

## The Determinants of Music Creation

Yaming Li

Department of Automation, North China Electric Power University (Baoding), Baoding130600, China.

73029283@163.com

### Abstract

The core of this article is to look for the factors that influence the development of music by studying the evolution and change of music. We established the TOPSIS analysis method based on the entropy method, using scores to characterize music influence. We use the score to represent the length of the directed line segment. Nodes are influencers and followers, connected by directed line segments. Draw a directed subnet. Through the analysis of the subnet diagram, it can be obtained: the larger the parameter, the greater the influence, and the influence of the followers will increase, thereby promoting the evolution and revolution of music.

### Keywords

TOPSIS; Directed Subnet; Parameter.

### 1. Introduction

Music plays an important role in the process of human development. Similarly, it will also change due to the combined effects of various factors. From ancient times to modern times, music genres have experienced early Gothic music, baroque music, classical music, blues to country music, jazz, rock, metal music, and electronic music. They have undergone significant changes and development. In addition to the creativity of musicians themselves, some classic concerts created by old artists can influence many generations of young artists. The psychological changes of the audience and changes in the social structure will affect the musicians' creations to a certain extent. The emergence of new musical instruments and more convenient exchanges between different cultures will also evolve genres. Our task is to study and measure the impact of previously produced music on new music and music artists.

### 2. TOPSIS model

#### 2.1 Model assumptions

- (1) Except for the artists given in the data sheet, the influence of other artists on them can be approximately ignored.
- (2) The All music.com and API of Spotify are large-scale music websites with many musicians and songs of various genres. The data given in the title is objective and inclusive.
- (3) We do not consider that new artists will affect old artists. When processing "influence\_data.csv", we found that only 2% of new artists will affect old artists, which is very rare, and this impact can be understood as new musical instruments and technological innovations on old artists Be inspired. Therefore, we classify time as a factor that only has a positive influence.

#### 2.2 Data Processing

For data analysis problems, there are usually some incomplete and abnormal data in a large amount of original data, which may seriously affect the efficiency of modeling and the accuracy of conclusions. Therefore, we first clean up and summarize the data to facilitate subsequent data use. The specific data processing process will be explained as the problem is solved.

### 2.3 Parameter Development

In a directed network, nodes represent artists, the lines connecting artists can represent the influence of the artist or the music he represents, and the number of lines represents the number of followers. Arrows are used to point from influencers to followers. Next, we develop a scoring model from multiple dimensions, comprehensively calculate the value, and use the length of the line to characterize it<sup>[1]</sup>.

### 2.4 Time

First, we process the data. The `influence_data.csv` provides a large amount of artist data, we choose to intercept a part for analysis. Time is an important factor affecting the development of music, so we use time to filter the data. First, filter out the data of `influencer_active_start` as 1930, calculate the difference  $t$  corresponding to each row of `follower_active_start` and `influencer_active_start`, and classify them with 30 years as the metric, that is,  $t=0\sim30$ ,  $t=30\sim60$ ,  $t=60\sim90$ . There are 3 categories (the number of parts with  $t<0$  is very small, and they are classified as 0 values, as explained in the hypothesis). Then, we counted the number of influencers, and deleted the relevant data with the number of influencers less than 100 (because through the drawing we found that the number of influencers less than 100 is too much, resulting in a decrease in image clarity)

we use the MATLAB to draw three-dimensional directed network, from influencers to followers. In the picture, the green node is  $t=0\sim30$ , the yellow node is  $t=30\sim60$ , and the blue node is  $t=60\sim90$ .

The directed network is as follows:

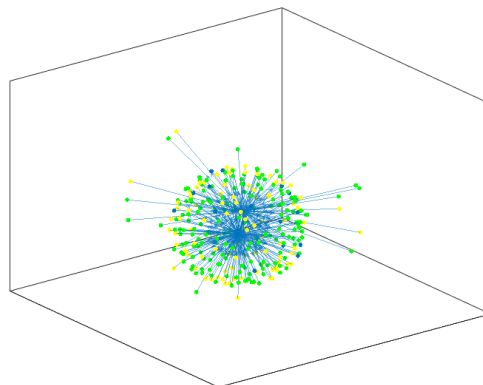


Figure 1. The directed network

The information in the picture shows that there is an obvious relationship between music creation and time. The number of green nodes is the largest, followed by the yellow nodes, and the blue nodes are very few and exist in the case of large number of followers. This shows that as the time span increases, the influence decreases but does not rule out a few very influential artists continue to have an impact on new artists. We will create a "degree" parameter below for comprehensive consideration. Therefore, time can be used as a parameter to measure the influence of music creation.

### 2.5 The "degree"

Next, we will re-process `influence_data.csv`. We will count the number of followers corresponding to each influencer, sort them in descending order, and get a new summary table excel 1. We named this parameter "degree", which directly reflects the influence and status of the influencer. When the "degree" is bigger, it means that the influencer's influence ability is stronger.

### 2.6 Genre

We observe that there is a genre indicator in the `influence_data.csv` and there are 19 genres in total. We divide all music types into three categories: A, B, and C according to indicators that have a greater impact on music genres such as energy, tempo, loudness, and key, and combined with the perception of listeners<sup>[2]</sup>.

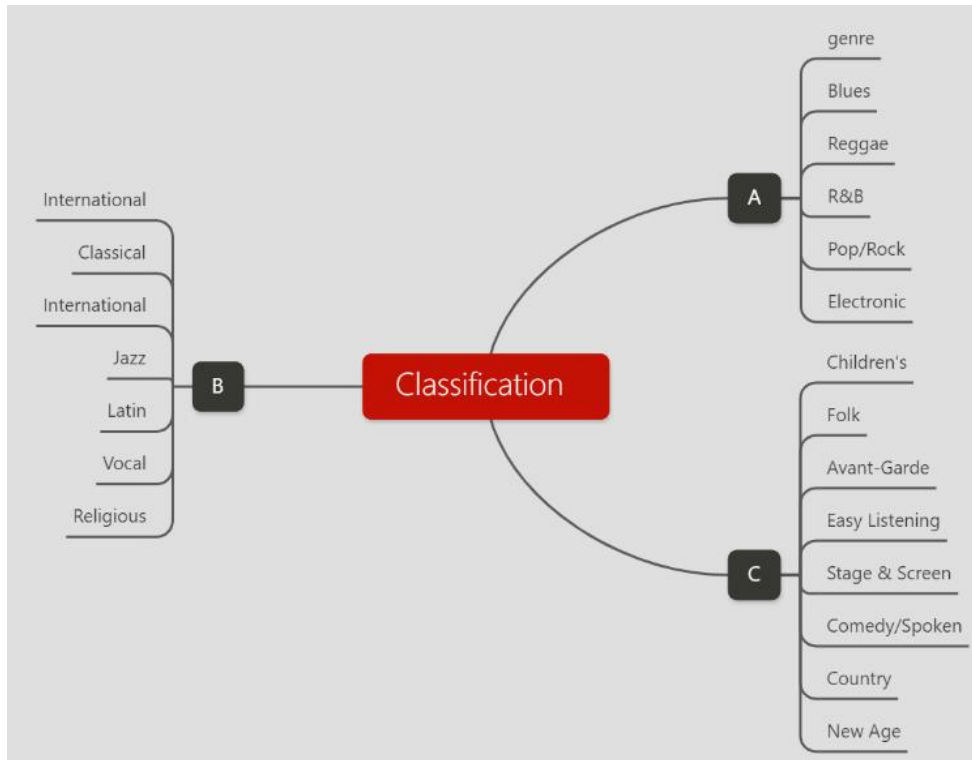


Figure 2. The classification of music

Next, count the number of followers corresponding to A, B, and C respectively, and normalize the data.

| Influencer | Follower | Processing      | Normalization |
|------------|----------|-----------------|---------------|
| A          | A1       | $A1/(A1+A2+A3)$ | 71.23%        |
|            | B1       | $B1/(B1+B2+B3)$ | 13.13%        |
|            | C1       | $C1/(C1+C2+C3)$ | 15.64%        |
| B          | A2       | $A2/(A1+A2+A3)$ | 5.57%         |
|            | B2       | $B2/(B1+B2+B3)$ | 88.48%        |
|            | C2       | $C2/(C1+C2+C3)$ | 5.95%         |
| C          | A3       | $A3/(A1+A2+A3)$ | 5.17%         |
|            | B3       | $B3/(B1+B2+B3)$ | 4.14%         |
|            | C3       | $C3/(C1+C2+C3)$ | 90.69%        |

Figure 3. Followers percentage

Show the proportion in the form of a pie chart, see the figure below:

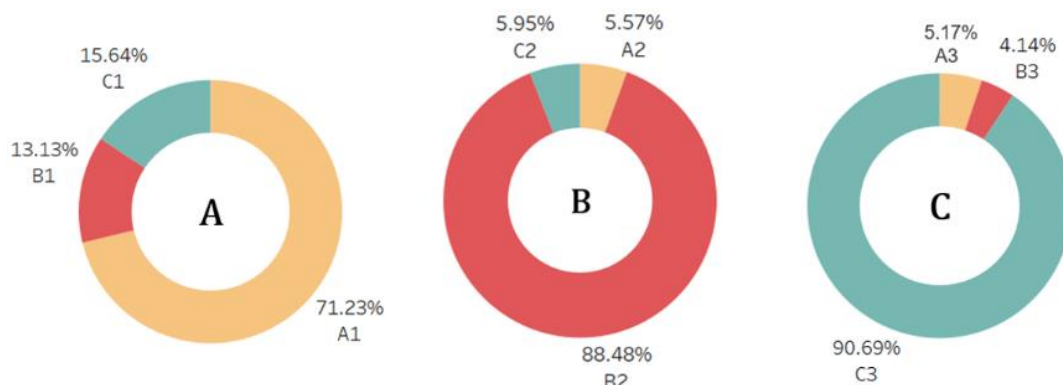


Figure 4. Pie chart

We can see that in general, influencers and followers belong to the same category. Of course, followers and influencers are not in the same category. This is because influencers have greater influence.

Next, we use the topsis method to find the side length<sup>[3]</sup>.

**2.6.1 Positive**

We use  $t_{max} - t$  to forward time to a maximum value parameter.

At this time, a normalized matrix is obtained

$$A = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

The meaning of n and m is: there are n objects to be evaluated and m evaluation indicators.

**2.6.2 Standardization**

In order to eliminate the influence of different dimensions of time, degree, and genre, it needs to be standardized. Let the standardized matrix be Z, Every element in Z,

$$z_{ij} = x_{ij} / \sqrt{\sum_{i=1}^n x_{ij}^2} \tag{1}$$

**2.6.3 Calculating weight  $w_j$  by entropy weight method**

Entropy method is an objective weighting method, and the data itself can tell us the weight. The smaller the degree of variation of the indicator, the smaller the amount of information reflected, and the lower the corresponding weight. Information entropy represents the amount of information, and its essence is the expected value of the amount of information<sup>[4]</sup>.

Calculating in Matlab, we get the weights of domain, degree, and time respectively as  $w_1 = 0.02324, w_2 = 0.61972, w_3 = 0.35703$ .

**2.6.4 Calculation**

Define maximum

$$Z^+ = (Z_1^+, Z_2^+, \dots, Z_m^+) \\ = (\max\{z_{11}, z_{21}, \dots, z_{n1}\}, \max\{z_{12}, z_{22}, \dots, z_{n2}\}, \dots, \max\{z_{1m}, z_{2m}, \dots, z_{nm}\})$$

Define minimum

$$Z^- = (Z_1^-, Z_2^-, \dots, Z_m^-) = (\min\{z_{11}, z_{21}, \dots, z_{n1}\}, \min\{z_{12}, z_{22}, \dots, z_{n2}\}, \dots, \min\{z_{1m}, z_{2m}, \dots, z_{nm}\})$$

Define the distance between the i-th evaluation object and the maximum value

$$D_i^+ = \sqrt{\sum_{j=1}^m w_j (Z_j^+ - z_{ij})^2} \tag{2}$$

Define the distance between the i-th evaluation object and the minimum value

$$D_i^- = \sqrt{\sum_{j=1}^m w_j (Z_j^- - z_{ij})^2} \tag{3}$$

Calculate the final score

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{4}$$

Obviously,  $0 \leq S_i \leq 1$ . And the larger the  $S_i$ , the higher the score, and the higher the comprehensive evaluation of the corresponding evaluation object.

**2.6.5 Normalization**

$$S_c = \frac{S_i}{\sum_{i=1}^n S_i} \tag{5}$$

At this point, we have obtained the score values of all influencers, and use this value as the length of the side to describe the influence of music. The longer the side, the greater the comprehensive

influence. Draw a new image in MATLAB. Due to the large amount of data, here we randomly selected 3000 influencers' data for drawing. Random selection can ensure that the image is clear and maintain the relationship between the original data.

The new directed subgraph obtained is as follows:

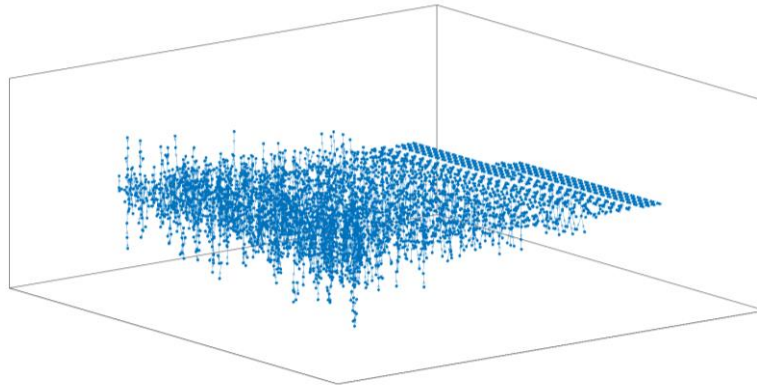


Figure 5. The new directed subgraph

We were pleasantly surprised to find that in a space with longer sides, the denser the artists, the more complicated the relationship between influencers and followers. This is a good confirmation of our model. We can conclude that time, genre and degree are the indicators of influence. The larger the parameter, the more influence he has, and the more influence his followers have.

### 3. Conclusions

In our directed network, artists with high influence will have more nodes connected, and the network of the same genre is denser, there will be directed line segments between similar genres. However, among the schools with big differences, the connection will be significantly reduced. The nodes can dynamically and flexibly affect each other through the line. The longer the directed line segment between nodes, the greater the influence. Changes in external social factors will change the length of some directed line segments at the same time, or directly cut off their connections. This effect is irresistible. The development and change of music is a process of continuous connection change or disappearance of the directed network, and the speed of the network change can reflect the development of music.

Hope our model and research can bring some help and inspiration to the research work of music to a certain extent. We believe that in the future, under the common development of more different disciplines, music research will be greatly promoted.

### References

- [1] [https://blog.csdn.net/weixin\\_42662955/article/details/89286893](https://blog.csdn.net/weixin_42662955/article/details/89286893).
- [2] Li Xin, Mi Hongjuan, Wu Xuejun. Classification comparison of multiple machine learning models of music genres [J]. Journal of Yibin University,2020,20(12):42-47.
- [3] [https://blog.csdn.net/qq\\_36384657/article/details/98188769](https://blog.csdn.net/qq_36384657/article/details/98188769).
- [4] <https://www.zhihu.com/question/357680646/answer/943628631>