Problem Chosen	2021 MCM/ICM Summary Sheet	Team Control Number
D		2100888

Clarify music's magic: Analysis on Network

Summary

In today's society, it has become more and more common to use "networks" connected by different nodes to analyze the relationship between things, and the rapid development of network science has also proved this. In order to measure the dynamic influence of music artists on the evolution of music, we used network science and technology to complete the following tasks related to musicians and music genres, and gave our conclusions on the evolution of music.

Firstly, We measure the influence of different musicians and genres through our self-built music network model. The musician is set as a node, and its influence on other musicians is taken as an edge, and three node indicators: degree, betweenness centrality and average clustering coefficient are selected to represent the musician's *"Music Influence"* from different dimensions. Then, the characteristics of all the artists of the genre are used to represent the genre, and the influence between the genres is measured in the same way with artists.

Secondly, we built a similarity scoring system to analyze the similarities between musicians and genres. Through the R-type clustering of the 12 effective musical characteristics of musicians, we have obtained 5 indicators: Musicality, Key, Randomness, Track duration, Acoustics. We use the **TOPSIS** entropy method to construct a similarity measurement system, so that we can calculate the similarity score between the two according to the musician's music index. When using this system to evaluate the similarity between genres, in addition to giving a similarity matrix between the genres, we also further use the **Graph Kernel** method to more accurately measure the similarity between and within the genres from the perspective of the graph.

Thirdly, we built a music evolution model, focusing on the development of Pop/Rock music. Through multiple linear regression, we found that the Musicality indicator has the highest appeal in the influence process. At the same time, we established a **sliding window** to do **Moveout correlation analysis**, and obtained a highly related relationship between the two genres. In order to further illustrate the evolution process, we determined that Popularity is the music characteristic that can best measure the development of a genre, and then based on the inflection point detection to explore the time period when the Popularity of Pop/Rock changed the most as 1954-1964, 2000-2010, which is a major change. The Beatles and Rihanna are their representatives. Going deeper, we used dynamic time warping to find the dynamic influencer indicator of the genre as Acoustics, and we found that its changes are highly consistent with the changes in the popular genre.

Finally, we explored the development of Pop/Rock music from different angles, combined with the historical background, and found that it was more obviously affected by external changes. In addition, our work also demonstrates the cultural impact of music on society, and distinguishes the effects of different external influences on music. At the end of the paper, we submitted a one-page report to ICM.

Key words: Music, Influence Network, TOPSIS, Moveout analysis

Contents

1 Introduction

1.1 Background

Music is one of the oldest art forms. From ancient times to the 21st century, it has been inseparable from human society. With the development of human civilization, a large number of talented artists emerged, which promoted continuous progress in the form and content of music. At the beginning of the 20th century, Schoenberg established the 12-tone system, which regulated the development of world music and promoted modernism. Music began to sprout, and new music techniques that broke the convention gradually entered people's vision. Compared with the previous impressionist music, the music of the 20th century showed the characteristics of large melody span and free rhythm organization. By the middle of the 20th century, due to the rapid advancement of technology, new genres such as technical music and electronic music appeared. Their birth also meant that the influence of folk-custom music was gradually weakened. Since the 1970s, human society has entered Beyond Modernism, romanticism has reappeared, and its influence has continued to these days.

Throughout the history of music development, we can find that earlier music has an impact on subsequent music, although the works of different music genres and musicians show different characteristics such as danceability and tempo. But the influence and promotion between them cannot be ignored. Under this circumstance, we must use network science to explore the influence, similarity and evolution of artists and genres.

1.2 Problem Restatement

After analyzing the subject requirements, we should complete the following tasks:

- Establish an influence network, and design parameters to capture *"music influence"* to analyze influence within and between artists and genres.
- Analyze similarities between artists and between genres respectively, and judge weather artists within genre more similar than artists between genres.Use the result of our analysis to identify the authenticity of above-mentioned influence.
- Explore the evolution of genres over time and find relations between genres. Then, Seek revolutions in development of one genre and its representatives. Furthermore, find out indicators of dynamic influencers.
- Extend our approach to social culture to reveal the influence of music on society and influence of society on music.

2 Assumptions and Notations

To simplify our model and eliminate the complexity, we make the following main assumptions in this literature. All assumptions will be re-emphasized once they are used in the construction of our model.

Assumption 1: Each artist only represents one genre and will not change.

Assumption 2: The influence among artists is directional.

Assumption 3: The genre of each work is the genre of its first artist.

Assumption 4: The average musical characteristics of each genre artist can represent the musical characteristics of the genre.

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

Table 1 Notations

Symbols	Definition
v_x	The node of network
e_{jk}	Edge of the network
ω_{jk}	The influence of musician j on musician k
$C_B(x)$	Betweenness centrality of Node x
$C(x)$	Average clustering coefficient of Node x and the nodes near it
$D_M(x,y)$	Manhattan distance between Node x and Node y
$D_{\text{Chebyshev}}(p,q)$	Chebyshev distance between Node p and Node q
$E_j(A,B)$	Tanimoto Coefficient between Node A and Node B

3 Influence Network Model

We established the Influence Network Model to represent the directed influence between artists and genres in the network, using three indicators, betweenness centrality, average clustering coefficient and degree, to quantitatively analyze *"Music influence"*.

3.1 Artist Influence Network

Since the influence among artists has been assumed to be directed, we created a directed Artist Influence Network. We use a node to represent an artist. The size of the node reflects the number of followers. The edges of the graph point from influencers to followers. We use ω_{jk} to represent the influence of musician j on musician k. By observing the entire network, we can see the influence of an artist.

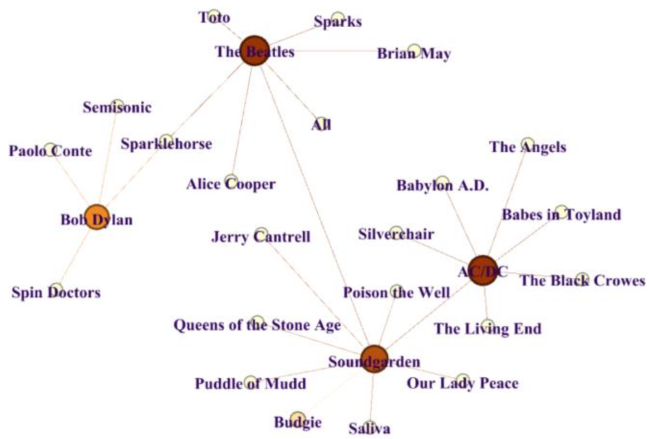

Figure 1 A Portion of Network of Artists by Gephi

As shown in the figure, the influence between artists is complicated. An artist influences many people, and followers also influence each other. Some artists are more influential, such as The Beatles, Bob Dylan.

3.2 Genre Influence Network

Based on the *Artist Influence Network*, we established the *Genre Influence Network*, so that a node represents all artists in a genre, and the size of the node represents the total number of followers that this genre influences. The number of influencers of genre i on genre j is regarded as the edge weight of genre i to genre j.

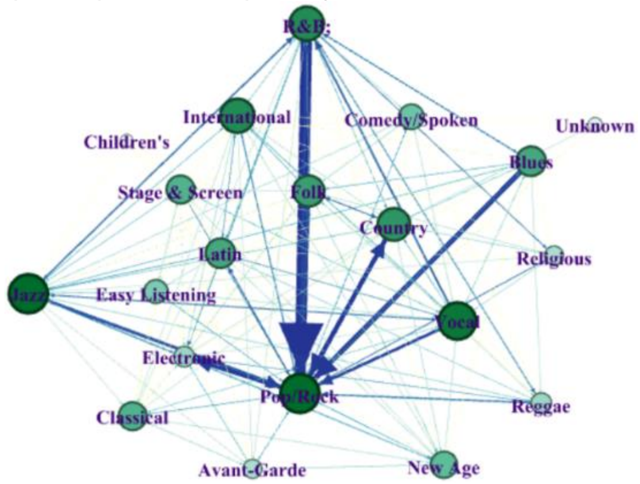

Figure 2 The Network of Genres by Gephi

We found that genres influence each other, and different genres have different influential power. R&B has the greatest influence on Pop/Rock, and Pop/Rock is most affected by other genres, indicating that its music is highly integrated.

In addition, it is not difficult to find that the influence within the genre is the influence among the artists in the genre.

3.3 Parameters of *"Music Influence"*

In order to measure "Music Influence", according to the characteristics of *Artist Influence Network*, we focus on the following three parameters from different dimensions: Degree, betweenness centrality, average clustering coefficient.

- **Degree**

In a directed graph, each node has two degrees: in-degree is the number of in-edges connected to the node, and out-degree is the number of out-edges. When there is an edge from node i to j, the corresponding element in the adjacency matrix A_{ij}= 1, then the in-degree and out-degree are recorded as:

$$k_j^{\text{in}} = \sum_{i=1}^{n} A_{ij}, \quad k_i^{\text{out}} = \sum_{j=1}^{n} A_{ij}$$

At the same time, considering that in a directed graph, the number of edges m is equal to the sum of the number of nodes of the in-edges and the sum of the number of nodes of the out-edges. There is:

$$m = \sum_{j=1}^{n} k_j^{\text{in}} = \sum_{i=1}^{n} k_i^{\text{out}} = \sum_{ij} A_{ij}$$

The average value of in-degree c_{in} and the average value of out-degree c_{out} of each directed graph are equal:

$$c_{\text{in}} = \frac{1}{n}\sum_{j=1}^{n} k_j^{\text{in}} = \frac{1}{n}\sum_{i=1}^{n} k_i^{\text{out}} = c_{\text{out}}$$

After simplification:

$$c = \frac{m}{n}$$

The greater the out-degree of artist node is, the more people the artist directly influences and the greater his influential power is. The greater the in-degree is, the greater the influence of the artist is subjected to and the higher the similarity with other artists.

Through statistical analysis of the entire network, we have found the top 10 artists and genres:

Table 2 Top 10 Artists & Genres with Highest Out-degree

Name	Outdegree	Indegree	Genre	Outdegree
The Beatles	615	31	R&B;	2146
Bob Dylan	389	29	Pop/Rock	2092
The Rolling Stones	319	39	Jazz	947
David Bowie	238	25	Vocal	877
Led Zeppelin	221	24	Blues	871
Jimi Hendrix	201	32	Country	799
The Kinks	192	8	Folk	739
The Beach Boys	186	13	Electronic	338
Hank Williams	184	3	International	247
The Velvet Underground	181	8	Reggae	222

- **Betweenness centrality $C_B(x)$**

The betweenness centrality C_B of a node shows how important the location of the node is in the network. Define g_{jk} as the number of shortest paths between node j and node k, $g_{jk}(x)$ is the number of shortest paths between node j and node k through node x, $(n-1)(n-2)/2$ is the maximum possible node betweenness. Then:

$$C_B(x) = \frac{2\sum_{j<k} g_{jk}(x)}{(n-1)(n-2)g_{jk}}$$

The greater the betweenness centrality of a node, the greater the importance of the information transmission of the artist or genre represented by this node in the upper and lower generations, and the greater the inheritance role it plays.

Taking Electronic music as an example, we have carried out betweenness centrality calculations in this genre and found that Aphex Twin is the artist with the largest betweenness centrality in this genre, which reaches 220.1952, indicating that he played an important role in the development of electronic music inheritance.

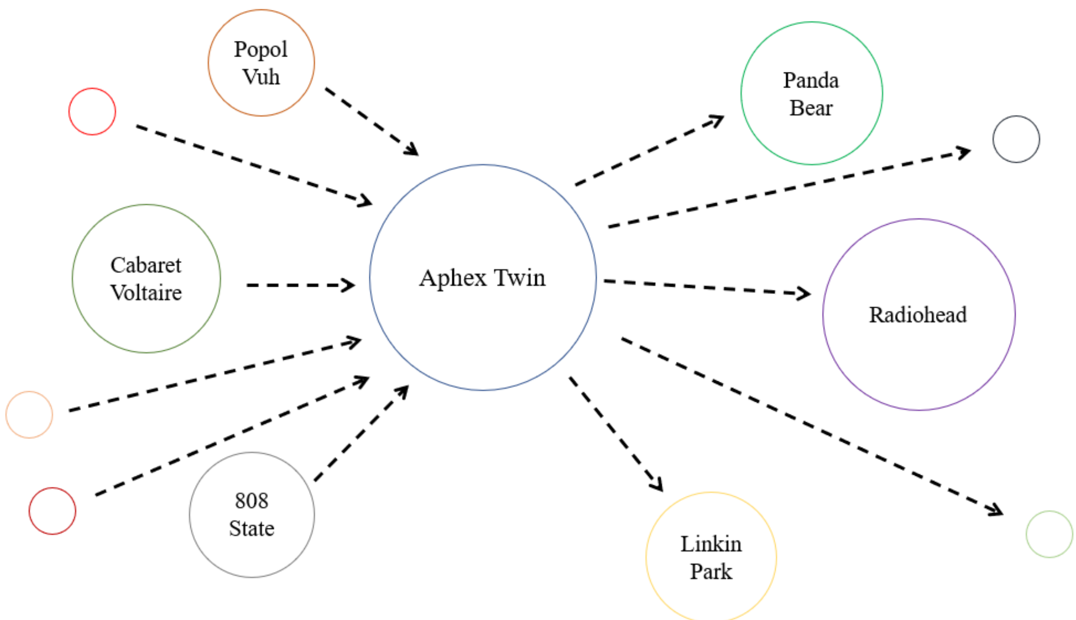

Figure 3 Aphex Twin's Network

- **Average Clustering Coefficient** $C(x)$

The average clustering coefficient is a parameter that describes the clustering degree of a node and its surrounding nodes. $C(x)$ is the ratio of the number of edges connecting the node v_x with surrounding nodes to the maximum number of edges that these nodes can connect. Since the network we assume is a directed network, e_{jk} is defined as the edge formed by the connection of node j and node k, and k_x is the total number of edges pointing to node v_x and edges pointing out from v_x. Then $k_x(k_x-1)$ is the maximum number of edges that the vertices around v_x can connect. Then:

$$C(i)=\frac{\left|\{e_{jk}: v_j, v_k \in L(i), e_{jk} \in E\}\right|}{k_i(k_i-1)}$$

The larger the average clustering coefficient in the Influence network, the closer the connection between musicians in this network, and the greater the influence of predecessors on the creation of younger generations.

We analyzed the average clustering coefficient of the 9 most influential music types, and the results are shown in the figure:

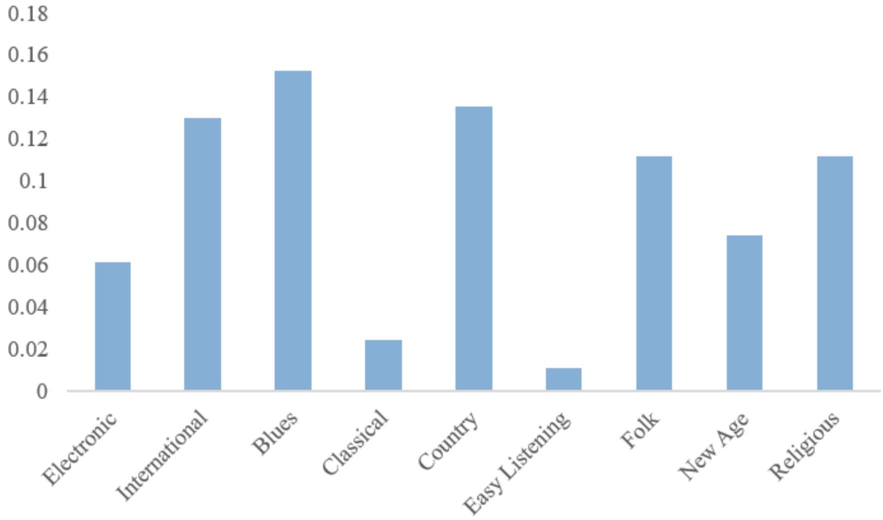

Figure 4 Average Clustering Coefficient of 9 Genres

Through calculating, we found that Blues artists often play a more important role in the network, and Country and International artists are not much important. But in comparison, the influence of Classical and Easy Learning is not so high.

4 Similarity Model

We choose to cluster the 12 music characteristics corresponding to each artist in *full_music_data*, and construct an evaluation model to quantify the similarity between two artists. Finally, we can get the difference between any two artists through similarity score. We apply the evaluation system to analyze the similarity of genres.

4.1 Cluster Analysis

Regarding the data in the file, since there are multiple indicators to measure the characteristics of songs, after removing irrelevant characteristics such as music release time, we noticed that both the Mode and Explicit are 0-1 variables, which cannot identify a high degree of distinction between different artists and between different genres, so we do not use this indicator as a measure of the similarity between two artists. We perform R-type clustering on the remaining indicators. R-type clustering can not only obtain the degree of intimacy between individual variables, but also the degree of intimacy between combinations of variables.

Furthermore, we can reveal the inner connection of a series of indicators such as danceability, energy, speechiness and so on. Use R-type clustering method to cluster all indicators.

First calculate the Pearson correlation coefficient between each index, the formula is:

$$r(X,Y)=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\cdot\mathrm{Var}(Y)}}$$

$Cov(X,Y)$ is the covariance of X and Y, and $\mathrm{Var}(X)$ is the variance of X.

After establishing the correlation coefficient matrix, we standardize the data of each indicator separately. The approximate measure between variables uses the correlation coefficient, and the calculation of the similarity measure between classes uses the class average method. The cluster tree map and cluster heat map are as follows:

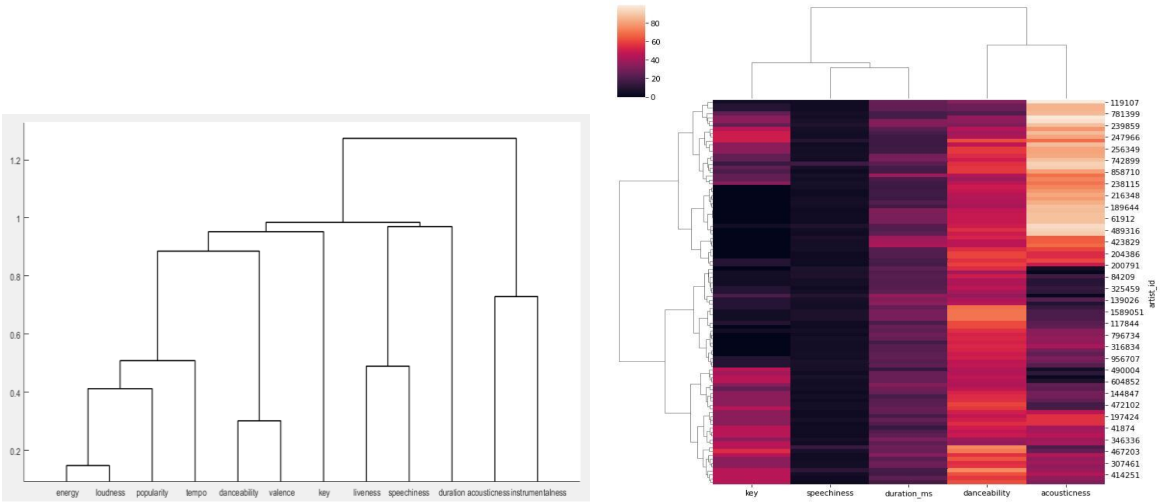

Figure 5 Cluster map

The final clustering results are shown in the following table:

Table 3 Species & Characteristics

Classfication	Characteristics					
Key	key					
Musicality	danceability	energy	valence	tempo	loudness	popularity
Track duration	duration_ms					
Randomness	liveness	speechiness				
Acoustics	acousticness	instrumentalness				

Taking into account that different characteristics have different attributes and dimensions, we define it as follows:

- Key: The estimated overall key of the track
- Musicality: It is one of the most essential characteristics of a song that can show the characteristics of the music attribute of a song.
- Track duration: The duration of the track in milliseconds.
- Randomness: A feature that can show the degree of freedom of musicians when singing or playing.
- Acoustics: Can show the characteristics of the acoustic properties of music.

4.2 Parameters of Similarity

In order to compare the similarity between two artists, after we analyze different types of characteristics, we select appropriate similarity metrics to analyze different types of them:

- *For Key & Track duration:* Since there is only one characteristic in the two categories, we use Manhattan distance and Tanimoto coefficient to describe the similarity.
- *For Musicality & Randomness & Acoustics:* There are multiple characteristics in these three categories. We treat each characteristic as a coordinate axis, and measure the similarity by calculating the Chebyshev distance and cosine similarity between two points.

4.2.1 Key & Track duration

- **Manhattan Distance $D_M(x,y)$**

Manhattan distance is a commonly used distance index, which is used to evaluate the similarity between data. The formula for Manhattan distance $D_M(x,y)$ is:

$$D_M(x,y) = |x-y|$$

The smaller Manhattan distance between the two groups of data is, the more similar the two sets of data are, and the corresponding two artists are more similar in the corresponding two sets of music characteristics.

- **Tanimoto Coefficient $E_j(A,B)$**

The Tanimoto coefficient is extended from the Jaccard coefficient and is also called the generalized Jaccard coefficient. Define A and B as two vectors, $A*B$ is the product of vectors, and $\| A \|^2$ is the modulus of A.

The formula is as follows:

$$E_j(A,B) = \frac{A*B}{\| A \|^2 + \| B \|^2 - A*B}$$

The greater Tanimoto coefficient of the two groups of data is, the greater the similarity of the two sets of data is, and the corresponding two artists are more similar in the corresponding music characteristics.

4.2.2 Musicality & Randomness & Acoustics:

- **Chebyshev Distance $D_{\text{Chebyshev}}(p,q)$**

Chebyshev distance is a metric in vector space, and the distance between two points is defined as the maximum value of the difference between its coordinates. Define p and q to be two vectors with coordinates p_i and q_i. Then the formula of the Chebyshev distance between the two is as follows:

$$D_{\text{Chebyshev}}(p,q) = \max_i(|p_i - q_i|)$$

The smaller Chebyshev distance between the two groups of data is ,the more similar the two sets of data are, and the corresponding two artists are more similar in the corresponding two groups of music characteristics.

- **Cosine similarity $\cos\theta$**

Cosine similarity is an index that evaluates the similarity of two vectors by calculating the cosine of the angle between them. The cosine similarity draws the vector into the vector space according to the coordinate value. For vectors A and B, in any dimensional space, the definition of cosine similarity has:

$$\cos\theta = \frac{A \cdot B}{\| A \| * \| B \|}$$

The greater cosine similarity of the two groups of data is, the greater the similarity of the two groups of data is, and the corresponding two artists are more similar in the corresponding music characteristics.

For each indicator, multiple data similarity analysis is required. We naturally think of the algorithm for face recognition in the project, which measures the weighted distance between each point and the standard model, and finally add the distance to form the final "score". From the divergence of this kind of thinking, we have established a similarity scoring system, and the range of scores is defined as the similarity criterion.

We use the four indicators we have obtained: Manhattan distance, Tanimoto coefficient, Chebyshev distance and cosine similarity as the distance, and use the entropy method to calculate the weight of each feature, and finally use the TOPSIS method to calculate the difference between the two musicians The similarity score between.

4.3 TOPSIS Method

- Use vector programming method to obtain the standard decision matrix. Suppose the decision matrix $A=(a_{ij})_{m\times n}$ of the multi-attribute decision-making problem, and the standardized decision matrix $\boldsymbol{B}=(b_{ij})_{m\times n}$. Then:

$$b_{ij}=a_{ij}/\sqrt{\sum_{i=1}^{m}a_{ij}^{2}}, i=1,2,\cdots,m; j=1,2,\cdots,n。$$

- Construct a weighted canonical matrix $\boldsymbol{C}=(c_{ij})_{m\times n}$. Calculate the weight vector of each attribute. Define x_{ij} the value of the j^{th} index of the i^{th} sample. Since the measurement units of each index are not uniform, we need to standardize first. We use different methods to standardize the positive and negative indicators, and record the standardized data as x_{ij}:

  Positive index：

$$x_{ij}=\frac{x_{ij}-min\{x_{1j},\ldots,x_{nj}\}}{max\{x_{1j},\ldots,x_{nj}\}-min\{x_{1j},\ldots,x_{nj}\}}$$

  Negative index：

$$x_{ij}=\frac{max\{x_{1j},\ldots,x_{nj}\}-x_{ij}}{max\{x_{1j},\ldots,x_{nj}\}-min\{x_{1j},\ldots,x_{nj}\}}$$

  Then the proportion of the index for the i^{th} sample station under the j^{th} index is:

$$p_{ij}=\frac{x_{ij}}{\sum_{i=1}^{n}x_{ij}}, i=1,\cdots,n, j=1,\cdots,m$$

  Then get the entropy of the j^{th} index:

$$e_j=-k\sum_{i=1}^{n}p_{ij}\ln(p_{ij}), j=1,\cdots,m \qquad k=1/ln(n)$$

  The weight of each indicator is:

$$w_j=\frac{d_j}{\sum_{j=1}^{m}d_j}, j=1,\cdots,m$$

  Then:

$$d_j=1-e_j, j=1,\cdots,m$$

  The Indicators are expressed as follows:

Table 4 Indicators' Abbreviation

Indicator	Description	Characteristic
KeyD	Manhattan Distance of Key	Negative
TdD	Manhattan Distance of Track Duration	Negative
KeyTC	Tanimoto Coefficient of Key	Positive
TdTC	Tanimoto Coefficient of Track Duration	Positive
MD	Chebyshev distance of Musicality	Negative
RD	Chebyshev distance of Randomness	Negative
AD	Chebyshev distance of Acoustics	Negative
MC	Cosine similarity of Musicality	Positive
RC	Cosine similarity of Randomness	Positive
AC	Cosine similarity of Acoustics	Positive

Then the index weights finally calculated according to the entropy weight method are shown in the following table:

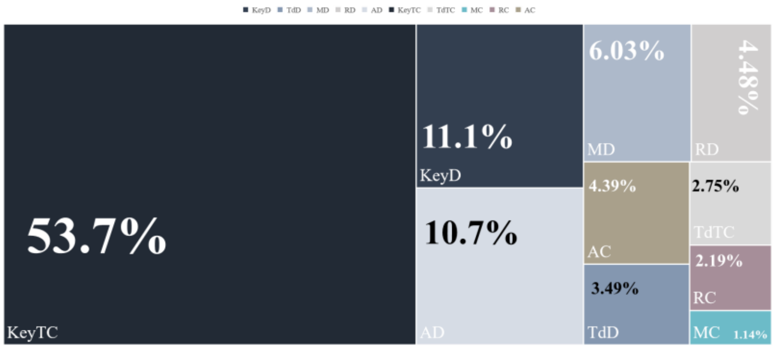

Figure 6 Weight Dendrogram

- Determine the positive ideal solution C^* and the negative ideal solution C^0. Suppose the j^{th} attribute value of the positive ideal solution C^* is c_j^*, and the j^{th} attribute value of the negative ideal solution C^0 is c_j^0, then

$$Positive\ \ ideal\ \ solution\ c_j^* = \begin{cases} \max\limits_i c_{ij}, j \text{ is the benefit attribute,} \\ \min\limits_i c_{ij}, j \text{ is the cost attribute ,} \end{cases} j = 1,2,\cdots,n\ ,$$

$$Negative\ \ ideal\ \ solution\ c_j^0 = \begin{cases} \min\limits_i c_{ij}, j \text{ is the benefit attribute,} \\ \max\limits_i c_{ij}, j \text{ is the cost attribute ,} \end{cases} j = 1,2,\cdots,n_0$$

- Calculate the distance from each plan to the positive ideal solution and the negative ideal solution. The distance from the alternative d_i to the positive ideal solution is

$$s_i^* = \sqrt{\sum_{j=1}^{n} \left(c_{ij} - c_j^*\right)^2}, i = 1,2,\cdots,m$$

The distance from alternative d_i to the negative ideal solution is

$$s_i^0 = \sqrt{\sum_{j=1}^{n} \left(c_{ij} - c_j^0\right)^2}, i = 1,2,\cdots,m$$

- Calculate the ranking index value of each plan, namely

$$f_i^* = s_i^0/(s_i^0 + s_i^*), i = 1,2,\cdots,m$$

- Arrange the pros and cons of the schemes in descending order of f_i^*.

4.4 Measure Artist Similarity

We selected different artists from different genres for similarity calculation, and the results are as follows:

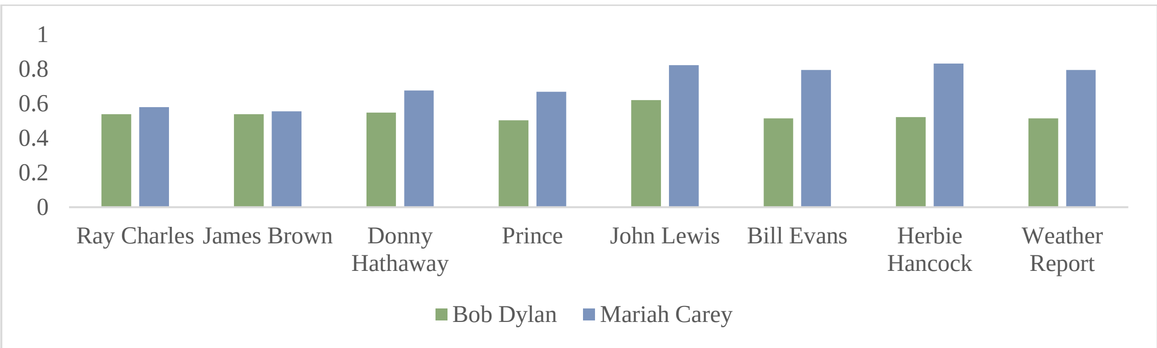

Figure 7 The Similarity of Bob Dylan & Mariah Carey

As can be seen from the figure, Bob Dylan is less similar to the other singers listed than Mariah Carey, indicating that Bob Dylan's innovation is more significant, while Mariah Carey interacts with other singers.

Considering the similarity of artists between genres and the similarity of musicians within the genre, we selected two representative artists from the four genres, and calculated the music of them and the genre or other genres respectively. The similarity between people, as shown in the figure:

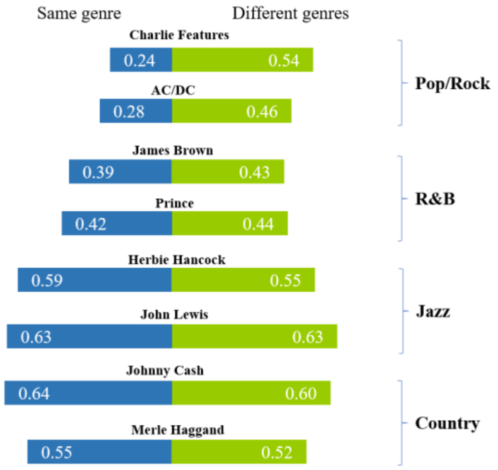

Figure 8 Similarity of 8 Artists from 4 Genres

We can observe that for Pop/Rock, the similarities between its artists are very small, but the similarity with artists outside the genre is very large, indicating that Pop/Rock is good at getting inspiration from other genres, and Pop/ Rock music combines a variety of music forms. The two data of R&B are almost the same, indicating that the genre is very innovative and adaptable. As for Jazz and Country, the musicians of both have learned a lot from the genre, and the two music types are relatively closed.

4.5 Measure Genre Similarity

4.5.1 Music Characteristics

We selected all musicians of each genre and averaged all their musical characteristics, so that the genre can be regarded as a musician with the characteristics of the genre. Using the similarity model, we calculated the difference between the genres. The similarity is as follows:

		AG	BL	CH	CL	CS	CO	EA	EL	FL	IN	JZ	LT	NA	PR	RB	RG	RE	SS	UN	VO
Avant-Garde	AG	1	0.501	0.618	0.253	0.647	0.564	0.12	0.583	0.525	0.443	0.121	0.526	0.227	0.573	0.559	0.56	0.596	0.115	0.574	0.574
Blues	BL	0.501	1	0.133	0.445	0.295	0.154	0.516	0.453	0.201	0.062	0.384	0.112	0.672	0.426	0.302	0.413	0.248	0.507	0.358	0.274
Children's	CH	0.618	0.133	1	0.57	0.253	0.281	0.623	0.572	0.122	0.202	0.513	0.243	0.742	0.545	0.431	0.531	0.368	0.615	0.489	0.171
Classical	CL	0.253	0.445	0.57	1	0.604	0.55	0.24	0.71	0.475	0.391	0.291	0.509	0.294	0.712	0.649	0.728	0.618	0.224	0.693	0.522
Comedy/Spoken	CS	0.647	0.295	0.253	0.604	1	0.417	0.691	0.672	0.269	0.293	0.55	0.376	0.847	0.654	0.532	0.62	0.477	0.679	0.587	0.294
Country	CO	0.564	0.154	0.281	0.55	0.417	1	0.581	0.399	0.37	0.218	0.465	0.094	0.72	0.288	0.167	0.273	0.138	0.575	0.218	0.439
Easy Listening	EA	0.12	0.516	0.623	0.24	0.691	0.581	1	0.617	0.547	0.487	0.188	0.557	0.241	0.589	0.595	0.589	0.634	0.107	0.616	0.586
Electronic	EL	0.583	0.453	0.572	0.71	0.672	0.399	0.617	1	0.635	0.497	0.562	0.361	0.654	0.321	0.398	0.363	0.434	0.634	0.429	0.686
Folk	FL	0.525	0.201	0.122	0.475	0.269	0.37	0.547	0.635	1	0.197	0.426	0.337	0.706	0.611	0.511	0.612	0.47	0.542	0.565	0.136
International	IN	0.443	0.062	0.202	0.391	0.293	0.218	0.487	0.497	0.197	1	0.345	0.193	0.65	0.483	0.361	0.472	0.31	0.477	0.423	0.267
Jazz	JZ	0.121	0.384	0.513	0.291	0.55	0.465	0.188	0.562	0.426	0.345	1	0.434	0.372	0.553	0.472	0.547	0.506	0.198	0.498	0.478
Latin	LT	0.526	0.112	0.243	0.509	0.376	0.094	0.557	0.361	0.337	0.193	0.434	1	0.707	0.344	0.228	0.329	0.191	0.551	0.28	0.413
New Age	NA	0.227	0.672	0.742	0.294	0.847	0.72	0.241	0.654	0.706	0.65	0.372	0.707	1	0.694	0.726	0.738	0.755	0.27	0.745	0.747
Pop/Rock	PR	0.573	0.426	0.545	0.712	0.654	0.288	0.589	0.321	0.611	0.483	0.553	0.344	0.694	1	0.193	0.16	0.247	0.62	0.185	0.665
R&B	RB	0.559	0.302	0.431	0.649	0.532	0.167	0.595	0.398	0.511	0.361	0.472	0.228	0.726	0.193	1	0.185	0.17	0.587	0.17	0.583
Reggae	RG	0.56	0.413	0.531	0.728	0.62	0.273	0.589	0.363	0.612	0.472	0.547	0.329	0.738	0.16	0.185	1	0.238	0.612	0.184	0.667
Religious	RE	0.596	0.248	0.368	0.618	0.477	0.138	0.634	0.434	0.47	0.31	0.506	0.191	0.755	0.247	0.17	0.238	1	0.636	0.219	0.53
Stage & Screen	SS	0.115	0.507	0.615	0.224	0.679	0.575	0.107	0.634	0.542	0.477	0.198	0.551	0.27	0.62	0.587	0.612	0.636	1	0.608	0.593
Unknown	UN	0.574	0.358	0.489	0.693	0.587	0.218	0.616	0.429	0.565	0.423	0.498	0.28	0.745	0.185	0.17	0.184	0.219	0.608	1	0.636
Vocal	VO	0.574	0.274	0.171	0.522	0.294	0.439	0.586	0.686	0.136	0.267	0.478	0.413	0.747	0.665	0.583	0.667	0.53	0.593	0.636	1

Figure 9 Similarity Scores between Genres

As shown in the figure, New Age music is very similar to other music, indicating that this kind of music born in the 20th century has widely absorbed the specialties of other music. We also found that R&B, Religious music is less similar to other music, indicating that it has maintained a better uniqueness in the development process.

4.5.2 Feature of Graph

In order to measure the similarity between the two genres, we generate sub-pictures according to the genres, and distinguish the differences between the genres by studying the differences between the pictures and the pictures,that is *Graph Kernel*.

A graph kernel $\boldsymbol{k}$: $\boldsymbol{G} \times \boldsymbol{G}$ is a kernel function over a set of graphs $\boldsymbol{G}$.

It's equivalent to an inner product of the embeddings $\phi: \mathcal{X} \to \mathbb{H}$ of a pair of graphs into a Hilbert space: $k(G_1, G_2) = \langle \phi(G_1), \phi(G_2) \rangle$. The similarity between two graphs can be obtained by the dot product operation in Hilbert space.

For a given graph G_1 and $G_2$and a graph decomposition method $\mathcal{F}$, The substructure after decomposition is:

$$\mathcal{F}(G_1) = \{S_{1,1}, S_{1,2}, \dots, S_{1,N_1}\} \mathcal{F}(G_2) = \{S_{2,1}, S_{2,2}, \dots, S_{2,N_2}\}$$

The kernel value of$G_1$and G_2 is:

$$k_R(G_1, G_2) = \sum_{n_1=1}^{N_1} \sum_{n_2=1}^{N_2} \delta(S_{1,n_1}, S_{2,n_2})$$

When S_{1,n_1} and S_{2,n_2} are isomorphic , $\delta(S_{1,n_1}, S_{2,n_2})$ is 1. When they are isomerism , $\delta(S_{1,n_1}, S_{2,n_2})$ is 0.

We decompose the subgraph and get the following results:

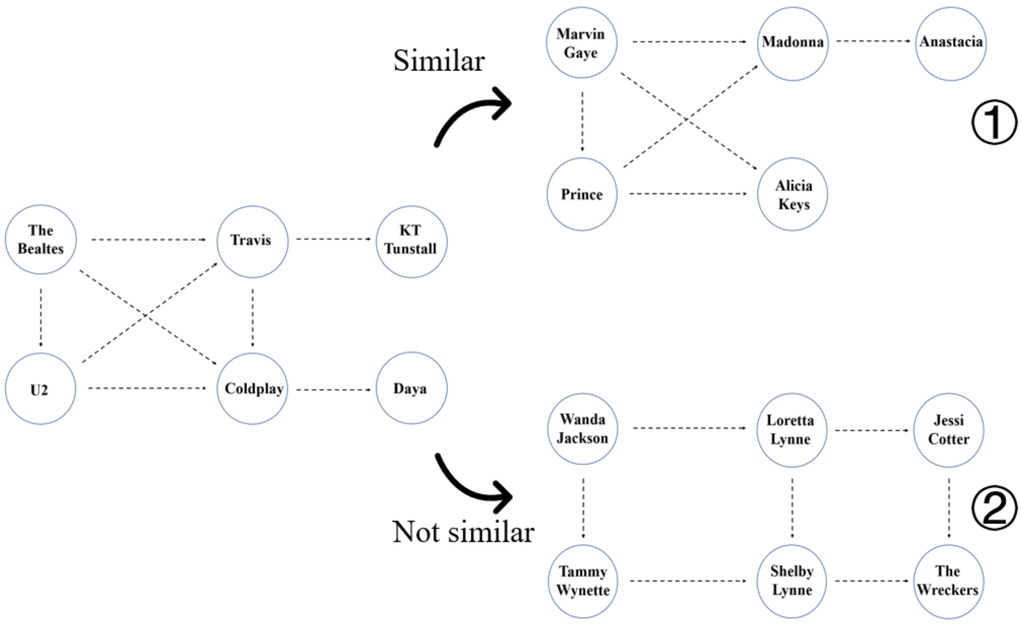

Figure 10 Similar and not Similar Subgrapghs

Taking the left picture as a benchmark, according to the definition of graph kernel, picture ① is similar to the left picture, while picture ② is not similar. This shows that the information transmission structure of the artists in the two decomposition diagrams is different from that of the artists in the left picture.

Through the calculation of the graph kernel, due to the different graph kernels between the subgraphs, the degree of difference in the local information transmission methods between different genres can be obtained.

4.6 Use the Similarity Model to Check the Authenticity

In order to explore whether influencers really have an influence on followers, we select musicians with influence relationships in *influence_data*, and perform similarity test on their data. If the similarity between the two is high, it can be considered as influences. In order to have a comparative effect, we also selected musicians who have no influence relationship to test the results.

Taking Whitney Houston as an example, our correlation test is shown in the figure:

Influencer	Follower/Other Artist	Name	Similarity
Whitney Houston	**Follower**	Jennifer Lopez	0.776179
		Beyonce	0.591545
		Donna Summer	0.816389
		Kelly Rowland	0.471733
		Britney Spears	0.419976
	Other Aritist	Leonard Cohen	0.339351
		Tanya Tucker	0.296828
		Adele	0.136963
		Bob Dylan	0.180159
		Albert King	0.447966

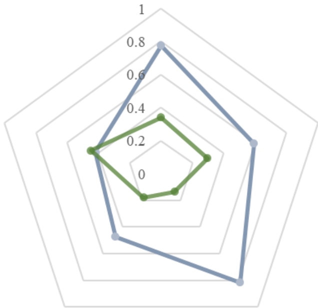

Figure 11 Similarity Scores of Related and Unrelated Artists

It can be seen that the similarities between influencers and followers that have been listed are still more significant, while the similarities between randomly selected musicians are relatively small. So we think that influencers really influence followers.

5 Evolution Model

First, we analyze the changes over time of the five major indicators of a genre, then do multiple linear regression with Popularity and we find that Musicality is the most contagious indicator. Moveout correlation analysis of the different time series of different genres proves the relationship between the two genres.

Through analysis, we found that Popularity is an indicator that can measure the evolution of music, and further identify the revolutionary time and influencers of major change in the evolution of music.

Throughout the history of music development from 1920 to 2020, we have found Acoustics as the indicator representing the dynamic influencer of music. Finally, we analyzed the development history of Pop/Rock music.

5.1 Changes of Five Major Indicators Over Time

In order to measure the changes of genres over time, we chose to use the previous method: treating the genre as an artist and analyzing the indicators of its music with the indicators of all genres. Let's take R&B music as an example, and its changes over time show such results (the time of vacancy is that no R&B song appeared in that year):

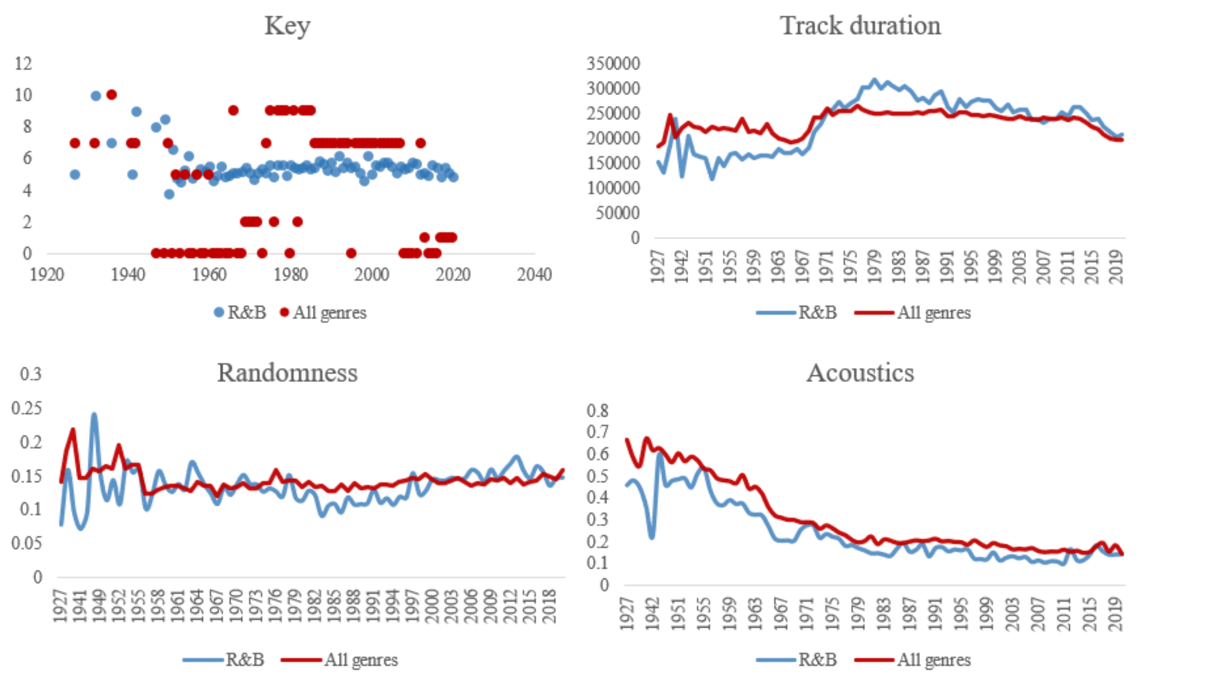

Figure 12 Four Major Indicators of R&B and All Genres

From the above figures, we can find that in terms of Key, the key of R&B music basically remains stable. In contrast, the Key of other genres fluctuates more obviously, indicating that over time, R&B's music recognition is still very high and has a strong representation. In terms of Randomness, the Randomness of R&B has not changed much, but the degree of fluctuation is still greater than the average level of the genre, which shows that R&B is more free. In Track Duration, R&B was lower than average at first, and then higher than average. It was not until around 2006 that the two recovered to basically the same level and began to decline, indicating that with the development of the times, R&B artists are more concerned about music Quality, not length. In terms of Acoustic indicators, the decline of the two acoustic indicators shows that with the development of music technology, music is more pursuing rhythm and melody and lyrics, and sound repairing technology is more perfect.

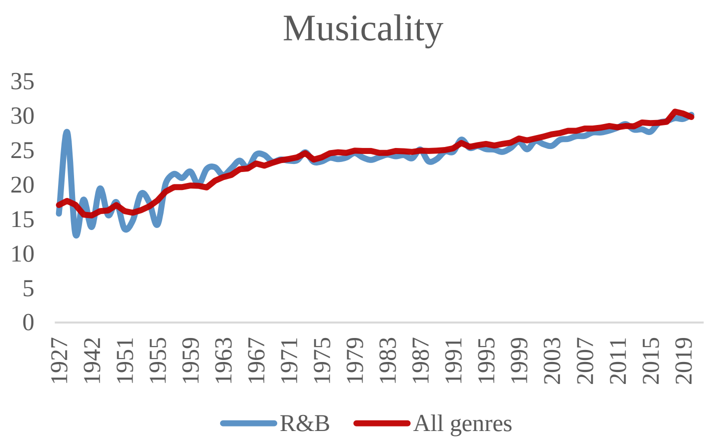

Figure 13 Musicality of R&B and All Genres

As for the Musicality of R&B, as one of the most essential characteristics of music, the average level of R&B and genre remains the same, and the overall trend is rising. It shows that the musicality of R&B has gradually improved over time, and the overall artistry has risen.

5.2 Compare the Infectiousness of Each Musical Characteristic

In order to test whether certain characteristics are more infectious or the influence of each characteristic on influence is the same in the influence of music, we carried out a multiple linear regression of the five major characteristics of musicians and the number of musicians they affected. The principle is as follows:

$$\textit{Influencing number} = \beta_0 + \beta_1 musicality + \beta_2 key + \beta_3 acoustics + \beta_4 randomness + \beta_5 duration\_ms + \varepsilon$$

The result is as follow:

Source	SS	df	MS			
				Number of obs	=	3,784
				F(5, 3778)	=	6.61
Model	17221.9287	5	3444.38574	Prob > F	=	0.0000
Residual	1970042.07	3,778	521.451051	R-squared	=	0.0087
				Adj R-squared	=	0.0074
Total	1987264	3,783	525.314301	Root MSE	=	22.835

influence_~m	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
musicality	-.2458278	.1051638	-2.34	0.019	-.4520111	-.0396445
key	-.5234896	.1057825	-4.95	0.000	-.730886	-.3160933
acoustics	-3.275377	2.237843	-1.46	0.143	-7.662875	1.11212
randomness	1.989674	4.91155	0.41	0.685	-7.639873	11.61922
duration_ms	-7.51e-06	4.39e-06	-1.71	0.087	-.0000161	1.09e-06
_cons	22.86843	3.261583	7.01	0.000	16.4738	29.26307

Figure 14 OLS result

As shown in the figure, we can see that the relationship of different indicators to the influence of musicians is different, among which Key and Musicality are more significant. It shows that these two indicators have played a greater role in the process of music influence transmission.

However, considering that Key has a small relationship with influence in reality, we finally chose to use Musicality as the most infectious indicator in the influence process.

At the same time, the F test result of the multiple linear regression is 0.0000, indicating that the overall significance of the regression is relatively high and has reference value.

5.3 Find the Relationship between Genres: Moveout Correlation Analysis

When measuring the influence between genres, since itis difficult for each music genre to respond synchronously to externalinfluences, we choose to use Moveout correlation analysis to consider the lagof the influence of one genre on another genre. Because the two columns of timeseries data are longer. We introduce a sliding window here to analyze thecorrelation between the two genres at different times.Sliding window: Frame the time series according tothe specified unit length to calculate the statistical indicators in the frame.It is equivalent to a slider with a specified length sliding on the scale, andthe data in the slider can be fed back every time it slides one unit.In order toillustrate whether there is a mutual influence relationship between the twoschools, we take R&B and Jazz as examples, and conduct a Pearson correlation coefficient test. For x_i and y_i, the formula forcalculating the Pearson correlation coefficient is:

$$r = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N\sum x_i^2 - (\sum x_i)^2}\sqrt{N\sum y_i^2 - (\sum y_i)^2}}, i \le N$$

The result is as shown:

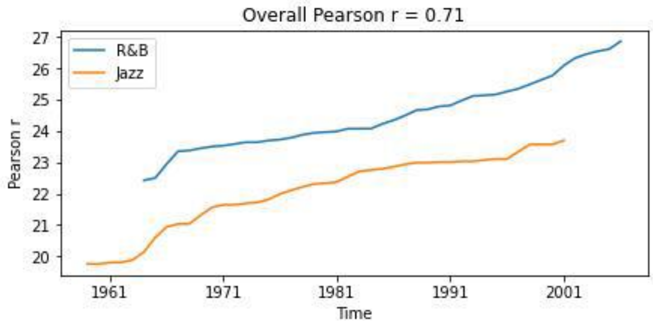

Figure 15 Musicality of R&B and Jazz Change over Time

It can be seen that the Pearson coefficient of the two is as high as 0.71, and the overall development trend is relatively consistent. In order to further study the correlation between the two, we consider using the time window to explore the correlation between the two and the time. Relationship between.

Firstly, we calculated the length of the time window, and the result of the calculation proved that the window length of 150 frames is more appropriate, and its impact under this condition is 5 years, that is, the time series of Jazz lagging 5 years has the highest correlation with R&B.

The final result is shown in the figure:

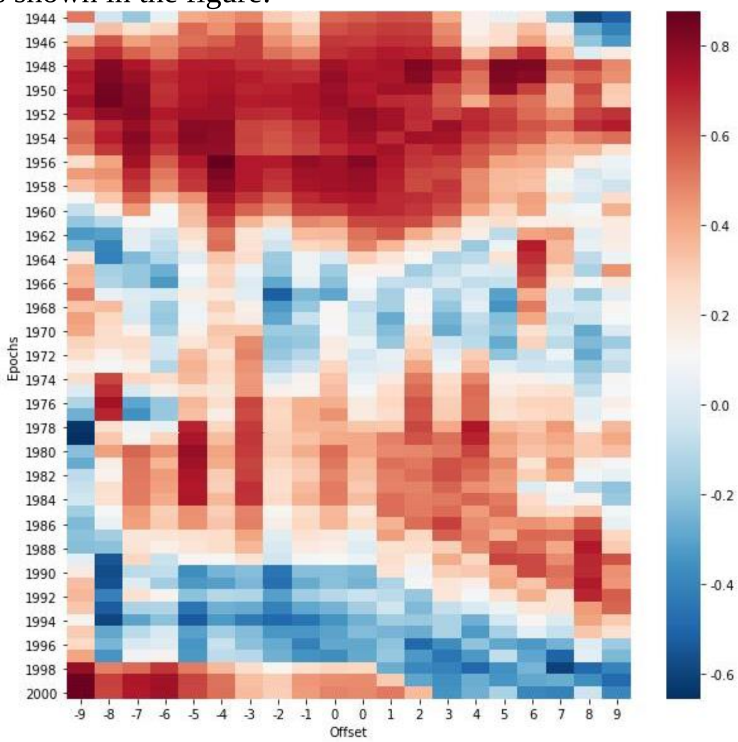

Figure 16 Heat Map of Moveout Correlation Analysis

We choose Jazz as the measurement standard. The correlation between R&B and Jazz changes over time as shown in the figure. It can be seen that around 1950, the correlation between the two was very strong, but in 1960, this correlation weakened, and it was not until the 1980s that it became short-lived again.

The results show that there is a correlation between two different genres, and certain genres will exert influence on other genres.

5.4 Music Revolution Analysis

Music is a constantly changing art. To measure major changes in music, we need to measure changes in music from multiple angles. However, it is not difficult to find that whether it is the external environment or the various indicators of the genre itself, the most intuitive reaction will eventually reflect the popularity of the music, and finally realize the transformation of the entire music genre. Therefore, we choose Popularity as a measure of music revolution.

We select the average of the popularity of all Pop/Rock songs as the y-axis, and time as the x-axis to draw the time series curve. It is not difficult to find that when the curve has an obvious turning point, it indicates that the genre has undergone major changes. According to the analysis of inflection point detection in book *Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior*, the inflection point of the Popularity curve can be obtained as shown in the figure below:

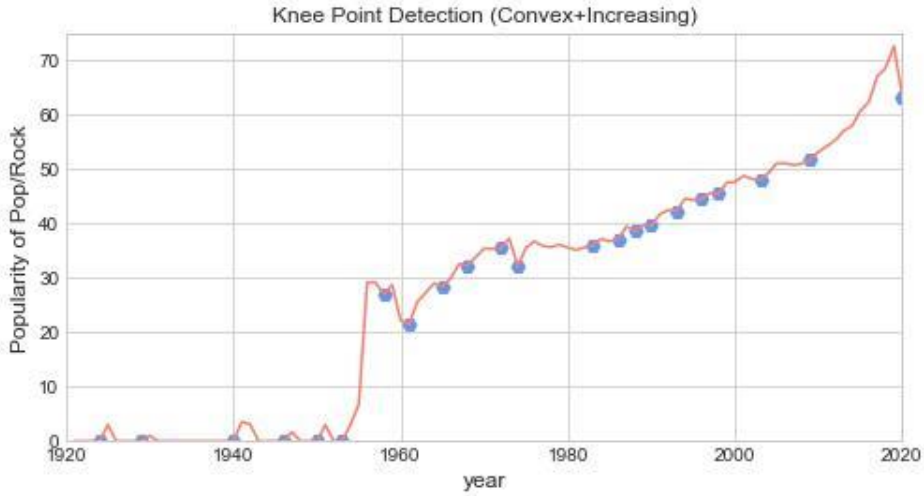

Figure 17 Original Knee Point Detection

Through our observations, a genre will not undergo a huge change in a single moment. Generally speaking, changes occur continuously in a time interval. So we combined the facts to filter the inflection points, and finally got the change of Popularity over time, and identified the time interval of major changes.

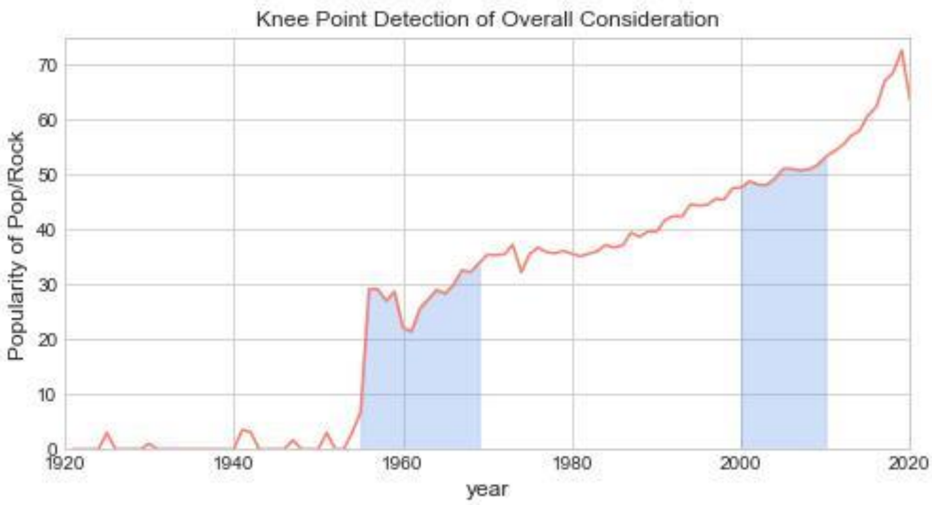

Figure 18 Knee Point Detection of Overall Consideration

In the era of these changes, as we can see is 1954-1964 and 2000-2010, there are often singers who has the most talents, representing a whole wave of change. We analyze the representatives of genre change from two aspects.

- On the one hand, when genre changes occur, there will be some creative and talented artists who will influence others while creating music. So we analyzed the number of artists in the time interval of 1954-1964 and 2000-2010. Use the number of people affected as a measure to judge whether there are artists representing genre change.
- On the other hand, considering that when the genre changes, in addition to thinking about the number of artists influencing, we have to analyze the similarity of the average characteristics between artists and genres. The aggregate scores of the two eras are as follows:

1960s			
Number of Followers		Similarity with whole genre	
The Beatles	615	The Kinks	1.846633686
Bob Dylan	389	The Beatles	1.642026825
The Rolling Stones	319	Bob Dylan	1.495623829
David Bowie	238	The Beach Boys	1.332659433
Jimi Hendrix	201	Elvis Presley	1.212673604
The Kinks	192	David Bowie	1.144277394
The Beach Boys	186	Jimi Hendrix	0.547548914
Elvis Presley	166	The Rolling Stones	0.515097318

Revolutionary of 1960s	
The Beatles	Bob Dylan

2000s			
Number of Followers		Similarity with whole genre	
Avril Lavigne	14	Fall out Boy	1.257599299
The All-American Rejects	11	Rihanna	0.705109543
Rihanna	9	My Chemical Romance	0.688589252
The Killers	8	Avril Lavigne	0.594415997
My Chemical Romance	8	The All-American Rejects	0.145345087
Kings of Leon	7	The Killers	0.034213283
Fall out Boy	6	Kings of Leon	-0.318154266
Katy Perry	5	Katy Perry	-0.680263852

Revolutionary of 2000s	
Rihanna	Avril Lavigne

Figure 19 Revolutionaries of 1960s and 2000s

5.5 Determination of Indicator of Dynamic Influencer

Based on the above results, we can clearly see that Pop/Rock has undergone major changes in the 1950s and 2000s.

The end of World War II in 1950, the rapid development of technology, followed by the popularization of television. Many singers perform on television, making music more and more popular . Around 1960, the advent of portable transistors made it possible for more and more people to listen to music at home. In addition, the characteristics of Pop/rock itself make this genre easier to be accepted by the public. On the other hand, a large number of talented artists have emerged during this period of time. Their own creativity in music and their superb musical skills have made Pop/Rock have undergone tremendous changes in both popularity and musicality. In the 2000s, with the popularization of the Internet, people's difficulty in obtaining music was greatly reduced. In addition, people's demand for spiritual culture is getting higher and higher, which makes the popularity of Pop/rock has a qualitative change.

In order to determine whether there is an indicator that the music genre is a dynamic indicator in the process of development and evolution. We decided to adopt Dynamic Time Warping(DTW) to measure the synchronization of various indicators and the entire genre in the process of change.Dynamic time warping (DTW) can find the best alignment between two given (time-correlated) sequences under certain constraints. Intuitively, the sequences are warped in a non-linear manner to match each other.

Taking the popularity of the genre as the x-axis, we have selected the five major indicators as the y-axis, and the white curve in the figure can reflect the degree of synchronization between

Popularity and other music indicators in the process of change. The closer the curve is to a straight line that crosses the origin and has a slope of 1, it means that the two are more synchronized in the changing process. We were pleasantly surprised to find that the degree of fit between Acoustics and popularity is very high. Comparing the degree of fit between other indicators and popularity, we can determine that Acoustics is a dynamic impact indicator.

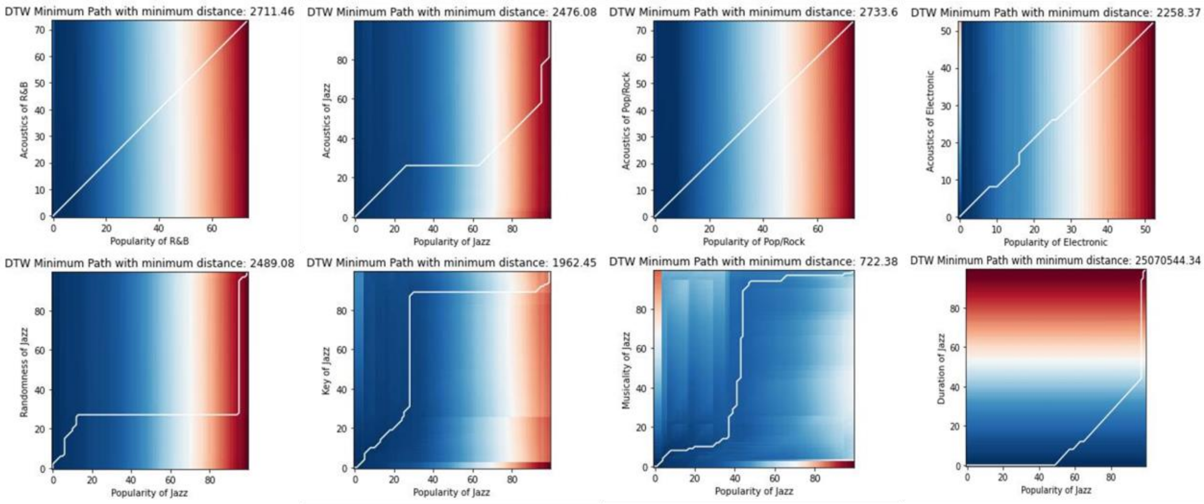

Figure 20 DTW Minimum Path of Popularity with Acoustics and Other Indicators

Since 1960, stereo and other technologies have gradually become popular, recording technology has developed rapidly, and the quality of sound-to-electric conversion, sound signal transmission and storage has gradually improved. Resulting in the gradual reduction of music requirements for singers' pronunciation skills, and artists can use it according to the taste of listeners. Devices create more popular music, which is why Acoustics is so relevant to popularity.

6 Extension of the Models

6.1 What's the cultural influence of music?

Music spreads its influence to the outside world all the time, and music can be seen everywhere in our life. We believe that when music influencers influence followers, the increase in the number of artists of a genre will have a similar-cultural impact on the society. When a genre has the largest proportion of artists, its cultural impact on society is the greatest.

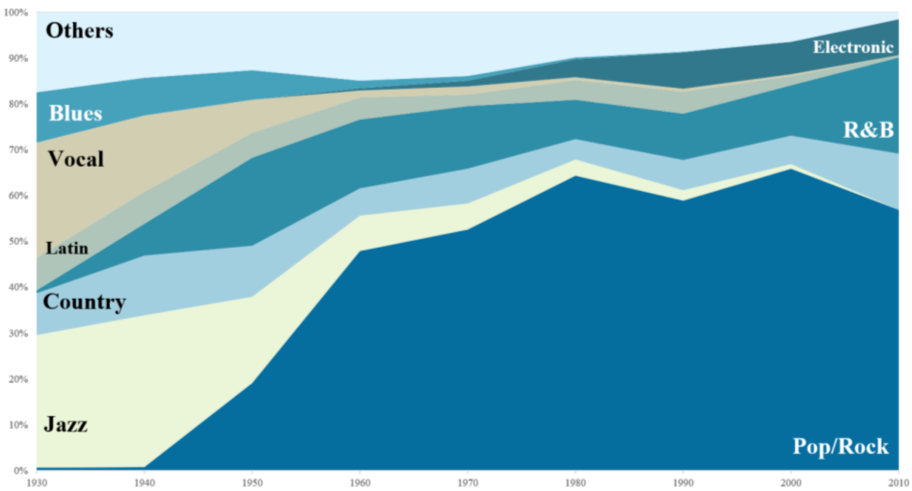

Figure 21 Number of Artists of Each Genre in Different Times

From the perspective of time, in the 1930s and 1940s, jazz music became popular and had the most artists. The cultural influence of jazz reached its peak at this time-the artistry of improvisation has affected a large number of artists and audiences. At the same time, people's thinking about race caused by jazz is also a new trend of thought. Similarly, in the 1960s, Pop/Rock music began to flourish, not only in terms of numbers, but also great artists such as The Beatles and Bob Dylan were born. Both of them, as the most influential artists, are two of the greatest artists all time. Their work has affected the cultural trend of the entire society and changed the view of life of many people.

“(The Beatles is) more popular than Jesus”. -John Lennon, 1966.

“(Bob Dylan) having created new poetic expressions within the great American song tradition.”-The Nobel Committee for Literature, 2016

Today, the charm of Pop/Rock has not diminished at all, and it still has the largest proportion of artists in the music industry. When Taylor Swift or Post Malone release new songs, they will be heatedly discussed in the community. People all want to hear their songs first. Although it's crazy, this is the charm of music, isn't it?

6.2 How does the sociey affect music?

Music is changeable, whether it is within or outside the genre, the various characteristics of music are always changing. So how to reflect the changes in the outside world-such as political, social and technological changes, on the impact of music is a question worthy of research.

Taking a music genre as an example, we consider from two perspectives:

- When the music characteristics of the genre change, we can observe the reason for the change through our genre influence and similarity model. If there is no strong correlation variable found in our model, that is, within the music, then we can think This change is due to changes in the outside world. For example, a sudden drop in the acoustics of a certain music in a certain period of time may be due to the emergence of new and better sound repair techniques.
- When the music characteristics of this genre have not changed significantly, but its influence on other genres and the number of artists have increased significantly, we think that this is a change in the overall society's preference for music style. Because the society’s preference for music styles is unpredictable by artists, this variable is strictly an external variable, a variable that is caused by changes in the political, social, and technological aspects discussed.

7 Sensitivity Analysis

- When using the sliding window for moveout correlation analysis in 5.3, the results of the moveout correlation analysis are also different due to the different sizes of the time windows. When the time window is set larger, the time difference analysis cannot reflect whether the specific time point is affected. When we set the window to 30, the accuracy of the data will be greatly reduced. It shows that the accuracy of the data is highly sensitive to changes in window size.
- When performing inflection point analysis in 5.4, the number of inflection points screened out is different due to the different sensitivity parameter settings during the analysis. When our sensitivity parameter is set to 1, there are 16 inflection points obtained. When we set the

sensitivity parameter to 10, the inflection point analysis changes to the following figure. It means that the number of inflection points displayed is more sensitive to changes in sensitivity

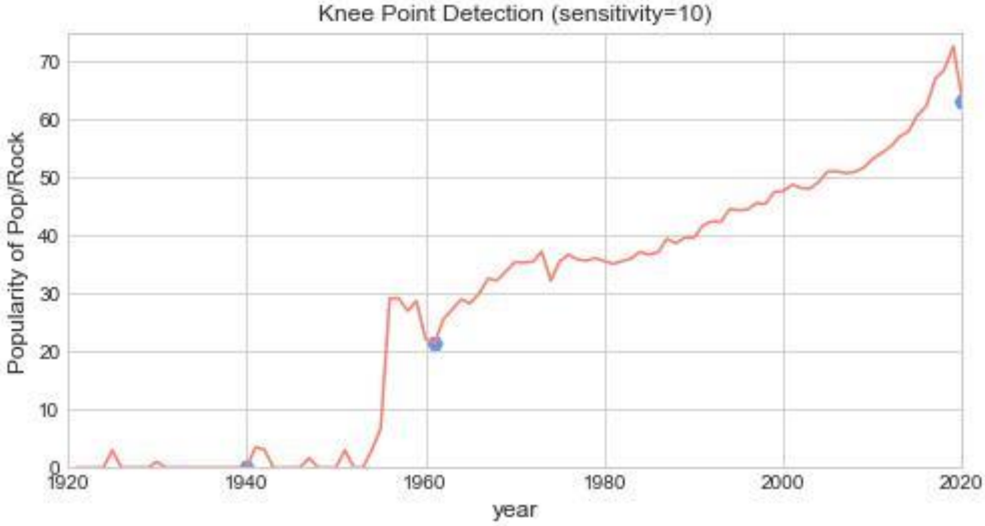

Figure 22 Knee Point Detection When Sensitivity=10

8 Strengths and Weaknesses

8.1 Strengths

- Our method is highly compatible with the problem, and we use various analysis flexibly to create a set of network scientific analysis models for music development.
- The selection of parameters is remarkably scientific, reflecting the influence of music and the correlation of music characteristics from multiple dimensions, and the results are true and credible.
- In this paper, inspired by the algorithm of face recognition in engineering, we solved the problem of distance and weight in the similarity calculation of various musical indicators, and the similarity degree of different artists is accurately quantified.
- We take the lag between the effects of music into account, and innovatively use a time series sliding window for time difference analysis, the analysis results accurately reflect the real changes in music genres.

8.2 Weaknesses

- We didn't analyze the evolution of each genre, and didn't consider the development of genres with fewer artists. The consideration of niche genres is relatively lacking.
- Only the average value is used to represent the musical characteristics of a genre, and no other processing methods have been tried.

References

[1]Kim J R, Kim J, Kwon Y K, et al. Reduction of complex signaling networks to a representative kernel[J]. Science signaling, 2011, 4(175): ra35-ra35.

[2]Lai Y J, Liu T Y, Hwang C L. Topsis for MODM[J]. European journal of operational research, 1994, 76(3): 486-500.

[3]West D B. Introduction to graph theory[M]. Upper Saddle River: Prentice hall, 2001.

[4]Xu L, Xu H, Yu J, et al. Linkage effects mining in stock market based on multi-resolution time series network[J]. Information, 2018, 9(11): 276.

[5]Satopaa V, Albrecht J, Irwin D, et al. Finding a" kneedle" in a haystack: Detecting knee points in system behavior[C]//2011 31st international conference on distributed computing systems workshops. IEEE, 2011: 166-171.

[6]Blanz V, Vetter T. Face recognition based on fitting a 3d morphable model[J]. IEEE Transactions on pattern analysis and machine intelligence, 2003, 25(9): 1063-1074.

[7]Müller M. Dynamic time warping[J]. Information retrieval for music and motion, 2007: 69-84.

[8]Jain A K, Li S Z. Handbook of face recognition[M]. New York: springer, 2011.

[9]Ward G C. Jazz: a history of America's music[M]. Knopf, 2000.

A report to ICM

After our analysis, we found that using network science to study music brought interesting results.

First, we analyzed 72446 pieces of music from more than 7,200 musicians belonging to 20 different genres and established a music network for them. Music network can directly reflect the influence, inheritance and closeness of different musicians and genres.

Through the similarity analysis in the artist network, we found that musicality, key, randomness, track duration, and acoustics are the five major indicators that reflect the characteristics of music, and the development of music is closely related to them. Through the similarity analysis in the genre network, we can have a more accurate grasp of the characteristics of each genre, which is conducive to academic research on the transformation of music genres. Simultaneously, it proves that progress in musical technology has a major impact on the field of art. More importantly, music network can also show the influence of social culture on music along with the reaction of music on society, which means it can complete the task of identifying external influences.

The above results come from the analysis of the given limited data set. If we can access richer music data, such as data about artists in the 19th century, minority genres in the 20th century, and the degree of influence among artists, then we can operate more precisely, we can further consider how the internal evolution of one genre leads to the rise of another genre, what are the differences in the pursuit of musical characteristics between different genres, and whether the influence of singers is more related to other indicators, etc. More significantly, the increase in data will make our model of influence and similarity more accurate and increase the reliability of our results.

As time changes, genres and artists never stop changing, and there are intricate influences between each other, so you need to further study this field. The application of music network to the music industry is effective and efficient. We recommend using the music network model for analysis. Music is a part of culture and at the same time a major influencer of society. In the context of the general trend of globalization, the world's music will have a greater integration, and its changes imply changes in world culture. Research in this field is of great significance. Hope you can carefully consider this approach of exploring music industry.

Yours sincerely,
Team #2100888