Summary

Music has always been an important component of human society, and it is constantly innovating with the development of the times. Musicians are often influenced by the previously produced music when they are composing, resulting in the creation of new music characteristics or even multi-genre music style. Our goal is to measure and quantify the musical influence, so as to better understanding the changes of music genres.

In Problem 1, we analyze music influence between influencers and followers and then we establish a **Grey Comprehensive Evaluation Model** which is able to measure musical influence.

In Problem 2, we use principal component analysis to reduce the dimension of music characteristics, and use cosine similarity method to establish a **Music Similarity Measurement Model**, and successfully quantity the similarities between musicians.

In Problem 3, we do further data mining to analyze the characteristics of all kinds of music genres, and succeed in finding the differences and similarities between and within music genres. And through the analysis of the characteristics of music genres over time, we can find out the changes of music genres.

In Problem 4, with the application of our music similarity measurement model, we can find it clear that influencers do affect the music created by the followers. Furthermore, We comprehend Contagiousness from two perspectives and confirm that certain musical features are more "Contagious" through radar maps.

In Problem 5, through the analysis of music popularity and other music characteristics, we take Pop-Rock as an example and find the revolutionaries in our "eyes".

In Problem 6, Through the analysis of the changes of music features over time, we take Vocal and Pop-Rock as examples to analyze the influence processes of musical evolution that occurred over time.

In Problem 7, in combination with the history of changes in music field, we establish **Social, Political and Technological Influence Model** and analyze the practical influences of music in time or environment.

Keywords: music influence; Grey Comprehensive Evaluation; PCA; cosine similarity

Contents

1 Introduction

1.1 Problem Background

Music has always been an important component of human society, and it is constantly innovating with the development of The Times. The creation of musicians depends not only on the creativity of the musicians themselves, but also on external factors such as social environment, new musical instrument, previously produced music, personal experience and so on. From the perspective of previously produced music, a musician's creation may be influenced by several other musicians in different musical genres. This kind of diversified and multi-genre art blend often breeds new music styles. When new musical styles are constantly produced and gradually become popular, together with the emergence of outstanding musical revolutionaries or social changes, new musical genres may gradually form.

The influence of music plays an important role in the study of the birth and evolution of musical genres, and this influence can be measured by the similarity of all kinds of characteristics of different songs. In order to better understand how music evolves over time, we need to build models to systematically and quantitatively study the influence of music.

1.2 Restatement of the Problem

Considering the background information and restricted conditions identified in the problem statement, we need to complete the following work:

- **Problem1:**
  We need to create a directed network of musical influence using the *influence_data* data set, in which influencers are connected to followers and develop parameters that capture 'music influence' in this network.
  Also, we need to create a subnetwork of our directed influencer network to explore the function of 'music influence'.

- **Problem2:**
  We need to develop measures of music similarity by analyzing *full_music_data* data sets of music characteristics and explore whether artists within genre are more similar than artists between genres.

- **Problem3:**
  We need to compare similarities and influences between and within genres and explore the differences and correlations between genres. Furthermore we need to analyze the evolution of music genres.

- **Problem4:**
  We need to explore the influence degree of influencers and followers, according to the similarity data in the *data_influence* data set. And we need to extract some most 'contagious' characteristics.

- **Problem5:**
  We need to identify the characteristics that might signify revolutions and find out possible revolutionaries in our network.

- **Problem6:**

We need to analyze the processes of musical evolution, and find out key indicators that can help explain the music genre changing process.

- **Problem7:**
  We need to express the influence of music on culture in time or circumstances, or explain how to identify the impact of social, political or technological changes (such as the Internet) in our network.

1.3 Our Work

- We analyze music influence between influencers and followers, then we build Influencers & Followers Network using Grey Comprehensive Evaluation, and successfully quantified the music influence.

- We use principal component analysis to reduce the dimension of music characteristics, and use cosine similarity method to establish the Music Similarity Measurement Model, and successfully quantified similarities between musicians and music genres.

- We do further data mining, solve the 7 problems, and we sum up the relationship between musicians, genres, and music characteristics. We summarize the relationship between musical characteristics and changes in musical genres and identify the revolutionaries.

- In combination with the history of changes in music field, we establish Social, Political and Technological Influence Model and analyze the practical influences of music in time or environment.

2 Assumptions and Justifications

1. **Assumption1: We assume that a musician composes songs that correspond to his genre. For example, a Pop-Rock musician only composes Pop-Rock songs.**
   In real life, a singer may play different genres of songs, but most of his music is in the field he majors in. Since our problems mainly focus on the study of the musical influence between musicians and the reasons for the emergence of new music genres, we can ignore the situation that a musician composes songs in different genres. In the *full_music_data* data set given in this problem, musicians usually have more than one songs, so we decide to assume that all the songs composed by one musician belong to the same music genre.

2. **Assumption2: We assume that two or more singers (bands) belong to the same music genre if they compose songs together.**
   In the *full_music_data* data set given in this problem, there is a certain proportion of songs that were composed by two more musicians. In real life, musicians who collaborate often belong to different genres, such as pop rock and rap. Since our problems mainly focus on the study of the musical influence between musicians and the reasons for the emergence of new music genres, and in our data set, only a tiny number of

songs are collaborated by musicians in different genres, we decide to assume that two or more singers (bands) belong to the same music genre if they compose songs together.

3. **Assumption3: We assume that in a certain period, the more songs a musician published, the more possibility he have to lead the music trend.**
   We think that if the songs have equal popularity index, the band that releases 100 songs will be far more famous than the band that releases one song. So the number of songs is the factor that affects the popularity, and further impact the music trend.

4. **Assumption4: We assume that the originators of a music genre have specially high music influence.**
   According to common sense, in a particular field of music, although the music style may change as time goes by, the general tone of the music genre is determined by the originator of the field, and later generations(followers) only innovate and improve on this basis.

3 Notations

The key mathematical notations used in this paper are listed in Table 1.

Table 1: Notations used in this paper

Symbol	Description
x_i	The length in the x direction
y_i	The length in the y direction
$\cos(\theta)$	Cosine similarity
U_{m*r}	The left singular value matrix
$\sum_{r*r}$	Singular value diagonal matrix
V_{r*n}^T	Right singular value matrix
C_j	Cluster partition
d_{ij}	The distance between the sample and the center of mass
μ_j	The center of mass
$Corr\_Coef[i]$	The correlation coefficient
$Score[i]$	Music Influence score

4 Influencers & Followers Network：

4.1 Grey Comprehensive Evaluation Model

Firstly, there is an *active_ year* column in the *influence_data* data set. According to common sense, in a particular field of music, although the music style may change as time goes by, the general tone of the music genre is determined by the originator of the genre, and later generations(followers) only innovate and improve on this basis. Therefore, we can trace the data to the source and find out the originators in each field.

The results are as follows:

Table2：Numbers of Originators in Different Music Genres

Music Genre	numbers	Music Genre	numbers
Pop/Rock	1117	Comedy/Spoken	53
Electronic	339	Classical	63
Reggae	98	Children's	4
R&B	412	Blues	103
Religious	135	Folk	199
Vocal	206	Easy Listening	38
Jazz	263	Avant-Garde	23
International	138	New Age	90
Latin	251	Stage & Screen	44
Country	297	Unknown	9

According to the results above, the number of musicians directly or indirectly influenced by each music artist can be obtained through reverse tracing, and this parameter will be one of the parameters for the final influence evaluation.

Then, since the number of artists in each genre is inequal, the influence of each musician should be discussed by genre. Moreover, the number of musicians competing in each era is different, which can be clearly seen from the line charts below(shown partly):
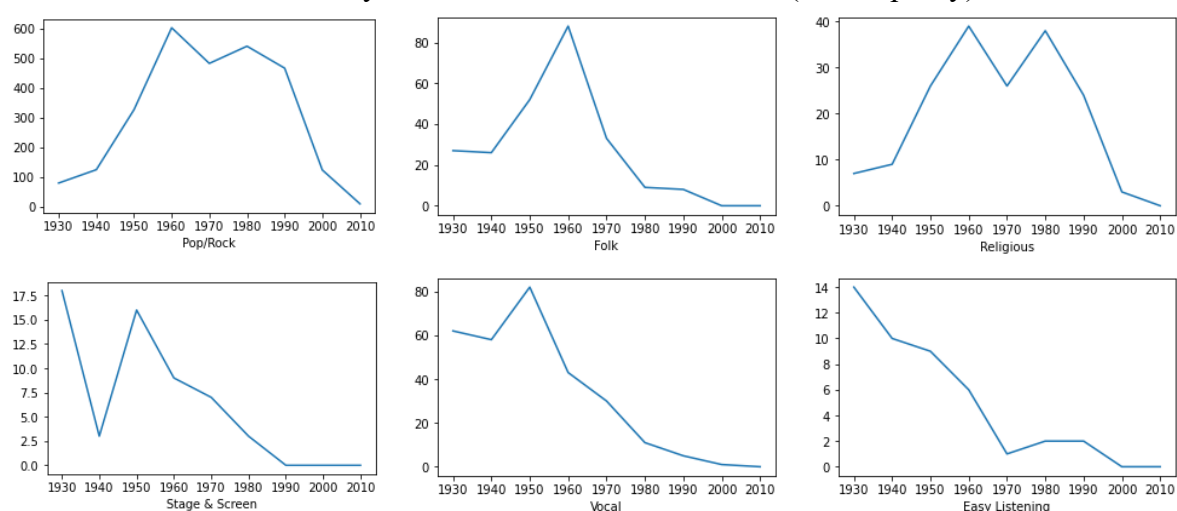
Therefore, the number of artists in the current era and in the current genre should also be discussed when considering their music influence. We used the rank of each musician in his or her era and genre as one of the final parameters for music influence evaluation.

Finally, we use the grey comprehensive evaluation method to evaluate the music influence of each musician. The evaluation parameters are as follows:

Table3: Evaluation Parameters and Descriptions

Parameters	Description
influence_num	Number of people he/she influenced
genre_num	The number of musicians in the genre
genre_year_num	The number of contemporaries in the genre
popularity_rank	Rank of the musician's influence in the genre & era
year_competitiveness	The level of competition in the current era

The calculation method of popularity_rank is as follows: by finding out all musicians in the music genre of the era, then calculate the relative ranking of all the selected musicians. Year_Competitiveness is the total number of musicians in the genre in the era.

4.2 Quantification of Music Influence

We take the ranking of artists in Pop-Rock music in the 1980s as an example, to show the extension of the Influencers & Followers Network.

For all indicators except popularity_rank, the maximum value should be the optimal value. For popularity_rank, the minimum value ,i.e. Rank =1, is the optimal value of the current indicator. Namely:

$$.A_1 = Optim(Parameter)\ for\ Parameters\ in$$
$$\{influence\_num, genre\_num, genre\_year\_num, popularity\_rank, year\_competitiveness\} \quad (1)$$

Table4: Optimal Values of Parameters

Parameters	Value
influence_num	53
genre_num	2751
genre_year_num	540
popularity_rank	1
year_competitiveness	540

Secondly we remove the influence of dimension by dividing the average:

$$Parameter[i] = Parameter[i]\ /\ Average(Parameter) \quad (2)$$

Results are as follows：

	influence_num	genre_num	genre_year_num	popularity_rank	year_competitiveness
0	0	1	1	1.01494	1
0	0	1	1	1.01494	1
0	0	1	1	1.01494	1
0	0.600346	1	1	1.01494	1
0	0	1	1	1.01494	1
...	...	...	...	...	...
0	0	1	1	1.01494	1
0	0	1	1	1.01494	1
0	1.20069	1	1	1.01494	1
0	0	1	1	1.01494	1
0	0	1	1	0.420651	1

Figure2:Result after Dimension Reduction

Then calculate the difference between it and the optimal value to get the maximum and minimum value of the difference to calculate the correlation coefficient. The formula for calculating the correlation coefficient is as follows:

$$Corr\_Coef[i] = (Min + 0.5 \times Max)\ /\ Parameter[i] + 0.5 \times Max \quad (3)$$

Results are as follows：

	influence_num	genre_num	genre_year_num	popularity_rank	year_competitiveness
0	0.986917	0.46796	0.822542	0.336347	0.822542
0	0.986917	0.46796	0.822542	0.336347	0.822542
0	0.986917	0.46796	0.822542	0.336347	0.822542
0	0.98716	0.46796	0.822542	0.336347	0.822542
0	0.986917	0.46796	0.822542	0.336347	0.822542
...	...	...	...	...	...
0	0.986917	0.46796	0.822542	0.336347	0.822542
0	0.986917	0.46796	0.822542	0.336347	0.822542
0	0.987404	0.46796	0.822542	0.336347	0.822542
0	0.986917	0.46796	0.822542	0.336347	0.822542
0	0.986917	0.46796	0.822542	0.336319	0.822542

Figure3:Results of Correlation Coefficient

The final comprehensive score is calculated by the sum of the above correlation coefficients, namely:

$$Score[i] = \Sigma Corr\_Coef[i] \tag{4}$$

The final score represents the final ranking of music influence, which is shown below (top 10) :

Table5: Most Influential Musicians

Artist ID	Synthesis Score	Rank
759491	0.68726147	1
91438	0.68726147	2
132710	0.68726147	3
755156	0.687310208	4
895136	0.68726147	5
899530	0.687310208	6
134729	0.68726147	7
446509	0.687358969	8
579839	0.687310208	9
741746	0.68726147	10

5 Music Similarity Measurement Model

5.1 Data Pre-processing

We add the music genres of the influencers and followers in the *influence_data* to the *full_music_data* data set, which can act as a verification index of similarity measurement.

We delete the *release_date* column, and retain the *year* column, and we delete the *song title* column, *artist_id* column, *artist_name* column and other objective indexes that do not actually affect the music similarity.

5.2 Establishment of Music Similarity Measurement Model

Music Similarity Measurement Model aims to calculate the degree of similarity between two musicians by analyzing the characteristics of different music. There are many music indicators, including danceability, energy, valence, et cetera, and different indicators have different influence on the similarity of music. Considering that some indicators do not differ greatly among music genres, we use principal component analysis (PCA) to reduce the dimension of the data.

PCA simplifies the coincidence variables and retain several key factors. We use singular value decomposition(SVD) covariance matrix to realize PCA, and obtain PCA dimension reduction in both row and column directions.

$$A_{m*n} = U_{m*r} \sum_{r*r} V_{r*n}^T \tag{5}$$

- *m is the number of rows in A and n is the number of columns in A*

After dimensionality reduction for the data set, we use the cosine similarity method to calculate the similarity degree between music according to the main characteristics of music. The main calculation formula of cosine similarity is as follows:

$$\cos(\theta) = \frac{\sum_{i=1}^n x_i \times y_i}{\sqrt{\sum_{i=1}^n x_i^2} \times \sqrt{\sum_{i=1}^n y_i^2}} \tag{6}$$

Then we use K-means method to carry out cluster analysis on musicians of different musical genres, so as to quantify the correlation between artists within and between genres.

The K-means method calculates the distance between the sample and each center of mass through iteration:

$$d_{ij} = \| x_i - y_i \|_2^2 \tag{7}$$

Then, we update C_j with the minimum d_{ij}, and we recalculate the center of mass:

$$\mu_j = \frac{1}{|C_j|} \sum_{x \in C_j} x \tag{8}$$

Repeat until all *k* centroid vectors have no changes, and we have the final clustering result.

5.3 Cosine Similarity Analysis & Clustering

We need to develop measures of music similarity by analyzing *full_music_data* data sets of music characteristics and explore whether artists within genre are more similar than artists between genres.

5.3.1 Principal Component Analysis（PCA）

We used PCA to reduce the dimensionality of 15 features of an artist's song. First we calculate the dimension after reduction[1]：

$$\frac{\frac{1}{m}\sum_{i=1}^m \left\| x^{(i)} - x_{approx}^{(i)} \right\|}{\frac{1}{m}\sum_{i=1}^m \|x^{(i)}\|^2} \leq t \tag{9}$$

- *t is the percentage of data loss, we assume that t=64.9%*

- $\frac{1}{m}\sum_{i=1}^{m}\left\|x^{(i)} - x_{approx}^{(i)}\right\|$ is average squared projection error, $x_{approx}^{(i)}$ is the mapping value of $x^{(i)}$.

- $\frac{1}{m}\sum_{i=1}^{m}\left\|x^{(i)}\right\|^{2}$ is the average variation in the data.

Calculating the above formula, we get that the data dimension is 2 after the dimensionality reduction.

After dimensionality reduction, the music characteristics are composed of only two parameters. The following table shows the music characteristics of some artists after dimensionality reduction.

Table6: Data after dimensionality reduction

Artist_id	Characteristic1	Characteristic2	Music genre
3413	0.22165447	1.26940619	Vocal
119058	-0.32609955	-0.04960572	Unknown
2130	-0.3477185	1.0988282	Stage & Screen
12802	-0.51848088	1.11196582	Religious
3627	0.37338373	0.72817909	Reggae
1190	0.69343678	0.26499097	R&B;
335	-0.07117314	-0.31168386	Pop/Rock
4458	1.98429797	-0.58044813	New Age
1097	0.63228317	-0.46329326	Latin
6143	1.2121273	0.90916983	Jazz
1266	-0.61934971	1.90675971	International
13377	-0.0517663	0.05737101	Folk
1611	-0.98913649	0.14885298	Electronic
2330	0.03705368	-0.25279089	Easy Listening
1433	-0.29354562	-0.16241488	Country
2273	-0.3752908	1.02861622	Comedy/Spoken
1589	-0.92296736	1.33931319	Classical
744969	-1.05594593	-0.04322702	Children's
3829	-0.2839349	0.70662882	Blues
186927	2.68547208	1.13801597	Avant-Garde

5.3.2 Cosine Similarity Analysis

Cosine similarity analysis is further carried out for the data after dimensionality reduction. For different artists, the cosine similarity is positive and the closer it is to 1, the greater the similarity between artists is. The cosine similarity is negative and the closer it is to -1, the smaller the similarity between artists is. The table below shows the validity of cosine similarity analysis by taking Anita O'Day, the originator of Vocal genre, as an example.

Table7: Cosine Similarity of Anita O'Day

Artist_id	Artist_name	Cosine_similarity	Music genre
29867	Shirley Bassey	0.9730792	Vocal
6334	Tony Bennett	0.99603395	Vocal
37197	Steve Tyrell	0.91402023	Vocal
38100	Pearl Bailey	0.993157	Vocal
38466	Andy Williams	0.99598096	Vocal
1416172	Jack Johnson	0.46040676	Religious
186927	Harold Budd	-0.52473244	Avant-Garde
355142	Nirvana	-0.26097036	Pop-Rock
12180	Trace Adkins	-0.09500609	Country
519715	Tommy Flanagan	-0.91466135	Jazz

Obviously, for other musicians in the Vocal genre, the cosine similarity is very high, reaching above 0.9 in most cases. Meanwhile, for artists of other musical genres, the cosine similarity is very low in most cases, a few are a little higher but not more than 0.9, and very few may reach more than 0.9.

Similarly, we take Billy Joel, the originator of pop-rock, as an example to show the feasibility of the similarity analysis model in the form of comparison.

Table8: Cosine Similarity of Billy Joel

Artist_id	Artist_name	Cosine_similarity	Music genre
105030	Coheed and Cambria	0.95098611	Pop-Rock
110161	Chicago	0.99999916	Pop-Rock
112462	Cream	0.98553317	Pop-Rock
773392	China Crisis	0.97009513	Pop-Rock
778655	Lacuna Coil	0.78713962	Pop-Rock
186927	Harold Budd	0.61253262	Avant-Garde
1416172	Jack Johnson	-0.55246053	Religious
12180	Trace Adkins	-0.01171003	Country
38466	Andy Williams	-0.98074708	Vocal
203711	Tuck & Patti	0.24099016	Jazz

We get the conclusion through our similarity analysis model that most musicians within genre are more similar than musicians between genres

5.3.3 K-means Clustering

In order to explore whether artists within genre are more similar than artists between genres, we use the K-means clustering method to classify the 20 music genres.

To calculate the distance between each point and the center of mass, we use cosine similarity as a measure of the magnitude of the difference between two individuals. Compared with the distance measure, cosine similarity focuses more on the difference between two vectors in direction, rather than in distance or length.
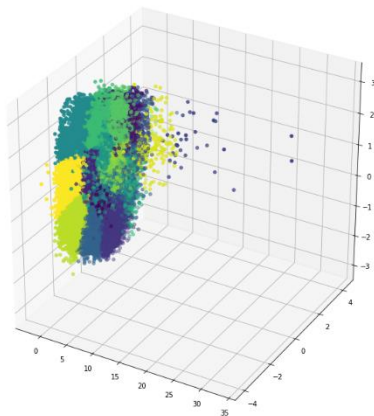
Figure4: Clustering of 3 dimensional data
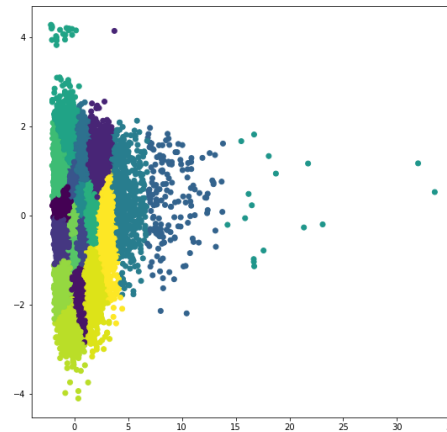
Figure5: Clustering of 2 dimensional data

In order to verify the effectiveness of K-means clustering, we add musicians of various genres to Figure5 in the form of red dots and observe their distribution. And for the sake of concise,2-dimension feature is taken as a demonstration here.

Take Pop Rock, Children's, Unknown, Easy Listening genres of music as examples, and their distributions are as follows
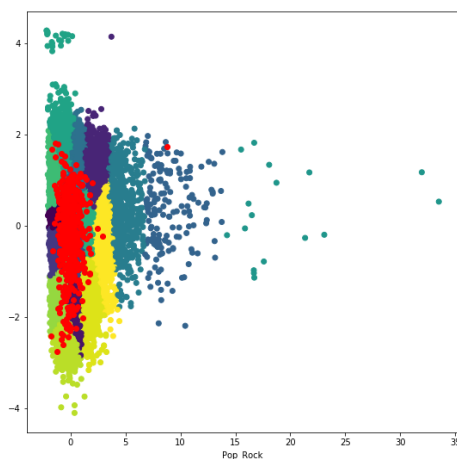
Figure6:Distribution of Pop Rock Genre
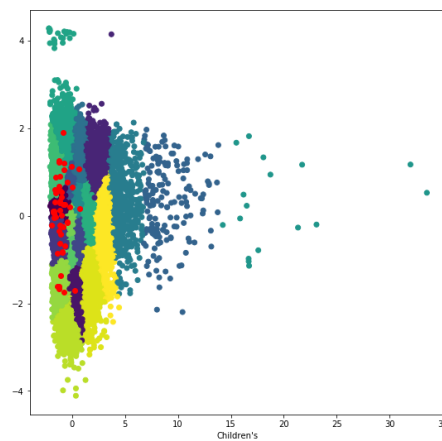
Figure7:Distribution of Children's Genre
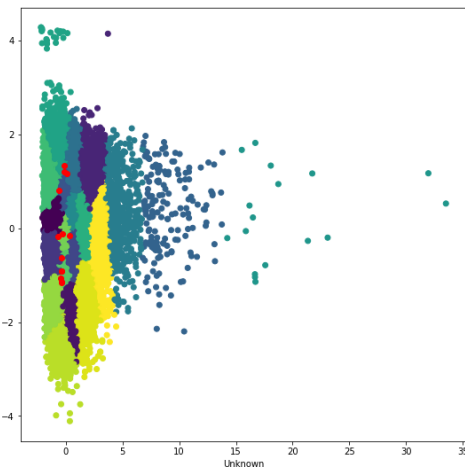
Figure8:Distribution of Unknown Genre
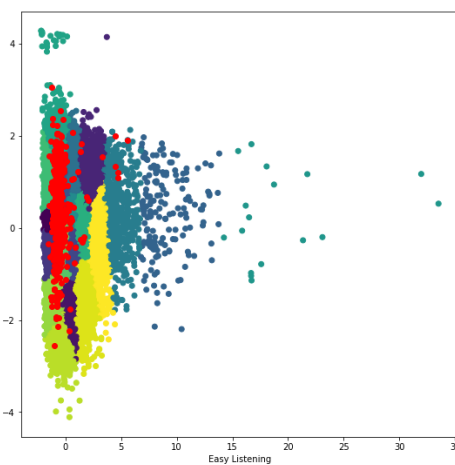
Figure9:Distribution of Easy Listening

Genres

5.4 Relationship between Music Characteristics and Genres

We need to compare similarities and influences between and within genres and explore the differences and correlations between genres. Furthermore we need to analyze the changes of music genres through time.

5.4.1 Data Normalization & Bar Diagrams

First, we normalize the data in *full_music_data*. Based on the characteristics of danceability, energy, valence, tempo, loudness, acousticness, instrumentalness, liveness, speechiness, duration_ms, the line chart below carries out the comparison between different music genres of music characteristics.

The normalized characteristics of all music genres can be shown by the line chart below:
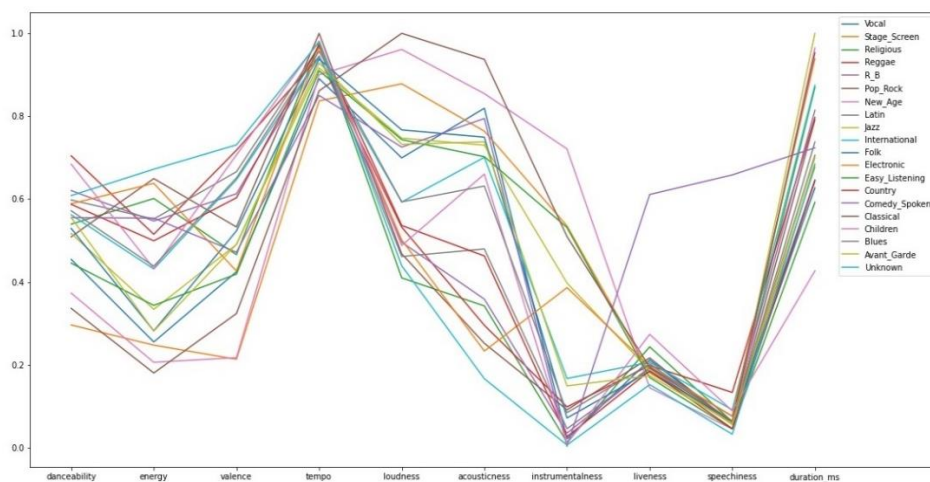
Figure10: Normalized characteristics of all music genres

From the figure above, we can intuitively see the values of normalized music characteristic of various kinds of music. We can find that there is little difference between music genres in tempo, liveness, speechness , and to some degree, these are correlations between genres. However, music genres differ greatly in characteristics like energy, loudness, acousticness and danceability.

We can choose some specific curves in the above diagram , make comparison and obtain the relationship between specific music genres. For example , when we plot the curves of Pop-rock, R&B, and Electronic music together, we can find it clear that all the three music genres are quite similar in all characteristics, except that Electronic is a little more instrumentalness.
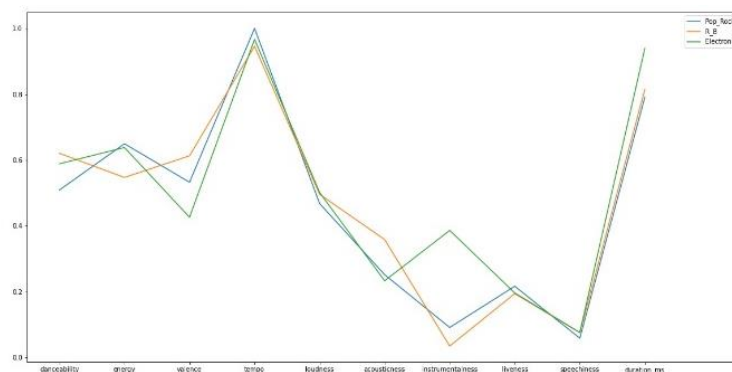
Figure11: Normalized characteristics of Pop-rock, R&B, and Electronic music

From the bar chart of all characteristics, we can also find the differences and correlations between genres. For example, in the liveness and speechiness bar chart, we can clearly and intuitively observe that the liveness and speechiness indexes of Comedy/Spoken music is particularly outstanding, which is obviously different from other music genres.
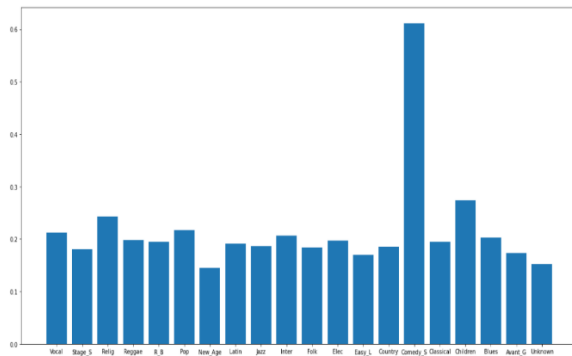
Figure12:Bar Diagram of Liveness   **Figure13: Bar Diagram of Speechiness**
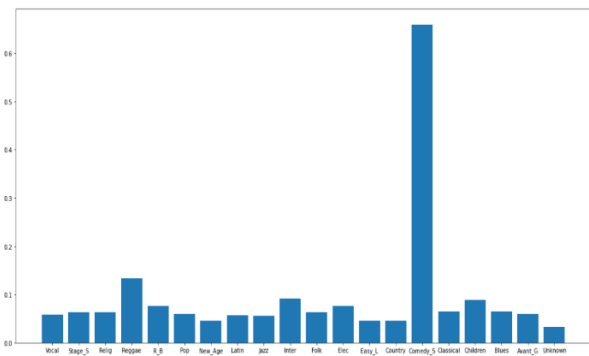
From the bar chart of Valence index, we can clearly and intuitively observe that the Valence indexes of Stage & Screen and New_age music genres are very similar, and also obviously lower than other kinds of music genres. At the same time, New_Age is very outstanding in terms of instrumentalness, which is higher than that of other musical genres, and can be used as an assistant indicator to distinguish New_Age from Stage & Screen.
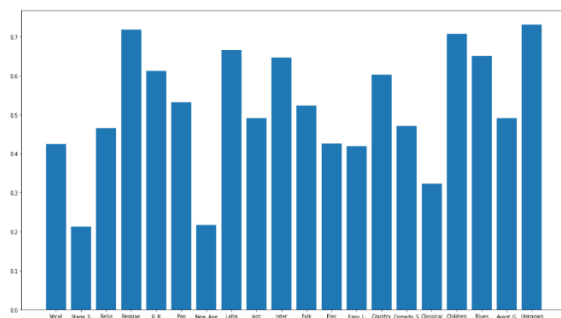

Figure14:Bar Diagram of Valence   **Figure15: Bar Diagram of Instrumental-ness**
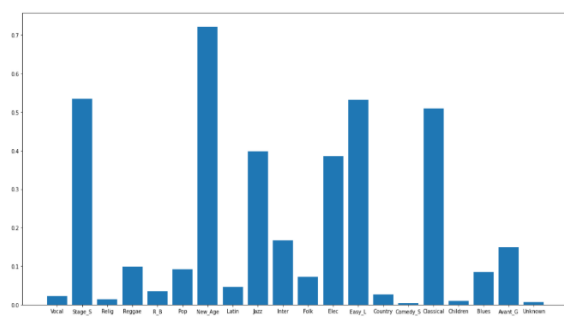
From the bar charts of all kinds of music characteristics after data normalization, we can clearly see the differences and similarities between different music genres.

5.4.2 Changes of Music over Time

The characteristics of music genres change over time. We take Pop-Rock genre as an example to discuss music changes over time.

Through the line chart which shows the changes in musical characteristics over time, we can analyze the changes of Pop-Rock music. The rising and falling trend of the line chart can reflect the development tendency of Pop-Rock music, and the key turning points are important criteria for us to judge the change of music style.
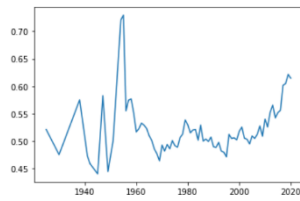
Figure16: danceability
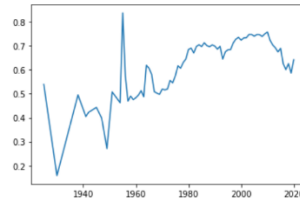
Figure17: energy

Figure18: valence
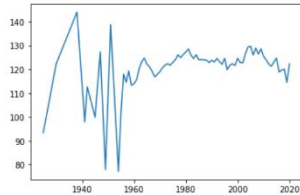
Figure19: tempo

Figure20: loudness
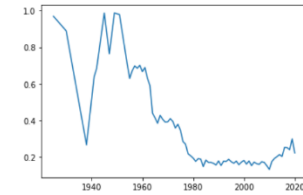
Figure21: acousticness
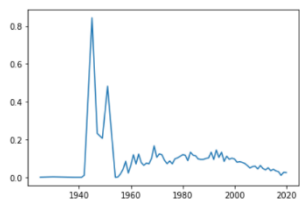
Figure23: instrumentalness
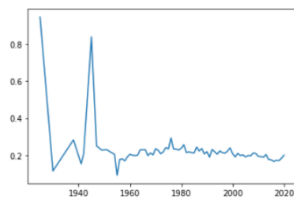
**Figure23:
liveness& speechiness**
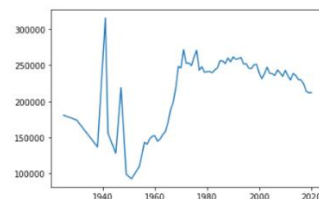
Figure24: duration_ms

The nine line charts above show how the characteristics of pop-rock have changed over time. The conclusions observed from the charts are as follows：

- Danceability and energy have very similar changes, both of them grew rapidly after 1950, peaked around 1955, and then declined. Both of them showed an upward trend after 2000.
- Valance is the positive level of music. Around 1940, the Valance of pop-rock reached its lowest point and the songs' negativity reached its maximum. Around 1955, the Valance reached its peak and the songs' positivity reached its maximum.

5.5 Influence on Creation & Contagious Characteristics

We need to explore the influence degree of influencers and followers, according to the similarity data in the *data_influence* data set. And we need to extract some most 'contagious' characteristics.

5.5.1 Verification that Influencers Affect Followers

In order to verify the influence of influencers on followers' music creation, we calculate the cosine similarity of their influencers and irrelevant musicians to form a bar chart for comparison.
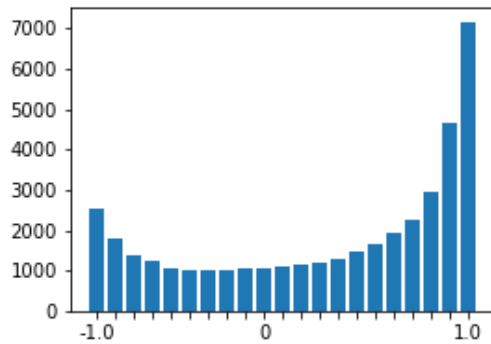
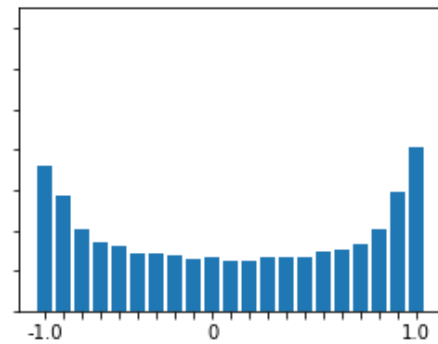**Figure25:Similarity of Music between influencers and followers**

Figure26:Similarity of Music with irrelevant musicians

The abscissa in the figure above is cosine similarity. The closer it is to 1.0, the more similar the music characteristics are. The Y-axis represents the number of people. From the figures above, we can find it clear that the number of followers and influencer with the cosine similarity close to 1.0 is significantly higher than the number of those irrelevant musicians. That means, the characteristics of followers' songs are more similar to their influencers', which directly proves that the influencers have significant influence on followers' music creation.

5.5.2 Contagiousness of Music Characters

We believe that some of those music characteristics are more "Contagious" than others. We think that "Contagiousness" can be analyzed from two aspects. First of all, "Contagiousness" can refer to the influence of a music characteristic on the popularity of a song. We can use correlation analysis to explore the correlation between the characteristics of various songs and their popularity. Higher correlation means more "Contagious".
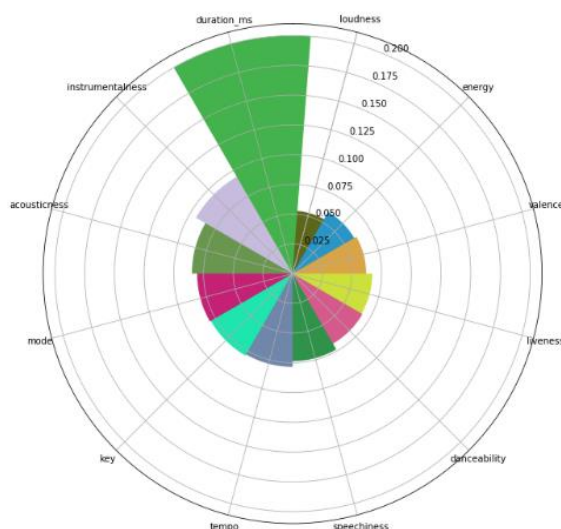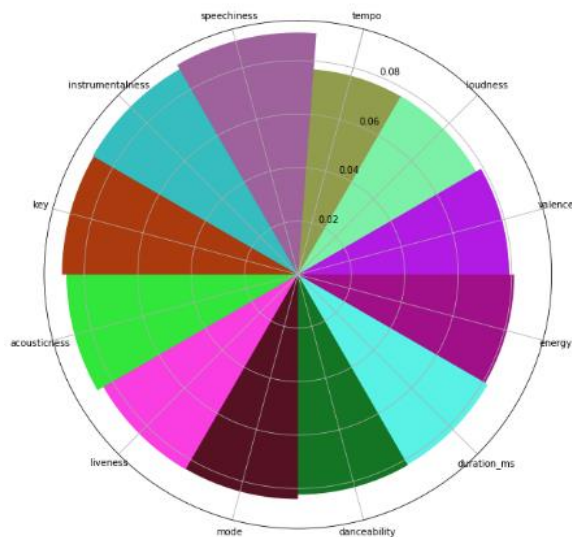

The chart on the left is the radar chart of music characteristics and music popularity. It can be clearly observed that *duration_ms* is far more correlated with popularity than any other indexes. In other words, *duration_ms* have greater influence on the popularity of a song to than any other indexes, and *duration_ms* is more contagious.

Figure27: Correlation between Characteristics and Popularity

Secondly, we suggest that contagiousness can be reflected by the similarity of this music feature between influencers and followers. Obviously, if a particular music characteristic is highly similar to its influencers', then we can argue that this kind of characteristic is highly

contagious. Therefore, we explore the similarity between influencers and followers in all kinds of music features and draw a radar chart .

The radar chart on the left shows the similarity between influencers and followers in all kinds of music features. As we can see, the indicators are fairly balanced, and none stand out. Among them, speechiness, intrumentalness, and key are more similar than others. To some extent, they are more contagious, but not very obvious.

Figure28: Similarity between Influencers and Followers

5.6 Revolutionary & Revolutionaries

We need to identify the characteristics that might signify revolutions and find out possible revolutionaries in our network.

Since Pop-rock music occupies the highest proportion in *full_music_data* and also has the most characteristic data, which is convenient for modeling and analyzing. Thus we shall take pop-rock music as an example in the solution of Problem5.

5.6.1 Popularity of Pop-Rock Music

According to the popularity and corresponding year of Pop-Rock music in the data set, we first draw the scatter diagram of Pop-Rock music's popularity over year.
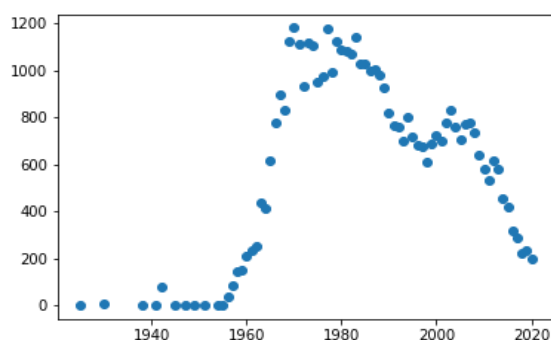
As it is shown in the graph on the left, Pop-Rock's popularity over time can be clearly seen from the scatter chart, even without the process of fitting, due to the large number of songs in Pop-Rock.

Figure29:Scatter Diagram of Popularity

As you can see from the chart, Pop-Rock was not very popular until 1960. Since then Pop-Rock quickly became popular, and reached the peak of the first trend in 1970-90. Pop-Rock declined in popularity after the 1990s, but began to rise again in the early 2000s, reaching the second peak in 2000-2008. Since 2010, the popularity of pop-rock Music has been declining.

5.6.2 Revolutionaries And Their Characters

During the first peak of pop-rock popularity, Queen released the most songs, 378. At the second peak of the popularity index, Linkin Park released the most songs, 89. Therefore, we take these two bands as examples to analyze music characteristics during the peaks.

First of all, we compare the musical characteristics of the two bands with the average of all the data, and draw the bar chart for intuitive comparison.
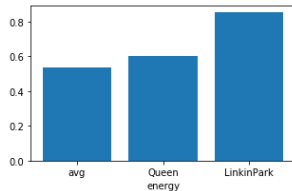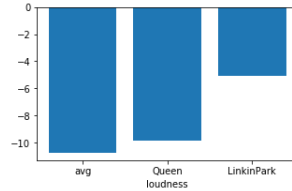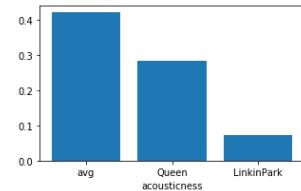

Figure30:Energy

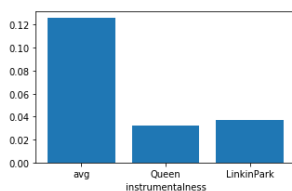**Figure31:Loudness**

Figure32:Acoustiness
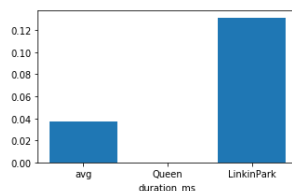

Figure33:Instrumentalness
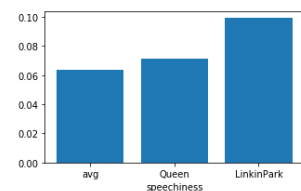

Figure34:Duration_ms

Figure35:Speechiness

From the six strikingly different bar charts above, we can find the following characteristics：

- In terms of energy and speechiness, Linkin Park is higher than Queen, and both two are above the average.
- In terms of instrumentalness, Queen and Linkin Park are both far below the average.
- In terms of duration_ms, Queen has the value of 0 and Linkin Park is much higher than the average, the two banks are quite different.
- In terms of loudness and acoustiness, Queen is close to the average, but Linkin Park is lower than the average

To further explore whether the two bands are revolutionists, we plot all the music characteristics of the two peaks of popularity as bar charts for intuitive horizontal comparison.
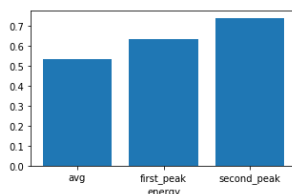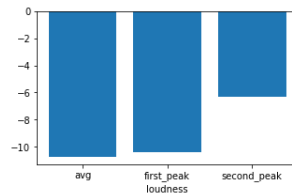

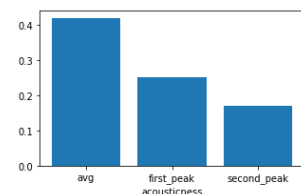**Figure36:Energy at Peaks**

Figure37:Loudness at Peaks
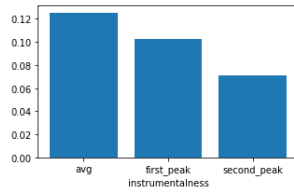

Figure38:Acoustiness at Peaks
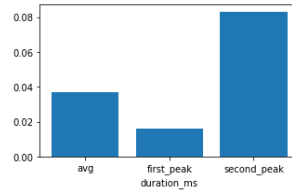
Figure39:Instrumentalness at Peaks

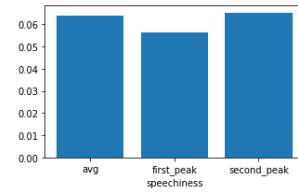**Figure40:Duration_ms at Peaks**

Figure41: Speechiness at Peaks

By comparing the six bar charts above with those of Queen and Linkin park, we can clearly find that the music characteristics of Queen and Linkin park are basically consistent with music features in the two peak periods. In a sense, Queen and Linkin park are the revolutionaries of the change of music style, who drives the trend of Pop-Rock music.

In addition, the following bands released more songs at the two peaks, and considering their popularity and number of songs, they're also called revolutionaries.

Table9: Revolutionaries in Pop-Rock Music

Name	Number of works	Name	Number of works
Billy Joel	172	Disturbed	47
Grateful Dead	175	Radiohead	51
David Bowie	180	Taylor Swift	51
Kiss	183	Coldplay	52
Talking Heads	186	blink-182	53
Bob Dylan	219	Slipknot	56
Elton John	241	System of a Down	57
Led Zeppelin,	245	Fall Out Boy	59
The Rolling Stones	245	John Mayer	60
Fleetwood Mac	312	Red Hot	68
Queen	378	Linkin Park	89

5.7 Changes of Musicians & Genres over Time

We need to analyze the processes of musical evolution, and find out key indicators that can help explain the music genre changing process.

Taking Pop/Rock as an example, we find that in terms of tempo and energy, Pop-Rock music is almost the highest compared with other genres. So Tempo and energy can be used as indicators of the Pop-Rock genre. Meanwhile, Comedy/Spoken is significantly more liveness and speechiness than other genres. Therefore, we can choose liveness and speechiness as the indicators of Comedy/Spoken.

By analyzing the changes in the characteristics of Pop/Rock genre over time, we find that from 1920 to 2020, its danceability index is basically stable, but there was a jump in the 1950s and 1960s. Meanwhile, energy index shows an overall rising trend, but it also experienced a huge change in the 1950s and 1960s. Through the comparison of other characteristics, we find that almost all characteristics of Pop-Rock had great changes before 1960s . In the 1960s and

1970s, the music features gradually stabilized, and as we've found in Problem3 that the popularity of the Pop-Rock reached the peak in the 70s and 80s. Therefore, we can assume that a stable musical genre tends to be more and more influential, and its music influence will gradually promote the stability of the genre's style.

We select Frank Sinatra as an example for further analysis, who has most songs in our data se, in order to get sufficient data. We plot the characteristics of the genre to which he belongs——Vocal. The result is as follows:
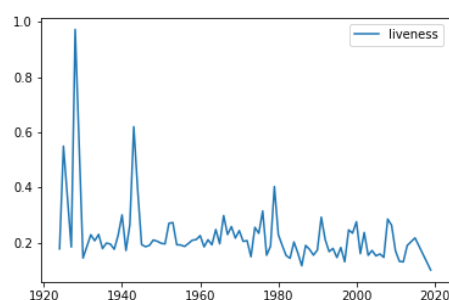

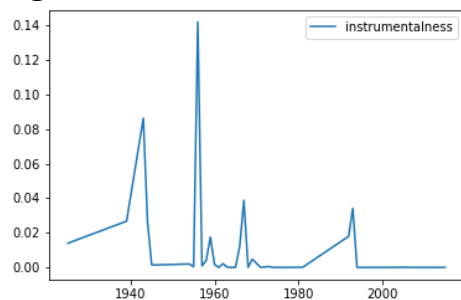**Figure42:liveness of Frank Sinatra**

Figure43:liveness of Vocal
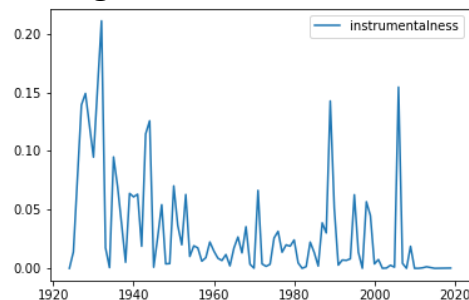

Figure44:Instrumentalness of Frank

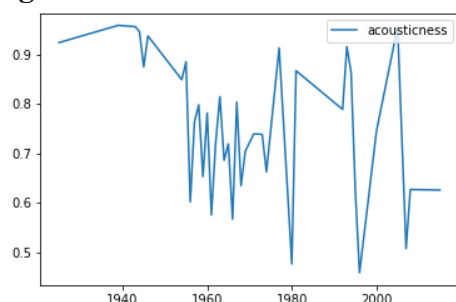**Figure45: Instrumentalness of Vocal**
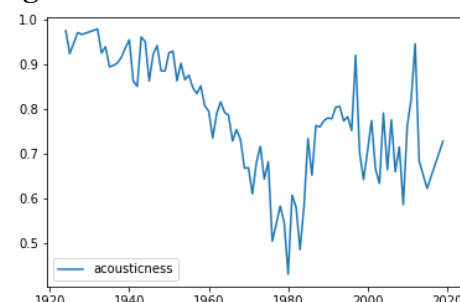

Figure46:Acousticness of Frank Sinatra

Figure47: Acousticness of Vocal

We can find in the diagrams above that the music features of the artist and Vocal genre present a certain degree of consistency. Over time, their instrumentalness and liveness gradually decreased. As an important feature of Vocal music, acousticness, was found decreased gradually before 1980 and then increased gradually after 1980. Through data search, we find that its popularity gradually decreases after 1980. Perhaps with the progress of science and technology, people's appreciation level gradually changes and they begin to turn to other types of music, so Vocal also gradually becomes a minority.

Based on the understanding of the background of Vocal music genre, its popularity gradually decreased after 1980. Perhaps with the progress of science and technology, people's appreciation level gradually changes and people prefer other genres of music. As is shown below,

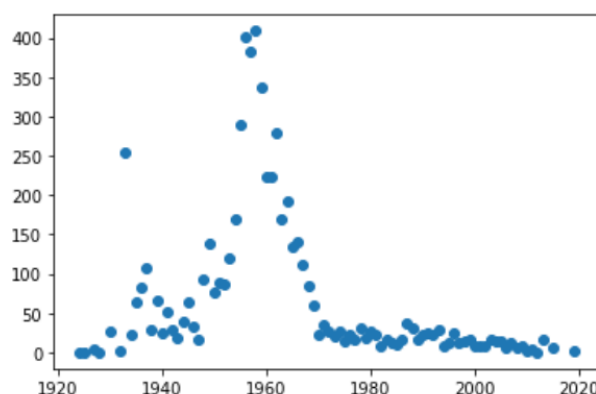the number of Vocal songs released per year reached the peak in the 1960s and has gradually become a minority.

Figure48: The number of Vocal songs released per year

6 Social, Political and Technological Influence Model

We need to express the influence of music on culture in time or circumstances, or explain how to identify the impact of social, political or technological changes (such as the Internet) in our network.

6.1 Estimation of Influence Model

In the previous analysis, we found that the change of musical style is related to social forms, political dynamics and technological changes to some extent. Take Pop-Rock genre as an example, we can see from **Figure16~24** that there was an obvious turning point in the 1950s and 1960s. After 1945, various genres and styles of experimentation became popular again and had greater development and wider influence. It is mainly reflected in the two types of absolute control and complete freedom. During this period, science and technology developed at a high level, industrialization and modernization reached an unprecedented scale, and new ideas, new ideas, new ways of life, new working conditions, new artistic tastes and so on emerged one after another. In music, he pursues new materials, new languages, new ideas and new techniques. There is an unprecedented degree of questioning of tradition. In most western countries the national character of music is no longer valued. At the same time, many composers became more and more inclined to pursue the form of their works, giving up thoughts and emotions; Music is no longer viewed as a tool for audience communication, but as a purely personal activity. So in terms of the expression of musical characteristics.

Western music after the 1970s still presents a pluralistic situation, and the climax of new music experiment has passed. On the one hand, there are still many composers who continue to write in the old musical language, and various genres and styles of music continue to exist. On the other hand, a new trend has emerged, that is, the so-called "return" phenomenon to a certain extent, and the tradition has been valued again. This return is the combination of the new and the old. With the development of science and technology, convenient transportation, quick consultation, and the formation of the concept of "global village", composers can absorb all kinds of nutrients on the basis of a wider range of musical factors in the past, and the distance

between the East and the West is shrinking. The combination of serious and popular music is also becoming increasingly common. The use of traditional and modern, east and west, serious and popular, science and technology and art makes music diverse and comprehensive. Minimalism, collage music, new romanticism.

6.2 Music Characteristics Change during Wars

At the same time, according to the history, we can know that the songs of the war years seem to be more exciting and have a sense of rhythm, and the songs represent the development requirements of social productivity, the fundamental interests of the people, and the forward direction of advanced culture. In different social stages, people have different levels of needs, so the resonant things are naturally different. In wartime, we need to survive. Naturally, the song conveys more fighting spirit. In times of peace, we need happiness, so songs are naturally more about love. This can also be reflected in the data characteristics of music.

6.3 Music Characteristics Change with Technology

Society influences music, and music influences society in turn. It was this initial need for music that drove the progress of the phonograph:

The 78R.P.M.Disc (78R.P.M.Disc) was the standard format throughout the recording industry during the period of electrical recording, and it remained in use until the end of World War II. The 78-rpm record has a high degree of sound fidelity, but because the speed of the record is too fast, each side can only record about 3 to 5 minutes of music, so it is only suitable for full recording of short songs and instrumental sketches[2]. When a piece of music is played for a little longer, the listening process has to be interrupted by the need to flip or change the record. In order to overcome this defect, it is necessary to develop longer playing time records. The record's length of time on each side is more than double the Standard Playing, or SP, of about three to five minutes on each side that was widely used before[3]. The appearance of 33 r PM records indicates that the subsequent Recording/playback time is getting longer and longer, the fidelity is getting higher and higher, and even the "Long Playing Recording" (LP) will emerge successively with the improvement and perfection of Recording technology[4]. According to the data, it is not difficult to find that the duration of music increased significantly in the 1930s and 1940s. At the same time of the rise of the long record, the tape recording technology also quietly entered the market, began its rival with the record process.

With the progress of science and technology and the development of sound retention equipment, we can find that more and more Pop- Rock; R & B; country; electronic music has been popular, and its popularity has been enduring since the 1970s.
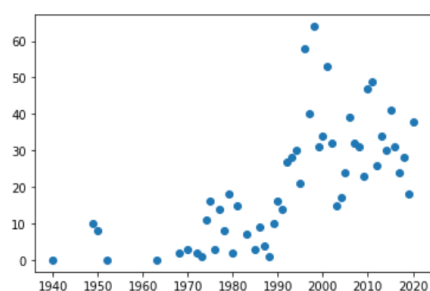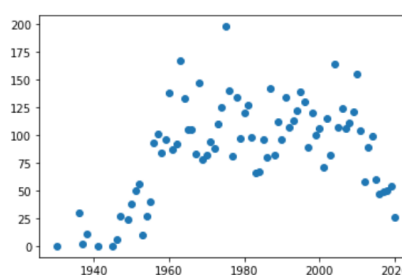

Figure49: Electronic songs per year

Figure50: Country songs per year

7 Model Evaluation

7.1 Strengths

- The selection of the network parameters of the Influencers & Followers Network is scientific and reasonable. Our network model takes all possible factors of music influence into account, which is close to reality and has high application value.

- Our Music Similarity Measurement Model is mature in theory and easy to realize. It can well satisfy our need in similarity analysis and has high calculation speed.

- In the process of establishing our Social, Political and Technological Influence Model, we comprehensively consider the influence of science, technology, politics, economy and social changes on music. And we verify the influence of external factors on music according to huge changes in music history.

7.2 Weaknesses

- Our model did not optimize PCA dimension reduction, the loss of music features data, is a little high ,resulting in poor clustering effect of the 20 music genres.

- Because of the length of the article, our exploration of the relationship between music characteristics and music genres mostly stays at the method level, and does not fully expand.

8 Conclusion

- We analyze music influence between influencers and followers, then we build Influencers & Followers Network using Grey Comprehensive Evaluation, and successfully quantified the music influence.

- We use principal component analysis to reduce the dimension of music characteristics, and use cosine similarity method to establish the Music Similarity Measurement Model, and successfully quantified similarities between musicians and music genres.

- We do further data mining, solve the 7 problems, and we sum up the relationship between musicians, genres, and music characteristics. We summarize the relationship between musical characteristics and changes in musical genres and identify the revolutionaries.

- In combination with the history of changes in music field, we establish Social, Political and Technological Influence Model and analyze the practical influences of music in time or environment.

References

[1] LeftNotEasy. Mathematics in Machine Learning (5)- Powerful Matrix Singular Value Decomposition (SVD) and Its Applications [EB／OL].
http://www.cnblogs.com/LeftNotEasy/archive/2011/01/19/svd-and-applications.html

[2] [3] [4]Graduate entrance examination observation room. The Great Changes of Western Music in the First Half of the 20th Century[EB／OL].
https://www.sohu.com/a/260238488_100264759

Documentation

Dear ICM leaders,

Our team establish an Influencers & Followers Network based on Grey Comprehensive Evaluation method and is able to measure music influence in the network. We sincerely recommend our network to you for the following reasons:

Firstly, our network model takes all possible factors of music influence into account. Our model will comprehensively consider the following four parameters: the number of musicians in the genre, the number of contemporaries in the genre, the rank of the musician's influence in the genre and era, the level of competition in the current era. Because of the completeness and correctness of the parameter selection, our model is of great accuracy and close to reality, which means it has high application value.

Secondly, our network trace the data to the source at the beginning and find out the originators in each genre. When it comes to more genres, huger data set and richer data ,we still first trace back to the source. At the algorithmic level, our model does not have to be modified, under such circumstances. However, for the music similarity measurement model, the data loss caused by PCA dimensionality reduction increases, results of clustering may become worse and the accuracy of cosine similarity may decrease sharply.

Thirdly, when it comes to more genres, huger data set and richer data, it is obvious that the characteristics between the music genres will no longer be clear. However, music genres are supported by culture, we can study the relationship between music genres and culture, and use culture as a medium to build a bridge between different music genres. And then our Social, Political and Technological Influence Model would help!

Yours sincerely ,

Team 2100291