A Quantitative Model of Music Influence Based on Similarity

Yuan QI[a], Xiatian JIANG[b], Xubin SU[c], Jie JIANG[a,1]

[a] *School of Material Science and Engineering, Xiangtan University, Xiangtan, China*
[b] *School of Mechanical Engineering, Xiangtan University, Xiangtan, China*
[c] *School of Automation and Electronic Information, Xiangtan University, Xiangtan, China*

Abstract. Music, as a kind of cultural heritage, has become an indispensable element in human life. How to intuitively understand the role of music in human collective has become a difficult issue in the field of music research. In this paper, a quantitative model of music influence, which is based on directed graph, Euclidean distance and correlation analysis, is proposed for the development of some genres. In this model, the influences of genre leaders on music development are obtained by evaluating the parameters of the directed graph among artists. Then, the similarity analysis of music within and between schools is con-ducted to get the musical characteristics of different schools. Then the correlation analysis of music features is carried out by SPSS to get the most "infectious" index of music. Finally, based on the time series, the visual analysis of music evolution is made to quantify the influence of music. In this paper, we've verified the model for many times accord-ing to the music development results, which proves that the model has strong guidance for the analysis of music development.

Keywords. Directed graph, Euclidean distance, Correlation analysis, Music influence.

1. Introduction

Music is one of the essential elements to represent culture.

Music, which doesn't develop in isolation, relies on other cultural fields to play their functions. And it always plays an important role in the inheritance, creation, enrichment and development of culture[1].

Nowadays, most of the research direction of papers is the influence of music on other things or other things on music, while ignoring the development and evolution of music itself. We hope to develop a way to quantify the evolution of music by measuring relationships between artists, influences within and between genres, and similarities between song characteristics, to explore the impact of classical music on new music and musical artists better and gain a deeper understanding of how music changes with society .Therefore, our team search a large number of artists and songs on Spotify and Applemusic, and establish a quantitative model of music influence based on directed graph, Euclidian distance and correlation analysis for the research of music development.

---

[1] Corresponding Author, Jie JIANG, School of Material Science and Engineering, Xiangtan University, Xiangtan, China; E-mail: jiangjie@xtu.edu.cn

2. Parameter Evaluation Based on Directed Graph

First, we explore the relationship between musical artists.

For many artists, some of them are the influencers of other artists, and they have effectual influence on the creation of other music artists. Meanwhile, these artists are impacted by the music of other artists. In order to explore the relationship between influencers and followers in music, this paper establishes a vector graph model, and takes the weight of leaders in the music field as the key factor to quantify the relationship between them.

According to the relationship between influencers and followers, we define our desired directed graph as G. G is a two-tuples, consisting of a non-empty finite set V and an ordered pair set A of influencers and followers in V:

$$G=(V,A) \tag{1}$$

Where: $V=\{v_1,v_2,\ldots,v_n\}$ is the set of vertices or nodes corresponding to the influencers and followers; $A=\{a_1,a_2,\ldots,a_m\}$ is an arc set of relationships between influencers and followers. If arc $a_k=v_iv_j$, we regard v_i as the tail of a_k and treat v_j as the head of a_k. In this paper, the influencer is the head of each arc and the follower is the end of each arc.

The Gephi software is used to draw directed graphs with the data of influencers as the source and the data of followers as the target. It is important to note that in the directed graph, the size of the point is the ratio of the number of people influenced by the influencer to the total number of musicians. The larger the point, the stronger the m usical influence of its influencer.
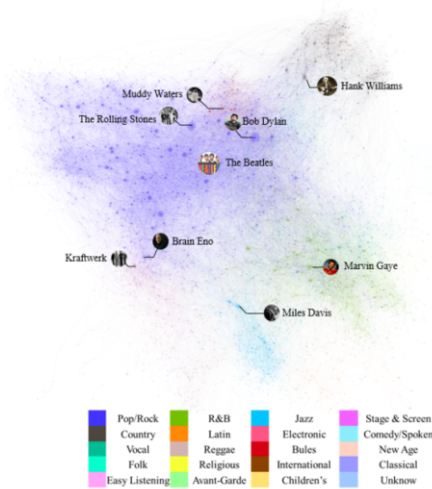

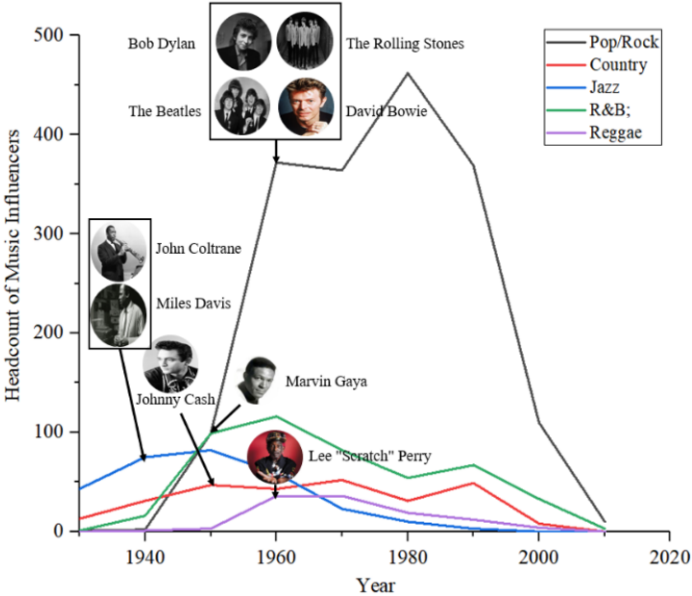**Figure 1.** Directed graph of several genres

Figure 2. Annual number of influencers change of genres

Through the figure 1, we find that the development of a musical genre is inseparable from its music leader. The weight of music leaders is also a matter of great concern, as well as an important index for assessing the development of a genre.

We make an evaluation analysis of the leaders of each genre using the following four indicators: number of artists influenced, total number of songs produced by the artist, number of artists in the genre in the year, and competitiveness in the year (total number of artists in all genres in the year).

The entropy weight method is availed to calculate the weights of corresponding indicators, and the TOPSIS model is used to calculate the scores of the top several leaders. The high-score leaders are shown in the table 1:

Table 1. Scores of Leaders of Genres

	Number of Artists Influenced	Counts of Songs by the Artist	Number of Artists in the Genre of the Year	Competitiveness of the Year	Scores
The Beatles	615	413	372	771	0.1198
Bob Dylan	389	553	372	771	0.1112
The Rolling Stones	319	524	372	771	0.1030
David Bowie	238	231	372	771	0.0852
Johnny Cash	112	588	47	482	0.0511
Miles Davis	160	456	75	230	0.0486
Marvin Gaya	169	292	99	482	0.0442
John Coltrane	118	262	75	230	0.0344
Lee 'Scratch' Perry	48	6	36	771	0.0199
Weigh	0.2202	0.2608	0.4408	0.0782	

At the same time, the number change chart of multiple genres in years is utilized to verify the in-fluence degree of weight. To a certain extent, the more music leaders, the bigger the genre. And the faster the number of music leaders grows, the faster the genre grows.

From figure 2, it is founded that the number of Pop/Rock influencers has increased significantly since 1940. Among them, the most rapid development was in 1950-1960. The Beatles, Bob Dylan, The Rolling Stones and David Bowie were among the top four artists in the Pop/Rock genre in 1960.In order to exclude contingency, we also selected several schools with more samples for verification. In order to exclude contingency, we also select a few genres, whose sample is sufficient, to verify.

First, the leader of the R&B genre was Marvin Gaya, who was active since 1950. At the same time, R&B grew at a relatively rapid pace during the period of 1940-1950.

In the Jazz genre, two most famous leaders, Miles Davis and John Coltrane, came out in 1940.At the same time, Jazz had the fastest growth during the period from 1930 to 1940.

In the Reggae genre, the development of the genre during the period of 1950-1960 was quite astonishing. Besides, it's calculated and analysed that Lee "Scratch" Perry, who has the largest weight in the Reggae genre, was active in 1960.

In the Country genre, the leader, Johnny Cash, was active in 1950.And in 1940-1950, the Country genre grew rapidly.

It can be seen from the above that the music leaders with large weight have played a large role in the development of music genres.

3. Euclidean Distance Model Based on Entropy Weight Method

At the same time, the similarity of artists within the genre and the similarity of artists between the genre is also a very noteworthy problem[2].

Through tabular data analysis, the corresponding index values of each genre are obtained.

Firstly, the index values corresponding to each genre are standardized to make them have the same dimension. Then, the entropy weight method is used to give the corresponding weight and vectorize to each index. Finally, the similarity between two groups within the school is calculated, and the results are obtained by analysing and comparing them. As for the similarity between genres, our team average the eigenvalues of each genre, and then solve the similarity by comparing the average value of the index values between two genres[3].
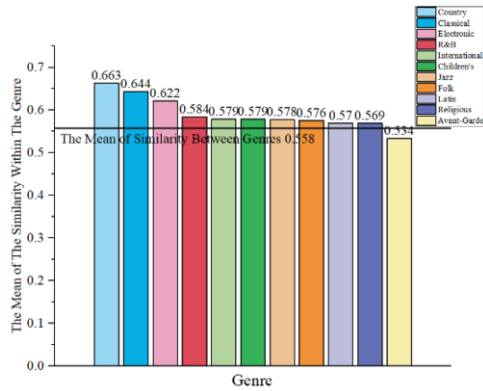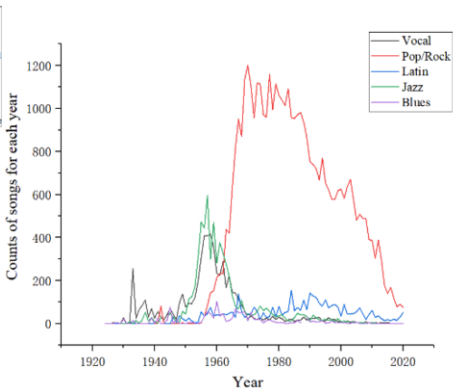
Figure 3. Similarity of artists within genres

Figure 4. The change of counts of songs in different genres

Here is the formula for the model:

$$d= \left[\sum_{n}^{i=1} w_i (x_i - y_i)^2\right]^{1/2} \tag{2}$$

$$S_{im}(x,y) = \frac{|d-d_{min}|}{|d-d_{min}|+|d-d_{max}|} \tag{3}$$

We calculate the similarity within 12 genres and the similarity between pairings of all schools. We take the average value of similarity between two genres, compare it with the internal similarity of the genres and visualize it, as shown in the figure 3:

It is concluded that the similarity of artists within a genre is mostly higher than that between genres and artists in the Pop/Rock genre are the most similar.

In order to explore the source of differences between schools, we compare the absolute values of the eigenvalue differences between genres. Results are shown in the table 2:

Table 2. The difference in eigenvalues between genres

Factor	danceability	energy	valence	tempo	mode
Difference	0.4781	0.5125	0.5306	0.5010	0.5129
Factor	acousticness	speechiness	liveness	count	key
Difference	0.5271	0.1007	0.1723	0.2625	0.6308
Factor	instrumentalness	popularity	duration_ms	loudness	
Difference	0.3188	0.4223	0.4946	0.6380	

It's found that the two largest parameters among all the characteristic values are loudness and key, whose values are 0.6380 and 0.6308 respectively. It shows that loudness and key are important factors influencing the difference between genres, which can be considered as the most essential difference between genres. Then, the difference values of valence, acousticness, energy and mode are 0.5306, 0.5271, 0.5125 and 0.5129, respectively, which can also help us distinguish genres.

4. The Analytical Model of Genre Development Based on Time Series

In order to understand how genres change along with time, we exploit a time series analysis of annual counts of songs for different genres to discover the change of genres.

From the figure 4, it is found that the counts of songs of Pop & Rock genre increased significantly in 1965, while Jazz genre and VOCAL genre increased slightly in 1955. We statistically analyse the popularity of songs of the corresponding genres in this period and find that the popularity value of Pop & Rock songs is 28.298, which is significantly increased compared with the previous 20 years and still on the rise. In 1955, the popularity of Jazz was 10.867 and the popularity of VOCAL was 14.182, which lasted for a period of time continued to decrease. In general, the Pop & Rock is the most competitive in the evolution of music. Through the figure 4, we analyse the relationship be-tween genres and come to the conclusion: From 1955, with the rise of Pop & Rock, Jazz and Vocal began to decline, which indicates that the former is in competition with the latter two, and the former has the greater competitiveness. The other genres compete with each other and always maintain the same level.

5. Comparison Model Based on Spearman Correlation Coefficient

Similarly, the influence of influencers on the musical style of their followers is also the part of our concern.

First of all, we count and analyse musical characteristics between influencers and their corresponding followers[4]. We respectively extract the musical features of the influencer and the follower, then use the Euclidean distance algorithm based on entropy weight method to calculate the similarity and arrange the results of each group in ascending order. It's obvious that the similarity values between influencers and followers are generally higher than the similarity values for either of any two in the entire sample. Due to the inability to display a large amount of data, the paper utilizes Monte Carlo algorithm to figure out influential artists and only shows these representative groups of mean values. The statistical table of the results and the comparison figure of the similarity with the overall sample are as follows.
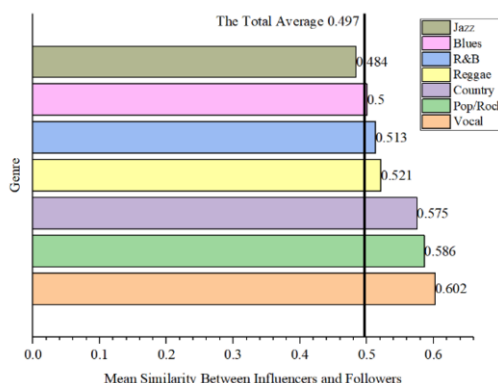

Figure 5. Similarity between influencers and followers

From the figure 5, it's clear that the similarity of musical characteristics between influencers and followers is higher than 0.5 except for Jazz, while the mean similarity between the total samples only 0.497, which strongly explains that influencers influence the works of their followers. Besides, it's concluded that the music of followers is really similar to the music of influencers.

We utilize Comparison Model based on spearman correlation coefficient to investigate the intensity of contagiousness of music features[5]. At the beginning, we calculate the correlation coefficient between each feature and the prevalence in the sample:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \tag{4}$$

Where, d_I is the rank difference between two indexes

Then, test the spearman correlation coefficient. Since spearman correlation coefficient is a large sample data, it fits:

$$r_s(n-1)^{1/2} \sim N(0,1) \tag{5}$$

The table 3-4 shows our statistics of the correlation coefficient of the eigenvalues in all the samples and the correlation coefficient of the eigenvalues of a certain artist.

Table 3. Correlation coefficient of eigenvalues of all samples

			Correlation			
			danceability	energy	valence	tempo
Spearman Rho	popularity	correlation coefficient	0.168**	.380**	0.005	0.093**
		significance	0.000	0.000	0.145	0.000
			loudness	mode	key	acousticness
Spearman Rho	popularity	correlation coefficient	0.446**	-0.037**	.015**	-0.457**
		significance	0.000	0.000	0.000	0.000
			instrumentalness	liveness	speechiness	
Spearman Rho	popularity	correlation coefficient	-0.181**	-0.071**	0.004	
		significance	0.000	0.000	0.175	
		**. P<0.01				

Table 4. Correlation coefficient between eigenvalues of both other artists and The Beatles

			Correlation			
			danceability	energy	valence	tempo
Spearman Rho	popularity	correlation coefficient	-0.023	0.085	0.171**	0.048
		significance	0.648	0.086	0.000	0.331
			loudness	mode	key	acousticness
Spearman Rho	popularity	correlation coefficient	0.179**	-0.044	0.065	-0.119*
		significance	0.000	0.370	0.187	0.015
			instrumentalness	liveness	speechiness	
Spearman Rho	popularity	correlation coefficient	-0.143**	-0.024	-0.309**	
		significance	0.004	0.633	0.000	
		**. P<0.01*. P<0.05				

A subsection From table 3 above, we employ the test values, which conform to the significant correlation coefficient, to obtain the results:

Firstly, no significant difference between the correlation coefficient of both Valence and Mode and 0 in musical features indicates that, relative to other features, Valence and Mode are not 'contagious', which reflects that music's Valence and Mode have no effect on the 'contagiousness' of mu-sic.

Then, compared the correlation coefficients, it's concluded that the correlation coefficients of acousticness, energy and loudness are larger, which indicates that energy and loudness are more "contagious" than other characteristics (except acousticness).

Next, from the analysis of table 4 (The particular artist we study is The Beatles), it is found that the correlation coefficient of valence is significantly different from 0. However, for The Beatles, the previous correlation coefficient, which is significantly different from zero, is not different from zero at this time (except loudness). So we analyse the correlations among several artists to discover that only the characteristic value of loudness has general correlation, which means that only the loudness of music features is strongly contagious.

We start with a cartogram of the evolution of music's popularity to seek the characteristics that were driving the revolution in music. Because every fluctuation in the data can indicate a major change in music during this period, we do not remove outliers from the data. Graph of the evolution of music popularity over time is as follow.
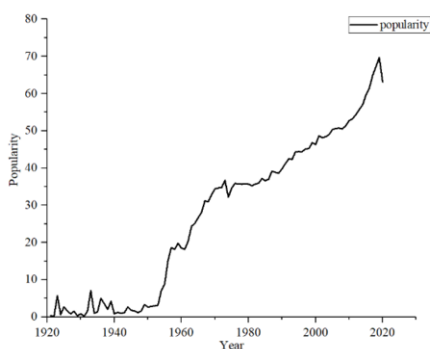

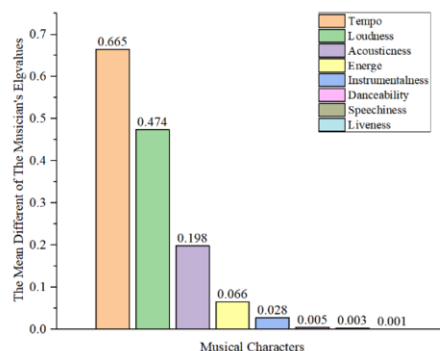**Figure 6.** The changing popularity of music   **Figure 7.** The difference between the artist's eigenvalues

From the figure 6, it is founded that before 1953, the popularity of music basically maintained at the same level, and after 1953, the popularity of music began to increase rapidly. We take 1953 as the time node, comparing the difference of artists who debuted in 1950 and 1960, to make an in-depth analysis. In this paper, our team exploit the mean value of all samples is exploited at a specific time point to reflect the eigenvalue at this time point, and compare the mean value at different time points to analyze the specific characteristics that affect the change of music. These results of our processing are as follows.

We analyse the differences in eigenvalues of artists and find that the following characteristics have biggish variation at this time point in 1953. We conduct visual processing on the difference between all the features as the figure 7. (The data has been standardized)

Our conclusion is that the variation of eigenvalues of 'acousticness', 'tempo' and 'loudness' marks the evolution of music. And the change of artist's eigenvalue 'tempo' is in possession of high probability to symbolize the evolution of music. To explore artist revolutionaries, we count the vari-ation of tempo in the musical characteristics of artists during this time period. It's found that artist Anita O'Day has a large change in the eigenvalue of Tempo, and it's believed that he has the greatest impact on the evolution of music.

Then, we continue to investigate the influencers with greater "musical influence" in our directed graph to reduce the one-sidedness of our results. To seek for more about the revolutionaries who have a significant impact on the music evolution, we find out all the artists who debuted after 1950 and had a high "musical influence" parameter (weight). Comparing the average characteristic values of all artists in 1950 and 1960, we find that The Beatles, Marvin Gaye and Muddy Waters have significant differences in some characteristics, such as 'acousticness', 'tempo' and 'loudness'. From the above results, it is concluded that these artists are of essential significance to the musical revolution and our judgment standard is correct.

At the same time, the evolution of music is something we should be concerned about. We choose Pop & Rock to investigate the influence process of its musical evolution. At the beginning, we calculate the standard deviation of whole music features in this genre to analyse the evolution process of this genre comprehensively. It is found that these indicators, 'acousticness', 'instrumentalness', 'speechness', fluctuate obviously. Then, we draw a time series diagram of them.
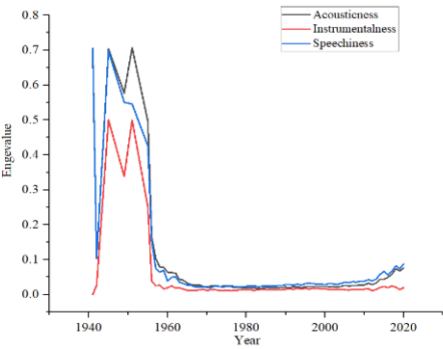

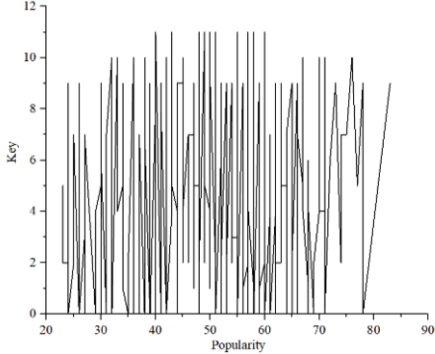**Figure 8.** Variation of musical characteristics

Figure 9. The variation of the Beatles' key

In this paper, we determine that the index of dynamic influencers is the popularity of artists.

In terms of influence, the popularity of an artist directly reflects "musical influence" of the artist.

We selected The Beatles, the most popular artist in Pop & Rock, to analyse the evolution process of his song characteristics, as the figure 9. Comparing with the change process of his popularity, we investigate the specific characteristics that influence his popularity.

It's found that among all musical characteristics, there is a corresponding relationship between key and the popularity. The key value decreases and the popularity of the artist increases, which signifies that the majority of the public probably had an affection on low-pitched music at this period. We analyse popularity changes of many artists, and find that people tend to prefer different kinds of music at different times. The popularity of most artists is determined by key, instrumentalness and valence, so we exploit changes in the three characteristics to judge the popularity variation of artists.

References

[1] Reinhard Strohm. Studies on a Global History of Music[M].Taylor and Francis:2018-03-22.

[2] Jie Zhou. Research on Practice Reform and Implementation of European Music History Course in Colleges and Universities Based on Music Feature Extraction Technology[J]. Journal of Physics: Conference Series,2020,1550(3):032040.

[3] Huirao Li. Improvement and application of K-means clustering method [D]. Northeast Agricultural University,2014.

[4] A comparative statistical analysis of music styles (seventeenth–nineteenth centuries) [J]. Interdisciplinary Science Reviews,2020,45(4): 581-594.

[5] Wenyao Zhang. Measuring the Degree Correlation of Network with Spearman Coefficient [D]. University of Science and Technology of China,2016.