Problem Chosen	2021 MCM/ICM Summary Sheet	Team Control Number
D		2107087

The Network Model of Music Influence Based on EWM and PCA

Summary

Music has a significant role in human culture, affecting the development of social process, thus, human beings need to understand the evolution of music in depth. We build a model of music influence network and introduce the entropy weight method, principal component analysis, afterwards a quantitative analysis of the influence of previous music on new music and artists is conducted, with music characteristics and other indicators taken into consideration.

In Task 1, we first use the social network model to establish a directed and weighed network of musical influence. Nodes represent genre categories and artists when arcs represent the influence relationship. After that the arcs is given weighs so that we can control the direction of the network indirectly. Then, we take each network of genres as a subnetwork, with the introduction of the degree centrality, betweenness centrality and closeness centrality. Taking betweenness centrality for instance, we analyze the impact and information exchange efficiency of the subnetwork and its core members. **In Task 2**, we take the music characteristics of each artist as a vector and introduce the Standardized Euclidean distance. Using this distance to measure similarity, there comes a conclusion that artists of the same genre are more similar than artists of different genre.

In Task 3, the entropy weight method (EWM) is put forward to calculate the weight of each music characteristic in different genres, then we study the fluctuation of its weight with different genres. At last, characteristics such as 'valence' and 'instrumentalness' which have great weight fluctuations are selected as the criteria to distinguish various genres. After that we calculate the weight of each music characteristic at different time periods. Combined with the music characteristic scores of each artist in Task 2, we use the Standardized Euclidean distance to measure the similarity, and get the trend of genre in every ten-year interval. In this way, we can analyze the process of evolution. **In Task 4**, taking 'folk' as an example, the EWM is also used to get the weight of music characteristics of each follower. With principal component analysis, we quantify the impact of music characteristics and find that some of them have stronger influence, such as 'valence' and 'danceability'.

In Task 5, we study the weight fluctuations of each music characteristics in different time periods, and then select characteristics with large weight fluctuations, such as 'energy' and 'tempo', as representatives of musical evolution. According to the indicators in Task 1, the artists are divided into high artistic influencers, inheritors and developers, and independent musicians, and we also figure out some important artists represented by Pete Seeger, the 'Father of American Folk Music'. **In Task 6**, we take 'Jazz' as a precedent to analyze its evolution process. Based on the above questions, we calculate the weight of jazz characteristics in different time periods and consider it as a dynamic influencing factor. With that the style changes of jazz at the interval of every decade are analyzed. **In Task 7**, we analyze the cultural influence of jazz in different times in combination with music characteristics. In addition, by calculating the subnetwork parameters of 'Electronic', we draw a conclusion that parameters such as 'pageranks' and 'betweenness centrality' are always increasing. The closer we get to the present, the faster they grow, which reveals the promoting effect of scientific and technological development on the development of music.

Finally， we conduct a sensitivity analysis by changing the genre variable, and it is pointed out that when we are in the face of abundant data, multi-dimensional analysis should be carried out in time stages from factors such as region and religion, so as to make the model more reasonable and have practical significance.

Keywords: Music Influence Network; Music Characteristic; EWM; PCA

Contents

1 Introduction

1.1 Problem Background

Music has been part of human societies since the beginning of time as an essential component of cultural heritage. We have been asked to develop a method to quantify musical evolution. Our goal is to understand and measure the influence of previously produced music on new music and musical artists.

Our team has been identified by the Integrative Collective Music Society to develop a model that measures musical influence. This problem asks us to examine evolutionary and revolutionary trends of artists and genres. Our teams need to explore the evolution of music through the influence across musical artists over time, by doing the following:

- Create a network of musical influence and develop parameters. Create a subnetwork of influencer network.
- Develop measures of music similarity and answer whether artists within genre are more similar than artists between genres.
- Compare similarities and influences between and within genres. Find the distinguishes of genres, the patterns over time, the connections between genres.
- Indicate whether the similarity data suggest that the identified influencers in fact influence the respective artists.
- Identify characteristics that signify revolutions in musical evolution and artists represent revolutionaries.
- Analyze the influence processes of musical evolution in one genre. Identify indicators and explain how the genre or artist changed.
- Express information about cultural influence of music. Explain the effects of social, political or technological changes identified within the network.

1.2 Our Work

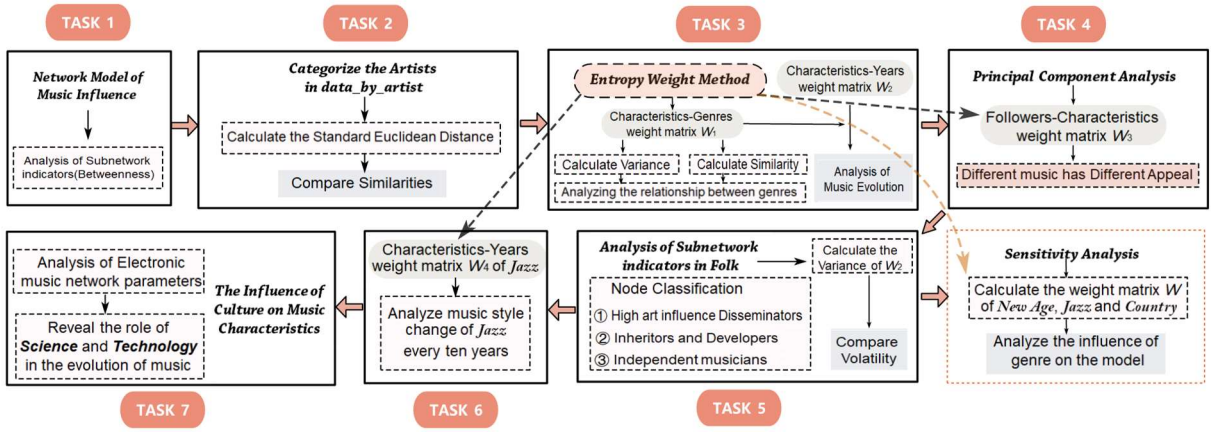


Figure 1: Flow chart of our work

Task 1： This question needs to create the music influence network and its sub-network. Based on influence_data, we create the network of musical influence by using social network model. Where, the node represents the artist and music genre, and the size of the genre node represents the number of artists in the genre. In addition, arcs are used to indicate liking. The direction of the arc can be controlled indirectly by assigning the weight to it. Then, we create subnetworks of the network and explore the influence of the corresponding core members by Betweenness Centrality.

Task 2： This question needs to develop measures of music similarity. Since music characteristics data of data_by_artist is highly volatile, we use standardized Euclidean distance to

measure the similarity between artists. The smaller the Euclidean distance, the greater the similarity. In this way, the similarity of artists can be compared.

Task 3： This question studies the relationship between different music genres and the changing rules of music genres over time. First, preprocess data_by_artist and use entropy weight method to give weight to the music characteristics of different genres. Based on the music similarity model, compare the similarity of different genres. Then, based on data_by_year, take 10-year time interval. We use the entropy weight method to give weight to the music characteristics in different time periods, calculate the similarity, and get the development of each genre in each time interval.

Task 4： This question mainly studies the contagiousness of each music characteristic. Combined with the influence_data and data_by_artist, we can obtain the comprehensive influence formula and the principal component expression, by using principal component analysis. According to the coefficient of comprehensive influence formula, the influence of music characteristics can be judged.

Task 5： This question needs to identify important characteristics and artists.We can judge the importance of a characteristic by how much it fluctuates from year to year. The larger the fluctuation, the more important the characteristic. Based on the three metrics established in Task 1, important artists can be grouped into three categories.

Task 6： Similar to Task 3, the difference is that this question studies a genre. Take Jazz as an example, the same method can be used to study the change rule.

Task 7： This question we mainly from the time background and technology these two aspects of music.

2 Assumptions and Justifications

- **Assumption 1:** All data in the list given by the question are accurate and valid. There are no model errors caused by objective statistical problems.
- **Assumption 2:** The number of artists in different years sorted out from the list given by the question can reflect the genre-population situation of the period, ignoring the fact that the number of some genres in the list is less than the actual situation.
- **Assumption 3:** Hypothesized that when Scat Sing appeared, there would be a decrease in ‘instrumentalness’ and/or ‘speechiness’ because the player imitated the instrument using human voice
- **Assumption 3:** Music characteristics can reflect the overall style of music in different genres and in different periods, so as to compare the similarity between genres and the popular trend in a certain time.

3 Notations

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

Table 1：Notations mentioned in this paper

Notation	Meaning
C_D	Degree Centrality
C_B	Betweenness Centrality
C_C	Closeness Centrality
sim	Similarity
S	Synthesis Score

4 Task 1：A Directed and Weighed Network of Musical Influence

In this chapter, based on the social network model [1], we create a directed and weighed network of musical influence and its subnetworks. Its role is to connect influencers and followers, and identify music genres at the same time. The identification of core members is an important function in complex network analysis. Typically, these members have broader relationships, which can affect more nodes. These extensive and dense relationships give them higher visibility in the network, so these nodes generally reside in the prominent position of the network, that is, the center. Based on this, we can use centrality indicators to analyze and discuss the influence of influencers (nodes or node groups) in the network.

4.1 The Establishment of Indicators

（1）Degree centrality： Degree centrality represents the number of nodes directly adjacent to the node in the network. For music influence networks, degree centrality represents the number of people affected. The degree centrality is defined as follows：

$$C_D(i) = d(i) = \sum_j^N x_{ij} = \sum_j^N x_{ji} \tag{1}$$

Where, i represents the node under study; j represents other nodes in the network; N represents the total number of nodes in the network; x_{ij} represents the number of connections between i and j .When there is a connection between i and j , x_{ij} =1 ; when there is no connection between i and j , x_{ij} =0 .

The degree centrality of a particular graph is limited by the size of its graph, that is, the number of nodes in the network. To compare the degree centrality of graphs of different sizes, we need to standardize it, in other words, the absolute node degree divided by the maximum possible value, the formula is as follows:

$$C_D^*(i) = \frac{d(i)}{N-1} \tag{2}$$

（2）Betweenness centrality： The measurement of network centrality evaluates the degree of relationship participation of nodes in the whole network through the structure of integrated nodes. Betweenness centrality is concerned with how a node adjusts or controls the relationship between two nodes that are not directly connected. **With betweenness centrality as the measurement index, it reflects the efficiency of information exchange between influencers and followers under the control of music influence network**, which is defined as follows：

$$C_B(i) = \sum_{j<k} \frac{d_{jk}(i)}{d_{jk}} \tag{3}$$

Where， d_{jk} represents the number of shortest paths between nodes j and k ; $d_{jk}(i)$ is the number of paths through i ; $C_B(i)$ represents the sum of the probabilities that node i is on the shortest path of any two nodes.

（3）Closeness centrality： It is used to measure the average distance from one node to another. When the distance between a node and other nodes is very short, the average geodesic distance of the node is very small. These nodes have a more direct impact on other nodes. In a music influence network, an influencer, whose average distance from other followers is small, will have an impact more quickly than an influencer, whose average distance from other followers is large. The reciprocal of degree centrality is defined as closeness centrality of node A, which is defined as follows：

$$C_C(i)=\frac{1}{d(i)}=\frac{N}{\sum_{j}^{N} d_{ij}} \tag{4}$$

Where, d_{ij} is the distance from node i to node j. The nodes with higher betweenness centrality can better control the transmission of information, while the nodes with higher closeness centrality can better observe the information flow.

4.2 Whole Network Model

In the music influence network, we will examine the relationship between influencers and followers and identify music genres. In this network, each artist or each music genre is a node, and the relationship between the three is represented by an arc. For illustration purposes, a brief relationship of the network is shown in Figure 2.

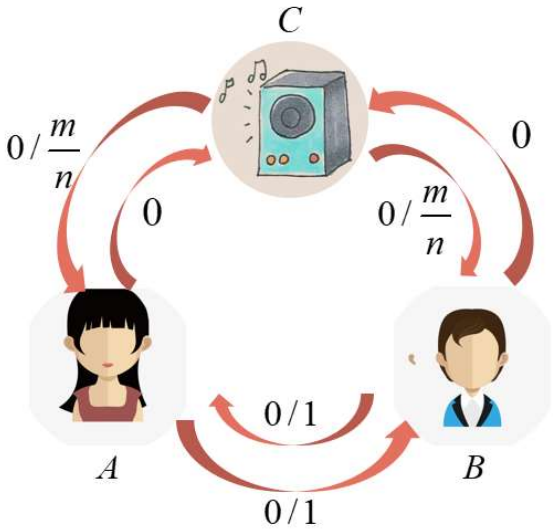


Figure 2: Concise diagram

In Figure 2, A and B represent any two artists (perhaps influencers or followers), and C represent music genres. There are two types of relationship arcs in a network: One is the arc between artists. If A likes B, then the arc is directed from B to A and the weight is 1, otherwise it is 0. The other is the arc between the artist and the music genre. If A likes n artists, and m of them belong to a certain music genre, then the weight of the arc is m/n. The weight of the arc from the artist to the genre is always 0. Thus, the network orientation can be realized indirectly by weighting the arc. When the weight of the arc is 0, it means there is no pointed direction arc. When the weight of the arc is non-0, it means there is a pointed direction arc. The weight of the arc is also used in the calculation of the index.

After processing the influence_data, it is found that there are 20 music genres in all the data. While each follower's influencers may belong to a different musical genre, most are consistent with their own. Therefore, the follower can be considered to belong to the genre indicated in the data. The classification results of music genres of artists in the directed network are shown in Figure 3.

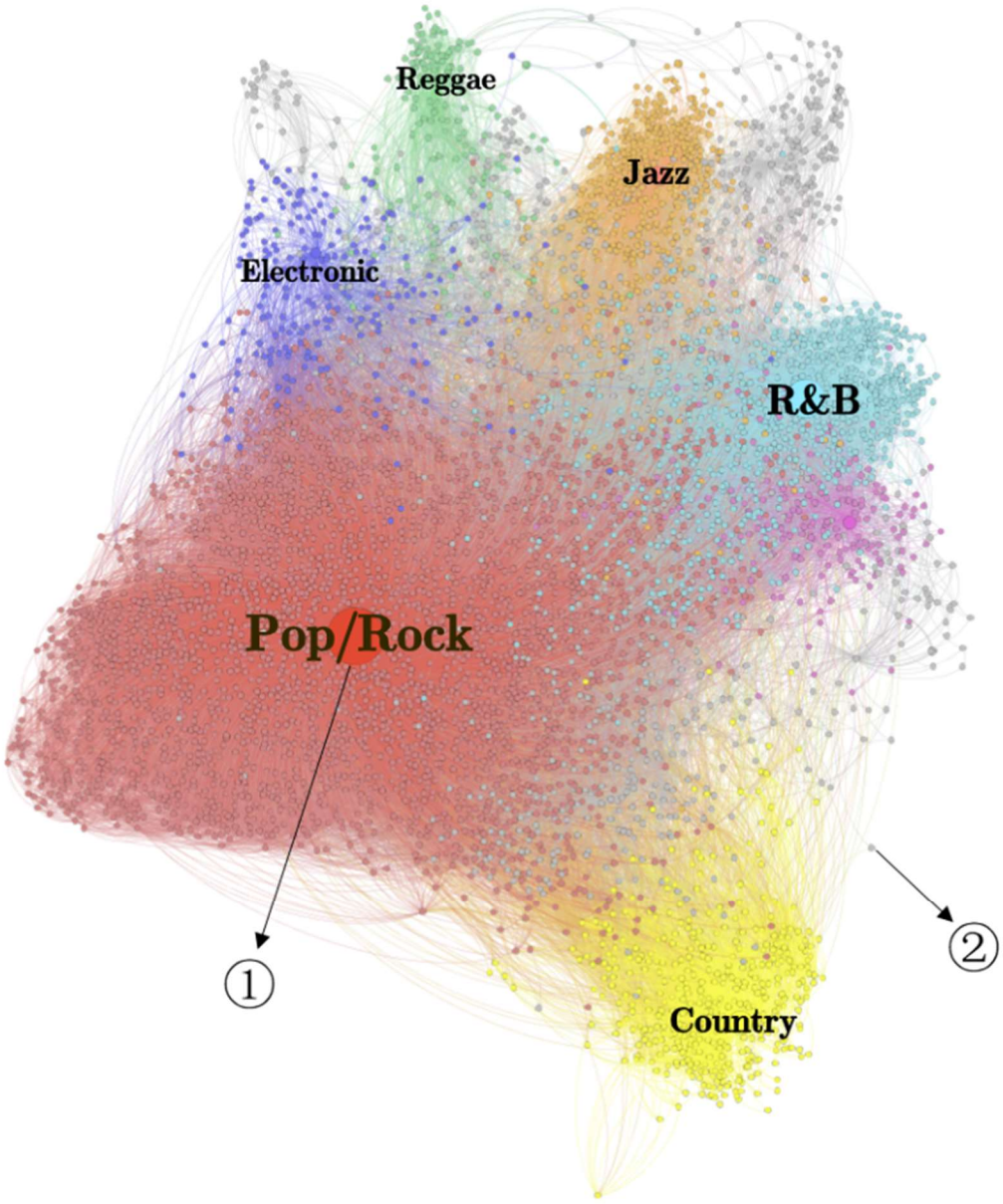


Figure 3：A directed and weighed network of musical influence

In Figure 3, each large node (e.g.①) represents a music genre, and the number of nodes represents the number of music genres (only some genres are marked in the figure). The size of a node depends on its degree centrality. The larger the degree centrality, the larger the node. Each small node (e.g.②) represents an artist. Different colored areas represent artists belonging to different genres, and the size of the area indicates the number of people belonging to that genre. The lines between the small nodes have the same meaning as the lines between the artists in Figure 2; The line between a small node and a large node has the same meaning as the line between an artist and a musical genre in Figure 2.

4.3 Subnetwork

In fact, each area with the same color in Figure 3 is a subnetwork, which can be regarded as a community. The subnetworks of some genres are shown in Figure 4.

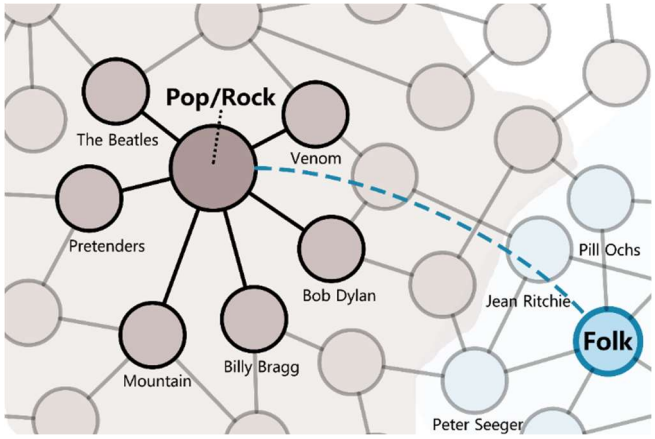


Figure 4: The subnetwork of partial genres

Three parameters degree centrality, betweenness centrality and closeness centrality, were proposed in the previous chapter. Because of their similarity, all of them can be used to evaluate the influence of core members. Here, we only take the betweenness centrality as an example to study the influence of core members.

We use 'Outdegree' to indicate the number of times each genre's artists have been identified as influencers. The more times, the greater the value of Outdegree.After calculation, the relationship between Outdegree and betweenness centrality can be obtained. Some results are shown in Figure 5.

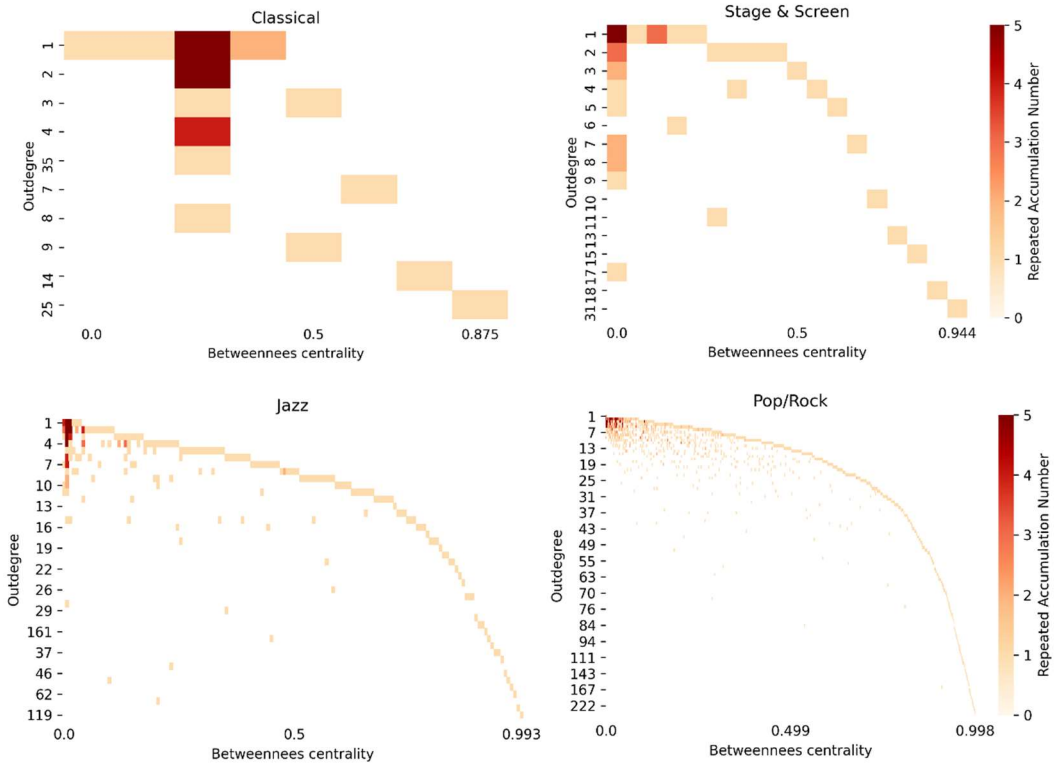


Figure 5: Outdegree-betweenness centrality heatmap of some music genres

In Figure 5, the number of squares represents the number of artists in the genre. The more squares, the greater the number of artists, the greater the genre. The colors of the squares represent the number of artists with the same betweenness centrality, and the darker the color, the greater the number of artists.

In addition, within each music genre, the more times an artist is regarded as an influencer, the greater his betweenness centrality will be and the more important the artist is, which is

consistent with the reality. And the larger the number of artists in a genre, the slower the curve drops. When Outdegree values are the same, the curve with slow decline has a higher value of Betweenness centrality. It shows that in the case of the same number of influencers, the greater the influence an artist has in a larger music genre, the more important it is. Moreover, it can be seen that the larger a genre will be, the greater its maximum Betweenness centrality is. Since the value of Betweenness represents the information exchange efficiency of the core members in the network, it indicates that the greater the genre, the higher the influence of the core members will be.

5 Correlation Analysis

5.1 Task 2：Music Similarity Measurement Model

Music similarity measurement is a measure to evaluate the degree of similarity between two artists comprehensively. The closer two artists are, the greater the similarity measure between them will be. In general, people usually use 'distance' to measure the degree of similarity between two samples, namely Euclidean distance. However, due to the large fluctuation range of the data given in the question, the results obtained by using simple Euclidean distance are not accurate. Therefore, we adopt standardized Euclidean distance for vectors with large differences in each dimension.

5.1.1 Data Preprocessing

（1）The division of music genres: After comparing data_by_artist with the influence_data of the known music genre in Task 1, we can divide the artists in data_by_artist into genres.

（2）Standardized processing of data: Since the distribution of music characteristics of each artist in data_by_artist is different, it is necessary to standardize each music characteristics first, and make its mean value and variance equal. Finally, we can obtain the standard component with mathematical expectation of 0 and variance of 1, and its formula is as follows:

$$X^* = \frac{X - m}{s} \tag{5}$$

Where, X^* represents the normalized value; X represents the value before standardization; m is the mean of the components; s is the standard deviation of the component.

5.1.2 Calculation of Standardized Euclidean Distance

In order to study the similarity of artists and the relationship between genres, we take the artist Billie Holiday in data_by_artist as an example. The artist belongs to the genre of 'Country'. Using the standardized Euclidean distance formula [2], the standardized Euclidean distance of two n-dimensional vectors $a(x_{11}, x_{12}, \cdots, x_{1n})$ and $b(x_{21}, x_{22}, \cdots x_{2n})$ can be expressed as:

$$d_{12} = \sqrt{\sum_{k=1}^{n} \left(\frac{x_{1k} - x_{2k}}{s_k} \right)^2} \tag{6}$$

The meaning is that the smaller the standardized Euclidean distance, the more similar the two artists will be.

Consider each music characteristics of an artist as a component of the above n-dimensional vector. We calculate the average standardized Euclidean distances between the vector of this artist and the vectors of other artists in the same genre, and the average standardized Euclidean distances between the vector of this artist and the vectors of artists in different genres. The results are shown in the following table:

Table 2: Average standardized Euclidean distance

Genre	Standardized Euclidean distance	Genre	Standardized Euclidean distance

Avant-Garde	1.54	Jazz	1.29
Blues	0.98	Latin	1.09
Children's	**0.89**	New Age	1.93
Classical	**1.79**	Pop/Rock	1.14
Comedy/Spoken	1.20	R&B	1.10
Country	**0.83**	Reggae	1.06
Easy Listening	1.16	Religious	1.20
Electronic	1.31	Stage & Screen	**1.76**
Folk	1.10	Vocal	1.24
International	1.14	Unknown	**0.85**

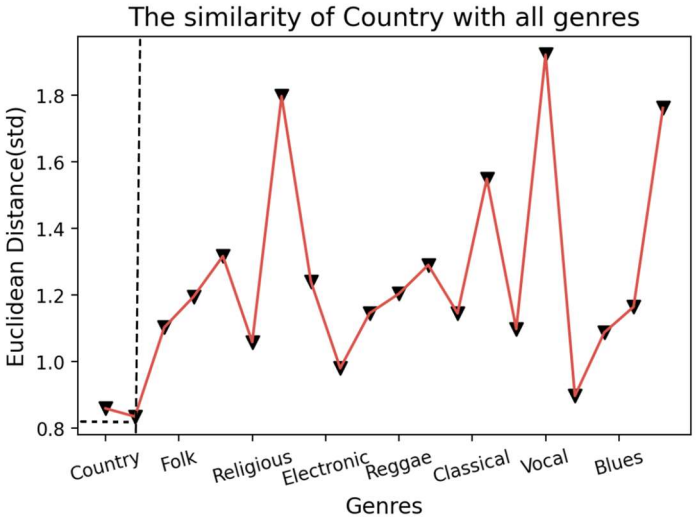


Figure 6: Line chart of average standardized Euclidean distance

Based on Table 2 and Figure 6, it can be seen that the standardized Euclidean distance between this artist and artists of his genre is the smallest, only 0.83, while the standardized Euclidean distance between this artist and artists of other genres is larger. The smaller the Euclidian distance, the greater the similarity of the artists. This explains that artists of the same genre are more similar than artists of different genres. In addition, the standardized Euclidean distance between the artist and artists of some genres is close, such as 1.79 of Classical genre and 1.76 of Stage & Screen genre, which indirectly indicates that artists of these two genres may have certain similarities.

5.2 Task 3：Associations Between Genres

5.2.1 Objective Weight Evaluation Method——Entropy Weight Method（EWM）

In order to study the factors that distinguish different genres, we first need to know how important different music characteristics are to different genres. Entropy weight method can be used to calculate the weight of different music characteristics in each genre.

Based on data_by_artist, taking genre of 'Country' as an example, the data of some of its music characteristics are shown in Table 3.

Table 3: Music characteristic data of some artists of Country genre

Artist_id	danceability	energy	valence	···	speechiness
816890	0.611846	0.448513	0.685619	···	0.080248
583959	0.547398	0.373049	0.494527	···	0.024065
⋮	⋮	⋮	⋮	⋮	⋮
3455945	0.529265	0.671686	0.267972	···	0.028241

The 10 music characteristics of all artists in this genre have been standardized in Section 5.1. Set the processed data as $y_{ij}\left(i=1,2,\cdots,402;\ j=1,2,\cdots,10\right)$, then the information entropy calculation formula of item j of music characteristics is：

$$E_j=-\frac{\sum_{i=1} p_{ij}\ \mathrm{In}\ p_{ij}}{\mathrm{In}\ n} \tag{7}$$

Where, $0<E_j<1$, $j=1,2,\cdots,10$; $p_{ij}=\frac{y_{ij}}{\sum_{i=1}^{n} y_{ij}}$. If $p_{ij}=0$， then define $p_{ij}\ \mathrm{In}\ p_{ij}=0$.

The larger E_j is, the larger the gap between this music characteristics among different artists will be, the more information this music characteristics will provide, and the greater the weight of the music characteristics will be. Calculate the weights of music characteristics of the 20 genres, and some results are shown in Table 4:

Table 4: The weight of music characteristics of some genres

Characteristic	Country	Electronic	New Age	···	Children's	Variance
danceability	0.004070	0.015131	0.073188	···	0.008360	0.000597
energy	0.016534	0.020073	0.120866	···	0.005556	**0.002289**
valence	0.011793	0.048123	0.196883	···	0.009926	**0.003396**
tempo	0.005749	0.009096	0.027599	···	0.002692	0.000041
loudness	0.001915	0.004165	0.013944	···	0.001389	0.000365
key	0.060510	0.098625	0.235119	···	0.320557	**0.004608**
acousticness	0.050097	0.297715	0.049754	···	0.014167	**0.007053**
instrumentalness	0.657804	0.228297	0.06013	···	0.477049	**0.042312**
liveness	0.031078	0.083048	0.167544	···	0.093747	0.000981
speechiness	0.160451	0.195727	0.054973	···	0.066557	**0.007786**

From Table 4, it can be seen that the six music characteristics of 'energy', 'valence', 'key', 'acousticness', 'instrumentalness' and 'speechiness' have larger variances. Variance is a measurement to measure the degree of dispersion of data. The greater the variance, the greater the volatility of data, that is, the greater the degree of differentiation of this music characteristics in different music genres. Therefore, the above six music characteristics can be used as factors to distinguish different genres. In addition, it can also be verified by horizontal comparison of various music characteristics.

For the genre of 'Country', characteristics such as 'instrumentalness' and 'speechiness' are more important, while characteristics such as 'danceability', 'tempo' and 'loudness' are less important. It shows that the genre has low dancability, slow rhythm and low loudness, which is consistent with the characteristics of stable rhythm, simple melody, cordial and enthusiastic without losing popular elements.

For the genre of 'Electronic', its 'acousticness', 'instrumentalness' and 'speechiness' are important characteristics, which match its usual use of a keyboard to simulate sound effects. Similarly, it can be seen that the 'valence' characteristic of New Age is significantly higher than that of other music genres. On the other hand, 'Children's' characteristics such as 'key' and 'instrumentalness' are very important, and they all match its playful features.

Horizontal comparison of music characteristics of different genres in Table 4 can also realize the comparison of similarities among genres. Similar to Task 2, consider the music

characteristics of each genre as the components of an n-dimensional vector. Calculate the standardized Euclidean distance between two of 20 genres, as shown in heatmap 7.

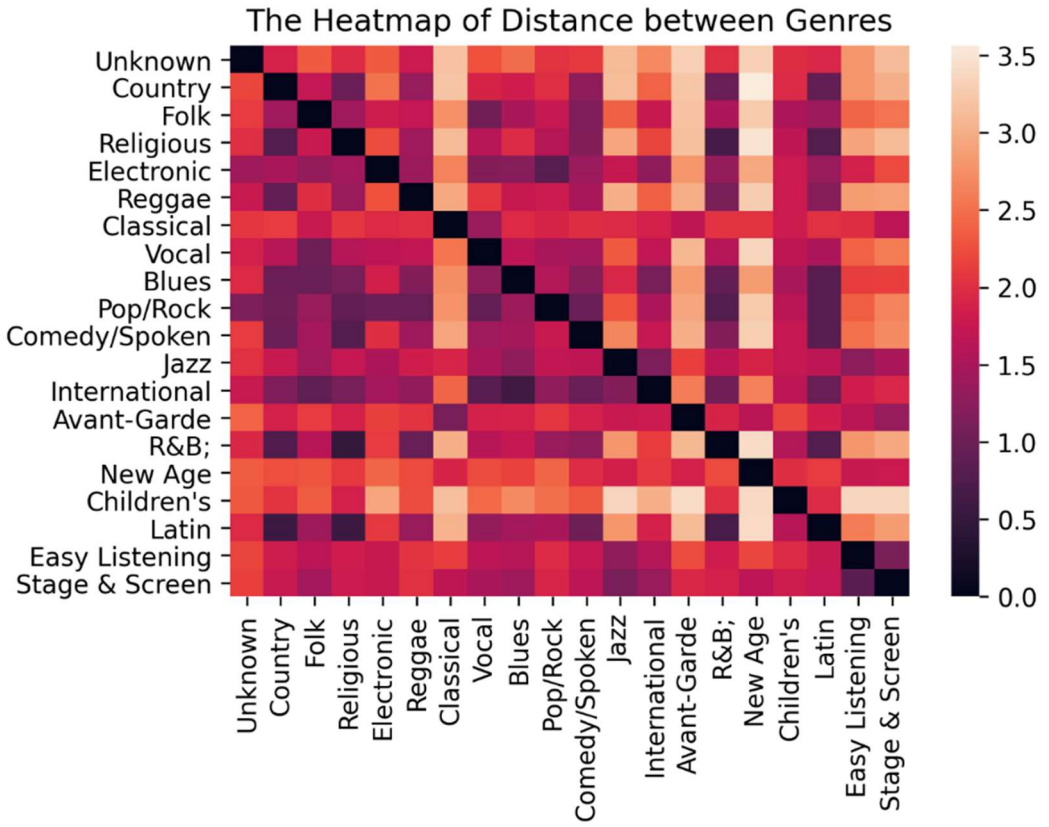


Figure 7: Standardized Euclidean Distance heatmap

In Figure 7, the horizontal and vertical coordinates are genres, and each genre intersects to form a color block. The gradation of the color block represents the size of the standardized Euclidean distance between the two genres of intersection. The darker the color is, the greater the standardized Euclidean distance and the greater the similarity between genres will be. Since each genre is always the most similar to itself, the blocks on the diagonal have the darkest color. In addition, it can be seen that Reggae has great similarities with R&B, Blues and Latin, and we speculate that Reggae may have originated from one of these three genres, or it may have been highly integrated with them in its later development； Jazz and Blues, R&B and Pop/Rock are also similar, so we can guess that there is a certain fusion between the two. Latin, on the other hand, is similar to Country, Religious and R&B, which may be due to the fact that the development of Latin is closely related to different geographical locations and Religious factors.

5.2.2 Changes of Genres Over Time

Based on data_by_year, take 10 years as the time interval, and there are ten 10-year intervals from 1921 to 2020. Therefore, a total of 10 music characteristic matrices can be obtained by drawing the music characteristics within each 10-year interval into a matrix. Taking 1920-1930 as an example, its matrix is shown in Table 5.

Table 5: The weight of music characteristic in 1921-1930

Year	danceability	energy	valence	···	liveness	speechiness
1921	0.425661	0.236784	0.425495	···	0.215814	0.077258
1922	0.480000	0.237026	0.534056	···	0.238647	0.115419
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1929	0.647005	0.240513	0.638208	···	0.235988	0.485278
1930	0.510382	0.327300	0.618458	···	0.217320	0.079702

Similar to 5.2.1, we use entropy weight method to calculate the weight of each music characteristic of these 10 matrices, and get the weight of each music characteristic in every 10-year interval. Some results are shown in Table 6.

Table 6: The weight of music characteristics in each time interval

Characteristic	1921-1930	1931-1940	···	2011-2020	Variance
danceability	0.063177	0.025419	···	0.009045	0.001194724
energy	0.139949	0.040351	···	0.001893	0.002788099
valence	0.048089	0.013809	···	0.019544	0.001074332
tempo	0.135921	0.032098	···	0.002251	0.00166656
loudness	0.099679	0.019430	···	0.000476	0.000938547
key	0.135030	0.554904	···	0.530807	0.036067749
acousticness	0.009797	0.003027	···	0.142596	0.011995657
instrumentalness	0.018729	0.043479	···	0.133456	0.002411082
liveness	0.093256	0.025996	···	0.127744	0.001844025
speechiness	0.256373	0.241488	···	0.032187	0.008789806

Based on Table 4 and the column vectors of music characteristics in Table 6, we use the similarity measurement model of Task 2 to calculate the standardized Euclidean distance between each genre and each time interval. The purpose is to study the similarities between the characteristics and genres of popular music in a certain decade, so as to infer the development of each genre. Some of the results are shown in Table 7.

Table 7: Genre - Time interval standard Euclidean distance

Genre	1921-1930	1931-1940	1941-1950	···	2011-2020
Country	3.404561	2.961812	3.464101	···	2.498856
Folk	2.790075	2.677342	3.184912	···	2.307504
Religious	3.228104	3.026091	3.458080	···	2.197868
Electronic	2.832337	2.491778	3.115647	···	1.858337
⋮	⋮	⋮	⋮	⋮	⋮
Stage&Screen	2.374462	1.817903	2.061407	···	2.275594

In order to better reflect the changes of each genre in each time interval, we adopt Euclidean Distance Similarity formula for mapping, and use standard Euclidean distance to define the Similarity. The formula is as follows:

$$sim(x,y)=\frac{1}{1+d(x,y)} \tag{8}$$

Where, $sim(x,y)\in[0,1]$; $d(x,y)$ is standard Euclidean distance. If the $d(x,y)$ of a certain genre is smaller in a certain time interval, the $sim(x,y)$ is larger. It shows that in this period of time, the development of this genre is better. In order to highlight the results, the similarity data is further standardized, and the formula is as follows:

$$sim^{*}=\frac{sim-sim_{\min}}{sim_{\max}-sim_{\min}} \tag{9}$$

Draw the radar map of the similarity between genres and years. As shown in Figure 8.

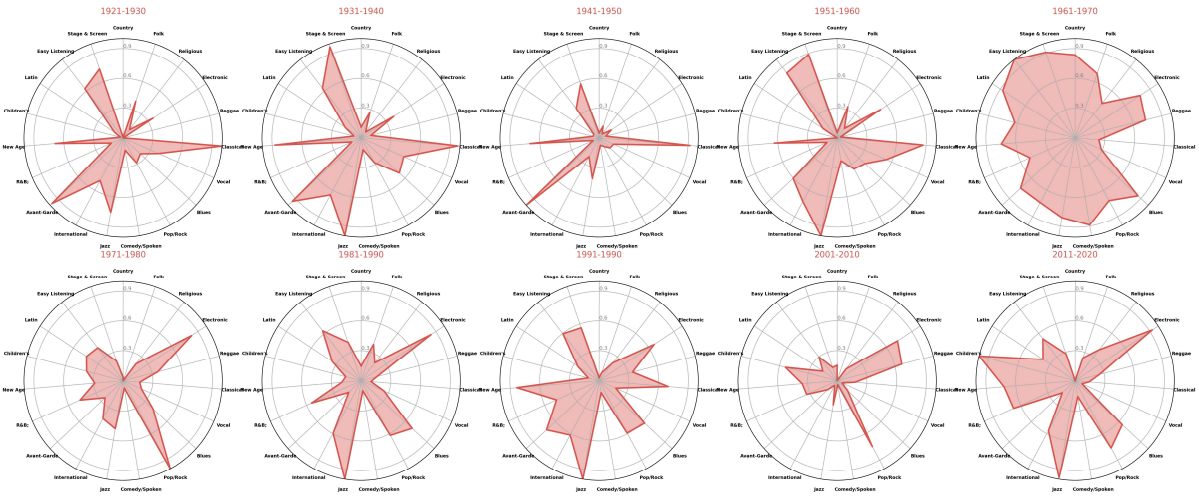


Figure 8: Radar map of the similarity between genres and years

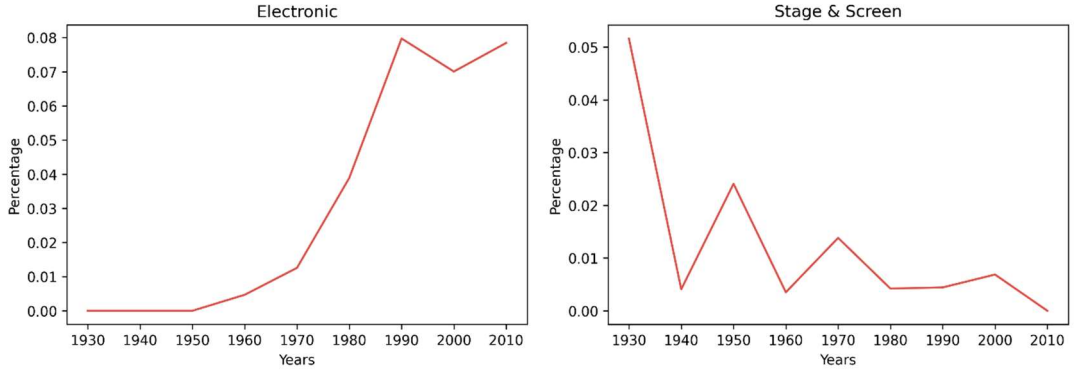
In the figure, if the similarity of a genre is closer to 1 within a time interval, the radar chart shows that there is a significant bulge on this axis, indicating the more development of the genre.

As can be seen from the figure, Classical flourished from 1920s to 1950s in the 20th century, and suddenly declined in 1960s.Avant-Garde flourished in the 1920s and 1940s, and then declined in the 1970s after experiencing fluctuations. Easy Listening and Stage & Screen both emerged in the 1920s and also declined in the 1970s.Jazz experienced a short decline in the 1940s, 1970s and 1980s, and maintained a good development trend in the other decades with high popularity.

It can be found that the radar map from the 1920s to the 1960s generally presents an iron triangle shape. The 1960s was the transition point. Since then, the lines of the radar map tended to be smooth, and the overall graph showed an outward expansion trend. It was no longer confined to the iron triangle shape in the map, but more inclined to the uniform distribution of circles. This shows that since the 1960s, various genres and styles of music began to rise and get greater development, and the music types tend to be diversified. Among them, R&B emerged in the 1960s and ushered in a new wave in the 2010s.

After the 1970s, Electronic, Pop/Rock, and the Blues became a craze that is still very popular today. The rest genres were only briefly popular in the 1960s and 1970s, and have not developed significantly since. In addition, it is worth noting that Children's began to rise in the 21st century and developed rapidly in the following two decades. This may have something to do with people starting to focus on education.

In order to further study the variation rule of genre over time, based on influence_data, we plot the proportion of the number of artists of each genre in different time periods to study the development of genre. The more people a genre has, the better development momentum of the genre in this time pired. We show only representative genres here, as shown in Figure 9.



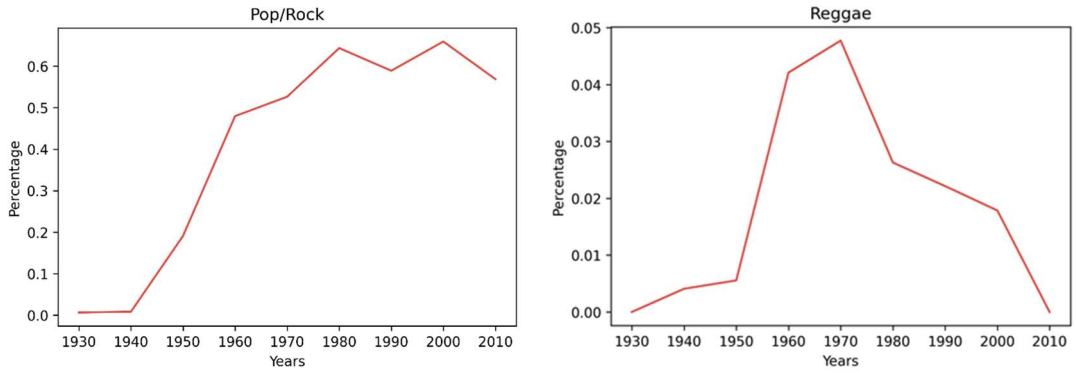


Figure 9: The proportion of each genre in different time periods

As can be seen from Figure 9, Electronic emerged in the 1970s and developed rapidly and maintained a good trend in the following 20 years. Stage & Screen rose in the 1920s and 1930s, and then declined in the 1970s after a period of fluctuations. Like Electronic, Pop/Rock started in the 1960s and 1970s and stayed popular; Reggae, on the other hand, had only a brief rise in popularity in the 1960s and 1970s, with no significant growth in the rest of the decade. The above conclusions are in high agreement with Figure 8, indicating the high accuracy of the research results.

6 Task 4：Principal Component Analysis（PCA）

In Task 1, we established a directed and weighed network of music influence and divided all the artists in the influence_data into music genres. Based on the music characteristics information of artists provided by data_by_artist. Task 2 studies the similarity between artists of the same genre and artists of different genres. The results show that artists of the same genre are more similar than artists of different genres. In addition, it has been found in the study of Task 1 that although the artists that the followers like belong to different genres, the genres of these artists are mostly the same as those of the followers, with only a few differences. **Therefore, the above studies all show that the identified influencers do influence the music created by the followers.**

In order to further study the influence of each music characteristic, we quantified it and used principal component analysis to calculate the comprehensive score of each music characteristic. Based on the classification of music genres of various artists in Task 1 influence_data, take the Folk genre as an example:

First, find out all the followers belonging to the Folk genre. Each follower has its own influencers. These influencers have corresponding music characteristics. Therefore, each follower has an influencer—the music characteristic matrix. Same as 5.2, we adopt the entropy weight method. After calculation, the weight of all the followers' music characteristics of their influences under the Folk genre is obtained. The results are shown in Table 8.

Table 8: The weight of music characteristics of Folk genre

follower	danceability	energy	valence	···	speechiness
1	0.003956	0.075844	0.029562	···	0.109423
2	0.011200	0.043848	0.016093	···	0.064657
3	0.007044	0.057961	0.013747		0.092055
⋮	⋮	⋮	⋮	⋮	⋮

Using Table 7, we extracted 5 principal components through principal component analysis. We use $y_k\left(k=0,1,\cdots 4\right)$ for each of the five principal components, $x_i\left(i=0,1,\cdots,9\right)$ for the nine

music characteristics of danceability, energy, valence, tempo, loudness, mode, key, acousticness, instrumentalness, liveness and speechiness. Combined with the coefficient scores of each principal component, we obtain the expression of each principal component as follows:

$$\begin{cases} y_0 = 0.5117x_0 + 0.0018x_1 + 0.4929x_2 + \cdots - 0.1153x_9 \\ y_1 = -0.0244x_0 + 0.0723x_1 - 0.0016x_2 + \cdots + 0.3102x_9 \\ y_2 = -0.0366x_0 + 0.6352x_1 + 0.0502x_2 + \cdots + 0.0364x_9 \\ y_3 = 0.0717x_0 - 0.4067x_1 + 0.0611x_2 + \cdots + 0.7912x_9 \\ y_4 = -0.0012x_0 - 0.1777x_1 + 0.1950x_2 + \cdots + 0.0402x_9 \end{cases} \tag{10}$$

After further calculate, we obtained the comprehensive influence expression of contagiousness, which is expressed as:

$$S = 0.3602y_0 + 0.1826y_1 + 0.1290y_2 + 0.1051y_3 + 0.1015y_4 \tag{11}$$

Substitute each principal component expression into S:

$$\begin{aligned} S=&0.1825x_0 + 0.0531x_1 + 0.2099x_2 + 0.1663x_3 + 0.1730x_4 \\ &- 0.1786x_5 + 0.1284x_6 - 0.0119x_7 + 0.0702x_8 + 0.1070x_9 \end{aligned} \tag{12}$$

Where，the magnitude of the absolute values of each coefficient represents the contribution of each music characteristic to the contagiousness contagion. The greater the absolute value of the coefficient is, the more contagious the music characteristic will be. Therefore, these seven music characteristics of danceability, valence, tempo, loudness, mode, key, acousticness and speechiness are more contagious. From largest to smallest, it can be ranked as: valence, danceability, key, loudness, tempo, acousticness, speechiness.

7 Task 5：Dentification of Key Factors

7.1 Important Music Characteristics

A characteristic that marks a revolution in the development of music, usually with large fluctuations in different time periods. The variance represents the degree of fluctuation of data. The greater the variance, the greater the degree of fluctuation of data. Therefore, on the basis of Table 6, we draw the variance histogram of the fluctuation of music characteristics over time, as shown in Figure 10.

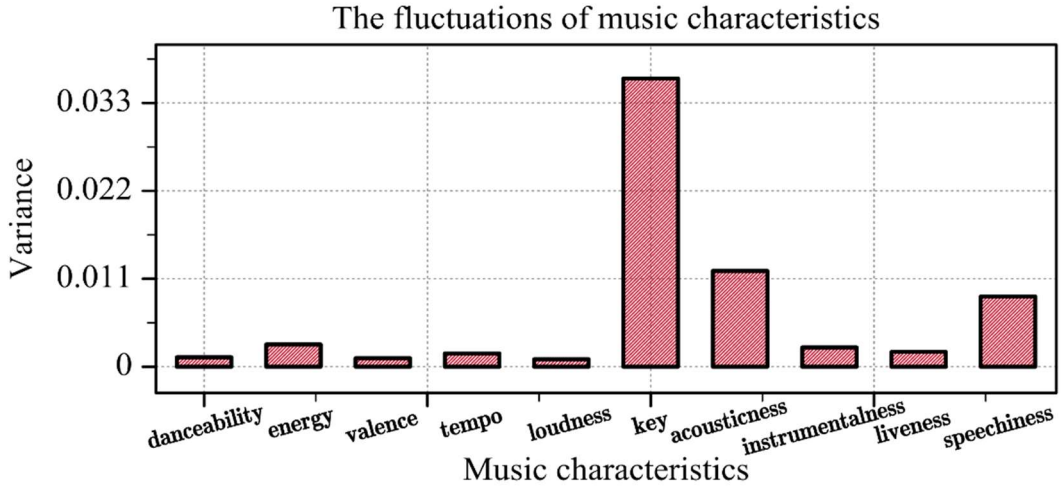


Figure 10: Histogram of variance

As can be seen from Figure 10, energy, tempo, key, acousticness, instrumentalness and liveness have large variances. This shows that these music characteristics change significantly over time, and these changes often represent significant changes and are therefore important.

7.2 Important Artists

Artists who represent major leaps are often important players in the network of musical influence, so it is critical to study them.

7.2.1 Classification of Important Artists

In Task 1, we propose three indicators to measure the influence of artists in the network, namely, degree centrality, betweenness centrality, and closeness centrality, which respectively have different meanings. After the above research on artists' data, according to the indicators, we find that the important artists in the music influence network can be roughly divided into three categories [3]:

（1）High artistic influencers:

Those with high artistic influence tend to have high degree centrality, betweenness centrality, and closeness centrality. They have many followers, which indirectly reflects the high degree of influence on art resources, and are often the core of a music genre in the music influence network.

（2）Inheritors and developers:

The inheritors and developers usually have high betweenness centrality, and closeness centrality. They are mostly cut points, located at the edge of the subnetwork, and are an important channel for a subnetwork to communicate with the outside world. They are closely connected with high artistic influencers and have high efficiency of information transmission.

（3）Independent musicians:

The independent musicians had higher closeness centrality, but lower degree centrality and betweenness centrality. They are directly related to and more influenced by high artistic influencers. However，they have little contact with other artists, and their artistic creation is mostly independent.

7.2.2 Important Artists of Music Influence Network

Take artists of the Folk genre for example. Substitute the values of music characteristics of artists belonging to Folk in data_by_artist into Formula (11). After calculating, we obtain the comprehensive influence score of each artist.The higher the comprehensive influence score, the more important the artist is. Then, select some of the artists with high score. Using the number and weight of each arc in Task 1, calculate the degree centrality, betweenness centrality, and closeness centrality. The results are shown in Table 9.

Table 9: Data of important artists of Folk genre

Artist	Degree Centrality	Betweenness Centrality	Closeness centrality	S
Jean Ritchie	17	0.000004	0.273584	0.1231
Shirley Collins	16	0.000004	0.170186	0.0544
Mike Seeger	14	0.000042	0.177547	0.0313
Tom Russell	7	0	1	0.0311
Kate Wolf	4	0.000004	0.172568	0.0256
Spirit of the West	9	0.000002	**0.666667**	0.0215
Michael Hurley	11	0.000002	0.230128	0.0194
Loudon Wainwright III	19	0.000066	0.170483	0.0181
Pete Seeger	**52**	**0.000277**	**0.291864**	0.0173
Damien Rice	23	**0.000148**	**0.647059**	0.0165

According to the classification of important artists in 7.2.1, compared with Table 9, it can be found that ‘Pete Seeger’ has high degree centrality, betweenness centrality, and closeness centrality; ‘Damien Rice’ has high betweenness centrality, and closeness centrality; ‘Spirit of

the West' only has a high degree of closeness centrality. Therefore, it can be inferred that in the music influence network, these three important artists are respectively high artistic influencers, inheritors and developers, and independent musicians. Taking Pete Seeger as an example, we map the relationship between music and influence. As shown in figure 11.

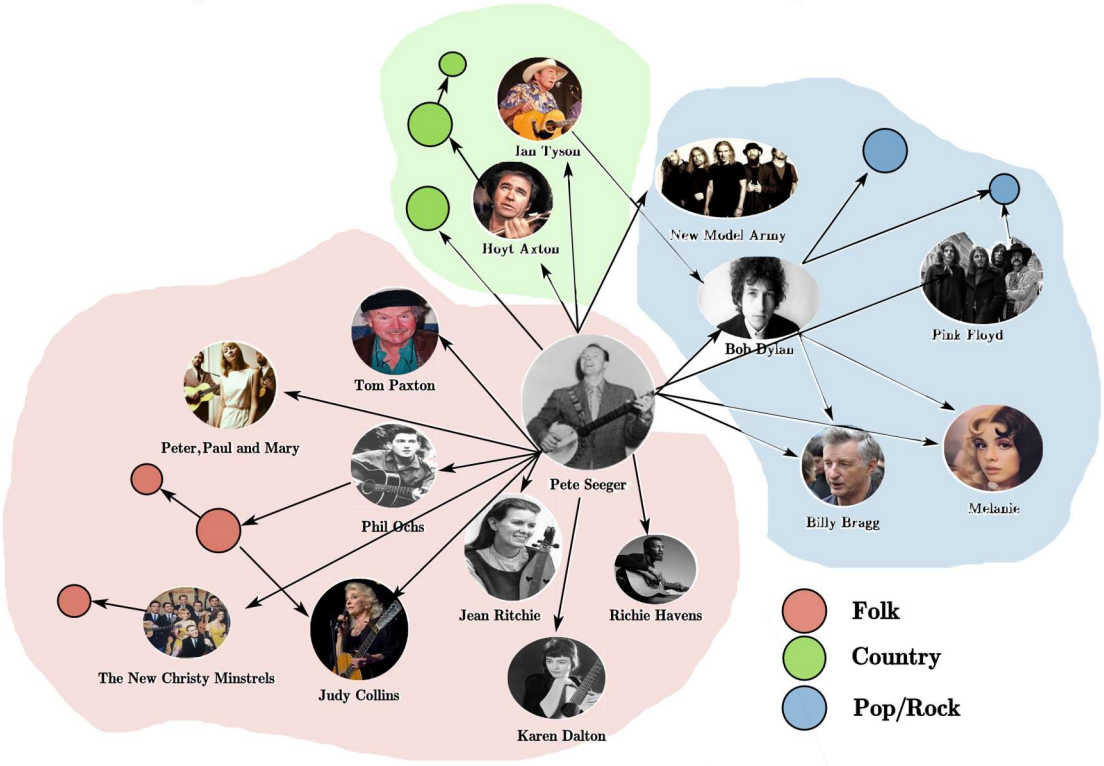


Figure 11: Pete Seeger's music influence diagram

In addition, by studying the full_music_data, we found that Pete Seeger published famous songs such as <*We Shall Overcome*> and <*Keep Your Eyes on the Prize BBB*>, which played a crucial role in the development history of Folk genre. Folk music regained its broad popularity and momentum, leading to the second revival movement [4]. This directly confirms that he is a high artistic influencer in the music influence network.

8 Task 6: Dynamic Indicators and Changes of Jazz Over Time

This question requires us to analyze the influence processes of musical evolution that occurred over time in one genre. Similar to Task 3, both questions need to study the process of genre(s) over time. The difference is that Task 3 studies the change of all genres over time, while this question studies only the change of one genre over time. Here, we take Jazz as an example to conduct related research.

Based on the classification of each genre of artists in influence_data, and synthesizing the music characteristics of artists belonging to Jazz genre in data_by_year, we take 10 years as the time interval in the same way. Then, plot the characteristics of each artist into a matrix at intervals of 10 years. Since there is no Jazz information for the years 1921-1930 and 2011-2020, we can only get a total of 8 music characteristic matrices. Adopt the entropy weight method similiarly. We calculated the weight of music characteristics in each time interval, and the result is shown in Figure 11.

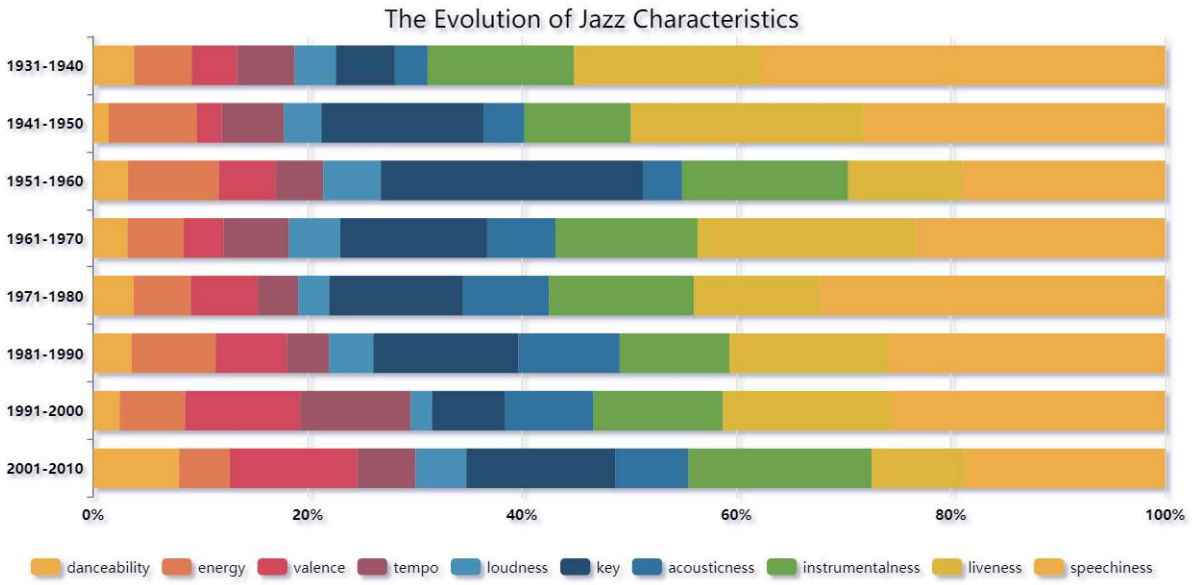


Figure 12: Weight of music characteristic

As can be seen from Figure 12, the weight of each music characteristic changes significantly with the passage of time. **Since these music characteristics carry information about changes in genres, we use the weight of each music characteristic as an indicator of dynamic influencers.**

By studying the change rule of the above dynamic indicators, we can get the change situation of Jazz genre over time. Further reference to the history data of Jazz for comparison, we come to the following conclusions:

- In the 1930s, danceability had a greater weight than other neighboring eras. At that time, big bands and swing music were popular, and the music had a stable rhythm, which was more suitable for dancing. Under this environmental background, Jazz tends to be strong in rhythm and strength, which is consistent with its strong characteristics of dance.
- In the 1940s, the weight of tempo and key increased compared with that of the 1930s, while the weight of danceability and instrumentalness decreased. At that time, 'Bebop' prevailed. This kind of Jazz is fast, twice as fast as easy swing, and is always in flux, not suitable for dancing, which coincides with this change in weight. In addition, as improvisation is one of the prominent features of this Jazz, 'loud refrain' is usually added to increase the rhythm's melodic quality, which is consistent with the changing trend of key weight. Since imitating instrumental sounds is unique to bebop, this leads to an increase in vocals. Instrumentalness has dropped in weight in the 1940s; Tempo's weight dropped in the 1950s compared to the 1940s, in line with the slow, melancholy 'Cold Jazz' of the time.
- In the 1960s, compared to the 1950s, the weight of instrumentalness is decreased, while that of speechiness and acousticness are increased. 'Free Jazz' and 'Fusion Jazz' [5] flourished at this time. The former is usually mixed with the wailing of the human voice and the wailing of the saxophone and other instruments, while the latter is more likely to use electronic synthesis. Its characteristic is consistent with the weight change.
- After the 1970s, the music characteristics of Jazz were more uncertain. At that time, the popular Jazz tended to be diversified with numerous styles. The mainstream forms of Jazz are collectively known as 'Contemporary Jazz'.

9 Task 7：Cultural Influence of Music

9.1 Music and Circumstance

Taking the development history of Jazz as an example, the music characteristics of Jazz in different periods are different, which reflects the phased genre evolution of Jazz, and the direction of evolution is related to the cultural background.

The emergence of big bands, swing music in the 1930s are different from traditional Jazz, with a very distinct spirit of resistance and struggle. On the one hand, this spirit is reflected in the resistance to old customs and ways of life. Young people are eager to pursue the liberation of individuality and a better life in the new world. On the other hand, the spirit of resistance is reflected in the defense of civil rights struggle, against racial discrimination and other areas of great political significance.

In the 1940s, Jazz gave way to 'Bebop Jazz'. In the past, jazz bands needed about 20 people to form. Due to the influence of World War II, big Jazz bands could not survive, so they were smaller than pop bands. At this time, the musicians also began to tire of the indistinguishable swing music, starting to pursue innovation. 'Cold Jazz' came into being in the 1950s. After World War II, people's material living standards gradually improved, and black people began to have more political freedom. They no longer relied on white people to make decisions for them, and were liberated from the feeling of slavery.

'Free Jazz' flourished in the 1960s, and blacks used it as a battle song for their civil rights. 'Fusion Jazz' also emerged in the 1960s, when the market for jazz was squeezed out by the new rock and roll genre, so artists introduced elements of new genres into jazz. After the 1970s, jazz became more diversified.

9.2 Music and Technology

From the data of data_by_artist and influence_data, it can be seen that there were artists in electronic music since 1960s. We created three networks using artists from electronic music genre of 1960-1979, 1980-1999, 2000-2020. Calculate degree centrality, betweenness centrality and pageranks of each point in the network and average them. The results are shown in Table 10.

Table 10: Average of indicators

Year	Pageranks	Betweenness Centrality	Outdegree	Degree Centrality
1960-1979	0.000093	0.000008	1.522388	4.880597
1980-1999	0.000125	0.000102	4.354331	10.826772
2000-2019	0.000141	0.000145	18.785714	24.000000

In Table 10, each indicator is increasing with each year. For electronic music networks, artists within the genre are more closely connected and information is exchanged more frequently. The influence of each person in the network and the influence of the network as a whole is constantly increasing, and the closer we get to the 21st century, the faster it grows. It shows that the development of music equipment has promoted the growth of electronic music genres, which is also promoted by the development of science and technology.

10 Sensitivity Analysis

In Task 4, we used entropy weight method and principal component analysis to study the influence of different music characteristics on followers in Folk. Considering that genre categories will have an impact on the result, we then choose New Age, Jazz and Country for comparative analysis. Substitute the principal component expression into the comprehensive evaluation score expression to obtain the coefficients of each music characteristic. Some results are

shown in Table 11.

Table 11: Coefficients of each music characteristic

Genre	x_0	x_1	x_2	···	x_9
New age	0.152647	0.115014	0.153734	···	-0.058180
Jazz	0.153021	0.183251	0.194708	···	-0.044100
Country	0.229863	0.003293	0.231687	···	-0.100580
Folk	0.182500	0.035100	0.209900	···	0.107000
New age	0.152647	0.115014	0.153734	···	-0.058180

At this point, the coefficient has changed. In order to make a more intuitive observation, we draw a line diagram, as shown in Figure 13:

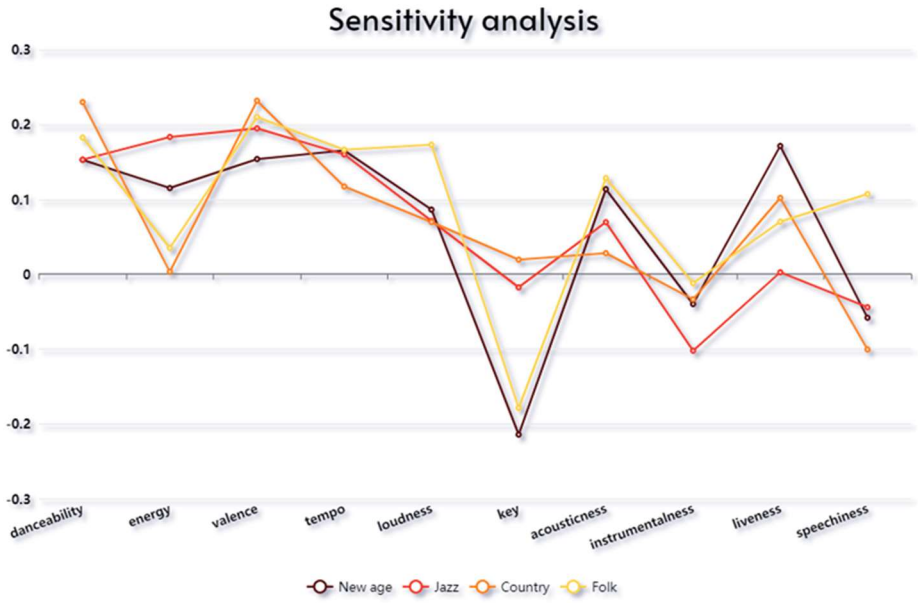


Figure 13: Line chart of coefficient change

The figure 13 shows that danceability, valence, tempo fluctuations small, and the coefficient of absolute value is larger. Thus, changing genres do not make much difference in assessing their contagiousness. However, energy, loudness, key, liveness, and speechiness show large fluctuations and different absolute values of coefficients, which indicate that genre categories are important for the assessment of their contagiousness. For instrumentalness, the waves are larger, but the absolute value of the coefficients is smaller overall. Therefore, it is of little significance to study the influence of genre and category on this basis.

11 Strengths and Weaknesses

Strengths

- Compared with the simple study of music characteristics, the network model reflects the attitudes of artists, the scale of their genres, and the connections between genres more intuitively.
- Degree centrality, betweenness centrality, closeness centrality and pageranks can quantitatively reflect the rate of information exchange in the network, the roles of different people in the network (high artistic influencers, inheritors and developers, independent musicians) and their importance.
- We divide the 100 years from 1921 to 2020 into 10 stages, calculate the objective weight of each music characteristic respectively, and then analyze the evolution of music genre in each stage, so as to make the model more accurate. In addition, we use radar charts to show periodic trends, which are clear and intuitive.

- Because the data transformation ranges of some indicators are different, we use standard Euclidean distance instead of Euclidean distance, so that the distribution of all dimensions of data meets the requirement of equal mean variance and avoids the influence of outliers and extreme values.

Weaknesses

- Network indicators only reflect the overall information exchange efficiency and influence of individuals and subnetworks, and there is no specific index to reflect the influence between subnetworks.
- This model is only established based on the given data, mainly considering time and cultural background, without considering more diverse factors such as region and religion.
- Considering the difference between the number of schools in the table and the actual situation, the accuracy of the information reflected by the number ratio needs to be improved.

12 Conclusion

In this paper, we build a model of music influence network and introduce the entropy weight method, principal component analysis, afterwards a quantitative analysis of the influence of previous music on new music and artists is conducted, with music characteristics and other indicators taken into consideration. Through the music influence network, we analyze the impact and information exchange efficiency of the subnetwork and its core members. By studying the characteristics of music, we have found the differences between the musical styles of different genres, the evolution of music genres in different eras, and the extent to which predecessors influenced later artists. Finally, the model is further extended by considering multidimensional factors such as geography and religion.

Our Document

From： Team#2107087
To： ICM Society
Date： February 8, 2021

Dear ICM Society，

We are very honored to be part of this study of the influence of music. The purpose of this document is to illustrate the value and improvement of the Music Influence Directional Power Network. We will start from two perspectives: the information provided by the Music Influence Network and the influence of music on culture.

The music influence network we built connects influencers, followers, and genres. Nodes in the network provide information about artists and genres, while arcs represent the relationships between the three. Through the network, we can qualitatively understand the type of genre, number and the relationship between artists. In addition, by setting indicators, we can identify key artists representing major leaps from the network, and divide them into three categories: high artistic influencers, inheritors and developers, and independent musicians. They are in different parts of the network and have different artistic resources.

Subnetworks of music influence networks divide genres into communities and provide information about each genre and its artists. Important artists of different genres have different efficiency of information exchange. The greater the genre, the greater the influence of the artist. With the help of mathematical model, the information and correlation of each subnetwork can be further studied. The results are as follows:

- Artists of the same genre are more similar than artists of different genres.
- The weight of each music characteristic is different in different genres. The greater the fluctuation in the weight of each genre, the more important the music characteristics are, which can be used as a factor to distinguish the genre. The association of genres can be judged based on music characteristics. It is worth noting that there are certain differences in the 'contagiousness' of different music characteristics.
- Different genres vary with time, and the years of their rise, popularity and decline are different.

In addition, in the face of richer data, such as the introduction of geographical, political, religious and other attributes, we should quantify the impact of music on culture from a more multi-dimensional perspective. We can consider the variation of music characteristics of geography-time, politico-time, religion-time first.

Since the characteristics of music reflect the spiritual outlook of people in different historical periods, music records the daily life and cultural background of people in different historical periods in this special way. That is to say, cultural development can be reflected from music characteristics. For example, in the era of war, the culture needed music with a stronger sense of strength and more passionate emotions, that is, the value of 'valence' might be on the high side. In the era of rapid economic development, people had more time to enjoy the music, and the tone of the music was more relaxing and happy, which was suitable for dancing. By integrating emotion into songs, music, as a carrier of culture, is branded with the traces of the times.

It is not comprehensive to study only from the perspective of two dimensions, because these factors influence each other. It can be increased to geographic-political-time, geographic-religion-time, and political-religion-time on a two-dimensional basis. We always need to take time into account, because geography, politics, time and other factors are always affected by time. To study the evolution of music characteristics from more dimensions will make our model more targeted and accurate.

References

[1] KANG L. Identification of Multi-influence Nodes in Complex Networks Based on Regional Density Curves [D]. Anhui University,2019.
[2] Ding Yi, Yang Jian. Comparison of Euclidean Distance and Standardized Euclidean Distance in K-Nearest Neighbor Algorithm [J]. SOFTWARE,2020,41(10):135-136+140.
[3] Liu Ziqing. Construction of a Complex Network and Text Topic Analysis of Deng Tietao's Academic Theory Literature Communication [D]. Guangzhou University of Chinese Medicine,2017.
[4] Liu Yunyan. The Main Schools of Western Music Emerging in the 20th Century [J]. Journal of Liaoning Education Institute,2001(07):87-90.
[5] Li Chenxi. Brief Introduction to the Characteristics of Music Language in the 20th Century [J]. Northern Music,2015,35(14):3-4.

Appendix

```
"""2021 ICM.ipynb
"""

import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
import seaborn
import numpy as np

influnence_data = pd.read_csv('/content/drive/MyDrive/2021_D_ICM/2021_ICM_Problem_D_Data/influence_data.csv')
inf_id = list(set(influnence_data['influencer_id'].values.tolist()))
fol_id = list(set(influnence_data['follower_id'].values.tolist()))
tol_id = list(set(inf_id+fol_id))
genre_list = list(set(influnence_data['influencer_main_genre'].values.tolist()))
# Source=fol_name , Target=inf_name ,Weight=0 or 1 ,Type=direct
person_relations = {i:dict() for i in tol_id}
genre_relations = {i:{j:0 for j in genre_list} for i in fol_id}
for index in range(len(influnence_data)):
  if influnence_data.loc[index].follower_id not in person_relations[influnence_data.loc[index].influencer_id]:
    person_relations[influnence_data.loc[index].influencer_id][influnence_data.loc[index].follower_id] = 1
  genre_relations[influnence_data.loc[index].follower_id][influnence_data.loc[index].influencer_main_genre] += 1
Source = []
Target = []
```

```
for i in person_relations.keys():

  if person_relations[i] == {}:
    continue
  for target in person_relations[i].keys():
    Source.append(i)
    Target.append(target)
Type = ['directed']*len(Source)
dataframe = pd.DataFrame({'Source':Source,'Target':Target,'Type':Type})
Source_g = []
Target_g = []
for i in range(len(influnence_data)):
  if influnence_data.loc[i].influencer_id not in Source_g:
    Source_g.append(influnence_data.loc[i].influencer_id)
    Target_g.append(influnence_data.loc[i].influencer_main_genre)
  if influnence_data.loc[i].follower_id not in Source_g:
    Source_g.append(influnence_data.loc[i].follower_id)
    Target_g.append(influnence_data.loc[i].follower_main_genre)
Type_g = ['directed']*len(Source_g)
dataframe2 = pd.DataFrame({'Source':Source_g,'Target':Tar-
get_g,'Type':Type_g})
```