

Big Data Analysis of Music Influencing Factors Based on Complex Network

Qirui Yu¹⁺, Wenshuo Zhai¹⁺, Rui Zhu¹, Yuanyuan Jiao^{1*}, Liang Bai¹ and Nanjun Li²

¹College of Systems Engineering, National University of Defense Technology, Changsha, China

²College of Electronic Science and Technology, National University of Defense Technology, Changsha, China

⁺These authors contributed equally to this work *Corresponding author

Keywords: Complex Network, Node Aggregation, Cosine Distance.

Abstract: Music plays an important role in human society. This paper attempts to explore the law of music development through the music style of artists over the years, the works created and the mutual influence between artists. Firstly, this paper adopts node aggregation algorithm to simplify the complex network of artists of different ages and genres into a simple sub-network. Then, based on the cosine distance, this paper analyzed musical characteristics of artists so as to work out the similarity degree between them. Finally, this paper combines the development of society and technology to comprehensively analyze the law of the development of music, and draws the conclusion that different music genres have different influences in different ages. Major social changes and major scientific and technological development related to music will also have a great impact on the development of music.

1 INTRODUCTION

As an important part of art, music plays an irreplaceable role in the development of human society. In order to better explore the development of music, we wanted to develop a method to quantify the development of music. There are many factors that influence a artist's work, such as personal experiences, the influence of other artists, social events, technological developments, and so on. And the song has rhythm, loudness, dance and other characteristics. In order to discover the law of the development of music, we should pay attention to the emergence of music genres and the development of different music genres in different periods from the perspective of the whole.

The Integrative Collective Music (ICM) Society aim to define a model for measuring musical influence. They launched a global challenge in 2021, hoping that teams would come up with a better solution to the problem. These data were scraped from AllMusic.com. This paper analyzed the music-related data given in the competition and the data collected from the Internet. In order to better explore the process of music evolution, looking for the law of music schools and branches, as well as the interaction between art-

ists, we use a variety of methods to analyze the similarity, characteristics, influence, genre evolution and other relevant aspects of music.

We built up the model of Node aggregation, fusing all artists of the same genre from the same era. We will get a new compressed network weighted by the sum of the out-degrees of the same kind of nodes in the original network, with the value of weight reflecting the influence of a particular genre in a certain year. We showed the measurement of the interaction between genres in some eras in the main body of our paper. And the influence of each artist is reflected by his genre and era. We defined seven intervals as the characteristic value of each genre according to the principle of Box Plot. The similarity between genres is determined by the coincidence degree of the intervals by contrast while the similarity within genres is ascertained by the variance degree of the intervals. We analyzed the degree of similarity between influencer and follower by straightly dividing the seven characteristic values of a follower by those of his influencers correspondingly and defining the similarity by the deviation degree between the ratio and 1. Through the fluctuation of each characteristic with time, we found that valence and energy functioned as the indicators that reveal the dynamic influencers. Combined with the figure of the amount of works

changing with time, we can explain the regulation of genres and artists changing with time.

In the end, the paper mentioned the method of combining musical revolutions with historical and cultural events, thus applying the idea of deep learning in order to dig out the effect laws of social, political or technological changes on music.

2 CONSTRUCTION AND OPTIMIZATION OF COMPLEX NETWORKS AMONG ARTISTS

2.1 The Construction of Complex Networks among Artists

We regard that only the amount of a specific artist's immediate followers counts when measuring his personal musical influence since if the followers of his followers were influenced by him, they must also be his followers according to the given information. To quantify the musical impact of every single genre, we consider conducting dynamic fusion of related nodes and measuring the influence between every two artists by the interacts of the genres they belong to.

Eventually, we are going to extract a sub-network from the original directed network, thus conducting micro analysis on three randomly-chosen nodes, making a comparison to find out whether the genre of a certain artist's main influencers is concurrently the main influential genre of the genre he belongs to. If so, we can conclude that our 'music influence' measures are revealed in this sub-network.

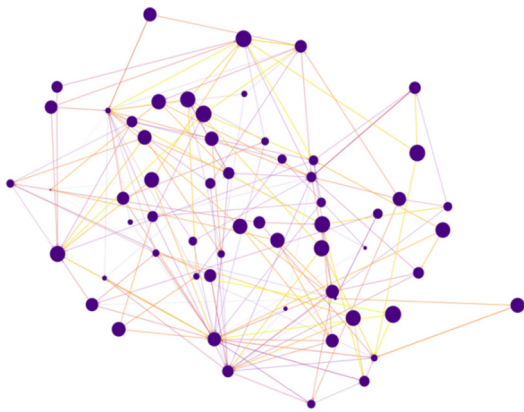


Figure 1. The directed network of musical influence drawn in python.

2.2 Simplify Complex Networks Based on Node Aggregation

Under the obtained network, we discovered that the in-degree of a certain node is exactly the number of its corresponding artist's influencers while its out-degree equivalent to the amount of his followers. Then, we fused the same type of music based on the year, that's, fusing the nodes according to both influencers' active starts and their main genres, at the same time accumulating both in-degree and out-degree of every single fused node in order to form the new in-degrees and out-degrees of the points after aggregation.

Eventually, extracted a sub-network from the original directed network, making a comparison to find out whether the genre of a certain artist's main influencers is concurrently the main influential genre of the genre he belongs to. If so, we can conclude that our 'music influence' measures are revealed in this sub-network.

2.3 Data Mining on the Network

Randomly choosing three nodes to test the accuracy of our results, corresponding to artists belonging to respectively 1980 Pop/Rock, 2000 Pop/Rock and 1960 Pop/Rock, we found that the largest proportion of their influencers are respectively 1960 Pop/Rock with the proportion 0.488372, 1960 Pop/Rock,1980 Pop/Rock, 1950 R&B and 1940 Vocal sharing the same proportion and 1970 Pop/Rock with the proportion 0.34. Comparing the results with the aggregated directed network, we found that the results are highly consistent, indicating that measuring the mutual influence of artists in different genres by the cross-influence of the genres they belong to embodies the idea does make sense.

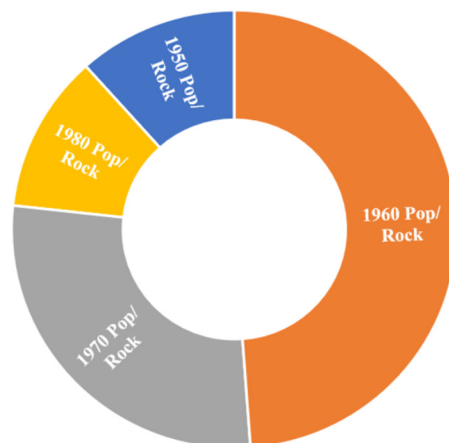


Figure 2. The directed network of musical influence.

Our model is based on the aggregation of networks. The fusion of genres of the same type by time scale reflects the influence of diverse genres in different periods. The subtlety of this model is that it possesses enough robustness to reflect the influence of a certain musical genre in each period reasonably and ingeniously since the musical influence of diverse genres differs greatly in distinct decades while a specific genre's influence on music of different ages in a particular year is also different. Our model realized the goal of conducting concrete analysis to specific problems. Therefore, it can be regarded more targeted in time, clearer and more quantifiable. Compared with the establishment of fuzzy comprehensive evaluation system, this model evaded the unnecessary trouble when conducting empowerment due to the abstractness of the research object and the lack of relevant data. However, we should admit that our model is limited in some respects. For example, the accuracy of the results largely depends on the soundness of the samples. If the number of chosen samples of artists in a certain genre or a certain era is too small, then the result of our solution wouldn't be representative enough, far from the actual situation.

3 ANALYSIS OF MUSICAL SIMILARITY

3.1 Similarity Analysis Based on Cosine Distance

We selected the corresponding data of seven music characteristics as valid data, thus generating a seven-dimensional characteristic vector for each song, each dimension corresponding to a characteristic index. Next, we identified the correspondence between artists and genres and classified the artists by the genre, that is, putting artists belonging to the same genre into a collection, while scouring off unclassified artist data. For artists of the same genre, we calculated the Cosine Distance between every two seven-dimensional characteristic vectors in the set. In conclusion, the similarity of a particular genre is the mean value of all these Cosine Distances.

$$\cos \theta = \frac{\sum_{i=1}^n x_i \times y_i}{\sqrt{\sum_{i=1}^n x_i^2} \times \sqrt{\sum_{i=1}^n y_i^2}} \quad (1)$$

And the Cosine Distance is defined as $\omega = 1 - \cos \theta$.

Set the Cosine Distance between the i^{th} artist and the j^{th} artist of the same genre as a_{ij} , and the sample size of the genre as n , thus the similarity of a certain genre can be expressed as the equation below:

$$W_1 = \frac{\sum_{0 \leq i < j \leq n} a_{ij}}{C_n^2} \quad (2)$$

For artists belonging to different genres, we thus calculated the mean value of cosine distance from all the seven-dimensional vectors in one genre to those in the other genre to work out the similarity between two genres. Set the Cosine Distance between the i^{th} artist in one genre and the j^{th} artist in the other genre as b_{ij} , and the sample size of the two genres as m and n , thus the similarity of two genres can be expressed as the equation below:

$$W_2 = \frac{\sum_{i=1}^m \sum_{j=1}^n b_{ij}}{mn} \quad (3)$$

The former figure is the original similarity data of 20 genres and the latter one is a more intuitive thermal graph presentation after grading and quantifying the similarity data.

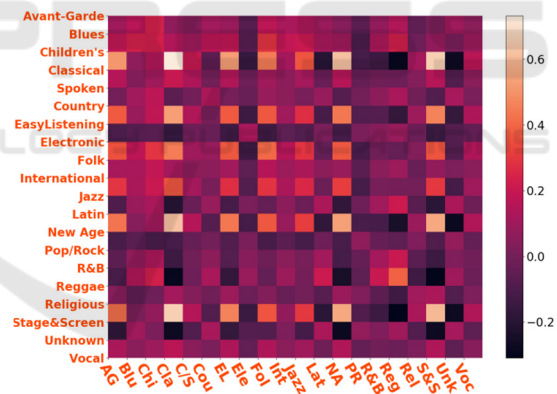


Figure 3. Thermal graph based on similarity

3.2 Comparing Similarities and Influences between Different Genres

For the sake of distinguishing a genre, we firstly put our hand to measuring the similarity between different genres. Due to the consideration of determining a parameter to characterize each characteristic, we defined the interval from one fourth to three fourth of all the songs' characteristic values in a certain genre (the characteristic values had been strictly sorted) as the parameter according to the data analysis method based on Box Plot.

As for comparing the similarity among different genres about a specific characteristic, we took turns contrasting the same music characteristic intervals of every two genres. And in order to unify the standard of measurement, we further defined the degree of similarity as the ratio of the obtained span and the span of the smaller parent interval.

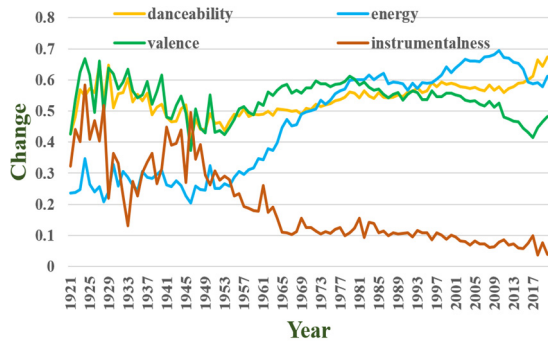


Figure 4. Line Chart of 4 Characters Changing with Time (data from <http://AllMusic.com>)

The area of each matching result was the total similarity between every single song and a particular genre. Thus, we defined the largest area of the matching result as the genre this song belongs to. In order to seek out the law of the change of genres over time, we started from both horizontal and vertical perspectives. From the angle of breadth, we calculated the variance of all elements in the song set corresponding to each characteristic for all the 20 genres. We found that most characteristics of the songs in each genre are not that similar, with only several of the seven ones having small variances. From the portrait angle, we classified and aggregated songs on a time scale, thus analyzing the changing trend of each characteristic with time.

As for judging whether some genres are related to others, we took both music influence and the similarity into consideration. If the similarity scaling between two genres belongs to close resemblance and their influencing scaling are over great impact, we then regard that these two genres are related to each other. For the sake of obtaining the result of how genres change over time, we took the genre Pop/Rock as an example.

4 ANALYSIS AND EXPLORATION OF MUSIC INFLUENCE

4.1 Analyzing the Influence Process of Musical Evolution

To determine whether the influencers actually affect the music created by the followers, we still started from analyzing the similarity of seven characteristic values. For each follower, we divided his values by those of the influencers correspondingly. Observing the data of proportion, we can conclude that the proportion of songs with extremely similar characteristics is not very large, indicating that the music styles of followers are not so affected by their influencers.

Next, we considered from the following two perspectives. On the one hand, we explored the influence of the same influencer on various followers. On the other hand, we took the distinct effects from various influencers on a specific follower into consideration as well. According to the similarity division basis above, we first calculated the average proportion of followers sharing high similarity for a particular characteristic with every single influencer. Subsequently, we worked out the average proportion of influencers sharing high similarity for a particular characteristic with every single follower utilizing the same method.

We concluded that the influencers actually affect the music created by the followers and there do exist some music characteristics more ‘contagious’ than others, they are Danceability and Tempo. In order to measure musical evolution of a genre over time, we transformed our focus on the works belonging to that genre. We took the average value of each characteristic of R&B songs in each year to reflect the overall situation of a specific characteristic of the genre in that year.

4.2 Discussing the Role of Cultural Factors Acting on the Field of Music

Our work expressed information about cultural influence of music in time or circumstances by building up the bridge between the alteration of characteristics and the time corresponding to its occurrence.

As for identifying the effects of social, political or technological changes on the music evolution, we come up with the idea of deep learning. After the previous data processing and analysis, we've already been equipped with the ability of quantifying both

music styles and genres by conducting data analysis of all the characteristic values.

In order to microcosmically measure a musical revolution during a historical period, we considered the musical characteristic values before the revolution as the input and those after the revolution as output. Through literature search, we were going to find out all the social, political and technological events during the period. Classifying these events strictly according to their categories, we would get three event sets covering Social Category, Political Category and Technological Category. Defining the law of action of the above three kinds of events on music respectively as equation $f(x)$, $g(x)$ and $h(x)$, we were enabled to build up the model of input, output and effect about deep learning. Through a large number of data training with events of music revolutions scattered in every year and every month, the effect laws of $f(x)$, $g(x)$ and $h(x)$ would be eventually dug out.

5 CONCLUSION AND DISCUSSION

Music plays a vital role in expressing people's thoughts and reflecting social life. In order to explore the law of music development and uncover the internal logic of music development, this paper constructs a complex network among artists, which includes the characteristics of artists, the characteristics of artists' works, social factors and technological factors, reflecting not only the cross-influence between different genres of music in different ages, but also the effect of a genre on itself during various periods. Then, we simplify the complex network based on node aggregation through taking genre of a specific era as the unit. Meanwhile, we macroscopically measure the mutual influence of artists of different genres by the mutual influence of their respective genres. In the analysis of the influence degree, this paper bases on the cosine distance for the correlation analysis.

Through our research, we find that the influence among artists, the major social events related to music creation and the development of technology all have a huge impact but with various degree of influence on the development of music. We have only made a preliminary study of the factors that influence the development of music. In order to better reveal the law of the development of music and make contributions to human art, let us all work together.

REFERENCES

- Anna J. Reisenweaver. Guido of Arezzo and His Influence on Music Learning [J]. *Musical Offerings*, 2012,3(1).
- Alberto Arroyo. The meta-instrument as compositional technique: Strategies of the current musical creation[J]. *Emille: The Journal of the Korean Electro-Acoustic Music Society*,2018,16.
- James Leonard, Jérôme Villeneuve, Alexandros Kontogorgakopoulos. Multisensory instrumental dynamics as an emergent paradigm for digital musical creation A retrospective and prospective of haptic-audio creation with physical models [J]. *Journal on Multimodal User Interfaces*, 2020,14(3).
- Jon M Sweeney. Review: Richard Wagner's immense influence on music (and history) [J]. *America*,2020,223(5).
- Kaskon W. Mindoti, Hellen Agak. Political Influence on Music Performance in Kenya between 1963-2002 [J]. *Bulletin of the Council for Research in Music Education*, 2004(161/162).
- Kong Wenting. Analysis of the Influence of Digital Media Technology on Music Communication. [J]. *Journal of Physics: Conference Series*, Volume 1453, 2019 2nd International Conference on Computer Information Science and Artificial Intelligence (CISAI 2019).
- Sunil Aryal, Kai Ming Ting, Takashi Washio, Gholamreza Haffari. Data-dependent dissimilarity measure: an effective alternative to geometric distance measures [J]. *Knowledge and Information Systems*,2017,53(2).
- Tao Wang, Hongjue Wang, Xiaoxia Wang. A novel cosine distance for detecting communities in complex networks [J]. *Physica A: Statistical Mechanics and its Applications*,2015,437.
- Trainor Laurel J. The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. [J]. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 2015,370(1664).
- Trainor Laurel J. The origins of music in auditory scene analysis and the roles of evolution and culture in musical creation. [J]. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*,2015,370(1664).
- W. Xiong, X. Yu, C. Liu, G. Wen and S. Wen, "Simplifying Complex Network Stability Analysis via Hierarchical Node Aggregation and Optimal Periodic Control," in *IEEE Transactions on Neural Networks and Learning Systems*.
- Xiaoyan Gao. P2P Network Node Aggregation Algorithm Research Based on Community [J]. *International Journal of Engineering Practical Research*, 2015,4(1).
- Yucheng Lin, Zhigang Chen, Jia Wu, Leilei Wang. An Opportunistic Network Routing Algorithm Based on Cosine Similarity of Data Packets between Nodes [J]. *Algorithms*, 2018,11(8).