Music Never Gets Old: Study On The Development And Evolution Of Music

Summary

Music does not evolve isolately, but is always influenced by other artists, other music, or by society and events. To quantify the musical influence, we developed four models based on building networks.

First, we established **Influence Evaluation Model Based on PageRank Algorithm** for problem 1. The influence relationship between follower and influencer is constructed into an artist influence directed network, where the influencers are the father nodes and the followers are the child nodes. We set sub network of artists by genre. Then, we improved the **PageRank algorithm** to calculate the absolute influence of each node in overall and sub network. Through this model, we sought for the nodes with high-influence, and got the top 20 influential artists as shown in the table 2.

Second, we established **Similarity Model Based on Distance of Vector** for problem 2. We developed the music characteristic vector of the artist, which is positively correlated with vector distance. By calculating the distance and drawing heatmap, we got the similarity between artists at each node. The results show that most of the artists had a high degree of similarity with artists of the same genre, but a few artists had a high degree of similarity with artists of other genres.

Third, we established **Hybrid Model of Influence and Similarity** for question 3 and question 4. We generalized the indexes of influence and similarity to describe genres. Based on this, we distinguished different genres and got the influence relation. We analyzed the changing ways of genres through line charts, completing question 3. Then, we established **Inheriting Degree Model of Music Characteristics**, combining the connection of network nodes with similarity to develop the inheriting degree index. According to the value of index, we completed question 4 and found that some musical characteristics are more "contagious" than others; the same musical characteristic has different influence in different genres.

Fourth, we established **Analysis Model of Music Evolution Process** for questions 5, 6 and 7. We selected musical characteristics with moderate inheriting degree, analyzing the major changes over time in music. Take the 1960's for example, we found that *the Beatles* were the revolutionaries of music, completing question 5. We averaged the whole genre of music, obtained genre characteristics and its change over time. We set music popularity as an indicator of dynamic influencers, completing question 6. We calculated the overall influence of music genres, observed its change over time, analyzing the influence of social environment on music, completing question 7.

Finally, we conducted a sensitivity analysis for our model, and the result of test goes well. We have optimized the model to meet larger amounts of data. We analyzed the strengths and weakness of the model. In the end, we introduced our model to ICM Society.

Keywords: Network, PageRank, vector distance, inheriting degree, influence index.

Contents

1 Introduction

1.1 Restatement of Problem

As a kind of art, music is an indispensable cultural element in the development of human civilization. **Integrative Collective Music (ICM) Society** invites our team to study the evolution and revolution process of music artists and music genres. Since the development of new music is bound to be affected by various factors, the study of the influence of music is the breakthrough point for our study. Therefore, we need to build models to measure musical influence.

Musical influence can be measured in three ways. First, artists are influenced by other artists. Second, new music are influenced by previously produced music, which can be seen from the similarities between previous and new music. Third, music changes over time because of social or technological events.

The model we are to build should solve the following problems:

1. Establish directed network and its subnetwork between music influencers and followers; find parameters that can describe the influence of influencers.

2. Establish music similarity measure; compare similarity between and within the genres.

3. Compare the "music similarity and artists influence" between and within genres; distinguish genres; describe change tendency and the relationship of the genre.

4. Describe efforts "influence" makes on "similarity"; find music characteristics that are more "contagious".

5. Find music characteristics and artists that represent revolution of music.

6. Further analyze music genres' and artists' change over time; find a dynamic influencers measure.

7. Describe how music is influenced by time and environment.

1.2 Our Work

After making reasonable assumptions to simplify the model, we built four models to solve the above seven problems, as shown in the flow chart, figure 1 below.

- First, we established the **Influence Evaluation Model Based on PageRank Algorithm**, to complete the question 1.

- Second, we established the **Similarity Model Based on Distance of Vector**, to solve the question 2.

- Third, we set up the **Hybrid Model of Influence and Similarity**, to analyze the question 3 and 4.

- Fourth, we set up the **Analysis Model of Music Evolution Process**, to make the analysis of question 5, 6 and 7.

In addition to the models, we have also done optimization and sensitivity analysis of our models. Finally, we wrote a one-page document for the ICM Society.
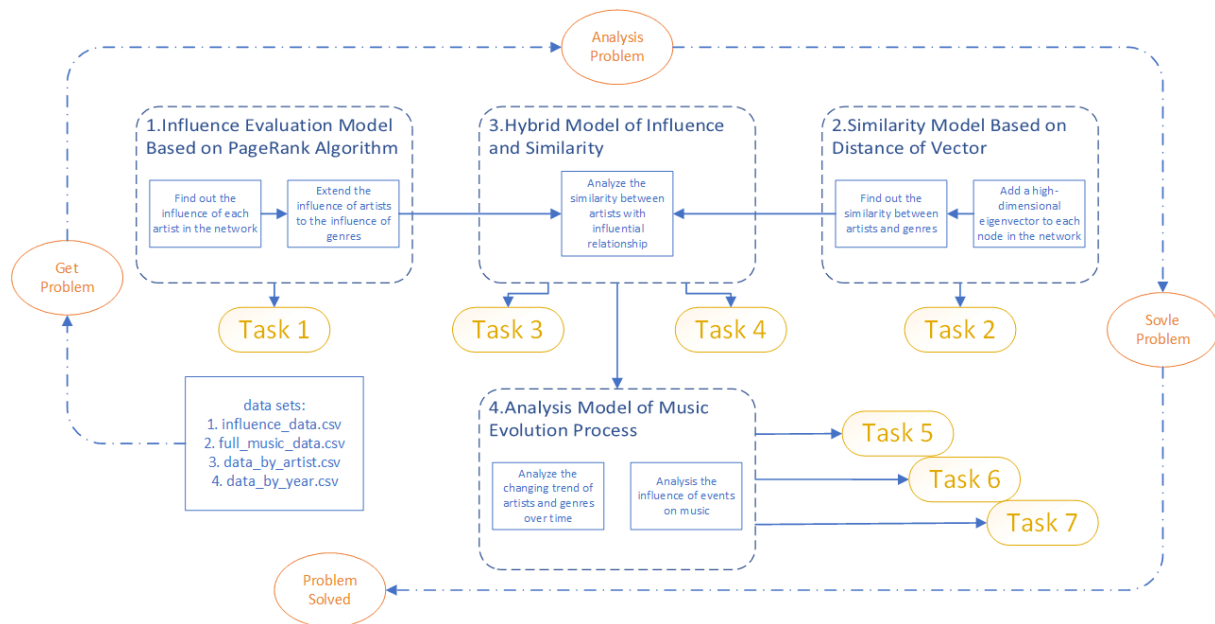

Figure 1: The flow chart of our models

2   Assumptions and Justifications

1. We assume that **the influence from influencer to every follower is the same.** Because it is not mentioned in the data set, that the influence weight of the influencers for one follower.

2. We assume that **the impact of a major change of music could be reflected in a year.**

3. We assume that **the inheritance of music characteristic is pure probability inheritance.**

4. We assume that **the most popular music of the year has a greater impact on the genre than any other music.** Because a hit song is often a sign of genre development.

3   Notations

The notation we used is shown in the table 1.

4   Influence Evaluation Model Based on PageRank Algorithm

This model is built for question 1. According to the requirement, we need to create a directed network graph connecting followers and influencers, with influencers as the

Table 1: Notations

Symbol	Definition
V_i	Nodes in the network (artists)
$L(V_i)$	The in-degree of node V_i
N	Number of nodes in the network
$F(V_i)$	The node set of the outgoing chain of V_i
PR	Influence coefficient of nodes in network
p	Attenuation coefficient in PageRank algorithm
C_i	Feature vector of music
x_{ik}	The k-th feature of music i
d	Distance between eigenvectors
sim	Similarity between music
a	The number of the follower with the same genre as V_i
b	The number of the follower with different genres from V_i
PR_{in}	The influence of the artists within the genre
PR_{outk}	The influence of the artists on the genre k
m_k	The number of artists and music in genre k
FI_k	The influence of the genre k on the inner
FO_{k1k2}	The influence of the genre $k1$ on the $k2$
SI_k	The similarity of the inner music in genre k
SO_{k1k2}	The similarity between genre $k1$ and genre $k2$
IH	The inheritance degree of the feature of the music
E_k	The average of the feature vector of music in genre
P_k	The most popular of the feature vector of music in genre
P_g	The sum of the influence of the genre musicians

father nodes and followers as the child nodes. We also set parameters that represent musical influence. In social networks, PageRank algorithm can be used to get the information of "which pages are more important" through the network connection of web pages. Inspired by PageRank[1], we calculate the influence of nodes and the relative influence among nodes.

In this model, the value of PR is used as a parameter to measure influence. It should be noted that in the original PageRank algorithm, the edge of the network graph points to the parent node. In our model, however, the edge points to the child node.

4.1 Establishment of Model

According to the dataset *influence-data*, we built a directed network graph with artists as the point and the influence relationship between artists as the edge. The direction of each edge is from influencer to follower. The network is shown in the figure 2 below. The larger the area of the node, the more outgoing links of the node.

In the figure 2, different colors represent different music genres. The corresponding relationship between color and music genre is shown in the following pie chart as figure 3.

The network can be divided into some sub networks. Artists of the same music genre form a sub network. For example, the blue part of the graph is the sub f Pop
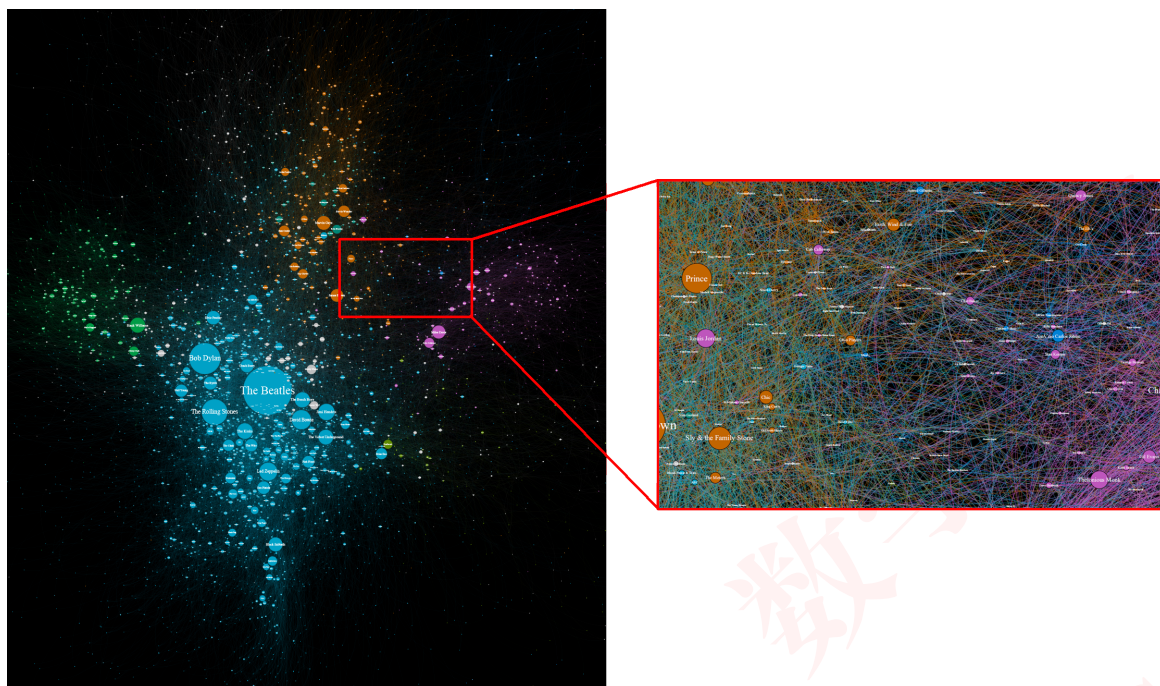
Figure 2: The influence network of artists

/ rock music artists, while the orange part of the graph is the subnet of R&B artists.

The advanced PageRank algorithm[2] is applied to our network. The usage and meaning of this algorithm are as follows:

- The goal of this algorithm is to find out the "importance" of each node in the network, that is, "influence".

- The father node has influence on the child node, and the child node reflects the influence of the parent node. The more child nodes a father node points to , the more the father node influences, and the greater the influence of the father node. In the network, the more the number of outgoing links of a node, the greater the influence of the node.

- Every child node is called "next generation" relative to its parent node. With the increase of generation, the influence of the child node on the parent node is less and less obvious. Therefore, we introduce an attenuation coefficient of p.

The idea of PageRank algorithm is iterative operation. Before the calculation, each node is given the same PR value, and each iteration updates, the node will evenly allocate the new PR value to its outgoing nodes. The iteration continues until the allocated value of PR is no longer updated.

In our model, in order to adapt to the problem we need to solve, we change the PageRank algorithm to: in the iterative update, all nodes will allocate the new PR value to their own in chain nodes. That is, the propagation order of the value of PR is from the child node to the father node.

In the iterative update phase, the specific process of allocating PR value from the child node to the father node is as follows. Figure 4 shows the process as well.

Assuming a network with only three nodes: A, B, C. If both B and C have edges
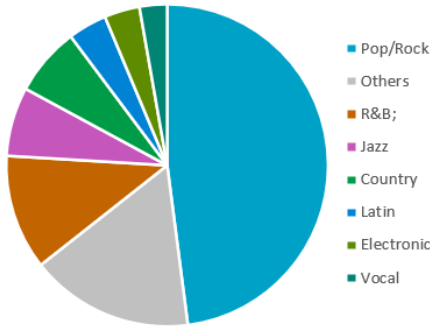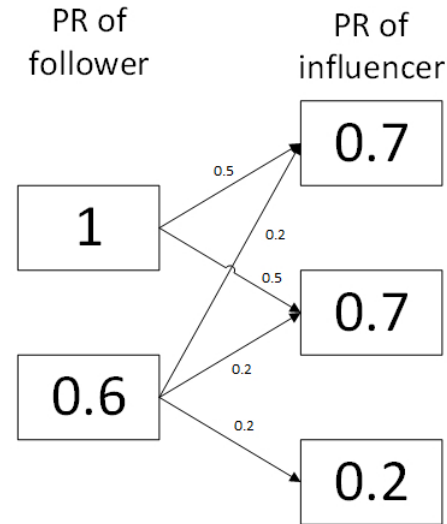
Figure 3: The corresponding relationship be- Figure 4: The process of allocating PR value
tween color and music genre

pointing to A, then the PR value of A will be the sum of B, C.

$$PR(A) = PR(B) + PR(C) \tag{1}$$

Continue assuming that B also has an edge pointing to C. The number of incoming links of B nodes is 2.

$$PR(A) = \frac{PR(B)}{2} + PR(C) \tag{2}$$

It is generalized to the general case,

$$PR(A) = \frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} \tag{3}$$

where $L(B)$ is the number of inbound links of node B.

As can be seen from the above figure 4, the child node will divide its own PR value equally, and then assign it to the father node. The PR value after allocation is the edge weight between two nodes. In this way, it forms a "music influence network". When the iteration stops, the edge weight between the two points means the influence of the parent node on the child node. At the same time, the calculation method of the PR value of any node V_i in the network is as follows:

$$PR(V_i) = \frac{1-p}{N} + p \sum_{V_j \in F(V_j)} \frac{PR(V_j)}{L(V_j)} \tag{4}$$

Where p represents the attenuation coefficient to reduce the influence of generations, generally 0.85. N represents the total number of nodes in the network. $F(V_i)$ represents the node set of V_i's outgoing chain, V_j is any point in this set. Obviously, the greater the PR value, the greater the influence of the artist.

4.2 Solution of Problem 1

The figure 2 shows *influence-data* set, an overall network of all artists. The blue part is a subnet of Pop/Rock artists. The figure 5 below shows the top 10 artists with

the value of PR in the overall network, in which the columns with different colors represent different genres. The following table 2 shows the top 20 artists with PR value.
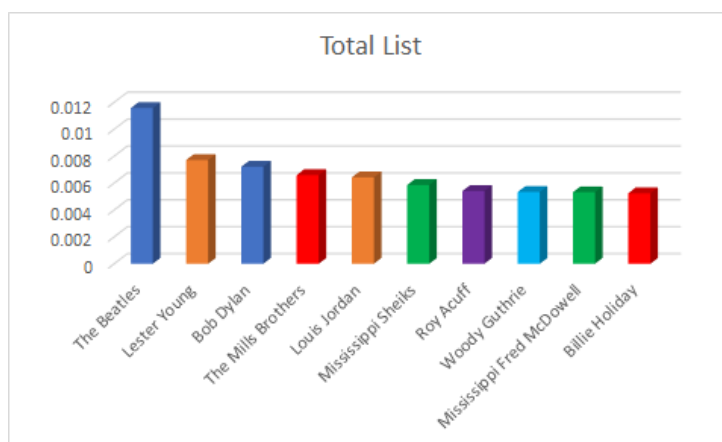

Figure 5: The top 10 artists in the overall network

We find that all the characters in the picture are legends of their genres. As an influence parameter, the value of PR is consistent with the reality and achieves our expected effect.

In order to prove the stability of our model, we apply our algorithm to Pop/Rock sub network and get the following results.
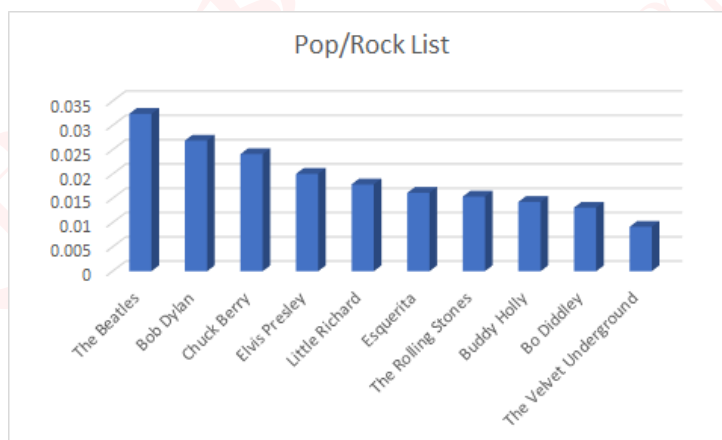

Figure 6: The top 10 artists in the Pop/Rock sub network

This is the top 10 most influential Pop/Rock artists. We find that the order of Pop/Rock artists in the general list is the same as that in the table 2, so we can prove that our model is stable. *The Beatles* is no doudt the most influencial band in the world.

Conclusion: Whether in the subnetwork or the overall network, influential artists will stand out with PR.

5   Similarity Model Based on Distance of Vector

This model is built for question 2. We use the characteristics of music to measure the similarity of music. The characteristics of music, as already given in the dataset, are

artist ID	artist name	artist genre	PR of artist
754032	The Beatles	Pop/Rock	0.011573239
259529	Lester Young	Jazz	0.007705564
66915	Bob Dylan	Pop/Rock	0.007210347
403120	The Mills Brothers	Vocal	0.006589893
287604	Louis Jordan	Jazz	0.006422992
898336	Mississippi Sheiks	Blues	0.005855095
848784	Roy Acuff	Country	0.00540526
577531	Woody Guthrie	Folk	0.005334219
898331	Mississippi Fred McDowell	Blues	0.005315981
79016	Billie Holiday	Vocal	0.00523804
894465	The Rolling Stones	Pop/Rock	0.004968142
317093	Nat King Cole	Jazz	0.004940834
549797	Hank Williams	Country	0.004919624
120521	Chuck Berry	Pop/Rock	0.004888935
608701	Muddy Waters	Blues	0.00452981
354105	Jimi Hendrix	Pop/Rock	0.004285491
128099	James Brown	R&B;	0.004214594
200775	Ernest Tubb	Country	0.004151536
180228	Elvis Presley	Pop/Rock	0.004072049
824022	Little Richard	Pop/Rock	0.003939197

Table 2: The top 20 artists in overall network

quantifiable. These characteristics include the following: danceability, energy, valence, tempo, loudness, and mode. We also consider the normalized human voice characteristics. After learning music theory, we think that key has little influence on music style, so we don't consider key in music characteristics.

The smaller the absolute value of the difference between the values of a certain characteristic of two music, the more similar the characteristics of two music. If we consider these characteristics as dimensions, then a music is a vector including these dimensions. The closer the two music vectors are, the higher the similarity between the two music vectors is.

For question 2, we need to use this model to calculate the music similarity between artists, and analyze whether the similarity between two musicians in different genres is higher than that between two artists in different genres. According to data set *data-by-artist*, we can get the average level of each artist's musical characteristics. In other words, we assume that each artist has only one piece of music, and we compare the similarity between two pieces of music, that is, we compare the similarity between two artists.

5.1 Establishment of Model

In order to make the data of music vector can be compared, we need to normalize the data. We normalized tempo and loudness in music characteristics, and other data have been normalized in data set.

Next, there are two ways to calculate the distance between two vectors. ^ssume

that the eigenvector of artist i is $C_i(x_{i1}, x_{i2}, ..., x_{im})$, and the eigenvector of artist j is $C_j(x_{j1}, x_{j2}, ..., x_{jm})$.

1. We use Minkowski Distance[3] to calculate the distance d between two vectors and transform it into similarity.

$$d(C_i, C_j)_p = \sqrt[p]{(x_{i1} - x_{j1})^p + (x_{i2} - x_{j2})^p + \cdots + (x_{im} - x_{jm})^p} \tag{5}$$

Let $p = 2$, that is, we choose Euclidean distance as the calculation of vector distance. We convert the vector distance d into similarity sim. The smaller the d is, the greater the similarity is, as shown below.

$$sim(C_i, C_j) = e^{-d(C_i, C_j)_2} \tag{6}$$

2. We use cosine distance[4] to calculate the similarity of two vectors, as shown below.

$$\cos \theta = \frac{\sum_{k=1}^{m} (x_{ik} \times x_{jk})}{\sqrt{\sum_{k=1}^{m} x_{ik}^2} \times \sqrt{\sum_{k=1}^{m} x_{jk}^2}} = \frac{C_i}{\|C_i\|} \cdot \frac{C_j}{\|C_j\|} \tag{7}$$

where θ is the angle between C_i and C_j. the smaller θ is, the greater $cos\theta$ is, the greater the similarity is. Therefore, the similarity sim is as follows.

$$sim(C_i, C_j) = \cos \theta \tag{8}$$

5.2 Solution of Problem 2

We've sorted more than 5000 artists by doing the following. First, all the genres are arranged in dictionary order, and then all the artists in a genre are arranged in their ID.

After sorting, we use the model to calculate the Minkowski distance similarities and cosine distance similarities of any two artists. The heatmap figure 7 shows the Minkowski distance similarity of all artists we get, and the heatmap figure 8 is the cosine distance similarity. Among them, as shown in the color bar, the color at the top of the color bar represents high similarity, while the color at the bottom of the color bar represents low similarity.
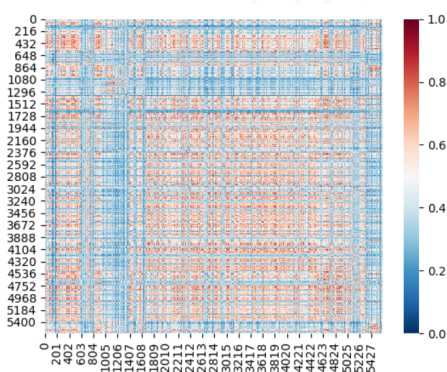

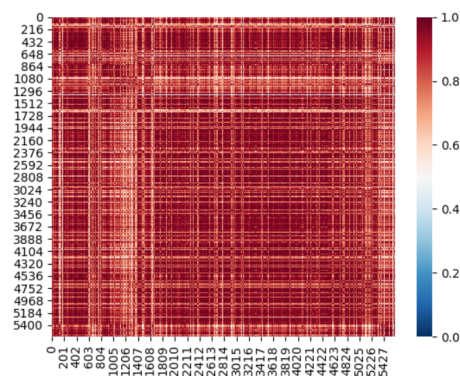
Figure 7: The heatmap of Minkowski distance similarity   Figure 8: The heatmap of cosine distance similarity

Conclusions:

- Although the colors of the two heatmaps are different, the shapes of the color blocks are basically the same. The two images are symmetrical about the main diagonal, and the effect is basically the same.

- On both sides of the main diagonal are red (the color with higher similarity) square areas. In the sorting process, we arrange the artists of the same music genre together, which means that the similarity of artists within the genre is higher.

- There are also light color areas on the main diagonal, which means that the similarity of very few artists in some genres is not high enough.

- In other areas, we can also see the similarities and differences between different artists through different color shades.

Similar to question 1, we compare the similarity of the top ten influencers of the overall network and Pop/Rock subnet (using Minkowski distance), as shown in the following figure 9 and figure 10.
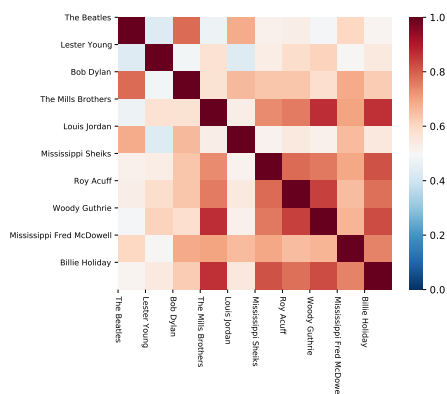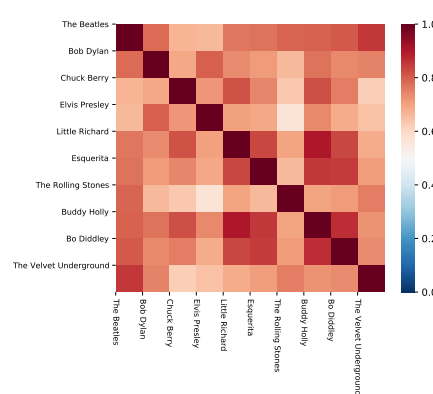

Figure 9: Similarity of overall network   Figure 10: Similarity of Pop/Rock network

Conclusions:

- In the overall network list, the similarity of artists of different genres is generally low.

- In the sub network list, the similarity of artists in one genre is high.

  We took *the Beatles*, *Bob Dylan* and *Lester Young* as examples, drew two radar maps of their musical characteristics, as shown in figure 11 and figure 12.

  Obviously, *the Beatles* and *Bob Dylan* belong to the same genre, their musical characteristics are indeed very similar; *The Beatles* and *Lester Young* belong to different genres, their musical characteristics are very different. This confirms our conclusions.

- Most of the artists have high similarity with the artists of the same genre, while very few artists have high similarity with the artists of other genres.

  We think this is reasonable. Music and art blend with each other. If an artist always studies one genre, he must have a higher similarity with the artists of that genre. But for some other artists, if they are involved in a wide range of genres or like to create experimental music, their works will integrate the characteristics of different genres of music.
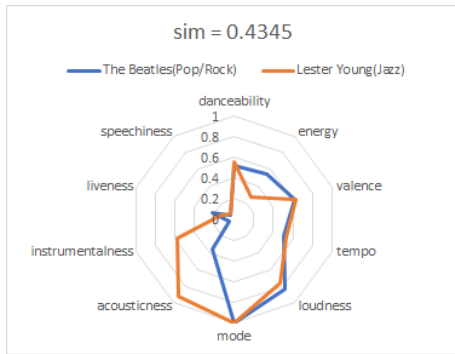
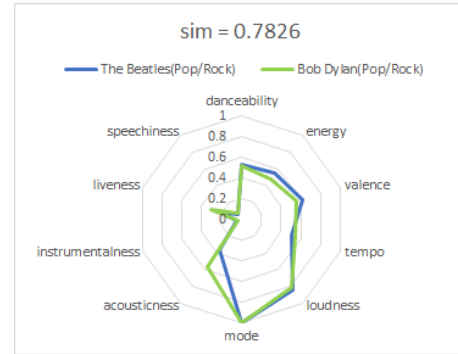Figure 11: Comparison of music characteristics of *the Beatles and Lester Young*

Figure 12: Comparison of music characteristics of *the Beatles and Bob Dylan*

6 Hybrid Model of Influence and Similarity

This model is built for questions 3 and 4. Some people think that an artist is influenced by other artists as well as other music. Ours "influence evaluation model" is used to measure personal influence of an artist. But after generalization, this model can be used to measure the music influence an artist receives. Ours "Similarity model" is used to measure the similarity of music. But after generalization, this model can be used to measure the similarity between music genres. In question 3, we will compare A with B, C with D. Among them, A is the similarity among genres, B is the similarity within genres, C is the influence among genres and D is the influence within genres. In question 4, one of our tasks is to prove whether "influence" works on "similarity". Therefore, this model needs to combine and compare the above two models.

For question 3, we generalize the "influence index" and "similarity index" from single artist to music genres, and find the way to distinguish genres. For question 4, we start from similarity, analyze the "infectivity" of the characteristics of music.

6.1 Generalization of The Two Indicators

6.1.1 Generalization of The Influence Indicator

In the previous model, we considered the absolute influence of artists themselves and their relative influence on their followers. In this model, we generalize the "influence index" to the relative influence of artists on genres.

In question 1, we got the PR value of all artists: $PR(V_i)$. The number of people influenced by artist V_i is $L(V_i)$. We assumed that among the artists influenced by V_i, there are a artists of the same genre as artist V_i, and b artists of different genres. therefore

$$L(V_i) = a + b \tag{9}$$

$$b = b_1 + b_2 + \cdots + b_k \tag{10}$$

The formula (10) indicates that b artists may come from k different genres.

So we get the influence of the artist V_i on the inner genre as follows:

$$PR_{in}(V_i) = \frac{1 - p}{N} + p\alpha_i \times PR(V_i) \tag{11}$$

The influence of artist V_i on genre k is:

$$PR_{outk}(V_i) = p\beta_{ik} \times PR(V_i) \tag{12}$$

where

$$\alpha_i = \frac{a}{L(V_i)} \tag{13}$$

$$\beta_{ik} = \frac{b_k}{L(V_i)} \tag{14}$$

Average the formula (11), the inner influence of genre k is

$$FI_k = \frac{\sum_{i=1}^{m_k} \dfrac{1-p}{N} + p\alpha_i \times PR(V_i)}{\sum_{i=1}^{m_k} PR(V_i)} \tag{15}$$

where m_k is the number of artists in the genre k. Average the formula (12), the influence of genre $k1$ on genre $k2$ is

$$FO_{k1k2} = \frac{\sum_{i=1}^{m_{k1}} p\beta_{ik2} \times PR(V_{k1i})}{\sum_{i=1}^{m_{k2}} PR(V_{k1i})} \tag{16}$$

6.1.2   Generalization of The similarity Indicator

It is known that the similarity between two pieces of music is $sim(V_i, V_j)$. The same as in the previous section, taking the average similarity within the genre, then the inner music similarity of genre k is

$$SI_k = \frac{\sum_{i=1}^{m_k} \sum_{j=1}^{m_k} sim(V_i, V_j)}{m_k^2} \tag{17}$$

The music similarity of genre $k1$ and genre $k2$ is

$$SO_{k1k2} = \frac{\sum_{i=1}^{m_{k1}} \sum_{j=1}^{m_{k2}} sim(V_{k1i}, V_{k2j})}{m_{k1} m_{k2}} \tag{18}$$

6.2   Inheriting Degree Model of Music Characteristics

When two artists have an influential relationship and their music has a certain similarity, we call one inheriting music characteristic from another one. Therefore, we mixed the two indexes and established a model to measure the inheriting degree.

We think that follower is influenced by many influencers, and it is impossible for him or her to completely inherit the musical characteristics of an artist. At the same time, artists have their own creativity, and their musical characteristics are not completely inherited.

Because the characteristics of different genres are different, the same kind of music characteristics may have different inheritance ability in different genres. Music characteristics inheriting degree can be divided into cross genre and within genre. Cross genre music characteristics inheriting degree is affected by music genre similarities,

which is difficult to measure quantitatively. Therefore, we think that the inheriting degree within the genre can better reflect the inheritance of music characteristics.

The artist's music feature vector is $C_i(x_{i1}, x_{i2}, ..., x_{i9})$. Setting Boolean variables G, which means whether a certain characteristic of two artists is similar in value, there is

$$G(k) = \begin{cases} 1, & x_{ik} \approx x_{jk} \\ 0, & x_{ik} \neq x_{jk} \end{cases} \tag{19}$$

We assume that, if $|x_{ik} - x_{jk}| \leq \pm 5\% x_{ik}$, then $x_{ik} \approx x_{jk}$. When $G(k) = 1$, We think the characteristic k has been successfully inherited.

Let $IH(x_k)$ be k^{th} of the 10 music characteristics in a genre with a total population of m, it is reflected in the ratio of "the number of successful inheritees of characteristic x_k" to "the total number of inheritees of characteristic x_k".

$$IH(x_k) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{L(V_i)} G(k)}{\sum_{i=1}^{m} L(V_i)} \tag{20}$$

The larger the index IH is, the easier it is to inherit the feature, and the more infectious it is.

6.3   Solution of Problem 3

According to formula (15) and formula (16), we draw the heatmap of genres influence, as shown in the figure 13 below.
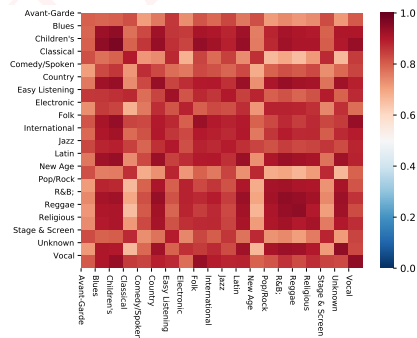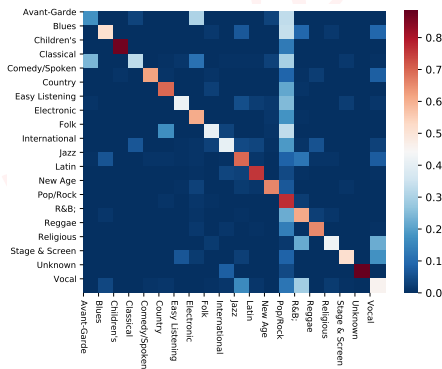

Figure 13: The heatmap of genres influence   Figure 14: The heatmap of genres similarity

In the heatmap, the diagonal blocks represent inner influence of the genres. The column of blocks represent the influence of other genres on diagonal music genre; the row of blocks represent the influence of diagonal music genre on other genres.

The heatmap directly reflects the influence degree of a certain genre on another genres. For example, the reddest block Pop/Rock is influenced by Blues, Country, Folk, Jazz, R&B (column) and so on. At the same time, Pop/Rock has a strong influence on Country and Electronic (row).

The rule of genres changing over time is: most genres have strong inheritance ability, and the larger the genre, the greater the inner influence. As shown in the heatmap, the color on the diagonal represents the degrees of the inner influence of genres, and it can be seen that most of them are clearly visible color blocks. After cons̲i̲d̲e̲r̲i̲n̲g̲ the

data, we found that our model results accord with the actual situation, so our influence model is successful.

According to formula (17) and formula (18), we draw the heatmap of similarity, as shown in the figure 14.

In the heatmap, the image is symmetrical about the diagonal, the diagonal blocks represent the average similarity within one genre, and the other blocks represent the average music similarity between two genres.

On the whole, genre similarity and artist similarity have the same features. The blocks on the diagonal show that the similarity within genres is greater than that between genres, showing the feature of centralization. What's more, Pop/Rock is more similar to R&B, Country, Blues, Reggae and Latin in this similarity heatmap, because they are all dark red, and in the influence heat map above, it shows that Pop/Rock is influenced by them. The two heatmaps show a certain consistency. Therefore, our similarity model is also successful.

Based on the above analysis, we draw the following conclusions.

Conclusions:

- The way of distinguishing different genres: music similarity model is used for clustering, and genre labels are added to the clustering results through the inner genre similarity.

- The way music genres change over time: The change of influence parameters FI and FO over time can reflect the dynamic influence and dynamic stability of a genre. Considering the increasing number of people in a genre, we can analyze the development of a genre. Take Pop/Rock as an example, its change of FI and FO over time is shown as figure 15.
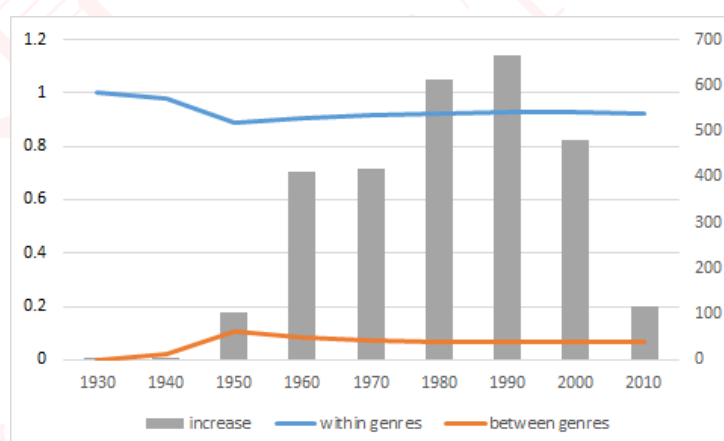

Figure 15: Change of FI and FO over time

Pop/Rock was born in 1930. During the period of 1930-1940, this genre had a great inner influence, indicating that it was accepting the influence of other genres and was in its stage of infancy. In 1950, the population of genre increased for the first time, and its between genres influence increased. This new genre had a great impact on the music circles. Representing Pop/Rock entered stage of development. In 1960, a large number of artists invested in Pop/Rock music creation, Pop/Rock's inner influence increased and maintain stability. Pop/Rock genre entered stage of mature. After 1960, $FI = 0.9$,

$FO = 0.1$, the number of Pop/Rock genre is still increasing, but the influence of the genre is basically shaped, it has stable self-development power, and music influence. By 2010, only 100 artists joined this genre, and there is no significant change of FI, FO. Therefore, from the perspective of population growth, Pop / Rock may be being surpassed by other genres.

6.4   Solution of Problem 4

According to the model, we got the inheriting degree heatmap of 10×20 by calculation, as shown in the figure 16 below.
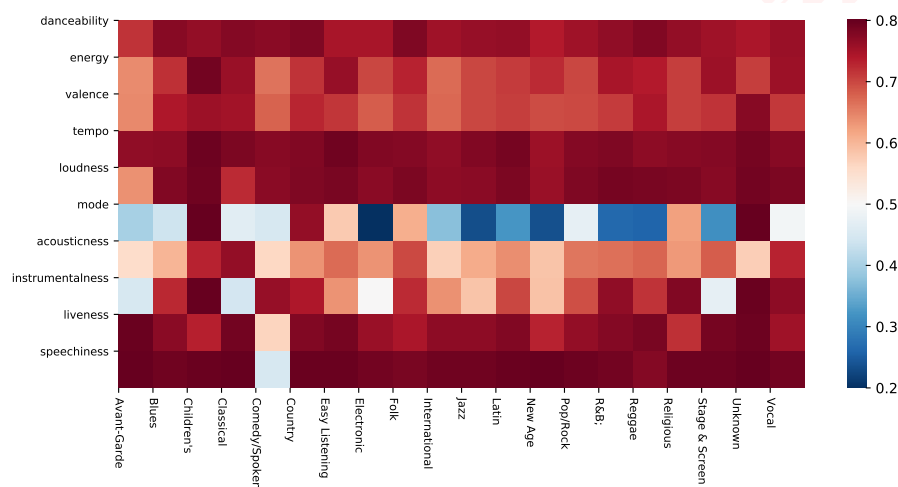

Figure 16: Inheriting degree heatmap

In different genres, different musical characteristics are inherited with different degrees of inheritance. We assumed that when $IH \geq 0.6$, musical characteristics are considered as good inheritance.

Among the 20 genres, at least four characteristics are inherited, and at most nine characteristics are inherited. For example, Avant-Garde inherits only tempo, dance ability, liveness and speech, while Children's inherits all the characteristics. If we further increase the IH, we can further screen representative music characteristics of different genres, and show the differences between genres.

Conclusions:

- The influencer does have influence on the followers, and the influence is reflected by the inheritance of music characteristics. For example, the characteristic speech of jazz genre has a 0.78 probability of being inherited in the process of influence between influencers and followers.

- Some music characteristics are more "contagious" than others. For example, the speech (0.78) of jazz is more inheritable than the valence (0.61).

- The same music characteristic has different influence ability in different genres. For example, the speech (0.78) of jazz genre is more inheritable than speech (0.44) of Comedy/Spoken genre.

- The inheritance abilities of some music characteristics in all genres are generally not weak, such as tempo (0.68-0.83).

- The inheritance abilities of some music characteristics in all genres are generally not strong, such as mode (0.16-0.77).

7   Analysis Model of Music Evolution Process

7.1   Establishment and Solution of Problem 5

We think that the stronger the inheriting degree of a certain musical characteristic is, the less the musical characteristic changes with time. Therefore, we take the mutation of musical characteristics with strong inheriting degree as an indicator, reflecting whether revolutionaries have taken place. However, if the inheriting degree is too strong, this characteristic will be too stable. If a strong inherited music characteristic in the data set *data-by-year* mutates in a certain year, according to the assumption, it shows that as a great revolutionaries, the artist played his role in that year.

We selected five music charataristics with moderate inheriting degree: energy, tempo, acousticness, instrumentalness, and speechiness. We drew a line chart, as shown in the figure 17.
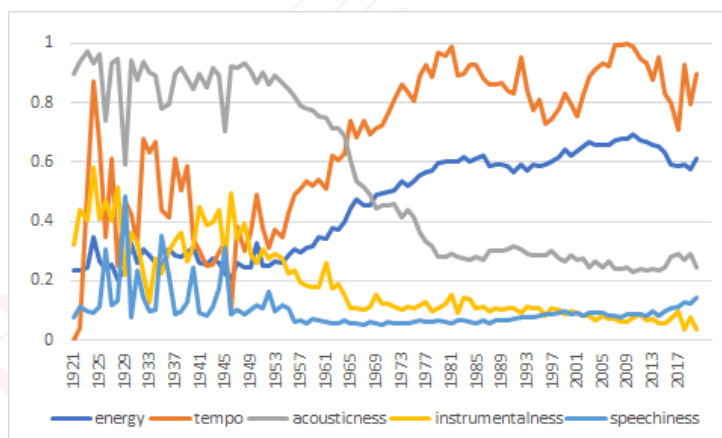

Figure 17: Music charateristics change over time

Conclusion: from 1964 to 1969, acousticness dropped sharply, and energy continued to rise. We found the active musicians and their albums in the corresponding period. We found that **the Beatles** were one of the most active influencers. Their following five songs had a very low level of acousticness and high energy, and were among the most popular in the same era.

The five songs are: **a.** Tomorrow Never Knows - Remastered ****; **b.** Sgt. Pepper's Lonely Hearts **** Band (Reprise) - Remix; **c.** Sgt. Pepper's Lonely Hearts **** Band - Reprise / Remastered ****; **d.**And Your Bird Can Sing - Remastered ****; **e.** Nowhere Man - Remastered ****. The information of the five songs is shown in the table 3.

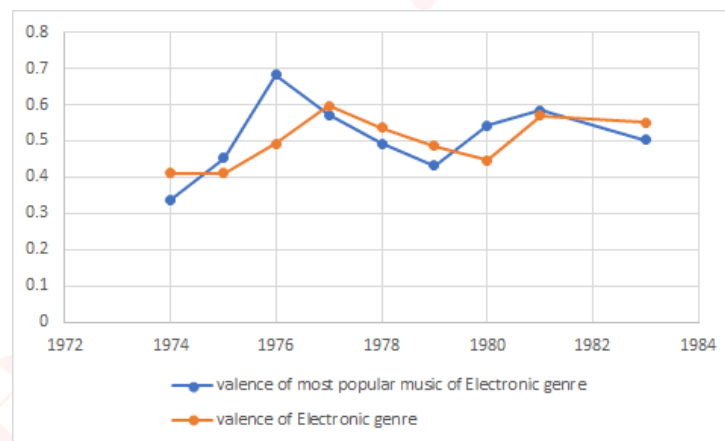## 7.2   Establishment and Solution of Problem 6

In question 6, we set up a model to represent the changes of music characteristics within the genre over time. We start with the dataset *full-music-data* to get the average value of the music characteristics of a certain genre every year, written as a vector is $E_k(ex_{t1}, ex_{t2}, ..., ex_{t10})$.

song	energy	tempo	acoustic	instrumental	speechi	popularity	year
a	0.829	125.887	8.40E-05	0.00208	0.0405	58	1966
b	0.951	118.769	0.000249	0.214	0.0521	40	1967
c	0.889	118.773	0.0061	0.589	0.0453	52	1967
d	0.65	132.59	0.0067	0.0491	0.0282	59	1966
e	0.624	121.402	0.00797	0	0.0462	64	1965

Table 3: The information of the five songs

We draw a time-dependent curve of each dimension e_{ti} to show the evolution trend of music characteristics of one genre changed over time. Then we take the most popular song of the genre every year, drew the same curve of each dimension P_{ki} of the music characteristics of the song. We compared the changes and differences between the two. Here, we take Electronic as an example. We only select one characteristic valence and chose the representative work, *Tangerine Dream*, that is, compared E_{ki} with P_{ki}, as shown in the figure 18 below.

Figure 18: comparison of E_{ki} with P_{ki} over time

As we can see from the figure, we draw the curve of E_{ki}, which reflected the overall evolution trend of the music characteristics of the Electronic genre. By comparing the evolution curve of the overall characteristics of genres, with the characteristic curve of the most popular music of genres, we draw conclusions.

Conclusions:

- The trend of the two curves is similar.

- The trend of E_{ki} is slightly behind that of P_{ki}.

- E_{ki} is more stable than P_{ki}.

Therefore, we infer that the most popular music in the genre influences the whole genre and promotes and guides the evolution of the genre, which is the law of genre change. For a piece of music, we examined two indicators: **popularity** and $\mathbf{P_k}$. Among them, popular determines whether a piece of music can be rated as "contagious", and P_k determines the direction of music's influence in the music circle.

We think that the artists who make these music are "dynamic influence"

7.3 Establishment and Solution of Problem 7

Any historical event is likely to influence on each genre and form a huge influence network. In order to simplify the model, we chose three genres (Blues, Electronic and Pop/Rock) and several historical events to illustrate the model.

We calculated the sum of artist influence of each genre in each era PG_{ikt}.

$$PG_{ikt} = \sum_{i=1}^{m_k} PR(V_{ikt}) \tag{21}$$

We drew a total influence curve $PG(t)$, as shown in the figure 19 below. Each influence curve is normalized into the same graph, and historical events are added to the time line for analysis.
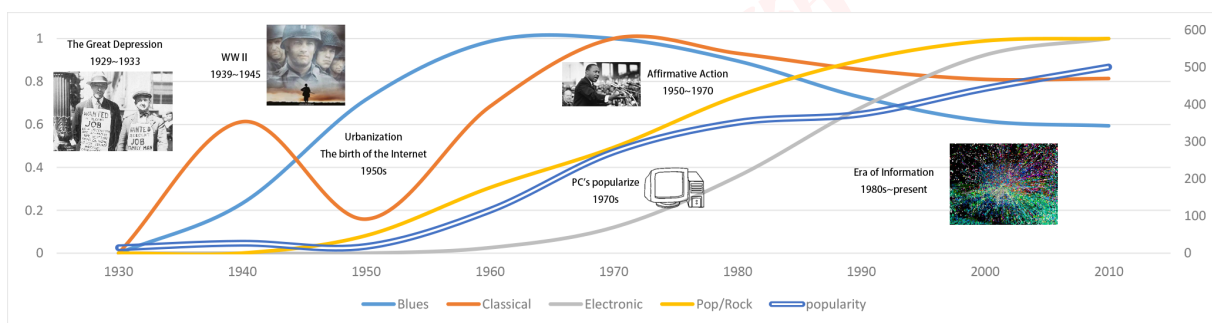

Figure 19: The total influence curve $PG(t)$

Combined with the actual history, the following conclusions can be drawn from the figure 19.

Conclusions:

- After the end of the economic depression, the music curve shows an upward trend, and it is speculated that it will be frustrated during the depression.

- In the Second World War, classical music suffered a serious setback[5]. Because the cost of classical music performance is high after the formation of the performance agent, the war leads to people's retrenchment, and financial institutions invest their funds in the war, and the lack of funds has a huge impact on it. With the economic recovery after the war, the influence of classical music recovered with the improvement of people's economic strength.

- The rise of American civil rights movement from 1950 to 1970 promoted the development of Blues: 1. Blues originated from black slaves in the past[6]. 2. In the early 1950s, the center of blue was in Chicago, USA. 3. The music of the civil rights movement and the freedom of speech movement in the United States led to the study of the early black American music.

- Since 1970, Moore's law has created affordable music technology, and the use of electronic music has become more and more popular. In the 1970s, personal computers began to be popular, corresponding to the acceleration of the development of electronic music.

- In the information age, the high-speed information flow keeps the influence of each school at a high level.

The change of $PG(t)$ curve perfectly reflects the change of music influence, can make timely response to historical events, and conforms to historical common sense. So, our $PG(t)$ curve is successful.

8   Sensitivity Analysis

Influence Evaluation Model and Similarity Model are our basic models, so we conducted sensitivity analysis on these two models.

- We changed the damping coefficient p in the PageRank algorithm to a range of 0.85 ± 0.15, and the error ratio of PR is shown in the figure 20 below. It can be seen that when the damping coefficient p changes, the error ratio is less than 5%, which meets our expectation. The model has passed the sensitivity analysis.

- We changed the dimension of vector with a range of 10 ± 3 dimensions, and the error ratio of the similarity is shown in the figure 21 below. It can be seen that when the dimension changes, the error ratio is less than 2%, which meets our expectation. The model has passed the sensitivity analysis.
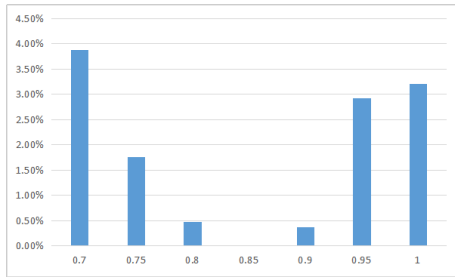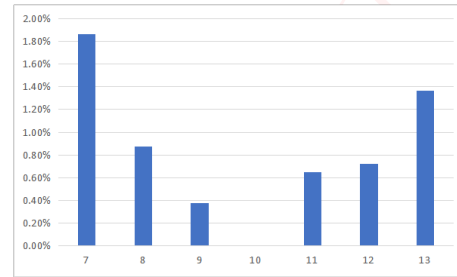

Figure 20: Error ratio of PR

Figure 21: Error ratio of the similarity

9   Optimization of Model

Since the data set given by ICM Society has certain limitation, we consider the optimization of our model when the data set becomes larger. When the generation involved in the data set increases, the time series becomes worse with the time span.

We introduced the time factor Δt and the outcome indicator W_i[7] as the influence coefficient, where W_i is from evaluation system. So we got

$$P_i = aW_i^b \times e^{-c\Delta t} \tag{22}$$

$$PR(V_i) = (1 - p)P_i + p \sum_{V_j \in F(V_j)} \frac{PR(V_j)}{L(V_j)} \tag{23}$$

where a, b and c are the parameters that need to be adjusted in the model. The PageRank algorithm itself is suitable for processing big data. After we add time factor, we can process data with long generation span better than the original algorithm.

10   Strengths and Weaknesses

10.1   Strengths

- Lightweight: Our models are not complicated in construction, and can use the least data to analyze the problem. The results are obvious. The time and space complexity of the algorithm is low.

- Creativity: In order to analyze the changes of evolution progress, we have created several indicators. The analysis results can be verified quantitatively, which also makes our model original.

- Accuracy: According to the results of quantitative analysis, our models are logically self-consistent and practical, making them successful and accurate.

- Stability: Our basic models have been tested by sensitivity analysis, and the error is acceptable, so the model is stable.

- Scalability: We optimize our model to handle much larger amounts of data.

10.2   Weakness

When it comes to problem 5, our model can only identify a few particular super influencers; For question 6, we chose the most popular music, ignoring the influence of other music on the list.

References

[1] Xu Hang, Zeng Wenhua, Zeng Xiangxiang, et al. An Evolutionary Algorithm Based on Minkowski Distance for Many-Objective Optimization.. 2019, 49(11):3968-3979.

[2] Zang Sisi. Research on Evaluation Index of Author Influence Based on Improved PageRank Algorithm [D]. Qufu Normal University,2020.

[3] Xu Hang, Zeng Wenhua, Zeng Xiangxiang, et al. An Evolutionary Algorithm Based on Minkowski Distance for Many-Objective Optimization.. 2019, 49(11):3968-3979.

[4] Grieb Nora, Oltrup Theo, Bende Thomas, et al. The Cosine Similarity Technique: A new method for smart EXCIMER laser control.. 2020, 30(4):253-258.

[5] John B. Strait. Geographical Study of American Blues Culture. 2010, 109(1):30-39.

[6] Lawrence Grossberg. Urban Rhythms: Pop Music and Popular Culture. 1987, 6(1):104-106.

[7] Li Chong, Wang Yuchen, Du Weijing, He Xiaotao, Liu Xuemin, Zhang Shibo, Li Shuren. PageRank Talent Mining Algorithm Based on Web of Science [J/OL]. Computer application: 1-7 [2021-02-08]. http://kns.cnki.net/kcms/detail/51.1307.TP.20201209.1623.020.html.

Document to ICM Society

To: ICM Society
From: Team 2100242
Date: February 8th, 2021
Subject: Network based on PageRank: A better approach to understanding music

Thank you so much for inviting our team to explore the evolution of music!

We propose the following solutions based on the influence network. I will introduce to you the purpose, advantages and conclusions of our modes.

First, what can our model do? As is known to all, music, as a cultural product, influences itself. In this way, artists form a network of influences, and so do music works. However, the network between works is not intuitive. So, our music network is on top of the network of artists. For the artist network, we built the first model, **Influence Evaluation Model Based on PageRank Algorithm**, to calculate the absolute and relative influence of artists. Then, we treat music as a vector, assigning properties to the network nodes. This is the second model, **Similarity Model Based on Distance of Vector**. We use it to calculate the similarity between two nodes. Next, we build the third model,**Hybrid Model of Influence and Similarity**. With the connections between nodes and their own properties. this model can calculate the similarity and influence of the whole network.The inheriting degree between two pieces of music, known as "contagious", can be sovled as well. Finally, the fourth model, **Analysis Model of Music Evolution Process** is built to analyze the variation of eigenvectors of each node.

Second, what are the advantages of our model? Our models are simple to use, highly original, and calculate the influence of music quantitatively. The models passed sensitivity analysis, and we optimized the models to accommodate larger data sets.

Thirdly, based on the data set, our model draws the following conclusions.

- The Beatles are undoubtedly the most influential artists, both absolutely and relatively. At the same time, they are the revolutionaries of our music network.

- In terms of genres, artists of most genres are similar in music characteristics. The more influential the genre is, the higher the internal similarity is, and the more stable the inheritance is.

- Music similarity is indeed influenced by the influencer. Some music characteristics are more "contagious" than others. The same music characteristics has different influence ability in different genres.

- In our network, the way we analyze social influences on music is by comparing changes in musical characteristics over time to historical axes of events.

As John Lennon said, "Music never gets old." Music is a bridge connecting human society, is a manifestation of human joys and sorrows. Music is always evolving, and artists are taking advantage of society and technology to add new vitality to the music. Artists learn from each other, making the development of music full of vitality. In this age of technology, artists, and ordinary people, have never been closer to art. It was, arguably, the best age for music.