ISSN 2618-1568 Vol. 3, Issue 2: 1-5, DOI: 10.25236/FAR.2021.030201

A complex directed network model by connecting followers and influencer

Shao Yang

School of Science, China University of Geosciences (Beijing), Beijing, 100083

Abstract: *Music is important to society, which is a reflection of people's spiritual world and affects the spiritual world of everyone. In our model, we will quantify the evolution of music based on the data provided. Then we get the reasonable evaluation result of the artists and reveal the evolution process of music. We constructed a complex directed network model by connecting followers and influencers with directed edges based on the influence data. This paper constructs the network structure index from several aspects. The relationship between artists is obtained by extracting subnetworks. Then, entropy method is adopted to determine the weight and construct the evaluation standard of artists' "music influence".*

Keywords: *network analysis, entropy method, breadth first search, cluster analysis, time series model*

1. Introduction

Music is an important part of the cultural heritage of human spirit, but also the crystallization of human thinking. Artists are influenced by their time, by their politics, by the rest of the artists, and the data-set gives the influencer of the artists, the characteristics of each song, the characteristics of each generation, and the characteristics of each artist.

2. Create Network

2.1 Why to generate

All the work that we're going to do is based on the relationship between the musicians. So out team created a fully directed network based on the analysis of influence data data-set, and there are 5,603 nodes in the network, which correspond to the musicians given by the data-set.

2.2 How to generate

The graph structure is generated as follows: each person is a separate node. If there is a relationship between two nodes (that is, the relationship between influencer and follower), a directed edge will be connected with influencer to follower. And our team counted the number of influencers connected to each follower here for subsequent processing. For the consideration of processing efficiency, our team used C++ here to generate network and process data. We noticed that the ID of musicians was not continuous and the maximum value were large. So we adopted discrete preprocessing here and hash the ID of all musicians to the range [1, 5603].

3. Model Development

3.1 Capture “"music influence"” in the network

In the network, we analyze the influence of a particular musicians through the following method: the musician is seen as a root node, using breadth-first search algorithm to find the musicians as directly or indirectly affect the person's subnetwork, we call on the musician's “influence subnetwork", we take the ID 5366 (Alfredo Gutierrez) of music for example, the analysis as shown in the Figure 1:

The figure above contains a subgraph of all nodes that we use for our analysis of a particular musician. When we consider the influence of a musician, the most intuitive thing is that when a musician influences

ISSN 2618-1568 Vol. 3, Issue 2: 1-5, DOI: 10.25236/FAR.2021.030201

more people, his *"music influence"* will be higher. We consider two evaluation methods: the size of the musician's "influence subnetwork" (the number of nodes in the subnetwork) is used as the index to evaluate the influence of the musician. The correctness of this method is obvious, and it is very easy for us to process the data. However, we further consider that if a follower follows multiple influencers, then the influence index of a follower following a particular influencer will decrease accordingly. Comparatively, the influence of the influencer should also decrease, so we define the *"music influence"* of a musician as Equation (1).

$$I_\alpha = \sum F_b \tag{1}$$

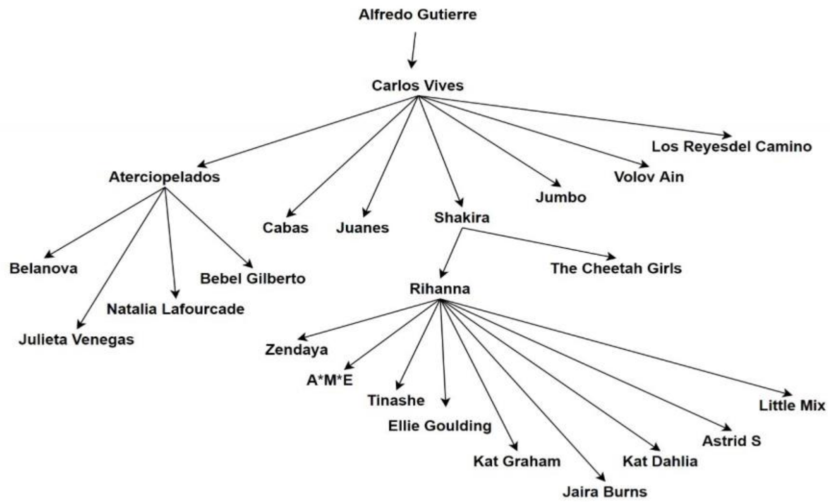

Figure 1: "Influence subnetwork" which rooted on 5366

Through this calculation method, the top 10 musicians of *"music influence"* are as Table 1:

Table 1: Top 10 influencer

rank	id	name	"music influence"
1	754032	The Beatles	1311.07
2	66915	Bob Dylan	1310.07
3	894465	The Rolling Stones	1309.07
4	531986	David Bowie	1307.57
5	139026	Led Zeppelin	1307.41
6	354105	Jimi Hendrix	1305.57
7	100160	The Kinks	1305.57
8	41874	The Beach Boys	1305.57
9	549797	Hank Williams	1305.37
10	840402	The Velvet Underground	1305.24

3.2 Analyse the artist on K-means

First, we add the mean data provided in the data sets *data_by_artisrt* and *data_by _year* as indicators to the data set *full_music_data* to make the data more comprehensive. To establish a music similarity measurement model, we define Euclidean distance as the distance between songs. That is, let $x_i = x_{i,1}, x_{i,2}, ..., xi, n$, $x_j = x_{j,1}, x_{j,2}, ..., x_{j,n}$ be song's coordinates, $x_{i,k}$, $x_{j,k}$, is the normalized value of each index, d is the Euclidean distance between songs, expressed as: $d_E = \sqrt{\sum_{k=1}^{u}(x_{i,k} - x_{j,k})^2}$. The distance between songs is used as a measure of music similarity. The smaller the distance, the higher the song similarity, and vice versa.

Next, we will compare the similarity between genres and genres. First, we use K-means clustering method to classify the songs in *full_music_data*. By comparing the aggregation of songs of each genre in the new subcategory, we choose Cluster the data into 5 categories. Then, the songs are classified according to the artist. According to the number of songs of each artist in the 5 categories, choose the category with the largest number of songs for comparison, and divide the songs in this category according to the same genre and different genres of this artist. According to the distance formula between songs,

ISSN 2618-1568 Vol. 3, Issue 2: 1-5, DOI: 10.25236/FAR.2021.030201

calculate the average distance between the songs of this artist in the category and the same genre and different genres. By comparing the mean value of the two distances, according to the rule that the smaller the distance between the two, the more similar, it can be concluded that each artist is highly similar to the same genre or different genres, and the comparison results obtained by 5854 artists are counted. Finally, the results show that artists within genres are more similar than artists between genres.

3.3 Analyse the characteristics of the genre

In order to find indicators for distinguishing genres, we divided the songs in the data set *full_music_data* according to music genres, and manually removed some of the indicators that were not conducive to subsequent processing and unnecessary according to the data, and then analyzed each of all songs in each genre. Calculate the mean and variance of the indicators respectively. The mean is used to compare the similarity of the indicators between the two genres to find the most similar characteristics between the two genres. We can calculate the difference between the 20 genres for each indicator mean value. And then compare the mean difference between the indicators. The indicator with the smallest mean difference shows that the two genres are most similar in this indicator. Variance is used to compare the similarity of different indicators of the same genre. The indicator with the smallest variance is the most similar feature in this genre. Through the above steps, we get the similar characteristics within each genre and the similar characteristics between each genre.

Based on the previous analysis, we propose the relationship between the degree of similarity and the degree of variation of genres. The more dissimilarity between two songs or genres, the greater the change or difference between them. Because we can find every genre index variance of all the songs of the maximum, the indicators represent the corresponding genre is the most representative change direction, so, we will each genre of songs according to s classification, respectively, for each s representative index to calculate the average of all the songs, and in accordance with the s to promote the index changes of images, so they can illustrate genre is how to change over time. We selected four of the images that clearly show the change of genre over time, as Figure 2, Figure 3, Figure 4, and Figure 5

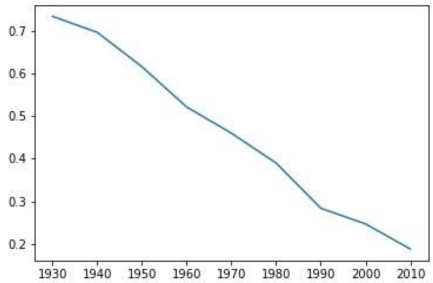

Figure 2

Figure 3

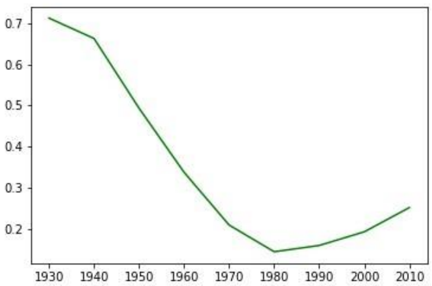

Figure 4

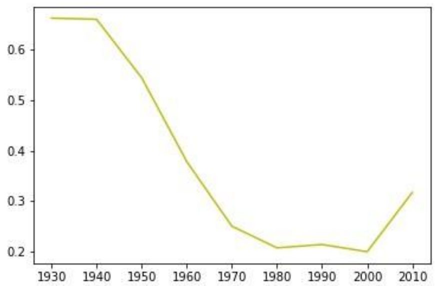

Figure 5

We can identify a genre to some extent by the similarities between the genres and the characteristics that each genre represents.

Exists between the genre and genre or strong or weak relationship, we applied before the directed network established between artists, artists will be carried out in accordance with the genre classification, the network connection between two points representing the effect between the two artists or contact,

ISSN 2618-1568 Vol. 3, Issue 2: 1-5, DOI: 10.25236/FAR.2021.030201

through the programming statistics under the two genres of the total number of connections between artists, 20 kinds of genres produce 190 kinds of connections. The strength of the influence or relationship between genres is clearly reflected by the comparison of the total number of connections.

3.4 Who is dynamic

We look for revolutionary features based on the directed network between artists. We will count all the complete chains in the directed network. After data removal and deduplication, we will calculate the difference between two adjacent artists in each chain. The cosine distance is calculated and used as a distance vector. At the same time, the same index of each artist on the chain is formed into a matrix and output. The distance vector and index matrix are calculated by the Pearson correlation coefficient, and the largest Pearson correlation coefficient can be found from this. This indicator is the most relevant indicator of this chain and distance. Next, we find the sub-chain with the largest distance interval in each chain, and make a distribution map of the largest distance sub-chain after deduplication, as shown in Figure 6.

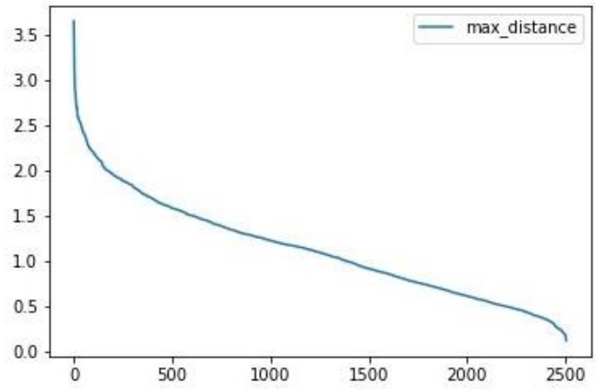

Figure 6: big distance

By observing the picture, because the curve rises fastest where the length of the sub-chain is larger, we select the top 10% of the sub-chain to count the most relevant features corresponding to it, and the order of the most relevant features obtained from it is It is possible to produce a ranking of major leap features. We then determine the influencers of major changes based on the positions of these sub-chains and rank them statistically.

4. Strengths and Weaknesses

4.1 Strengths

• We use C++ to process the network and data, so the processing is very fast. We can process the network with **10,000** nodes in 1 second.

• We use Euclidean distance to evaluate the similarity between songs, which is reasonable and accurate.

• Our model processes a very reasonable ranking of music influence, with well-known artists at the top.

• Our models are basically based on the network of relationships we established at the beginning, so subsequent models can reflect the relationship between artists and music.

4.2 Weaknesses

• Our model does not further consider the influence of the “influencer’s influencer” on followers.

• In the initial "music influence", we only averaged the influence of different influencers on followers without further consideration.

• Our model is not very sensitive to the development of music, and focuses more on the analysis of the relationship between music genres and its own characteristics.

ISSN 2618-1568 Vol. 3, Issue 2: 1-5, DOI: 10.25236/FAR.2021.030201

5. Conclusion

In our work, wo generate the directed network to describe the connections of the artists. By calculating the number of followers of a person as the person's "music influence", we get the most influential artist - "The Beatles". Then, with the help of network and clustering algorithm, analyzes the similarity between musicians and among genres. Through the in-depth excavation and analysis of the directed network features among artists, through the screening and sorting of the features, based on the complex relevance of network nodes and the mutual influence between genres, the outstanding potential revolutionary features, reformers and their influencers are found out.

References

[1] Sobia Zahra, et al. "Novel centroid selection approaches for K-Means-clustering based recommender systems." Information Sciences 320. (2015): Doi: 10.1016/j.ins.2015.03.062.

[2] Zhuo Ming Ren, et al. "Structure-oriented prediction in complex networks. "Physics Reports. (2018): Doi: 10.1016/j.physrep.2018.05.002.

[3] Assal Habibi, et al. "Changes in auditory cortical thickness following music training in children: converging longitudinal and cross-sectional results." Brain Structure and Function 225. Republish (2020): Doi: 10.1007/s00429-020-02135-1.