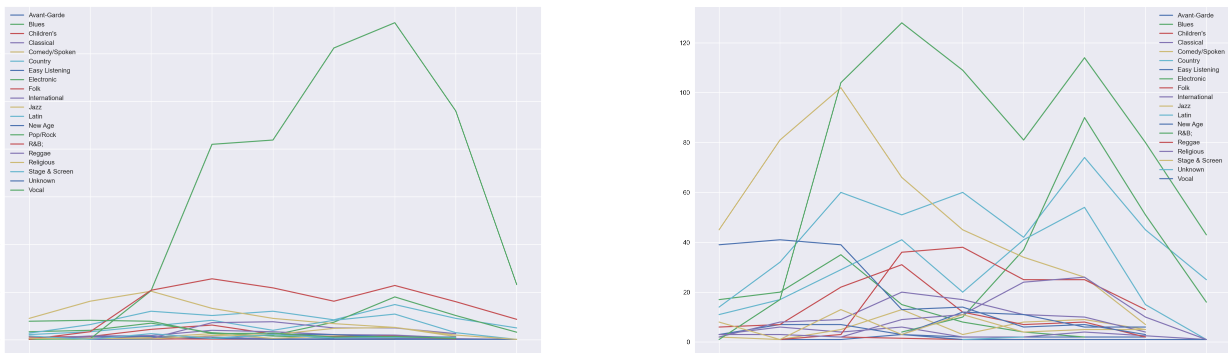
- Speechiness describes the type of vocals. It detects the presence of spoken words in a track. For example, Classical music may hardly contain spoken words, while Pop/Rock usually has it, so it may well distinguish genres.

To conclude, the five factors ('Acousticness', 'Energy','Danceability','Instrumentalness', and 'Speechiness') well, distinguish the genre.

3.4.3 Development of genres over time

After analyses on what distinguishes genre, findings also come that a genre changes over time, which is important for understanding the development of music.

To quantify the change of genre, we use the increasing number of each genre per year (Δ) as an indicator and compare this indicator and the time to show the changes. In Figure.15, due to Pop/Rock music's much higher increase, a subfigure shows the changes without considering Pop/Rock music.



(a) With Pop/Rock music (b) Without Pop/Rock music

Figure 15: Increasing Number in each Genre per year

As we can see in Figure.15, two major revolutions can be concluded:

- For nearly all genres, the increase in 1970s is much slower than in 1960s
  Based on the analyses in section 3.2, part of the reason may be that the leading figure at that time, The Beatles', activeness start in 1962 and disbanded in 1970, which shadowed negative influence in 1970s' music field[8].

- In the 1990s, genres including Pop/Rock, RB, Blues, Jazz, and Country reach their highest increasing speed
  The 1990s is a special period that marks the expansion of the Internet and information revolution, which provides various media for the popularity of these genres[9].

More detailed analyses combining the genre changes over time and its determining characteristics can be found in section 3.6.

3.4.4 Relationship between genres

Although the features of similarities and influence networks mainly showed the consistency inside genres, from the similar changing track and similarity analyses, we can find that there may exist a relationship between two different genres. These relationships can be concluded into three kinds by their causation:

I. Fusion and development: Blues, R&B, and Pop/Rock

Followers in the genre of Pop/Rock showed high similarity with their influencers in the genre of R&B and Blues.

Blues, R&B, and Pop/Rock all showed Black underclass culture. Pop/Rock combined Rap into R&B,

which makes this style more powerful and casual. Similarly, Blues can be viewed as a predecessor of R&B. Comparing with Blues, R&B simplified the instrumentality of pure Blues and focus more on the singer[10].

II. Simultaneously stimulated by the war background: Pop/Rock and Folk music

Pop/Rock and Folk music both showed a rapid increase after the 1960s.

From 1955-1975 in the US, during the Vietnam War, the 'Anti-war music', represented by Pop/Rock and Country music, gain its popularity. For example, the Folk singer *Joan Baez* write in her song: "How many dead men will it take to build a dike that will not break". On the other hand, the economic depression caused by the Vietnam War set the stage for the explosive growth of the counterculture, which also fueled the growth of Pop/Rock music [11].

III. Simultaneously supported by technology development: R&B and Electronic

R&B and Electronic music both showed a rapid increase from the 1970s.

Early in the 1970s, artists started to find that the new timbre development is restricted by the limited physical material, so they started to turn to the newly emergent electronic technology for help. By the development of editing and recording technology, like the invention of tape, electric guitar, and music synthesizer, Electronic, and R&B music attract more attention [12].

3.5 Inherent Correlation of Similarities and Influence Network

After building the network of influence as well as the model for measuring similarities, we also find the inherent correlation of these two models.

3.5.1 Check of consistency of similarities and influence connection

Artists influenced by one another do share more characteristics in their music.

To investigate this problem, we calculated the similarity between all the 5,560 artists, using the measure introduced in section 3.3.2, and thus attain over three million pieces of data. In the matrix $\begin{pmatrix} 1 & \cdots & a_{1,5560} \\ \vdots & \ddots & \vdots \\ a_{5560,1} & \cdots & 1 \end{pmatrix}$, A_{ij} stands for the similarity between the i^{th} artist and the j^{th} artist.

Then we pick out the similarity data for artists with influence relationship and compare them with that in the whole data set. Due to the feature of cosine similarity S, we only consider those values close to a positive one as the interpretation of “being similar”. To avoid the impact of negative values which implies negative similarity, we calculate the proportion of large positive values P (see Figure.16).

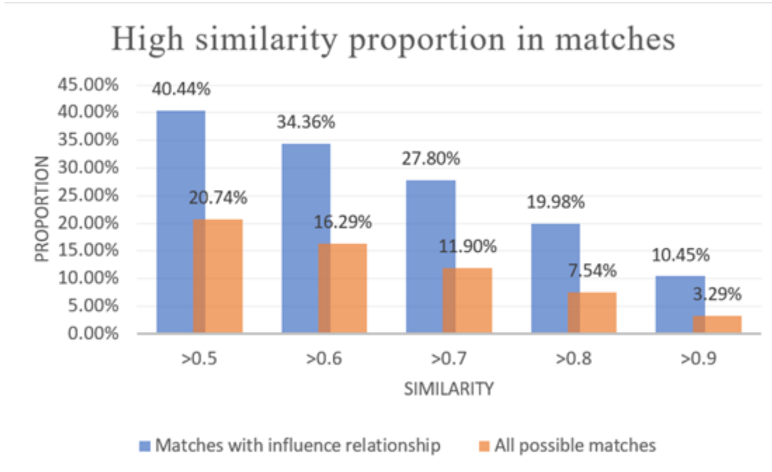


Figure 16: High Similarity Proportion in Matches

From the data, we clearly see that artists with influence relationships share more characteristics in music. Influencers do affect followers.

3.5.2 Most contagious characteristics

Based on this fact, we want to find out whether there are characteristics more contagious than others.

Whether the followers imitate certain musical characteristics of influencers? The problem turns to find out the similarity of each characteristic between the influencer and the follower. We pick out all influencers who have more than 20 followers with high similarity ($S > 0.8$). We apply the filter because these followers are more likely to imitate their influencers' musical characteristics. Then for each characteristic, we calculate the average deviations (δ^2) between the follower and his corresponding influencer with the following formula:

$$\delta^2 = \frac{1}{n} \sum\nolimits_{i \in influencers} \sum\nolimits_{f \in followers\ of\ i} (c_f - c_i)^2 \tag{12}$$

where c is the value representing the characteristic and n is the total number of followers of all influencers.

The calculation gives result as the Table.2 below:

Danceability	Energy	Valence	Tempo	Loudness
0.035896	0.028694	0.051050	0.045759	0.069324
Acousticness	**Instrumentalness**	**Speechiness**	**Liveness**	**Duration**
0.019481	0.181011	0.122464	0.106668	0.078727

Table 2: Average Deviations (δ^2)

Since δ^2 is the sum of squares of deviations, the smaller it is, the more similarity is in the characteristic between the influencer and the follower. From the chart, we attain that the acousticness in followers' music is very similar to that in the influencers' music. Followers tend to imitate influencers' music in this aspect. In other words, acousticness is more contagious than other characteristics.

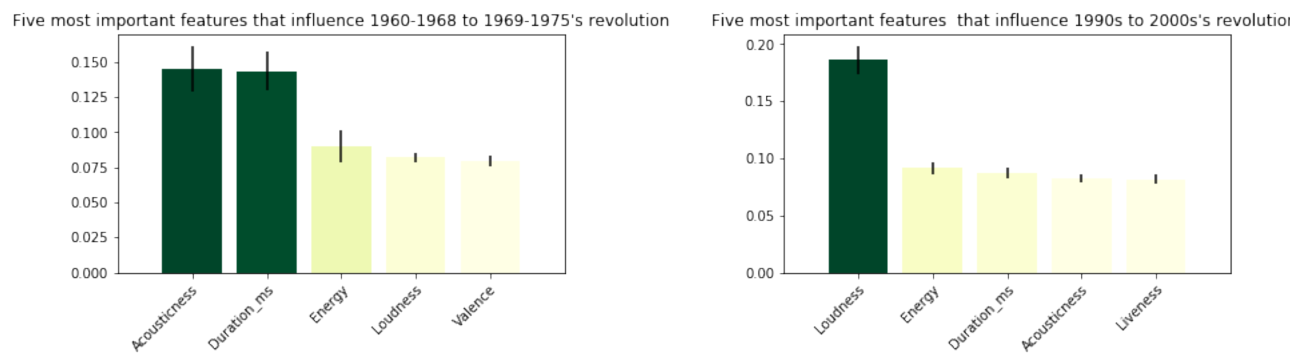
The result also corresponds with that in 3.4.2, which argues that acousticness is the most distinguishing feature between genres. It is sensible because most of the influence is within the genre, and acousticness is more contagious than other musical characteristics, so the acousticness within one genre would be inevitably more convergent.

3.6 Revolutions in the Musical Filed

Based on the data, major leaps in musical evolution represent revolutions in the musical field. Here we want to find that what Characteristics can signify the revolutions and what artists can represent them.

3.6.1 Characteristics signify the revolutions

To quantify the influence of each characteristic on the revolutions, we adopt a random forest algorithm to measure the contribution of each characteristic W to predict the period. If a characteristic attribute to the prediction far greater than others, this characteristic will signify the major leaps in musical evolution. Figure.17 shows W of the five most important characteristics.



(a) (b)

Figure 17: Characteristic that Signify the Revolution

As can be seen from the graph, Acousticness features between 1965 and 1970 are distinguishable. After the 1970s, music styles became more diverse, moving towards avant-garde and extreme, manifested in improvisation, emphasis on the use of electronic means, emphasis on instrument and effect manipulation technology, and increasingly obvious commercialization tendency [13].

After 1990s progressive Rock prosperity quickly rise again, the establishment of Musea in France, Cyclops in the UK, Magna Carta in the United States, Mellow in Italy, Brazil's Rock Symphony, and Mylodon in Chile provides a good stage for pre-rock bands all over the world to show themselves [14]. At the same time, some artists are dedicated to re-releasing the best works of progressive rock's golden age and declining era [14]. As for the 2000s, Pop singers are more inclined to lyrical music, hip-hop, and rap music, so that the loudness can be a main characteristic the signify the revolution [15].

3.6.2 Artists represent the revolutions

To determine whether an artist represents revolution, first, we filter the data set to pick out the artist who has more than 70 new-start followers in the revolution period.

Follower count C	Infulencer Name	Follower Debut Year
195	The Beatles	1960s
115	Bob Dylan	1960s
106	The Rolling Stones	1960s
72	Nirvana	1990s

Table 3: Artists Influence the Pop Revolution

To make further analysis, we take The Beatles in the 1960s as an example, In the 1960s, 195 artists declare that they create music under the Beatles' inspiration. Among them, 79 artists passed our similarity model that their music composition's characteristics are highly similar with the Beatles's which verify that the Beatles indeed impact on 1960s' revolution. To quantify each characteristic's influence, we proceed with calculating the average squares of deviations of each characteristic and find the top 3 characteristics that were highly shared with Beatles and its followers: Energy, Valence, and Loudness. This three characteristic is consistent by the typical features of revolutionary music type psychedelic rock music and hard rock in the 1960s.

3.7 Example: Pop/Rock Music in the Revolution

After generally analyzing the revolutions in the musical field, here we have an example of Pop/Rock music in the revolution of 1960-1975 and 1990-2010. To indicate the change of Pop/Rock music over the revolution, we compare the features of music by three main influencers before and after the revolution. The three main influencers are determined by the number of followers they directly influenced during the revolution period.

3.7.1 Dynamic influencers of Pop/Rock in the 1960s and 1970s

In the revolution from the 1960s to 1970s, Pop/Rock music experienced rapid development and tend to have more energy

After ranking the influencers' number of followers whose activities started in the 1960s and 1970s, the three most important influencers in the 1960s and 1970s are shown in the Table.4:

Time	Most influential	2^{nd} most influential	3^{nd} most influential
1960-1968	The Beatles	Bob Dylan	The Rolling Stones
1969-1975	The Beatles	The Rolling Stones	Sex Pistols

Table 4: Three most Important Influencers in the 1960s and 1970s

From Table.4, we can find that the most influential influencers are dynamic, which partly represent the revolution in this genre, and these three influencers' work (including choral songs and re-edited songs) are believed to be representative of the period. The two time period shared two same influencers, so to further compare the differences of features, a radar map of these three main influencers' features of music is shown in Figure.18

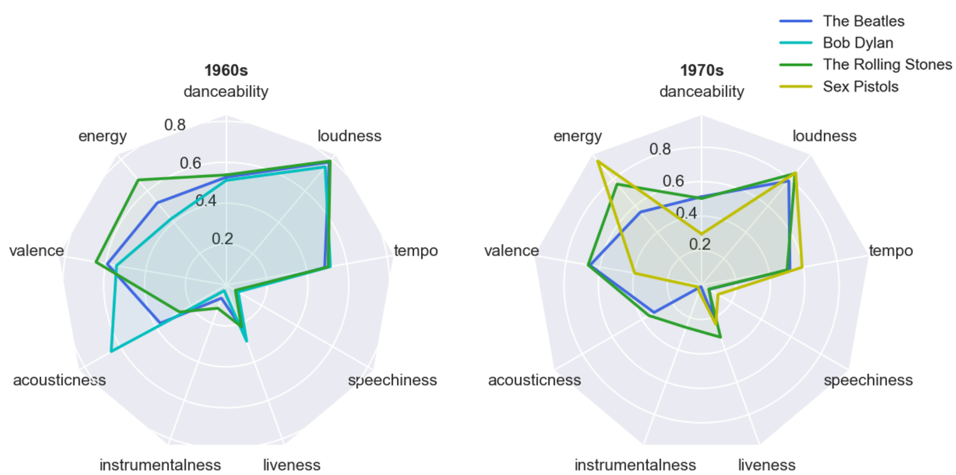


Figure 18: Five most Important Features Determining Genres in the 1960s and 1970s

From Figure.18, we can find that the features of one influencer's work are also dynamic, which represent the revolution in this genre. For example, The Beatles' work becomes more acoustic after the revolution. On the other hand, the popularity of Sex Pistols suggests that artists in Rock/Pop genre tend to prefer music with more energy, which represents the perception of intensity and activity. This tendency may result from the needs in the wartime and Pop/Rock music's combination with Blues at the beginning of the 1970s [8, 10].

3.7.2 Dynamic influencers of Pop/Rock in the 1990s and 2000s

In the revolution from the 1990s to 2000s, Pop/Rock music experienced even more rapid development and tend to have more energy and valence.

Similarly, the three most important influencers in the 1990s and 2000s are shown in the Table.5:

Time	Most influential	2^{nd} most influential	3^{nd} most influential
1990-2000	Nirvana	The Beatles	Neil Young
2000-2010	The Beatles	Bob Dylan	David Bowie

Table 5: Three most Important Influencers in the 1990s and 2000s

From Table.5, we can also find that the most influential influencers are dynamic. The newly emerged artist David Bowie once said 'The impact that the Internet is about to have on society is unimaginably large and both good and bad, and the Rock music has come of age to joining the new internet world', so his joining represents Pop/Rock music finds its way to enter the online social network [16]. To further compare the differences of features, a radar map is shown in Figure.19

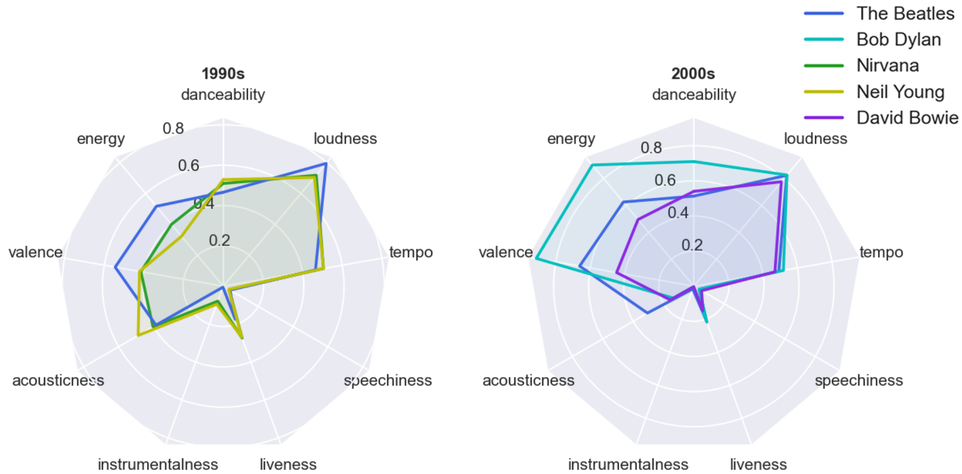


Figure 19: Five most Important Features Determining Genres in the 1990s and 2000s

From Figure.19, we can find that Pop/Rock music tend to gain more energy and valence. This tendency may result from the development of the internet, which result in more chase for cheerfulness via music [12].

3.8 Effects of Social, Political or Technological Changes on the Model

Our work also shows information about cultural influence of music in time or circumstances: how the influence indicator model and the data network help investigate such topics, and the impact of society, technology, and politics on music can also be drawn from the data network.

Music as a cultural form must express some ideas and aesthetic orientation. Any popular music inevitably expresses the mainstream thinking or mainstream aspirations of an era so that many fans like or follow it. The listener empathizes with what the music is trying to say, and the music gives vent to their emotions or strongly conveys value judgments. In other words, music can somewhat represent an era. In the network, we can take a peek at such society representation by looking into the most influential artists in a time, using the influence indicator model and other querying methods introduced in 3.6 to analyze musical revolutions.

There are several examples. WWII dealt a huge blow to the emerging supernation, America. In the aftermath of the war, however, what arrived was the Cold War and Vietnam Wat, instead of the peace in expectation. In the 1960s, the peace movement prevails in America. Under such circumstances, Bob Dylan composed a large number of anti-war songs and gained great popularity, such as Blowin' in The Wind and Knockin' on Heaven's Door (with the popularity being 73). Similarly, the punk band Sex Pistols with 153 followers, who led the rise of punk music, also represents the mainstream of thinking in 1970s England. Their album "There Is No Future" almost became the appealing slogan of all punk people. Sex Pistols expressed the confusion, anger, and disappointment in governments for teenagers at that time, who were a generation of disenfranchised youth amidst the declining economic situation.

However, music is not only a recorder of the times but also a participant and promoter of the times.

Still using the examples drawn above, the anti-war music composed by Bob Dylan was widely used in peace movements and he was considered one of the most iconic protest musicians and a political dissenter [17]. The folk band Peter, Paul, and Mary who was famous for covering Dylan's songs also involved in many political movements and greatly advocated those antiwar movements.

Sex Pistols became the icon for Anarchism. As the critics said, Sex Pistols were defined by ambitions that went well beyond the musical. Greil Marcus directly put it that "in the mind of their chief terrorist and propagandist, anarchist veteran...and Situational artist McLaren, the Sex Pistols were meant to be a force that would set the world on its ear...and finally unite music and politics" [18].

Speaking of social influence, The Beatles can never be ignored. As a musical and cultural revolutionary, their cultural influence has long exceeded that of their art itself. Their anti-upper-class, anti-artistic style songs, and their challenging demeanor on stage have conquered young people from all over the world. John Lennon stated that, "the 60s saw a revolution ... in a whole way of thinking. The Beatles were part of the revolution, which is really an evolution and is continuing. We were all on this ship – a ship going to discover the New World. And the Beatles were in the crow's nest" [19]. historian Michael James Roberts also argued that the band represented "cultural change and the oppositional stance of the youth culture against the establishment" [20].

On the other hand, changes in society, technology, and politics also affect music. As explained in 3.4.4, electronic music rose due to the development of technology. For the political effects, folk music, which is more peaceful, flourished when antiwar movements prevailed. In the 1970s, rock music took off because the sluggish economy made the public prefer more intense and extreme expressions.

The rapid development of the Internet in recent times also has a great impact on the spread of music. As drawn from the full music data, almost all songs with the highest popularity are released after 2010. Specifically, among all the 203 songs with a popularity greater than 80, only 29 of them are released before 2010. Music is more easily and widely spread, which makes the cultural influence of music significant.

4 Evaluation and Promotion

4.1 Strengths

Our model gains many strengths:

- Well processed data of large amount. The database is large and contains various data type and some unreasonable values, so we combined pandas in Python and SQL, efficiently finishing the data mining;
- Simplified but fully considered the model for evaluation of influence indicator (I). When evaluating I, we used 4 well-organized factors instead of iterating all nodes, which is wasteful.;
- Clear visualization. Combining Gephi and Matplotlib in Python, graphs containing bountiful information are used for visualization;
- Well supported result by previous research and facts. Nearly all results generated from data analyses can be supported by literature review;
- Potential for generalization. Our model can be popularized in many other fields for evaluation of influence networks.

4.2 Weaknesses

Our model may have the following Weaknesses:

- Subjectively setting coefficient in the evaluation of influence factor (I). Though the coefficients setting best satisfied our model, little literatures can be found to support it. In future researches, other coefficient-setting methods can be tried and compared.
- In the database provided, the influence of the Beatles is much higher than all the other artists in many dimensions, suggesting the potential existence of bias. In future research, more data can be combined from different sources to generate more plausible conclusions.

5 A Letter to the ICM Society

Dear ICM Soceity:

According to your requirements, we analyze the influence of the music through the network. We first form the music influence indicator model and the characteristic similarity model to quantify the influence measurement and similarity measurement. Based on the previous two models, we check the consistency of the influence and similarity index between artists.

For the music influence indicator model, we get a simplified but fully considered model for evaluation of influence indicator. we consider four factors to evaluate the quality of the influencer (the number of influencers the follower claims to have; whether the follower chooses the same genre with the influencer; the time gap between the influencer and the follower to become active; The influence of the follower) instead of iterating all nodes. We assign each factor a parameter to represent and construct the mathematical model for measuring the influence indicator. We take the Beatles as an example and use our model to give a quantified index on its influence indicator.

To quantify the similarity between two artists, we construct the second model. We use the principal component analysis to reduce the dimension of characteristic information. We select 5 principal components that can have an 88.9 % cumulative contribution rate which can retain nearly 90 % percent of the original information. We calculated cosine similarity S based on the first five principal components of the model to construct the similarity model between different observations and find that artists within a genre are more similar than artists crossing genres, especially when the genre has some distinguishable characteristics.

Based on the previous model, we pick out the similarity data for artists with an influential relationship, calculate the proportion of similar pairs and compare it with the proportion of similar pairs in unconnected artist groups. We find the inherent correlation of similarities and influence networks. Artists with the influential network are much more likely to be similar to each other which is a well-supported result by previous research and facts.

In our model and solution, we always take the Beatles and the decade 1960s to give an example since the influence of the Beatles is much higher than all the other artists in many dimensions and the main genre pop/rock get its booming development in 1960s. If we can get reach of the richer data, our model can be applied to each genre, and parameters in the influence model can be better selected to eliminate bias. The current data set is restricted to the artists that existed in both data sets. If we get larger data set with artists from all kinds of genres and years, we can obtain more normally distributed data which reduces the defect of the principal component analysis algorithm to some extent.

Music, as part of the art, is a form of social consciousness that reflects social life vividly and concretely. In music education, one of the basic requirements for creators is to integrate their own feelings, experiences, or thoughts into music works. However, no matter how it is feeling, experience or thinking, what it reveals is the phenomenon of society and the culture, and the historical development also fully proves this point. The influence relationship between musical artists can also be seen as the culture transmission. For further research, we recommend you to research the relationship between music and social mainstream cultural preferences and how artist value can be converted into cultural influence.

6 Conclusions

To conclude, based on the data of features of music and influence relationship among artists from 1930s to 2010s, our team builds a model to measure the influence of previously produced music on new music and musical artists. Our model mainly contains two parts: a network to show the reported connection between influencers and followers and a quantitative model to evaluate the similarities between two artists' work. After quantitative analyses, we determined the consistency of these two models. Then we use these two basic models to explore different music genres and musical evolution/revolutions.

References

[1] Schäfer, T., Sedlmeier, P., Städtler, C., & Huron, D. (2013). The psychological functions of music listening. Frontiers in Psychology, 4, 511–511.

[2] Nikolsky, A., & Perlovsky, L. (2020). Editorial: The Evolution of Music. Frontiers in Psychology, 11, 595517–595517.

[3] Tekelioğlu, O. (2001). Modernizing reforms and Turkish music in the 1930s. Turkish Studies, 2(1), 93-108.

[4] Samson, J. (2001). Genre. Oxford University Press. https://doi.org/10.1093/ gmo/9781561592630 .article.40599

[5] Laurie, Timothy (2014). "Music Genre as Method". Cultural Studies Review. 20 (2). doi:10.5130/csr.v20i2.4149.

[6] "Musical genres are out of date – but this new system explains why you might like both jazz and hip hop". Econotimes. August 3, 2016.

[7] Kobourov, Stephen G. (2012), Spring Embedders and Force-Directed Graph Drawing Algorithms

[8] Lewisohn, M. (2013). The Beatles: All These Years (Vol. 1). Three Rivers Press.

[9] Robertson, D. S. (1990). The information revolution. Communication Research, 17(2), 235–254.

[10] Ripani, R. J. (2006). The new blue music: changes in rhythm & blues, 1950-1999. Univ. Press of Mississippi.

[11] Regev, M. (2013). Pop-rock music: Aesthetic cosmopolitanism in late modernity. John Wiley.

[12] Collins, N., Collins, N., Schedel, M., & Wilson, S. (2013). Electronic music. Cambridge University Press.

[13] Mullen, J. (2018). What has popular music ever done for us? pleasure, identity and role play in UK pop music in the 1970s. Études Anglaises, 71(1), 14-28,122.

[14] Hawkins, S. (1996). Perspectives in Popular Musicology: Music, Lennox, and Meaning in 1990s Pop. Popular Music, 15(1), 17-36. Retrieved from http://www.jstor.org/stable/931202

[15] Seago, A. (2000). "Where Hamburgers Sizzle on an Open Grill Night and Day"(?): Global Pop Music and Americanization in the Year 2000. American Studies, 41(2/3), 119-136. Retrieved from http://www.jstor.org/stable/40643233

[16] Buckley, D. (2012). Strange fascination: David Bowie: the definitive story. Random House.

[17] Kraaijvanger. (2017). Bob Dylan as a Political Dissenter

[18] Hatch, David, and Stephen Millward, From Blues to Rock, p.170.

[19] Nash, Alanna (2014)."The Beatles: 7 Ways the Fab Four Changed America".The Fiscal Times.

[20] Roberts, Michael James (2014).Tell Tchaikovsky the News: Rock 'n' Roll, the Labor Question, and the Musicians' Union, 1942–1968. Durham, NC: Duke University Press.ISBN978-0-8223-5475-8.

7 Appendix

The source of code could be found at https://github.com/alphabet-he/ICM2021.