

The Development Trend of Musicians' Influence and Music Genres of Big Data

Zhang Xi chun^{1,a}, He Li^{1,b}, Wang Jing yi^{1,c}, Liu Wei^{1,d*}

¹SETTING: Shan Dong Normal University, Ji Nan, China

Abstract. This paper uses the data crawled from the AllMusic website to establish a directional network of followers and influences of music genre artists, analyzes the music influence influenced by genre. The Beatles had the greatest influence from 1950 to 2010, and promoted the development of Pop/Rock and Country music genres. In addition, it was found that “influencers” would actually influence the music created by followers. Based on the music feature data set of 91719 songs provided by Spotify’s API, drawing the correlation heat map and making the measurement of music similarity, it is found that the songs of artists of the same genre are more similar. For the similarity between different genres, by selecting the representative music in the genre and using the music characteristics to analyze their correlation, it is found that Folk and Avant-Garde, New Age and Stage & Screen all have high similarity, reaching 0.97. In addition, songs can also be classified into genres according to music characteristics. For example, if a genre has high performance in livability, speech and explicit attributes, it can be considered as Comedy/Spoken. Finally, combined with the historical reality, it is found that there may be characteristics and music revolutionaries^[1] that mark the great revolution of music development.

1 Introduction

Since ancient times, music has become a part of human society and an important part of cultural heritage. Based on the relevant data provided by the Integrated Collective Music (ICM) Association, this study develops a model to measure the influence of music and explores the evolution and revolutionary trends of artists and genres.

2 Establishing a targeted network of followers and influences of music genre artists.

First, we calculate the number of people affected by the influencers. “Impact” includes direct impact and indirect impact. Indirect influence means that assuming that a, b and c are three music artists, a affects b, b affects c, and a affects c indirectly. Since indirect influence may cause A to affect B, and B affects A in turn, it cannot be counted at this time, so we only consider direct influence and get the following table of the number of people affected (due to the large number of artists, we only show the top 10 here):

Table 1: Number of Influencers

Influencer	Number of people affecting	Number of people affected
The Beatles	615	31

Bob Dylan	389	29
The Rolling Stones	319	39
David Bowie	238	25
Led Zeppelin	221	24
Jimi Hendrix	201	32
The Kinks	192	8
The Beach Boy	186	13
Hank Williams	184	3
The Velvet Underground	181	8

As can be seen from the Table 1. The Beatles has the largest number of people, up to 615. However, it is inaccurate to determine the music influence only according to the number of people affected. Therefore, we choose the top six music artists to further observe the years and genres of their influence and draw sub-network diagrams and music influence comparison diagrams accordingly.

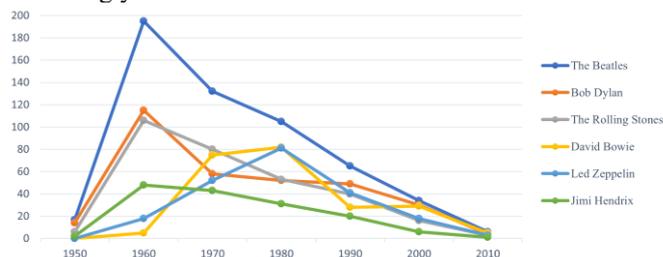


Figure 1: Comparison of the Influence of Six Music Artists
It can be seen from Figure 1 that The Beatles had the

^ae-mail: 1359540160@qq.com ^be-mail: 931727403@qq.com ^ce-mail: 2240867642@qq.com
^d*Correspondence: Liu Wei: e-mail: 191880166@qq.com

greatest influence from 1950 to 2010, but many other music artists such as David Bowie and Led Zeppelin also had outstanding influence and made many contributions to the development of music from 1970 to 1990. It can be seen from the sub-network diagram of Figure 2 that these six music artists have great influence on Pop/Rock and Country and promoted the development of these two genres.

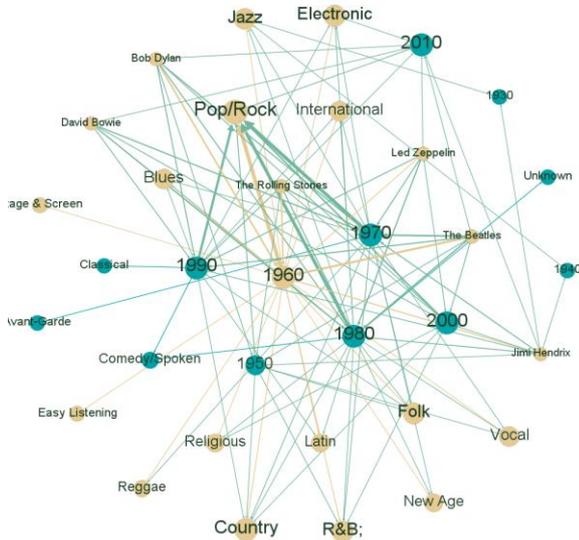


Figure 2: A subnetwork diagram of six music artists

We use the genre data of followers of each *influence_data*, take influencers and followers as nodes, and make network diagrams for different genres in different colors. Only the Beatles and Miles Davis network diagrams are shown here.

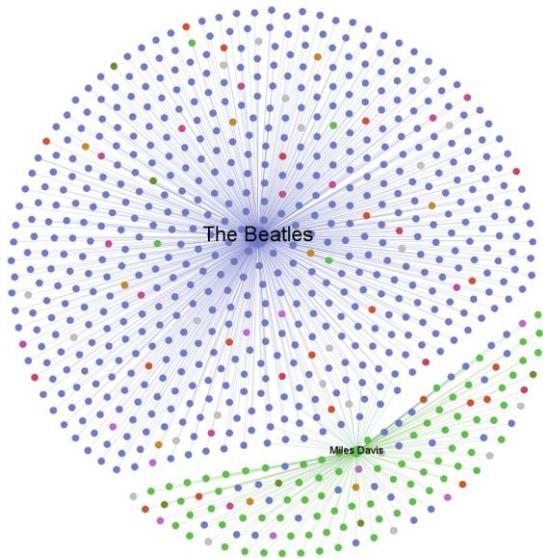


Figure 3: The Beatles and Miles Davis Followers Genre Map

As can be seen from Figure 3, most follower nodes have the same color as influencers. Therefore, we think that “influencers” will actually influence the music created by followers.

3 Measurement of musical similarity

3.1 Data preprocessing

In order to clarify the genre of songs, we connect *full_music_data* with *influence_data* data sets through *artist_id* index. In the process, it is found that the authors of some songs in *textfull_music_data* do not appear in the *influence_data* data set, which leads to the inability to judge the faction of these songs. Therefore, we removed this part of the songs and used the remaining 91,719 songs for research.

3.2 Decrease the Dimension of Music Index

For songs, because there are many music indicators and different indicators are often correlated, we choose principal component analysis to reduce the dimension of the indicators. The steps are as follows:

3.2.1 In order to eliminate the influence of different variable dimensions, it is necessary to standardize the variables first.

There are 15 indicators involved in this part, and there are 91,719 songs. F_{ij} is the i -th index of the j -th song, and each standardized value is standardized as follows to F'_{ij} :

$$F'_{ij} = \frac{F_{ij} - \bar{F}_i}{s_i} \quad (1)$$

Where \bar{F}_i and s_i are the mean and standard deviation of i -th index respectively. The purpose of standard deviation is to eliminate the influence of different variable dimensions, and the standardized transformation will not change the correlation coefficient of variables.

3.2.2 The correlation coefficient matrix of standardized data is calculated, and the eigenvalues and eigenvectors of the correlation coefficient matrix are obtained.

The correlation coefficient between the i -th index and the i' index is $r_{ii'}$, and its calculation method is:

$$r_{ii'} = \frac{\sum_{k=1}^{15} F'_{ik} F'_{i'k}}{15-1} \quad (2)$$

Then the correlation coefficient matrix is $R = (r_{ii'})_{15 \times 15}$, where $r_{ii} = 1$, $r_{ii'} = r_{i'i}$.

3.2.3 Calculate eigenvalues and eigenvectors.

Calculate the eigenvalue $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{15} \geq 0$ of the correlation coefficient matrix R and its corresponding eigenvector $\gamma_1, \gamma_2, \dots, \gamma_{15}$. Where $\gamma_1 = (\mu_{11}, \mu_{21}, \dots, \mu_{151})^T$, the eigenvectors form fifteen new metric variables:

$$\begin{cases} Y_1 = \mu_{1,1}F'_1 + \mu_{2,1}F'_2 + \dots + \mu_{15,1}F'_{15} \\ Y_2 = \mu_{1,2}F'_1 + \mu_{2,2}F'_2 + \dots + \mu_{15,2}F'_{15} \\ \dots \\ Y_{15} = \mu_{1,15}F'_1 + \mu_{2,15}F'_2 + \dots + \mu_{15,15}F'_{15} \end{cases} \quad (3)$$

3.2.4 Determine P principal components and carry out statistical analysis.

According to the above steps, we use SPSS statistical software to first obtain the correlation coefficient table of each index. From the table, we can find that some indexes have strong correlation. Principal component analysis can transform multiple indicators into a few unrelated comprehensive indicators. The comprehensive indicators obtained are as follows:

Table 2: Table of Proportion of Each Indicator to Comprehensive Indicator

	Composition(10 ³)				
	1	2	3	4	5
Danceability	960	-388	302	302	006
Energy	228	001	182	-138	150
Valence	100	-427	054	-005	177
Tempo	074	000	217	-346	161
Loudness	221	-022	071	-125	029
Mode	-012	085	002	-406	503
Key	012	005	027	277	440
Acousticness	-220	-074	-008	112	440
Instrumentalness	-112	168	000	012	258
Liveness	016	079	542	-030	-044
Speechiness	036	-010	447	390	-213
Explicit	074	138	096	410	-330
Duration_ms	004	293	004	005	301
Popularity	187	184	-278	076	-034
Year	201	219	-214	079	-030

As can be seen from the Table 2, a total of five principal components are obtained. The following five principal components are used to study music similarity.

3.3 The Pearson correlation coefficient is calculated to study the musical similarity.

According to the five principal component data obtained in the previous step, the Pearson correlation coefficient R formula for every two songs is as follows:

$$r = \frac{\sum(x_{ij} - \bar{x}_j)(x_{i,j+1} - \bar{x}_{j+1})}{\sqrt{\sum(x_{ij} - \bar{x}_j)^2 \sum(x_{i,j+1} - \bar{x}_{j+1})^2}} \quad (4)$$

Where x_{ij} is the i -th index data of the j -th song, \bar{x}_i is the average value of all index data of the j -th song, $i = 1, 2, \dots, 5, j = 1, 2, \dots, 91719$.

Because the problem requires the similarity between genres, but the number of genres is large, the whole analysis is too long, so we choose Pop/Rock, R&B two genres as an example. Each genre randomly selects any song of five music artists, standardizes their Pearson correlation coefficient, and makes the heat map as follows.

As can be seen from the Figure 4 heat map, the colors of the five squares in the upper left corner and the five squares in the lower right corner are generally darker and have high similarity. The upper left corner and the lower right corner are the songs of music artists of Pop/Rock and R&B genre respectively. According to this, artists of the same genre are more similar than artists of different genres.



Figure 4: Song Similarity Heat Map

4 Similarity and Influence between Different Genres

4.1 Studying the similarity between different genres

Due to our lack of index data on the genre itself. However, for songs, we can often identify the genre of songs according to their characteristics^[2]. Therefore, this topic uses the index data of 91,719 songs so that it represents the genre to study the similarity. The specific steps are as follows:

4.1.1 The average value of the song index is processed, and the Pearson correlation coefficient is obtained.

We first standardized the index data of these songs. Then, for each genre, the average value of index data of all songs is obtained, and the average value of index data is taken as index data of each genre. Then calculate the Pearson correlation coefficient of every two genres, and the formula is the same as Equation (4).

4.1.2 Standardize the Pearson correlation coefficient and make the thermal diagram as follows:

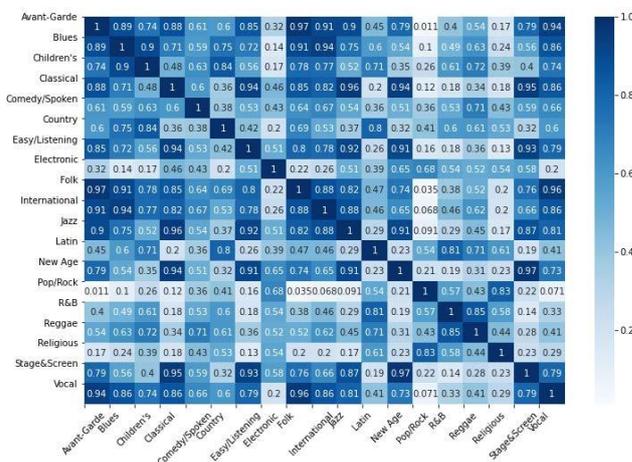


Figure 5: Heat Map of Similarity between Genres
The similarity of different genres can be seen from

Figure 5 thermal diagram. For example, Folk and Avant-Garde, New Age and Stage & Screen all have high similarity, reaching 0.97.

5 Symbols of Music Revolution and Music Revolutionists

In order to find a major turning point in the development of music, we draw a line chart of year and music index. The specific steps are as follows:

5.1 Data preprocessing

Since we need to draw a line chart of the year and the index, and the years of different songs in *full music data* will be repeated, we average the index data of the same year and use the average data to represent the index data of the year.

5.2 Dimension reduction of music indicators

Since there are multiple music indicators, in order to obtain line charts, we use principal component analysis^[3] to reduce music indicators to one dimension.

The dimension reduction step is the same as above, except that the number of principal components generated needs to be additionally set to one, which will not be repeated here.

5.3 Draw a line chart

The line chart of the year and music index drawn is as follows:

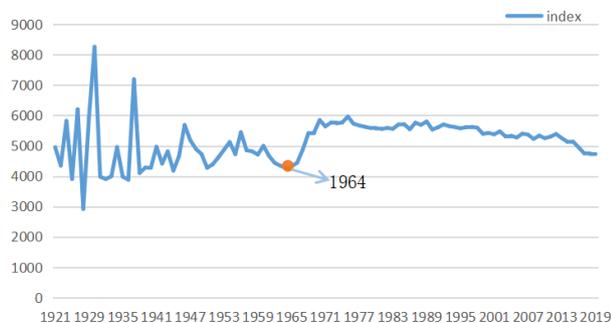


Figure 6: Music Index Change with Year

According to the line chart and historical background, it is analyzed: from 1921 to 1935, because it was the early stage of music development and the music market was unstable, the line fluctuated greatly^[4]. From 1936 to 1950s, after the end of World War II, people life tends to be stable, the demand for music increases, and many new genres appear, so the broken line still fluctuates greatly. As the curve went down in 1964 and then gradually rose and stabilized in the future, it can be seen that around 1964 was a major turning point. It should be that many influential music artists appeared to promote the rapid development of music and stabilize in the future.

According to the line chart above, we can look for influential music artists who appeared around 1964. They promoted the development of music and belonged to revolutionaries. Therefore, we get artists with more

followers around 1964, such as The Beatles, Bob Dylan and The Rolling Stones.

6 Conclusion

Through the establishment of network relationship diagram and data visualization analysis, we can get the influence of music in time or environment: one influencer will influence followers of multiple times at the same time. One genre will affect the emergence and development of many genres. Genres will change with the background of the times when they develop. Using data combined with historical background can integrate the rise and fall history of different genres and revolutionaries appearing at major turning points.

Acknowledgment

Undergraduate Innovation and Entrepreneurship Training Program of Shandong Normal University

References

1. Wang min. (2005) American music history. Shanghai Music Publishing House, Shanghai.
2. Hu Wei. (2012) On the styles, genres and characteristics of jazz. Jin Tian, 000 (008):71.
3. Li Yanshuang, Zeng Zhenxiang. (1999) Application of principal component analysis in multi index comprehensive evaluation method [J]. Journal of Hebei University of technology, (1): 94-97.
4. Zhang Kuo, Kang Ning. (2014) "Soul music" -- a glimpse of the history and current situation of African American music [J]. Literary contending, 000 (005): 184-187.