Problem Chosen	2023 MCM/ICM Summary Sheet	Team Control Number
D		880

Has the influence of musicians influenced the evolution of music?

The interplay of musicians from different genres played a key role in the evolution of music. In this paper, the social network analysis model is established to analyze the mutual influence among musicians, and then the index to measure the similarity of music is introduced, and on this basis, the style change of music on the time scale is analyzed.

For problem 1, we first build a **directed graph** with musicians as nodes and influence relationships between musicians as edges. The weights of the edges in the directed graph are determined by combining the time span and genre span of influence. Then we calculate the point degree centrality and finally derive the influence of musicians. Finally, the **Pagerank algorithm** is used to modify the eigenvector centrality to evaluate the influence of musicians. It is found that the **Beatles, the Rolling Stones, Bob Dylan and other 7 musicians have a high influence in the music industry**. Through the distribution of musicians' influence, we also find that the social network satisfies the **scale-free network** condition, which means that the influencers and followers satisfy Pareto's Principle, that is, a smaller number of musicians influence the majority of musicians.

For problems 2 and 3, we use Euclidean distance and cosine distance to determine the index of **global similarity** of music, and traverse the most dominant combination of features from all 13 music features to determine the index of **partial similarity** of music, and define the influence factor of popular trend to reduce the influence of popular trend on musician's style to obtain the index of similarity of music, MS. Finally, using Pop /Rock and R&B genres as examples, 200 songs are randomly selected from the two genres to calculate MS one by one, and the similarity density distribution is plotted. It is found that **the music within genres is more similar than that between genres.** In addition, we find that "loudness" and "instrumentalness" are the main factors influencing the genre of Pop/Rock.

For question 4, a **sliding window-based long-short term impact evaluation model** is developed to explore whether influencers actually influence the music of their followers. We sample Bob Dylan and his three followers as example and find that the **influencer's musical style does have either a long-term or short-term influence on the followers.** Finally, we modify the model in question 2 by comparing musical similarities year by year and find that the **music scene underwent significant changes in seven years, including 1929, 1938, 1945, and 1949.**

Keywords: Directed graph, Scale-free Network Pagerank, Sliding Window Method

Contents

1 Introduction

1.1 Background and Problem Restatement

Music has witnessed the development of human history for thousands of years. At the same time, it has undergone continuous and profound evolution with social changes. Perhaps we can analyze its development and shifting trends in terms of the visible characteristics contained in music, even though it is commonly considered an activity to be appreciated with the ear and mind.The development of computer technology has made it possible to extract and analyze crucial factors influencing the development process of music by establishing influence networks among musicians. In this paper, the following goals need to be achieved:

- Establish a network describing the influence among musicians, determine the parameters to capture the "music influence" in the network, then select a subnetwork, explore the phenomenon explained by the subnetwork.
- Develop a model to measure musical similarity and use it as a basis to explore the connections and interactions between different musicians and genres.
- Identify the significant revolutions that occurred in certain music genres from the data given and identify the prime influencers of these major changes.
- Select a musical genre as an example to reveal significant indicators of change in the dynamic evolution of musical genres, and try to identify and demonstrate the external factors that influence changes in musical genres.

1.2 Our Work

2 Assumptions and Explanations

Considering that practical problems always contain many complex factors, first of all, we need to make reasonable assumptions to simplify the model, and each hypothesis is closely followed by its corresponding explanation:

Assumption 1: **Assuming that the data is accurate and there is no falsified data.**
Explanation: Only when the data is accurate can our analysis fit the facts

Assumption 2: **Assuming that followers can only be influenced by songs released before the current year, and that songs after the current year have no effect on the follower's for that year.**
*Explanation:*Because of the sequence of releases, musicians are unlikely to be influenced by music in future.

Assumption 3: **Assuming that the decade when the musician begin his carrer is the year when they debut.**
*Explanation:*It is difficult to know the exact debut year of each musician, since media technology for most eras in the data set is underdeveloped.

Assumption 4: **Assuming that the musician's genre is limited to the genre provided by the data, and have no genre changes, that all his/her songs are in**

the style of his or her genre
*Explanation:*If a musician is active in multiple genres, it is difficult to identify the genre to which the song belongs only from the provided song feature data, making the model no longer accurately reflect genre information.

Additional assumptions are made to simplify analysis for individual sections. These assumptions will be discussed at the appropriate locations.

3 Preparation of the Models

3.1 Notations

Table 1: Parameter Settings

parameter	description
T	time-span influence factor
D_{Ci}	degree centrality
E_{Ci}	eigenvector centrality
MI	music influence
S_{global}	global similarity
S_{part}	partial similarity
$V^{(i)}$	music feature vector
MS	multi-similarity
SMS	short-term similarity
LMS	long-term similarity

3.2 Data Preprocessing

Since the characteristic data of choral (more than one author) musical works cannot precisely reflect the musical characteristics of each of the choristers, all musical works with plural authors are removed from the data processing.

Table 2: Data discard rate for different tables

full_music_data	**data_by_year**	**influence_data**
0.0448	0.0521	0.0621

4 Musician Influence Model Based on Social Network Analysis

Since there is a directed relationship between followers and influencers, directed graphs can be used to describe the mutual influence relationship between musicians.In the directed graph, musicians are nodes. When there exists a following relationship between two musicians, there will be an edge from the follower to the influencer.

4.1 The Determination of Indicators

By analyzing the given dataset, it can be determined that the influence of a musician can be affected by the following factors:

- The time span of a musician's influence.
- The number of people influenced by a musician.
- Whether the influence of a musician breaks the genre boundary.

4.1.1 Time Span Impact Factor

Assuming that the existing musician i influences musician j, the time difference of the influence is defined as follows:

$$\Delta t = t_j - t_i \tag{1}$$

As an objective rule, predecessors who were active earlier are more likely to influence the musical style of their successors, and the greater the time span, the more influential the predecessor tends to be in his or her genre. However, if a predecessor's musical style is influenced by a successor, this indicates that the successor is highly influential.Therefore, we define the time span impact factor T as follows:

$$T = \begin{cases} -\beta \Delta t, & \text{if } \Delta t < 0 \\ \Delta t, & \text{if } \Delta t \geq 0. \end{cases} \tag{2}$$

Where, β (β>1) is the time span impact factor.

4.1.2 Genre Span Impact Factor

In general, musicians of the same genre are more likely to influence each other, and they make their music better by learning from each other. The cultural and technical differences between different genres make it difficult for musicians to influence each other. However, if a musician is able to influence musicians from other genres, he or she has a great deal of influence.The cross-genre influence factor between influencer i and influencer j is defined as follows:

$$a_{ij} = \begin{cases} 1, & \text{cross-genre} \\ 0, & \text{none cross-genre} \end{cases} \tag{3}$$

4.1.3 Edge Weights of a Directed Graph

With the above two parameters, we can define the edge weights of the directed digraph as follows:

$$w_{ij} = (1 + a_{ij})\, e^{\frac{T}{80}} \tag{4}$$

where w_{ij} is the weight of the directed edge between musician node i to node j. Since the time span is a parameter with units, it is normalized with the maximum time span of 80 years. w_{ij} also represents the intensity of the influence of musician i on musician j. The larger the value, the stronger the influence intensity.

4.2 Influence Evaluation Model Based on SNA

We have obtained the weight calculation method of edges in the directed graph, but it is still needed to define the evaluation method of influence weight of musician node itself. By calculating the degree of the nodes and using the Pagerank algorithm, it is possible to calculate the node weights (i.e. musician influence).

Step1: **Calculation of node degree centrality** In a directed digraph, the in-degree of a node can represent the number of people affected by a musician. The larger the in-degree of a node is, the more people a musician has influenced ,reflecting a greater influence.

Therefore, the degree centrality of node i is:

$$D_{Ci} = \sum_{j=1}^{n} w_{ij} d_{ij} \tag{5}$$

Where n represents the total number of nodes in the network, and d_{ij} represents whether there is a directed edge between node i and node j. The definition is as follows:

$$d_{ij} = \begin{cases} 1, & \text{there is directed edge from} i \text{ to } j \\ 0, & \text{no directed edge from } i \text{ to } j \end{cases} \tag{6}$$

Step2:Calculation of centrality of eigenvector

The influence of a musician is not only related to the number of his followers, but also to the influence of his followers. If a musician has influential musicians among his followers, it means that the musician is very influential.

We define the centrality of the eigenvector E_{Ci} to measure the macro-importance of node i, as shown in the formula:

$$E_{Ci} = \lambda \sum_{j=1}^{n} d_{ij} E_{Cj} \tag{7}$$

Where λ is the proportionality constant and the initial value of E_{Ci} is D_{Ci}

The Pagerank algorithm can be used to synthesize the importance of nodes and the importance of the connected edges. In this way, a more comprehensive evaluation model of musician influence can be established.The revised influence evaluation model can be expressed as:

$$E_{Ci} = \lambda \sum_{j=1}^{n} d_{ij} E_{Cj} + (1-\lambda) \sum_{j=1}^{n} w_{ij} d_{ij}, j = 1 \sim n \tag{8}$$

After several iterations, E_{ij} can finally converge to a stable value, and musician influence is defined as follows:

$$MI_i = \frac{E'_{Ci}}{\max\left\{E'_{Cj}, j = 1, 2, \ldots, n\right\}} \tag{9}$$

Table 3: Top 10 influential musicians

Musician_name	MI	Musician_name	MI
The Beatles	1	Chuck Berry	0.344699
The Rolling Stones	0.586876	Little Richard	0.336405
Bob Dylan	0.54064	Elvis Presley	0.322267
Hank Williams	0.352528	Jimi Hendrix	0.313048
The Velvet Underground	0.34826	The Who	0.308027

4.3 The Result and Analysis of Musician Influence Model

All the data of influence_weight are input into the SNA model, and after nearly 3000 iterations of the Pagerank algorithm, the final results are obtained as following:

As can be seen from The table above, **The Beatles** has the greatest influence, far ahead of **The Rolling Stones** and **Bob Dylan**. The gap between the 4th and 10th most influential musicians is not such huge.

The most influential musicians obtained by our analysis are highly similar to the 20 most representative musicians in the world jointly selected by CNN and Songlines, a world-class music publication. It can be considered that the establishment and analysis process of the model is accurate, reasonable and feasible.

4.3.1 The Analysis of Our Result

In reality, the probability that musicians interact with each other is completely random is very small. Generally, a small number of high-influence musicians in the music industry influence most people, which is consistent with the conclusion of the pareto principle.

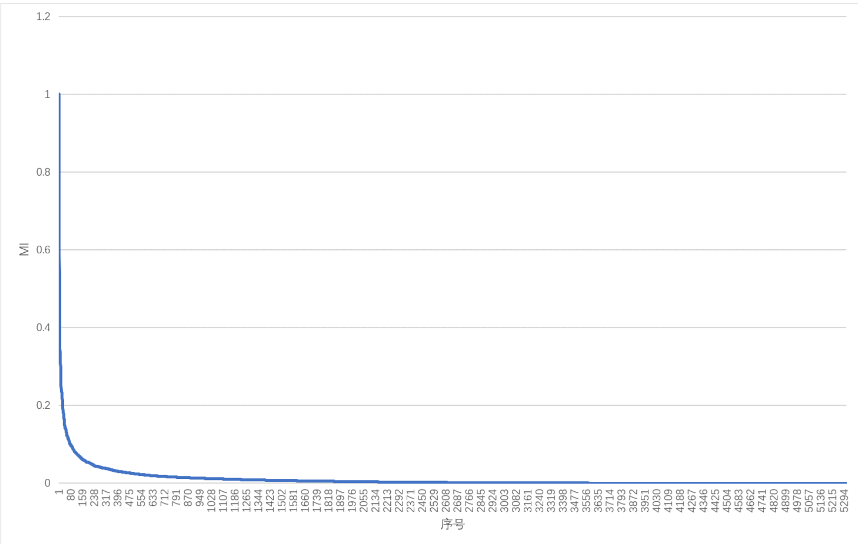
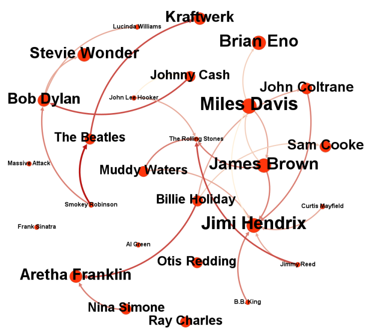


Figure 1: MI distribution diagram

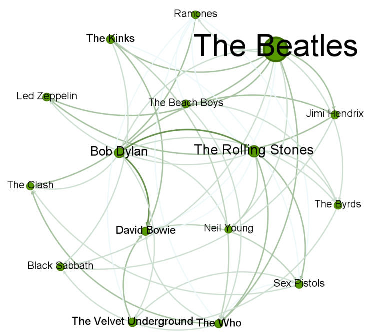
The approximate power-law distribution can be obtained by ranking the musicians from the most influential to the least influential, and the network can be preliminarily considered as a scale-free network.

4.3.2 Analysis of Musician Influence Subnetworks

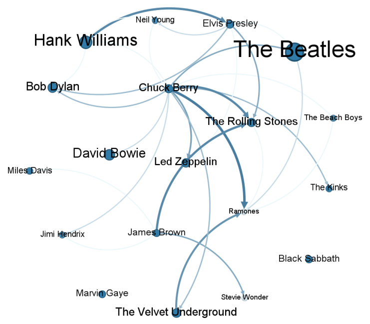
After the establishment of the musician influence model, we divided the data set into four categories, and analyzed the influence of musicians within genres, cross-genre influences, the influence of seniors on juniors, and the influence of juniors on seniors.



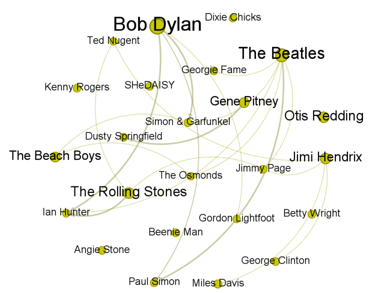
(a) Cross-genre influence



(b) Intra-genre influence



(c) Predecessor influence junior



(d) Junior influence predecessor

Figure 2: Sub-network of Musical Influence

In the figure above, the size of the nodes in each subnetwork represents the MI of the singer's corresponding type, while the thickness of the directed curve represents the weight of influence. As can be seen from the figure above, intra-genre influences occur more frequently than inter-genre influences, and older generations' influences on younger generations occur more frequently than younger generations' influences on older generations. It can be believed that in the interaction of musicians, musicians are more likely to influence musicians in the same school, and musicians are more likely to influence the younger generation rather than the predecessors.

Specific to certain musicians, it can be found that the Beatles has a very strong influence within the genre with as predecessors. Bob Dylan, however, has less influence than the Beatles within the genre, but as a junior has a notable impact on his predecessors, which is an uncommon situation in the music industry.

5 Music Similarity Evaluation Model

5.1 Establishment of Multi-dimensional Music Similarity Evaluation Model

By using all 12 features in the $full\_music\_data$ file, the overall similarity between different music can be obtained. This similarity can be realized by abstracting the features into 12-dimensional vectors.

5.1.1 Global similarity

The global similarity can be expressed by the Euclidean similarity and cosine similarity between two vectors:

Euclidean similarity:

The Euclidean similarity measure is based on the absolute distance between the feature vectors and is used to reflect the absolute difference in the feature values between the two songs. Then the distance between the eigenvectors of music A and music B $distance_{AB}$ is defined as follows:

$$distance_{AB} = \sqrt{\sum_{i=1}^{n}(A_i - B_i)^2} \tag{10}$$

Then the Euclidean similarity between music AB $S_E(\mathbf{A}, \mathbf{B})$ is defined by the following equation:

$$S_E(\mathbf{A}, \mathbf{B}) = \frac{1}{1 + distance_{AB}} \tag{11}$$

Cosine similarity:

Unlike Euclidean similarity, cosine similarity is more concerned with the angle between two vectors than the distance. We define cosine similarity as follows:

$$S_{cos}(\mathbf{A}, \mathbf{B}) = 1 + \frac{cos\theta}{2} = 1 + \frac{\mathbf{A} \cdot \mathbf{B}}{2 \cdot ||\mathbf{A}|| \cdot ||\mathbf{B}||} = 1 + \frac{1}{2} \cdot \frac{\sum\limits_{i=1}^{n} A_i \cdot B_i}{\sqrt{\sum\limits_{i=1}^{n}(A_i^2)} + \sqrt{\sum\limits_{i=1}^{n}(B_i^2)}} \tag{12}$$

Then the global similarity can be expressed as:

$$S_{global}(\mathbf{A}, \mathbf{B}) = S_E(\mathbf{A}, \mathbf{B}) + S_{cos}(\mathbf{A}, \mathbf{B}) \tag{13}$$

The global similarity uses two measures of similarity to take into account the important attributes between two vectors in terms of musical similarity.

5.1.2 Partial similarity

If only global similarity is used, some music with low global similarity but which exhibits high similarity in some of these features cannot be identified. For example, some music is similar in energy, rhythm and danceability, but the global similarity of their feature vectors is low. Although we believe that these music are similar, our model cannot show them.

To deal with this issue, we introduce partial similarity.

We randomly select two, three, and four features from the key features of music to form different feature vectors $\mathbf{V}$. There are a total of N new feature vectors composed of the extracted features.

$$N = C_{12}^{2} + C_{12}^{3} + C_{12}^{4} + C_{12}^{5} = 1573$$

The new feature vectors are $V^{(1)}, V^{(2)}, \cdots, V^{(1573)}$, respectively. For music i and music j, we first calculate the S_{global} of each combination of their features and select the highest 50 S_{global} to calculate the average value as the partial similarity ($S_{part}(i,j)$) between music i and music j. $S_{part}(i,j)$ is defined in the equation(14).

$$S_{part}(i,j) = \frac{\sum_{k=1}^{50} S_{global}(\mathbf{V_i^{(k)}}, \mathbf{V_j^{(k)}})}{50} \tag{14}$$

5.1.3 Fashion Trends Impact Factor

Considering the social and historical nature of music creation, musicians are usually influenced by the current trend of music in their creation. If two pieces of music have different trends in the year they were released, they will naturally be less similar. However, this influence cannot reflect the similarity of a musician's own creative style well, so it is necessary to define and weaken the influence of popular trends on musician's style. Therefore, we define the influence factors of fashion trend as follows:

$$S_{time} = S_{global}(Year_a, Year_b) \tag{15}$$

Where $Year_a$ and $Year_b$ are the music style attribute vectors of year a and b, respectively.

In summary, we establish a multidimensional music similarity (MS) evaluation model considering the influence of popular trends, defined as follows, β is a parameter defined later (0<δ<1).

$$MS = ((1-\delta)S_{global} + \delta S_{part}) \cdot (1 + \frac{0.1}{S_{time}}) \tag{16}$$

It can be seen that when the trends of two songs differ greatly in the year of their release, their similarity gains a large coefficient correction.

5.2 Data dimension reduction

We find that some of the features given in the data set cannot reflect the similarity of musical styles well (such as popularity). The existence of such features not only increases the calculation cost, but also causes deviation in the measurement of musical similarity. Therefore, the dimension of the characteristic data of music can be reduced to make the characteristics of the input model more representative.

Principal component analysis (PCA) is a commonly used method for dimensionality reduction. The main idea of principal component analysis is to reconstruct the N-dimensional characteristics into K-dimensional characteristics on the basis of guaranteeing the explanatory property of variables (k< n).Before using PCA, it is necessary to analyze whether there is sufficient correlation between variables to reduce the dimension of high-dimensional matrix. The commonly used methods are to conduct KMO test and Bartley test for dimensionality reduction data:

Table 4: KMO test and Bartlett test

KMO value		0.634
Bartlett sphericity test	Approximate chi-square	325804.748
	df	91
	P	0.000

KMO value>0.6 indicates that the principal component analysis is effective, and Bartley spherical test indicates that there is a strong correlation between the input data, which can be processed for dimensionality reduction.

Finally, through PCA method, we reduced the 13-dimensional vector data to 4-dimensional, and marked these features as:

$$F_i = (F_{i1}, F_{i2}, F_{i3}, F_{i4}) \tag{17}$$

5.3 Result Analysis of Music Similarity

After the establishment of the similarity evaluation model for music features, we take pop and rock as examples and extract musicians of these two genre for analysis.

Firstly, we calculate the MS of two musicians in two genres respectively, then we select two musicians from each of the two genres and calculate the MS of musicians in different genres, and finally plot the three sets of calculation results into a heat map as shown below:

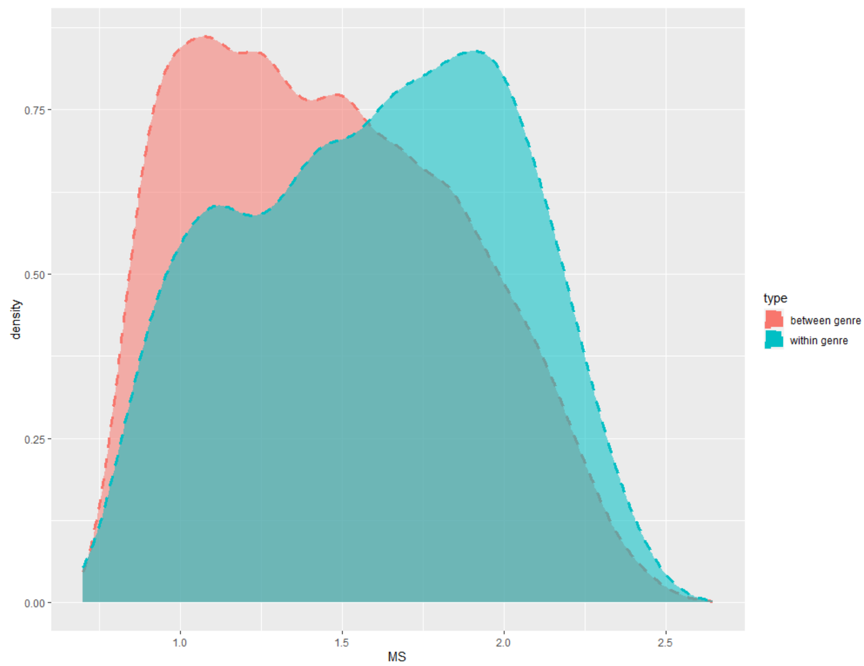


Figure 3: Density distribution of song similarity between two musical genres

The graph shows that the similarity of musicians within the pop genre is significantly higher compared to the similarity between pop and rock musicians, as is the case for rock music.

Bringing in more music genre musician data into the model using this method also leads to the same conclusion. This suggests that there is a high degree of similarity between the works of musicians within the same genre, while in comparison, there is a low degree of similarity between the musical works of different musicians.

In order to explore the main factors that distinguish a genre, we put all 13 musical characteristics into the model, calculate the overall similarity, and plot the density of MS. Then one feature is removed, MS is recalculated and density map is drawn. Finally, the density distribution of MS changes greatly after which feature is removed, so as to determine the most important characteristic in the genre.

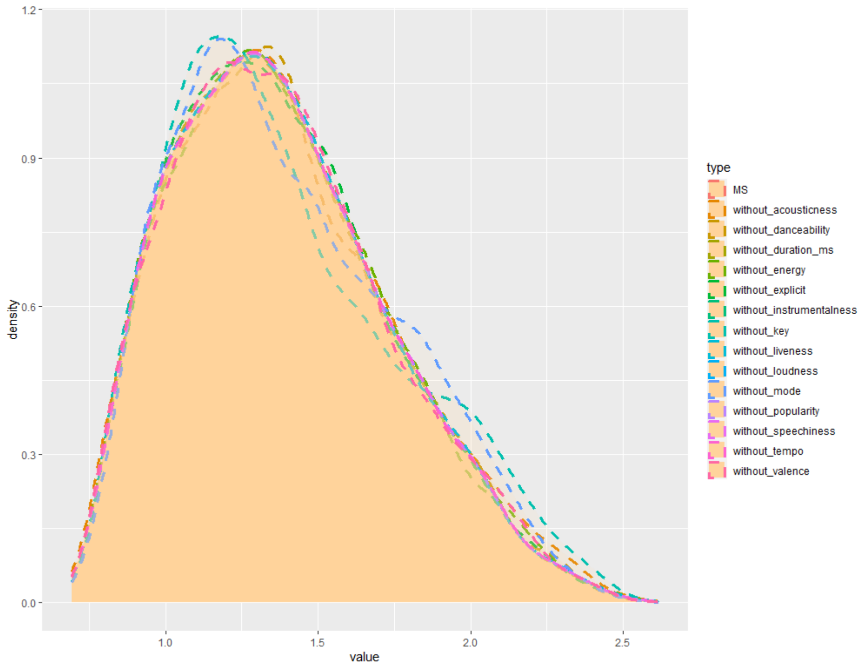


Figure 4: Comparison of key characteristics of genre music

As can be seen from the figure, when the feature "loudness" and "instrumentalness" are removed, the overall MS density map of musicians within the genre tends to shift to the left, while when other variables are removed, the MS density map does not change significantly. This indicates that the "loudness" and "instrumentalness" are the core of this type of music, and when the feature "key" no longer exists, the common feature of this type of music will be weakened.

6 Long-Short Term Influence Evaluation Model Based on Sliding Window

6.1 Establishment of Sliding Window Model

We divide the influence of influencers on followers into two categories: long-term and short-term, and capture the two kinds of influence by sliding the window model:

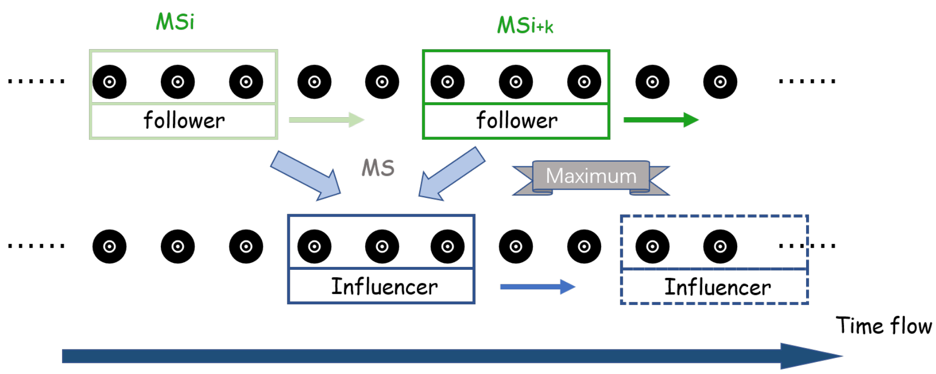


Figure 5: The slide window model

We define the similarity of the song in the i th follower window to the influencer (MS_i):

$$MS_i = max\{MS(Fl_i, In_j), \quad j = 1\ to\ n\} \tag{18}$$

Fl_i and In_j are the music in Follower window i and influencer window j, respectively. $MS(Fl_i, In_i)$ represents the short-term similarity between works in influencer window j and works in follower window i. In order to satisfy that the music release time in the influencer window is earlier than that of followers, n will change with the change of i.

Then define the magnitude of similarity change $(\Delta MS)_i$ for the follower's timeline window:

$$(\Delta MS)_i = \frac{MS_i - MS_{i-1}}{MS_{i-1}} \tag{19}$$

The short-term similarity(STS) and long-term similiarity(LTS) can be defined as follows:

$$\begin{cases} STS = max\{(\Delta MS)_i\} \\ LTS = \sum(\Delta MS)_i \end{cases} \tag{20}$$

Since the variation scale of similarity is mostly 10^{-2}, it can be considered that when STS>0.01 or LTS is greater than 0, the influencer can be considered to have a short-term or long-term influence on the followers, respectively. Define impact score IDS, if IDS>0, the influencer can be considered to have influenced the followers:

$$IDS = max(0, LTS) + max(0, STS - 0.01) \tag{21}$$

6.2 Result of the Established Model

We set the window size to 5 and the move step to 3 for window sliding detection. The IDS between all influencers and followers are calculated using the model developed above and find that 86.08% of the IDS are greater than 0. It can be concluded that in general influencers do influence the music created by their followers.

We choose the influencer: $BobDyaln$, and his followers: PJ Harvey, Travis and Spin Doctors as an example, the results are shown in the Table(3).

Table 5: Long-term and short-term influence factors of influencers on followers

Influencer	Follower	LTS	STS	IDS
"Bob Dylan"	"PJ Harvey"	0.054481751	0.003709955	0.054481751
"Bob Dylan"	"Travis"	0.106512407	-0.011158731	0.106512407
"Bob Dylan"	"Spin Doctors"	0.039947312	0.049023517	0.07897083

It can be seen that Bob Dylan has a significant long-term influence on PJ, but not only has a long-term influence on Travis, but also is influenced by Travis in the short term. As a loyal follower of Bob Dylan, Spin Doctors' music style has been significantly influenced by Bob Dylan in both the long and short term.

When exploring the possible changes of music genres, we used the data in "data by year" to calculate the similarity MS between the NTH year and the n+1 year respectively, and drew a line chart. It can be seen that the year in which MS fluctuates more is more likely to have significant changes:

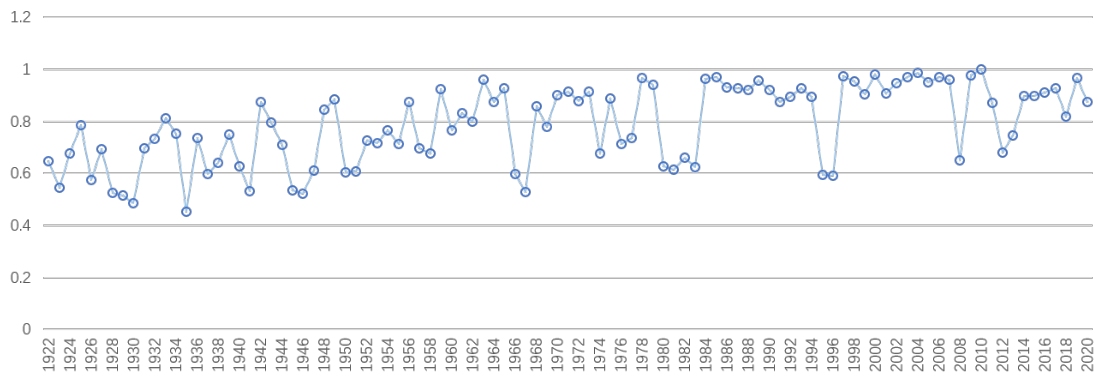


Figure 6: The similarity of music changes by years

From the timeline we can conclude that most of the time the evolution of music is quite smooth, with the similarity between two years above 0.8. However, there are some dramatic decreases that have been marked red near the year 1929, 1938, 1945, 1949, 1966, 1974, 1980. Also, there are some small dips in the 21st century. These years which have a lower similarity with the year before can be seen as a period of music revolution, resulting in major leaps in music characteristics and significant reduction of similarity. We find that there are less revolution periods in recent years. Considering the diversity of music genres nowadays and the comparably peaceful world situation, we think this result is quite reasonable.

References