Problem Chosen	2021	Team Control Number
D	MCM/ICM	2125244
	Summary Sheet	

Music Influence Evaluation Model Baced on Directed Network and Correlation Coefficients

Abstract

In this text, our team try to understand and measure the influence of previously produced music on new music and musical artists. To achieve this goal, our team develop influence model and similarity model.

Influence model is a directed network used to measure the influence between artists. Through this model we have found that the influence reflects the rise and fall of genres and relates the evolution of musical characteristics to real events and artists' previous creations.

Similarity model is used to measure the similarity between artists. We select some representative characteristics to compute the pearson correlation coeffcient. Through this model we have learnd the similarity between artists or genres, which helps us to measure the actual influence between artists or genres.

Then we combine two models mentioned above to analyze how musical evolution happened and what changes it brought about on genres' characteristics. Through the development of a specific artist or a genre, we identify the characteristics that signify major leaps in musical evolution and find out the leader in those major leaps.

In the end, we analyze our model by viewing it from the culture and reality point. And we find out that it does reflect the influence of social, political and technology development on music.

However, our model still has its weakness. We also discuss how to improve it in the future and how to apply it in real life. We hope this model can be used to analyze the development of music in the next 10 years and offer some advice to new artists .

Keywords: Music; Directed networks; Correlation coefficients.

Contents

1 Introduction

1.1 Background

Music has played an important role in human societies as an essential component of cultural heritage. It can be influenced by characteristics of the musical artists, access to new instruments or tools, or current social or political events. For example, in the 1950s, improved microphone design as well as inexpensive and more durable 45 rpm records for singles ”revolutionized the manner in which pop has been disseminated” and accelerated the wide spread of pop music. Besides, artists can also affect each other. These effects can be measured by the degree of similarity between music characteristics, such as tempos, lyrics, or instrumentalness.

There are sometimes revolutionary shifts in music, offering new sounds or rhythm, such as when a new genre emerges, or there is a reinvention of an existing genre (e.g. classical, pop/rock, jazz, etc.). This can be due to a sequence of small changes, a series of influential artists, or a shift within society. To study the musical evolutionary and revolutionary trends, we analyze the data that records the musical features and some other descriptions of the songs from 1930 to 2020. We created a model to quantify the influence of previouly produced music on new music and musicians to understand the interaction between different artists. We also developed measures of music similarities. In this paper, we analyze the influece of music and give some conclusions based on our models and music history.

1.2 Our Task

- Construct a directed network of musical influence.
- Develop measures of music similarity.
- Use models to compare similarities and influences between and within genres, and then

identify characteristics that distinguish genres.

- Study the consistence of influence model and similarity model.
- Analyze the development of music. Find revolutions in musical evolutions and identify characteristics that signify it and representative artists. Explain the cultural influence of music.
- Evaluate the performance of models with richer data.

1.3 Assumputions

- The influence of music can be measured by the similarity of musical characteristics.
- The influence of an artist on another artist from a different genre is strongger.
- Revolutions happened in the years when the number of released songs surged.

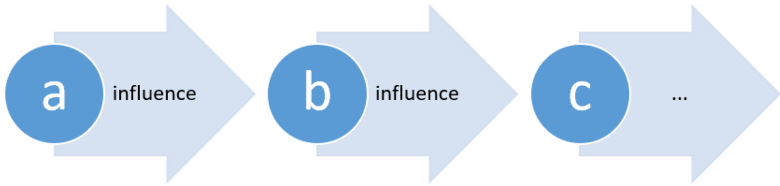
2 Models

2.1 Directed Network of Musical Influence

To understand and measure the influence of previous artists, we created a directed network, where influencers are connected with their followers. Each node represents an artist and has attributes including the year when the artist started to be active, the artist ID number, and corresponding genre. Directed edges point from influencers to followers and they have different weights to quantify the influence. To get the weights, we first define the influence of the influencer a who belongs to genre g_a:

$$\mathbb{I}_a = R_a \times \sum_{b} (\lambda_1 P_b + \sum_{c_b} \lambda_2 P_{c_b}), \tag{1}$$

where $P_b = \frac{1}{n_b}$, $R_a = \frac{N_a}{N}$,
b represents the artists who are influenced by a,
c_b represents the artists who are influenced by b,
n_b is the number of artists who influence b,
N_a is the number of artists who start to be active in genre g_a in the same decade,
and N is the number of artists who start to be active in all the genres in the same decade.



λ_1 and λ_2 are constants. In our model, we set λ_1 to 1 and λ_2 to 0.5, which shows indirect influence of a on c is weaker than that on b. R_a is used to quantify the current influence of genre g_a. The bigger the influence of g_a is, the bigger the influence of a is. The definition of P_b indicates that the influence of a on b will lessen if b is affected by more artists.

Then we define the influence of a on b:

$$W_{a,b} = \frac{\mathbb{I}_a}{\log_2(n_b + 1)} \times C^{I(g_a \neq g_b)} \tag{2}$$

where I is the indicating function and C is a constant. In our model, we assume that if a influence b from another genre, then this impact should be strongger. Here we set C to 1.2.

Then we will select a subnetwork and visualize it. First, select the node X_1 with the greatest influence, and all its outgoing edges are sorted in descending order according to the weight. What' s more, 10 edges are selected at equal spacing from the sorted outgoing edges, and then select the corresponding incoming edges at equal spacing for the vertex pointed by each of the 10 edges above. Similarily, the incoming edges are also sorted in descending order according to the weight, and x edges are selected according to the in-degree ratio. For example, the node with maximum in-degree (it' s about 48) will get $x = 16$, and the node with in-degree N will get $x = N * 16/48$ edges. Second, select node X_2, which has the second largest influence. Do the same thing as X_1, but the number of edges extracted initially is not 10, but $\frac{10 \times X_2\text{'s } out{-}degree}{X_1\text{'s } out{-}degree}$.

According to the subnetwork ($figure$ 1), the darker the edge color, the greater the weight ,which reflects that the influence of the corresponding influencer on followers is greater. Similarly, The darker the color of a node, the more influential the artist for that node is. It' s clear that the darker the node, the bigger the out-degree. Furthermore, the influence of a on other points c is closely related to the in-degree of c. The greater the in-degree of c, the

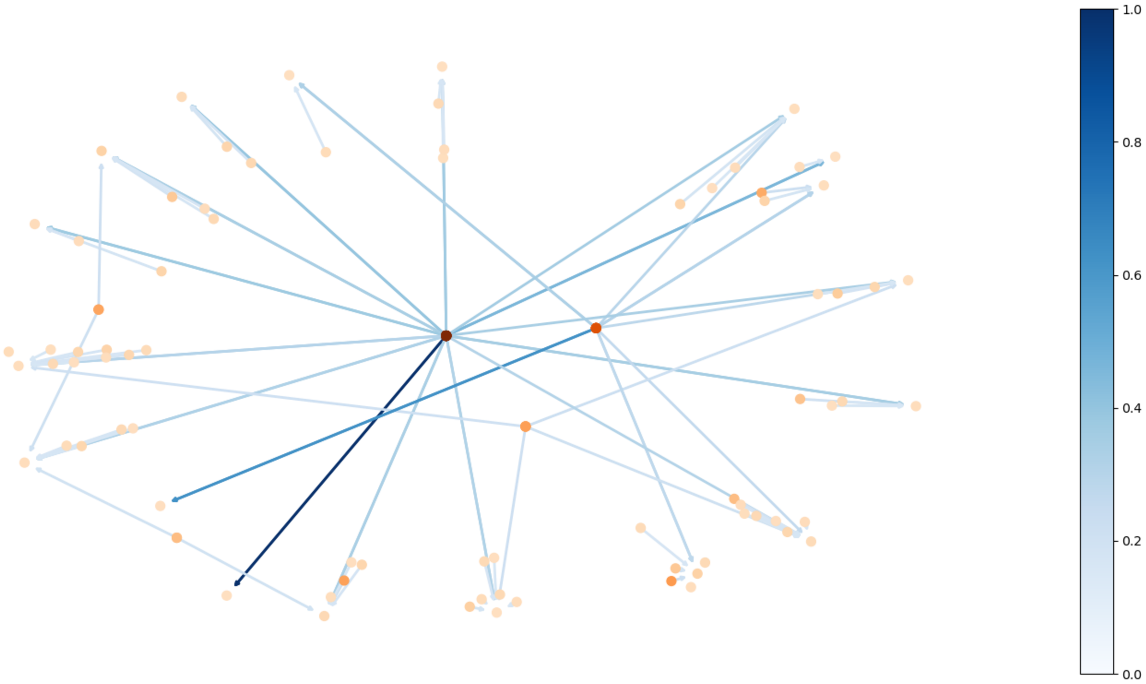


Figure 1: Subnetwork selected from the directed network

smaller the influence a can exert on c. This can be seen from the figure. The in-degree of the point pointed to by the darkest edge is 0.

2.2 Measure of Smilarity

It's suggested that the influece of music can be measured by the degree of similarity. The larger the influence of an artist on the other is, the more similar their musical pieces are. To quantify the music similarity, we first selected characteristics that made greater contributions to the music style than others. We used Random Forest Classifier that operate by constructing a multitude of decision trees. Here the genres of music is regarded as the dependent variable and all the musical characteristics are independent variables. Features selected by the classifier play an important part in music classfication. Danceability, energy, tempo, loudness, acousticness, instrumentalness, and duration time are selected charateristics.

Then we used Pearson Correlation Coefficient to construct the measure of music similarity after standardizing the characteristics. The degree of similarity of music X and Y is defined as:

$$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \tag{3}$$

where $X = (x_1, x_2, ..., x_7)^T$, $Y = (y_1, y_2, ..., y_7)^T$, and $\{x_i\}_{i=1}^{7}$, $\{y_i\}_{i=1}^{7}$ are selected features. As for the degree of similarity of different artists, we just substituded X with the average of selected charateristics of the artist.

3 Results and Conclusions

3.1 Similarities and Influences Between and Within Genres

3.1.1 Similarities Between and Within Genres

In this part, we compare the similarities between and within genres by comparing the similarities between artists from several genres. Take genres Jazz, Country, R&B, Stage & Screen as an example. We chose three artists from each genre as representatives. Then we used the degree of similarity between artists to draw a heat map to visualize the correlation

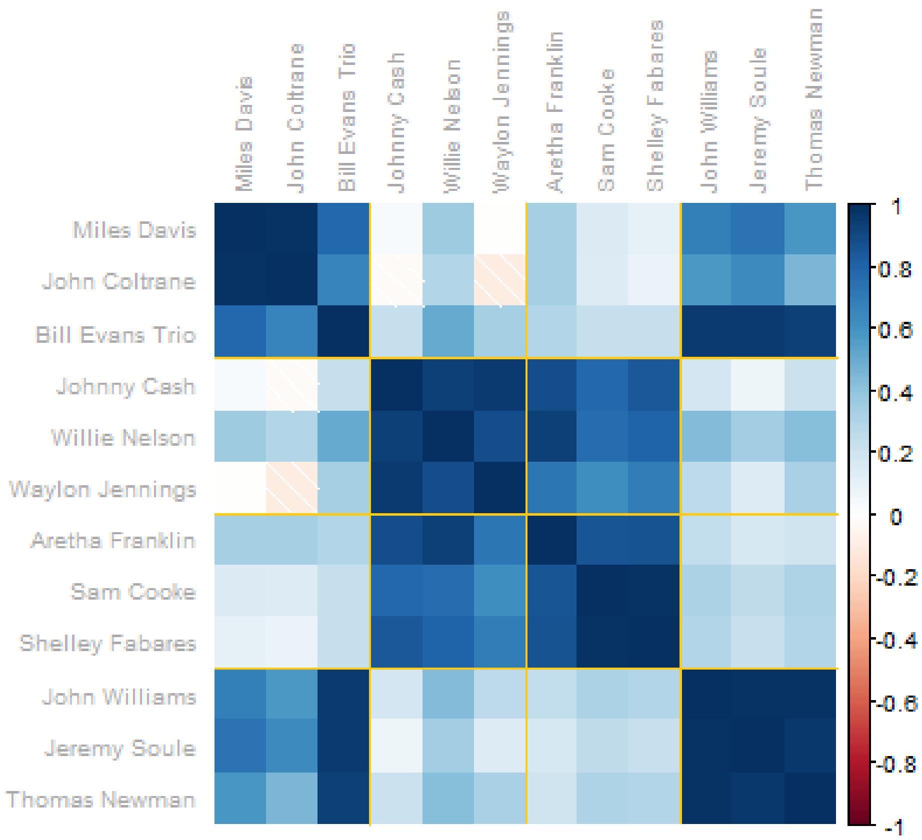


Figure 2: Correlation Heat Map of Artists from Jazz, Country music, R&B, and Stage & Screen music

between artists.

Fighue 2 shows the music similarities of 12 artists. Every three artists belong to the same genre. The genres of artists are Jazz, Country, R&B, Stage & Screen from left to right. We can see that musicians from the same genre are highly correlated. The similarities within genres are higher than that between genres. We find that artists from Jazz, Stage & Screen are hardly similar to Country music and R&B. We also notice that Country music musicians are quite similar to R&B musicians. This could be because country music is influenced by R&B and absorb its musical features, especially in the 21th century. The 2010s witnessed an increasing number of mainstream country acts collaborate with R&B acts. Stage & Screen music, such as movie backgroud music, has similarity to Jazz. One of the reasons of this observation could be that some film music is just a piece of Jazz music. For example, the music City of Stars and The Complete Musical Experience in the movie La La Land (2016) are exactly famous Jazz music.

3.1.2 Difference Between Genres

To have a better comprehension of the similarities and difference between differnt genres, we used the hierarchical cluster analysis method, where vectors with shorter Euclidean distance will be classified to one genre, to classify twenty music genres into several clusters. The result is showed in *Figure* 3. We can see that the genre Children's has great difference from other genres. Country and folk music are highly correlated. Blues, pop and rock music have similar characteristics. These results are consistent with the reality.

We hope to identify characteristics that can distinguish a genre. First, we define the feature F_i of the genre X as:

$$F_{i,X} = \frac{\sum_{j \in X} f_{i,j} Pop_j}{\sum_{j \in X} Pop_j} \tag{4}$$

where j is the music piece belonging to the genre X, $f_{i,j}$ is the feature F_i of the j, and Pop_j is the popularity of j. Here we regard that more popular songs contribute more to the whole style of the genre. After calculating the characteristics of all the genres, we computed the standard deviation of each feature between different genres. Resultes showed that tempo, loudness, key, acousticness, instrumentalness and duration time have larger standard deviation than others, which indicates these characteristics have larger difference between genres. Hence they can be used to distinguish a genre.

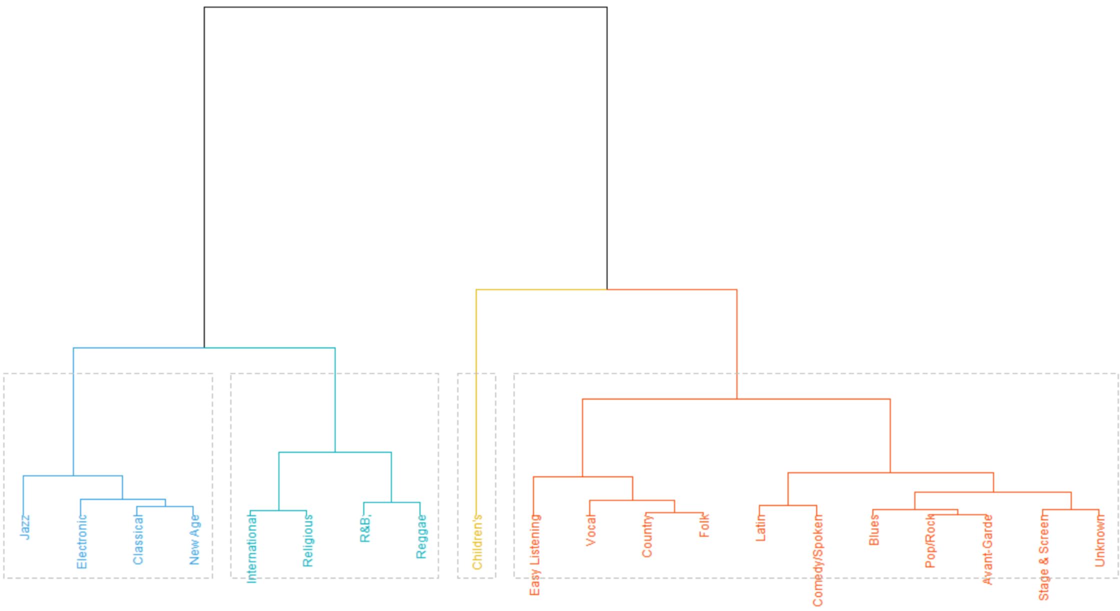


Figure 3: Clusters of music genres

3.1.3 Influences Between and Within Genres

We used the directed network constructed before to calculate the influence between and within music genres. Results are showed in *Table* 1. The influence data at the diagonal represents the influence within each genre. The influence data at the coordinate (i, j) represents the influence of i on j. Apparently, we can see that influence within genres is always larger than that between genres. We notice that the influence of Pop/Rock on other genres is quite remarkable. R&B has a notable influence on Country music, which accords with our previous results.

Table 1: Part of influence between and within Genres

Genre	Blues	Classical	Country	Electronic	Folk	Jazz	Pop/Rock	R&B	Stage & Screen	Vocal
Blues	663.50	0.00	5.05	1.16	42.05	27.14	1213.59	180.03	0.00	6.49
Classical	0.00	1.74	0.01	2.71	0.00	0.02	6.03	0.00	0.68	0.02
Country	14.98	0.00	2423.55	4.69	142.90	9.50	1149.63	7.17	0.00	14.21
Electronic	0.00	0.01	0.00	51.63	0.00	0.00	66.84	2.01	0.02	0.00
Folk	0.78	0.01	56.75	1.08	149.36	0.16	185.51	2.16	0.00	3.35
Jazz	327.89	11.22	103.31	165.91	57.39	5855.89	1664.44	985.43	34.38	613.79
Pop/Rock	546.34	110.89	4645.34	3880.50	3479.97	1271.93	194703.86	3919.00	152.77	642.55
R&B	159.29	8.96	144.96	254.91	27.71	185.49	4624.13	7761.89	0.00	99.39
Stage & Screen	0.00	0.01	0.01	1.45	0.08	0.95	6.30	0.00	9.88	1.91
Vocal	51.87	36.50	74.01	2.11	56.84	261.31	789.17	860.91	13.82	1319.90

The influence of each genres changes as time goes. *Figure* 4 shows counts of songs per year for each genre and *figure* 5 shows part of the influence of each genre per year. We notice that the influence and genre size, which is measured by the number of released music at that time, both surged. Furthermore, we can identify that around 1950 Jazz and Vocal became mainstream genre. Pop/Rock has been central to pop music since 1970s.Through the perspective of Jazz (*figure* 6), we can see how a genre change over time. After it became popular, its tempo, danceability, valence, and liveness became stable. Therefore, it might means a genre become mature. With the development of technology, music’s influence was growing, and the duration also became longer. After entering the 21st century, Jazz was influenced by more new kind of music, its characteristic has further development.

3.2 Consistence of Similarities and Influences Between Different Artists

To some extent，the influencers identified by influence_data in fact influence the respective artists’s music. We assume that influence can be measured by the similarity of song characteristics. Therefore, we compare the similarity of song features among artists with the influence of artists which is computed using the directed network. We find that the similarity between artists increases as the influence of one artist on the other raises when artists belong to different genres. Hence we can draw the conclusion that influencers really affect their followers to some extent.

According to the data set, the influence and song similarity between artists in the same genre is not highly related, while the influence and song similarity between artists from different genres has closer relationship. It should be because the similarity of music in the same genre is higher than that between genres. The influence between artists in the same genre doesn't have an obvious relationship with their similarities. However, influences from other genres would lead to obvious changes in the characteristics of songs of follower. In return, the follower’s characteristics of music tend to be similar to the influencer’s.

We select artists, who have great influence and are influenced by more than 30 artists, as representatives. For example, we consider the artists who influence the Pop/Rock musician Bob Dylan. *Table* 2 shows the influence of these artists on Bob Dylan, their similarities and the genres each artist belongs to. We find out that when the genre is fixed, as one artist's impact on the other increases, the correlation between them becomes higher. Therefore, it confirms that Bob Dylan actually influence his follower’s music.

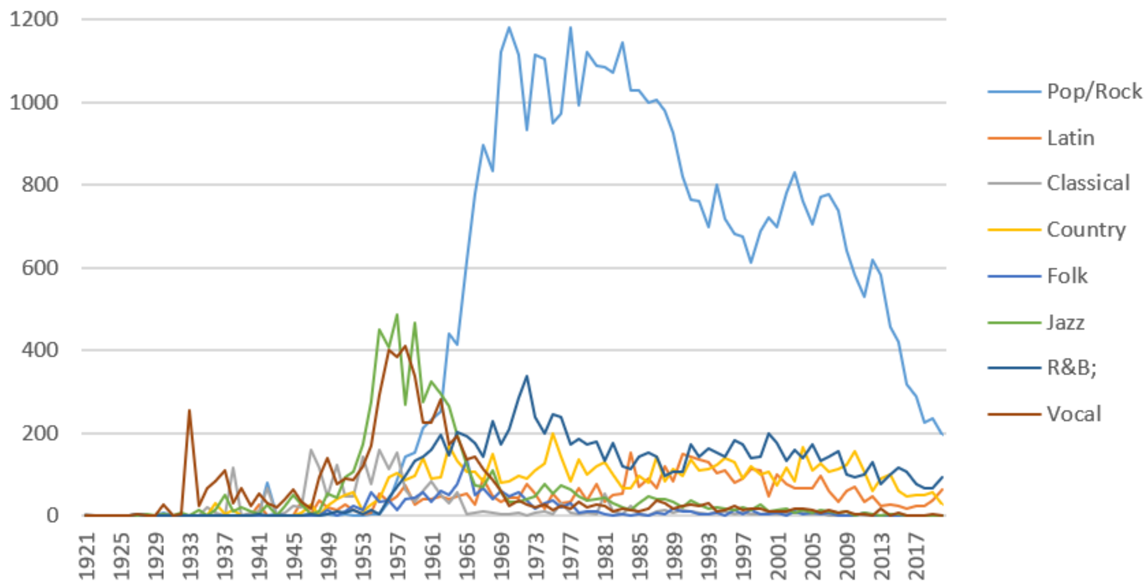


Figure 4: Counts of songs per year for each genre

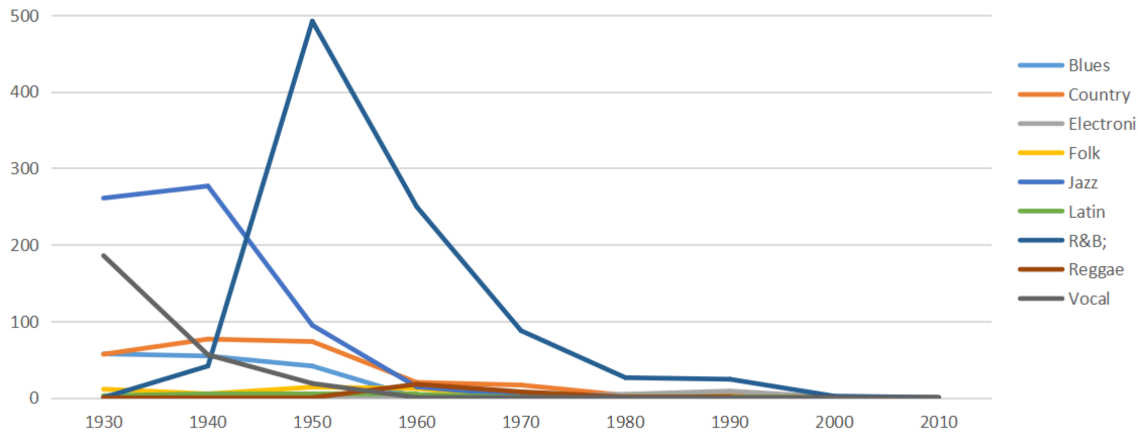


Figure 5: Influence of each genre per year

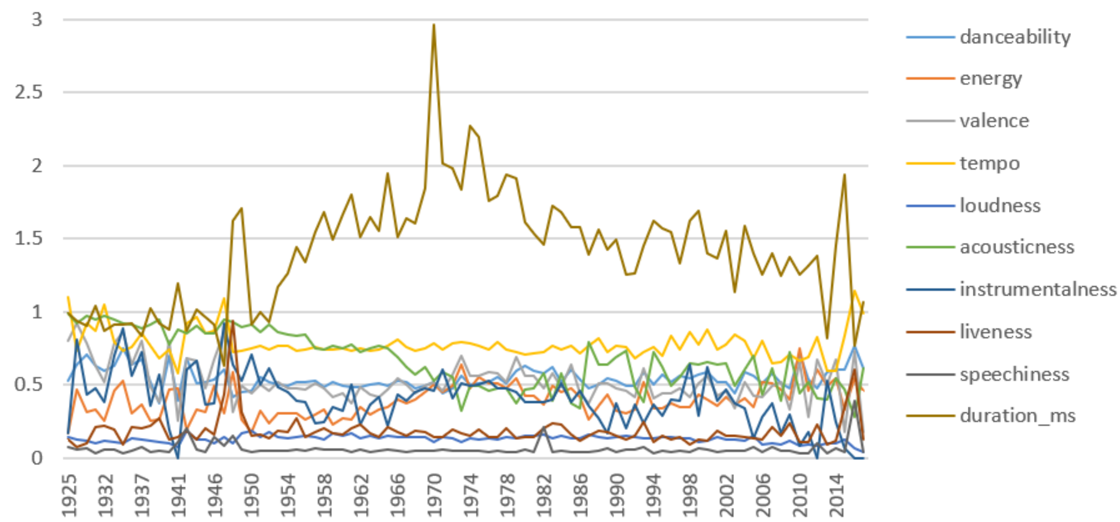


Figure 6: Characteristics of Jazz per year

However, we also find some anomalies with more data. *Table* 3 shows the influence of these artists on Norah Jones (Pop/Rock), their similarities and the genres each artist belongs to. According to *Table* 3 there is no obvious relationship between influence and correlation of artist in Pop/Rock. The reason might be that Pop/Rock is a big genre itself. From the perspective of music system, Pop music is developed on the basis of American popular music such as Blues, Jazz, Rock and R&B. For instance, a large proportion of Michael Jackson's (R&B) songs can be classified as Pop/Rock.

Table 2: The influence of influencers on Bob Dylan, their correlation coefficients and genres of influencers

Influencer name	Influence	Similarities	Genre
Sonny Terry	0.419970162	0.564698856	Blues
Mississippi Sheiks	0.766864316	0.619248512	Blues
Blind Boy Fuller	1.082601149	0.809011495	Blues
Hank Snow	1.75092982	0.555799846	Country
Johnny Cash	2.836068086	0.546448631	Country
Bill Monroe	3.032812353	0.630900184	Country
Hank Williams	4.99948725	0.78083013	Country
Cisco Houston	0.207653621	0.699211838	Folk
John Jacob Niles	0.268171062	0.7069571	Folk
Doc Watson	0.270404483	0.762647247	Folk
Odetta	0.275298517	0.818197602	Folk
Alan Lomax	0.275743036	0.786005169	Folk
Ewan MacColl	0.285640008	0.880937936	Folk
Billy Lee Riley	1.007297408	0.137111484	Pop/Rock
Buddy Holly	7.079100426	0.238907941	Pop/Rock
Little Richard	8.94174187	0.323498559	Pop/Rock
Elvis Presley	9.740188266	0.677291639	Pop/Rock

In addition, we notice that there is no obvious relationship between the influence of Vocal artists on Pop/Rock artists and the correlation. However, the similarity between them are generally high. This may be related to the vague positioning of Vocal singers, refers to the singer who takes his own singing talent and singing skills as the selling point. So in fact,

Table 3: The influence of influencers on Norah Jones, their correlation coefficients, and genres of influencers

Influencer name	Influence	Correlation	Genre
Phoebe Snow	0.599773244	0.858490347	Pop/Rock
Annie Lennox	0.681619165	-0.23855874	Pop/Rock
Sarah McLachlan	0.75203365	0.76736194	Pop/Rock
Rickie Lee Jones	0.843534107	0.820600617	Pop/Rock
Bonnie Raitt	0.870107477	0.934304358	Pop/Rock
Carole King	2.980210576	0.916031613	Pop/Rock
Elton John	5.561115162	-0.114147621	Pop/Rock
Joni Mitchell	6.480264695	0.943932196	Pop/Rock

it is more appropriate to Pop/Rock genre, and the overall similarity between is very high.

3.3 ”Contagious” Characteristics

In this part, we hope to identify musical features that can be easily affected by other music or musicians. Again we take Bob Dylan as an example. Bill Monroe (Country), Hank Williams (Country), and Woody Guthrie (Folk) are Bob Dylan’s influencers. If we take 0.1 for gap threshold, we can calculate the absolute value of difference for each characteristics between the two artists (data have been normalized). If the value is lower than 0.1, we consider the two artists relatively close in that characteristics, indicating that the influencer actually influences the music created by the follower in that characteristics. Therefore, such characteristics can be considered as more “contagious”.

We choose the features whose absolute value of difference is less than 0.1 to draw the radar maps, which are showed in $figure\ 7-9$. Then we can find that loudness, tempo and danceability are more ”contagious”, while energy, valence and key have greater difference, so they are considered to be less ”contagious”.

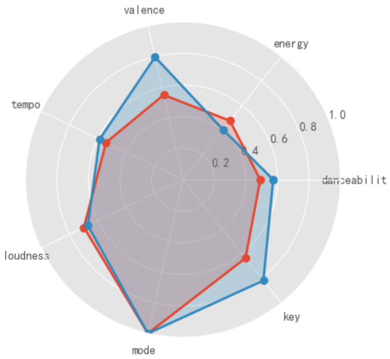


Figure 7: Radar map of Bob Dylan (red) and Bill Monroe (blue)

Figure 8: Radar map of Bob Dylan (red) and Hank Williams (blue)

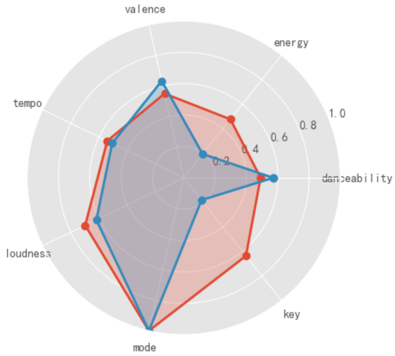


Figure 9: Radar map of Bob Dylan (red) and Woody Guthrie (blue)

4 Analysis of the Development of Music

4.1 Indicators and Representative Artists of Revolutions

There are sometimes revolutionary shifts in music, offering new sounds or tempos and changing previous music style. We assumed that revolutions happened in the years when the number of released songs surged. In this part, we take the genre Vocal as an example. $Figure$ 10 shows the counts of songs from all the genres per year and $figure$ 11 shows the counts of Vocal songs per year. According to $Figure$ 4, we can identify that different genres take their major leap at different times. For example, Pop/Rock became popular in 1960s, while Vocal and Jazz reach the peak in 1950s. What's more, the curve of Vocal and Jazz are of similar type. In consideration of Pearson correlation coefficient, Jazz is similar to Vocal. In addition, many active vocal singers in the 1950s have close links with Jazz. For example, Ella Fitzgerald was called Queen of Jazz, while Frank Sinatra was regarded as an idol by many jazz singers.

From $figure$ 11, it is clear that Vocal took its major leap around 1950. $Figure$ 12 and $figure$ 13 show how the main features of Vocal change every year. According to $figure$ 12, the attribute acousticness dropped in 1950s, which indicates the spread of microphone and the improvement of recording technology, when Vocal became popular. For singers who take his own singing talent and singing skills as the selling point, the change of acousticness offers Vocal singers much room to exercise their talent in developing a softer, more personal and more subtle style. In return, Vocal came into fashion.

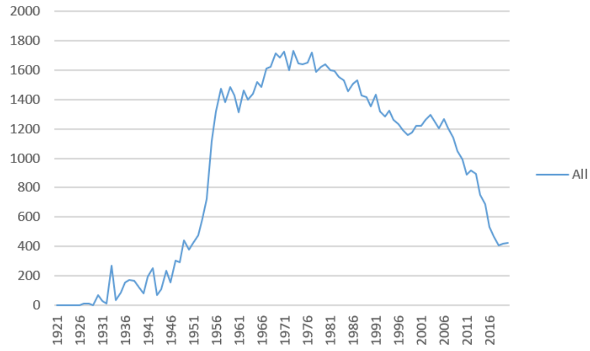


Figure 10: Counts of songs from all the genres per year

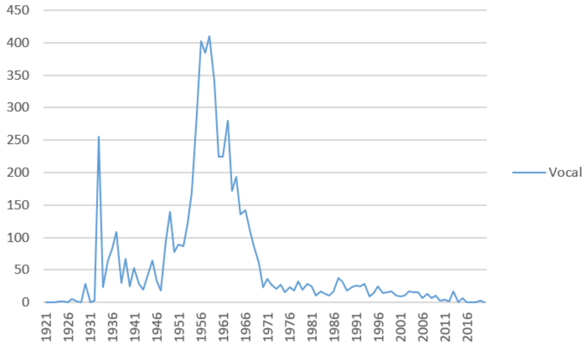


Figure 11: Counts of Vocal songs per year

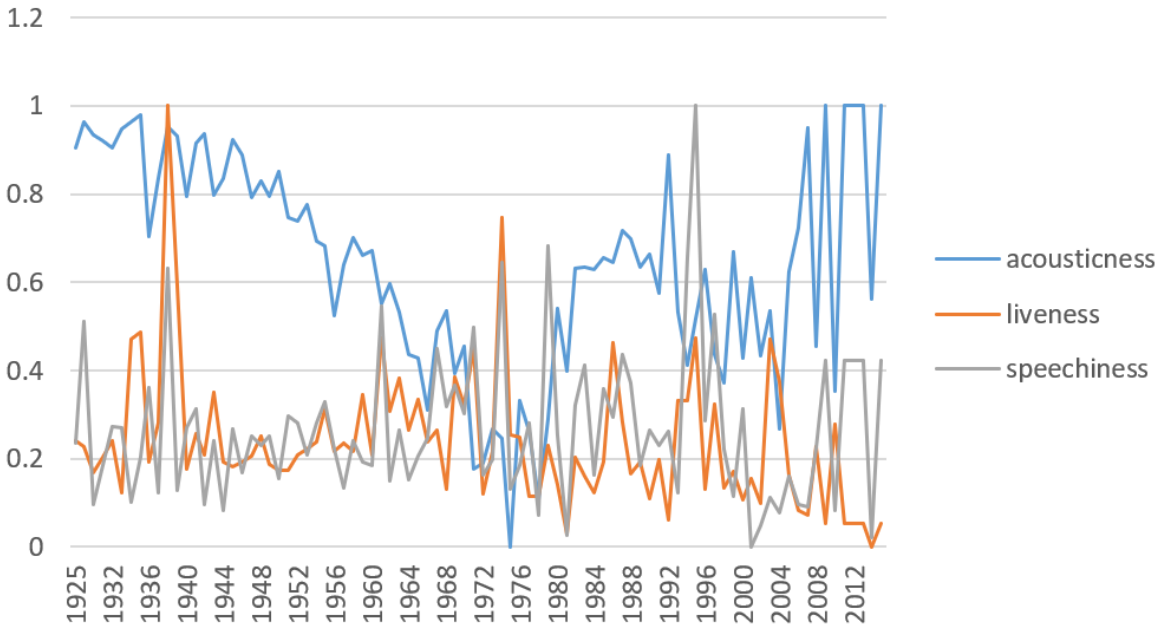


Figure 12: Some Characteristics of Vocal per year (1)

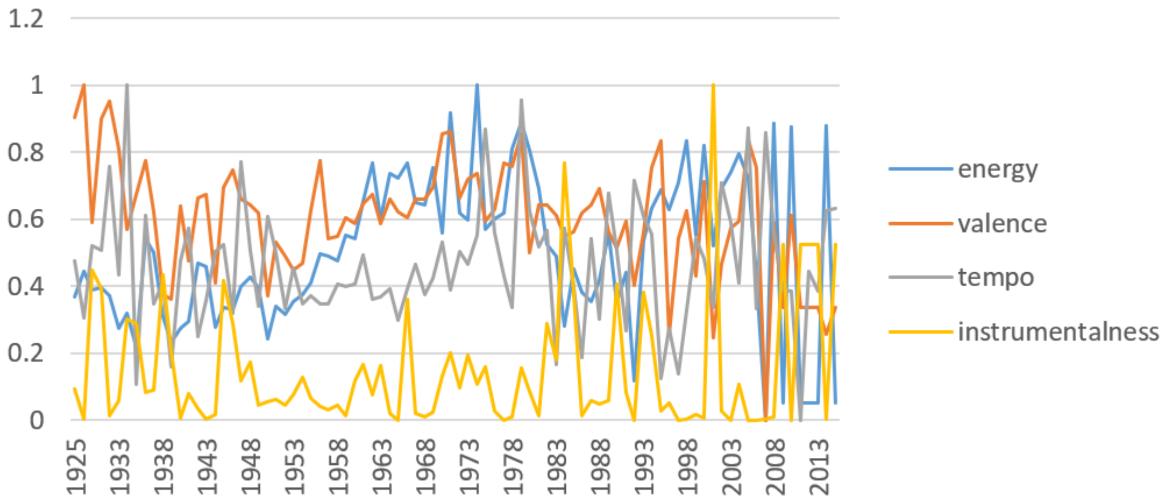


Figure 13: Some Characteristics of Vocal per year (2)

To find the representative singer in 1950s, we consider some influential Vocal singers and count their songs every year. Then we have $Figure$ 14 which represents counts of different Vocal singers per year. We can identify that Frank Sinatra is the representative of Vocal. $Figure-15$ is the subnetwork that is centered at Frank Sinatra. We find that Frank Sinatra has a great impact on those Vocal artists. During his peak period, he not only made a great success, but also influenced the development of Vocal. Therefore, Frank Sinatra represented the revolution in the network.

4.2 Changes of Genres and Artists Over Time

We continue the example mentioned before. According to $figure$ 12, the liveness and speechiness became lower in the 1950s to 1960s. And most of his songs were published in this period. Therefore, the level of liveness and speechiness might be the indicators that reveal a singer' s number of publications. What' s more, there was a high correlation between energy, valence, and instrumentalness, which reveal the artist' s style. Therefore, we can identify that, in the 1950s to 1960s, Frank Sinatra spent more time on recording and his style became more active.

From $figure$ 16, we also find that Frank Sinatra' s characteristics changed a lot after 1973. The liveness, energy, acousticness, and speechiness reached the peak. Because in 1973, Sinatra came out of his short-lived retirement. And he initially developed problems with his vocal cords during the comeback due to a prolonged period without singing. Therefore, his style changed greatly. From the correlation between his songs before and after that, we can identify the problem even made him look like a new person. Therefore, the characteristics mentioned above can be indicators that reveal the dynamic influencers.

4.3 Cultural Influence of Music

In this part, we take Jazz as an example. Through some research on history we can find that the outbreak of World War II (1939-1945) marked a turning point for jazz. In the United States, the war presented difficulties for previous jazz form. The military's need for shellac limited record production , which caused small groups of young musicians developed a more uptempo, improvisational style of jazz known as Bop. In this stage, the instruments are louder, the energy is stronger, and the music is played in a more complex pattern and at a faster tempo. $Figure\ 17-18$ show that some music characteristics such as energy and

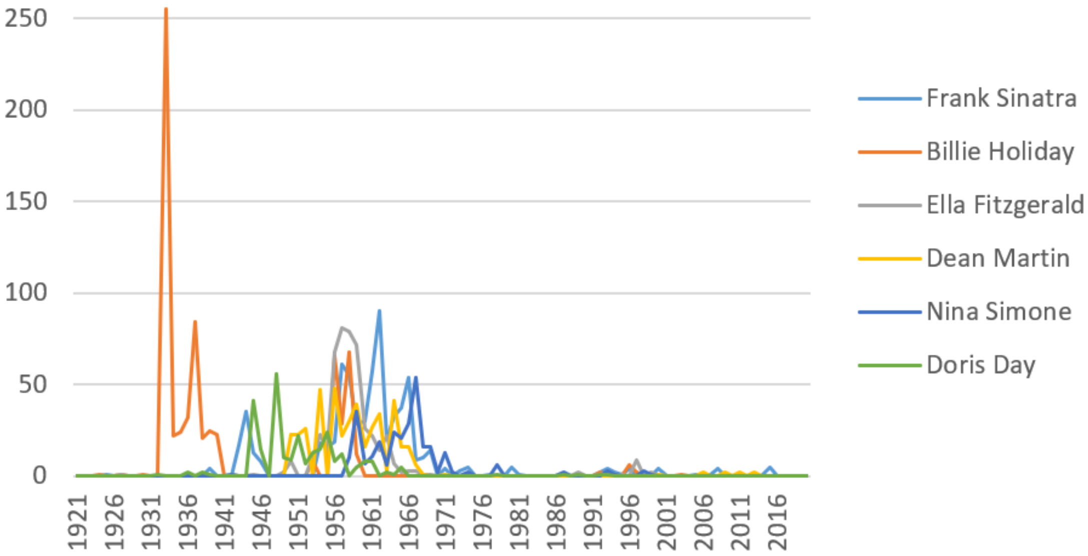


Figure 14: Counts of different Vocal singers per year

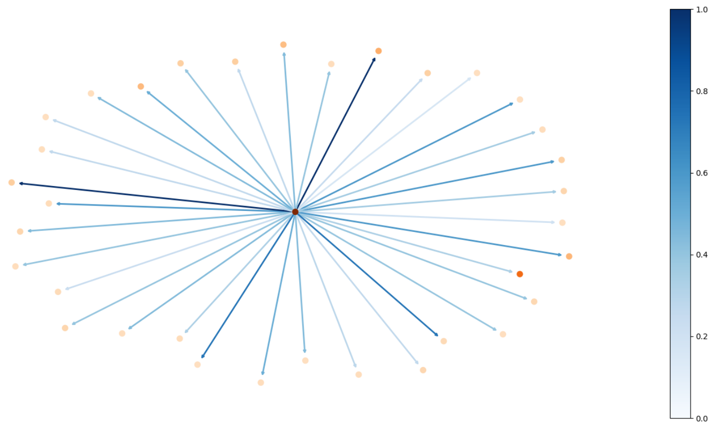


Figure 15: The subnetwork which is centered at Frank Sinatra

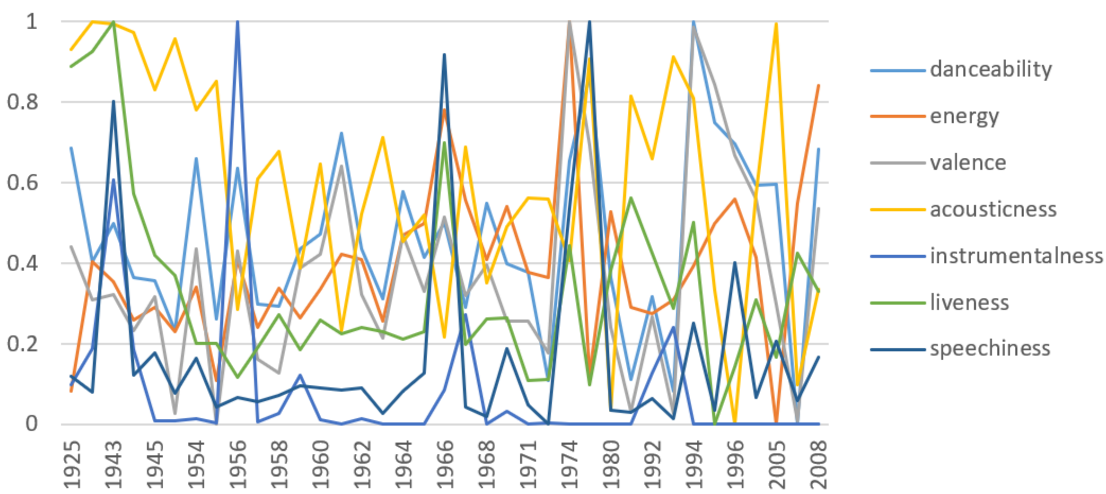


Figure 16: Characteristics of Frank Sinatra per year

tempo actually rose up during 1940-1947, which fits in with reality.

In the late 1960s, young listeners liked rock music, while older fans were no longer fascinated by modern Jazz, which was mostly abstract and emotionless. To make a living, Jazz musicians realized that they needed to absorb some elements of pop music from jazz, so Fusion Jazz began to emerge. Since then, jazz has been fully integrated with various genres, resulting in the classification of many songs not Jazz but others. Therefore, we can see the influence of Jazz computed by the directed network actually decreased. This trend is showed in *figure* 19. This also shows that the influence model we have established can indeed correspond to the development and evolution of music.

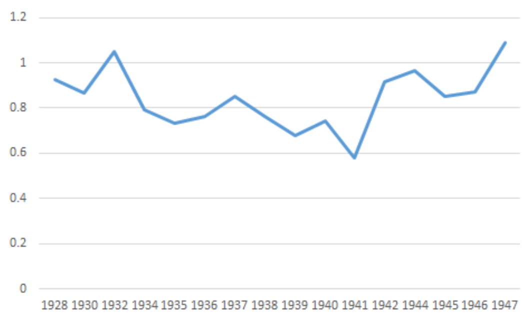


Figure 17: Tempo of jazz per year

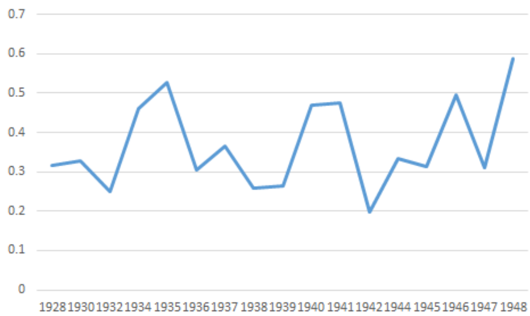


Figure 18: Energy of jazz per year

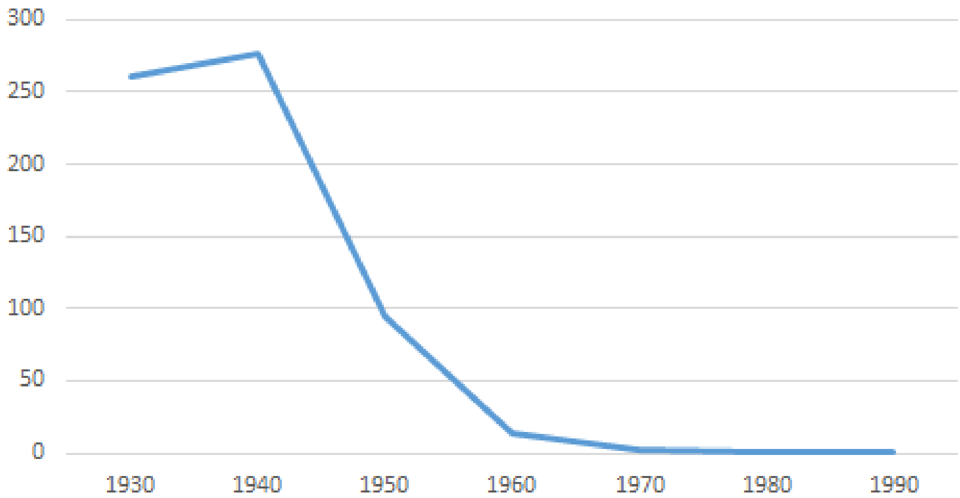


Figure 19: Influence of Jazz per year

5 Evaluation of the Model

5.1 Strengths

- Full Use of Given Data

  When calculating node' s influence, we make use of the multi-dimensional feature in the data, and fully mines the data information in the influence_data. In addition to using the in-degree and out-degree information in the figure, the ratio of the number of people in corresponding genre in the active_start year is also taken into account (R_a, refer to formula (1)). When calculating the edge weight, the influence between the same genre is different with the influence between different genres.

- Consistence of different models

  The results of our influence and similarity models are consistent. Especially after we considered the indirect influence in the directed netwok, the influence model accord better with the similarity model.

- Simplicity and efficiency of computation

5.2 Weakness

- Lack of weights in similarity model

  As defined above, the similarity model uses features, such as danceability, tempo, and valence, to measure the similarity between music and artists. However, in this model, all attribute plays the same role. In other word, the weight of different attribute is the same value. In the further study, we would improve our model by giving different weight value. Then the model will show that some features have more important impact on similarity.

6 Document to ICM Society

Our team have developed a model to understand the influence of music. This model is based on a directed network. Through the network, we can identify how the genres and artists affect each other, how genres' influence develop over time .What' s more, we also invent a method to measure the similarity between songs, artists, or genres. This has contribute to understand if the influencer really affect the follower.

Up to now, through the analysis on the given data set, Our model has got some satisfactory results. In terms of influence, the influence of mainstream artists on other artists is more consistent with the actual situation. It also reflects the influence of time dimension and the difference of influence between different genres and the same genre. What' s more, we even get influence of the artists themselves. In terms of similarity, by extracting the main features, we find some pairs of artists that influence each other and become more similar to each other. In addition, we also find some genres similar to each other, like Vocal and R&B, Pop/Rock and Avant-Garde. After analyzing the historical facts, the accuracy of our model has been confirmed.

In the future, we will apply the model to the analysis of current music trends and characteristics. Considering that our model is based on the given data set from 1921 to 2020, many genres have not been considered in the model. Therefore, before applying to the current analysis, we need to make some adjustments to our model, like considering the impact of the development of the Internet and so on. With more or richer data , maybe there are many artists which are not reported influence relationships, so we can consider generating the influence relationship in the network through the similarity of their songs' features. We hope that after the adjustment, the model can predict the evolution trend or characteristics changes of some genres in the next 10 years and provide development suggestions for some artists.

In terms of further study of music and its effect on culture , it is highly recommended that we can find some representative artists with great influence ,and relates their characteristics to the local cultural features. For example , Bob Dylan is a very representative artist who had great effects on the whole world's culture.

Thank you for reading our report. Your idea and suggestion are highly appreciated .

7 Reference

[1] M.K. Shan, F.F. Kuo and M.F. Chen, ”Music Style Mining and Classification by Melody”, Proc. of IEEE: ICME02, 2002.

[2] Matityaho and M. Furst, ”Neural Network Based Model for Classification of Music Type”, Proc. of 18th Conv. Electrical and Electronic Engineers in Israel, pp. 1-5, 1995.

[3] Bagci, Erzin, ”Boosting Classifiers for Music Genre Classification”, Signal Processing and Communications Applications 2006 IEEE 14th, pp. 1-3, 2006.

[4] M. Slaney and W. White, ”Similarity based on rating data”, Proc. Int. Symp. Music Information Retrieval (ISMIR’07), 2007.

[5] D. Bogdanov, J. Serr, N. Wack and P. Herrera, ”From low-level to high-level: Comparative study of music similarity measures”, Proc. IEEE Int. Symp. Multimedia (ISM’09). Int. Workshop Advances in Music Information Research (AdMIRe’09), pp. 453-458, 2009.

[6] K. West and P. Lamere, ”A model-based approach to constructing music similarity functions”, EURASIP J. Adv. Signal Process., vol. 2007, pp. 149, 2007.

[7] E. Pampalk, Computational models of music similarity and their application in music information retrieval, 2006.

[8] M. Slaney, K. Weinberger and W. White, ”Learning a metric for music similarity”, Proc. Int. Symp. Music Information Retrieval (ISMIR’08), pp. 313-318, 2008.

[9] Leo Breiman. Random Forests. Machine Learning,2001,45(1).

[10] Mohammad Jafarzadegan, Faramarz Safi-Esfahani, Zahra Beheshti. Combining hierarchical clustering approaches using the PCA method. 2019, 137:1-10.

Appendices

Here is a part of Python codes we used to construct the directed network.

```
import numpy as np
import pandas as pd
import networkx as nx
```

```
import matplotlib.pyplot as plt
import csv
import matplotlib as mpl

genre=["Avant-Garde","Blues","Children's","Classical","Comedy/Spoken","Country
                                        ","Easy Listening",
"Electronic","Folk","International","Jazz","Latin","New Age","Pop/Rock","R&B;"
                                        ,"Reggae","Religious",
"Stage & Screen","Unknown","Vocal"]

G=nx.DiGraph()

def initialization(): #initialization of the influence model
  influence_data=pd.read_csv("./2021_ICM_Problem_D_Data/influence_data.csv",
                                        encoding='unicode_escape')
  for index, row in influence_data.iterrows():
  #the establishment of the network
    influencer_id = row['influencer_id']
    influencer_main_genre = row['influencer_main_genre']
    influencer_active_start = row['influencer_active_start']
    follower_id = row['follower_id']
    follower_main_genre = row['follower_main_genre']
    follower_active_start = row['follower_active_start']

    G.add_edge(int(influencer_id),int(follower_id))
    G.nodes[int(influencer_id)]['id']=int(influencer_id)
    G.nodes[int(influencer_id)]['main_genre']=influencer_main_genre
    G.nodes[int(influencer_id)]['active_start']=int(influencer_active_start)
    G.nodes[int(follower_id)]['id']=int(follower_id)
    G.nodes[int(follower_id)]['main_genre']=follower_main_genre
    G.nodes[int(follower_id)]['active_start']=int(follower_active_start)

  for node in list(G.nodes()):
  #Calculate the influence of each node and it's outgoing edges
    lambda1=1
    lambda2=0.5
    C=1.2
    Attc_sum=0
    for adj in list(G.adj[node]):
```

```
      Attc_sum += lambda1 * 1 / G.in_degree(adj)
      bc_sum=0
    for adj1 in list(G.adj[adj]):
      bc_sum += 1/G.in_degree(adj1)
      Attc_sum += lambda2*bc_sum

    node_genre=G.nodes[node]['main_genre']
    node_year=G.nodes[node]['active_start']

    all_people_num=0
    genre_people_num=0
    rank_in_genre=[]

    for item in list(G.nodes()):
      if G.nodes[item]['active_start']==node_year:
        all_people_num += 1
      if G.nodes[item]['main_genre']==node_genre:
        genre_people_num += 1
        rank_in_genre.append([G.out_degree(item),item])

    rank_in_genre = sorted(rank_in_genre,reverse=True)
    index = rank_in_genre.index([G.out_degree(node),node])

    G.nodes[node]['influence'] = Attc_sum * genre_people_num / all_people_num
    G.nodes[node]['attc_sum']=Attc_sum
    G.nodes[node]['index']=index
    G.nodes[node]['ratio']=genre_people_num / all_people_num

    for adj in list(G.adj[node]):
      G[node][adj]['weight'] = G.nodes[node]['influence'] * (1/np.log2(G.
                                    in_degree(adj)+1)) * (C if G.nodes[node
                                    ]['main_genre']!=G.nodes[adj]['
                                    main_genre'] else 1)

def influence_matrix():  #Calculate the influence matrix between genres
    influence_frame=pd.DataFrame({a:[0.0]*20 for a in genre} ,index=genre)
    for (u,v,wt) in G.edges.data('weight'):
      influence_frame[G.nodes[v]["main_genre"]][G.nodes[u]["main_genre"]] +=
                                    wt
```

```
    influence_frame.to_csv('./csv1/influence.csv',index = True)

def draw_subnetwork():  #draw the subnetwork
    draw_list=[754032,66915]
    G_new = nx.DiGraph()
    edge_list=[]
    node_list=[]
    edge_colors=[]
    node_colors=[]
    for max_node in draw_list:
      node_list.append(max_node)
      node_colors.append(G.nodes[max_node]["influence"])
      G_new.add_node(max_node)
      edge_sift=[]
      for adj in list(G.adj[max_node]):
        edge_sift.append([G[max_node][adj]['weight'],adj])
      edge_sift=sorted(edge_sift,reverse=True)

      temp=[]
      for k in range(0,len(edge_sift),60):
        node = edge_sift[k][1]
        node_list.append(node)
        node_colors.append(G.nodes[node]['influence'])
        edge_list.append((max_node,node))
        edge_colors.append(edge_sift[k][0])
        G_new.add_edge(max_node,node)
        temp.append(node)

      for node in temp:
        edge_sift=[]
        for pre in list(G.predecessors(node)):
          edge_sift.append([G[pre][node]['weight'],pre])
          edge_sift=sorted(edge_sift,reverse=True)
        for k in range(0,len(edge_sift),3):
          pre = edge_sift[k][1]
          node_list.append(pre)
          node_colors.append(G.nodes[pre]['influence'])
          edge_list.append((pre,node))
          edge_colors.append(edge_sift[k][0])
```

```
            G_new.add_edge(pre,node)

    edge_color=[]
    node_color=[]
    edge_colors=np.asarray(edge_colors)
    node_colors=np.asarray(node_colors)
    min1=np.min(edge_colors)
    max1=np.max(edge_colors)
    min2= np.min(node_colors)
    max2=np.max(node_colors)
    for x in edge_colors:
      edge_color.append(float(x - min1)/(max1- min1))
    for x in node_colors:
      node_color.append(float(x - min2)/(max2- min2))

    print(node_color)
    print(edge_color)
    options = {"node_size": 50, "alpha": 1}
    pos = nx.layout.spring_layout(G_new)
    nodes = nx.draw_networkx_nodes(G_new, pos, nodelist=node_list,node_color=
                                    node_color,cmap=plt.cm.Oranges,vmin=-0.
                                    2,vmax=1.0,**options)
    edges = nx.draw_networkx_edges(G_new,pos,edgelist=edge_list,arrowstyle="->",
                                    arrowsize=6,edge_color=edge_color,
                                    edge_cmap=plt.cm.Blues,alpha=1,width=2,
                                    edge_vmin=-0.2,edge_vmax=1.0,)
    pc = mpl.collections.PatchCollection(edges, cmap=plt.cm.Blues)
    pc.set_array(edge_color)
    plt.colorbar(pc)
    ax = plt.gca()
    ax.set_axis_off()
    plt.show()

if __name__ == '__main__':
    initialization()
    influence_matrix()
    draw_subnetwork()
```