IV. Influence Network Model

4.1 Building musicians' influence network

Are there any data problems with the data\_influence dataset?

Considering that influencers' influence relations on followers are directed, they can be described by the adjacency matrix in the data structure [1]. To generate an adjacency matrix for network analysis, the conditions characterizing the connections must be defined. In the influence\_data dataset, a total of 5603 musicians are assigned the roles of influencers and followers. Musicians as network nodes, which we will define as influencers Having an influence relationship on followers, the adjacency matrix between musicians can be expressed as



Then, we import the data and visualize it. Due to the large number of nodes and connected edges, the nodes are jumbled together and poorly visualized. Multidimensional scaling (MDS) converts a set of pairwise dissimilarities among objects into coordinates that can be used to visualize the objects in a low-dimensional space [MDS used to visualize the objects in a low-dimensional space) [this is the translation], and the reduced-dimensional samples are centralized because only the distance information is retained [zh1 13], facilitating the embedding of their nodes into two dimensions for visualization. Therefore, we perform graphical visualization with the help of Gephi using the MDS layout approach, from which the musician influence network is generated.

图 ：MDS layout of the musician influence network, having 5603 nodes and 42770 edges. 754032 has the highest degree, far left, is far away from other points.

In the musician influence network, we use the term musician to represent each musician, and we design 20 nodes with four colors and five shapes to distinguish the different factions of musicians. If the influencer has influence on the followers, a directed edge is generated from the direction.

4.2 Musicians' personal musical influence

To determine the impact indicators of musician influence, we imported the data into UCINET for individual network analysis of musician influence networks, where the indicator with the most significant differences was the degree of nodes in the network.

The degree of influence a musician has on other musicians is characterized by the musician's presence.



The use of the term "musician's access" represents the extent to which a musician incorporates the strengths of other musicians when composing.



If a musician has too many references and relatively little influence, it can be considered that the musician's music is weak in originality, and from a practical point of view, such artists generally do not have far-reaching influence; while the music masters of the last century, there is only out-degree without in-degree, because they have left far-reaching music through their continuous research, and such musicians are highly influential. Based on the study of the nature of the network and the degree of nodes, we define the individual musical influence of musicians .



4.3 Subnetworks of Music Impact Research

Based on the study of individual musicians, we clustered musicians belonging to the same faction in the musicians' network, which resulted in twenty sub-networks of different factions, interconnected within the sub-networks and connected between different sub-networks.

Figure Sub-networks of the Musician Influence Network

Based on the faction sub-network, we evolved the musician influence network into the faction influence network. Music influence is defined for each music faction as the main body, i.e., the magnitude of the music influence of each faction is calculated. The nodes are each music faction, denoted by If there are musicians in the faction who have influence relations on the faction, then a directed edge pointing to the faction is generated, and the weight of the edge is the total number of influencers in the faction who influence the followers of the faction. All nodes of each faction sub-network, the self-loop, the outgoing edges pointing to each faction, and the incoming edges pointing to each faction are fused separately by gephi. The number of musicians in the sub-network determines the size of the node, and the weights of the outgoing and incoming edges of each node determine the thickness of the outgoing edge, respectively, thus generating the faction music influence network.

4.4 Faction Music Influence

Since the faction influence network evolves from the musician influence network, where the nodes are the fusion of musician nodes of the same faction, the degree in the faction influence network is still the key indicator of influence. We still extend the ratio of out-degree to in-degree to characterize the original propagation power of musicians, then the original propagation power of music in factions can be calculated by summing this property of all musicians in the faction, thus we define the individual original propagation negative indicator.



On the basis of degree analysis, cluster analysis of faction influence network according to **distance** or **year** (← must be bolded), we can find that there are groups with closer mutual influence between factions, and factions that link groups internally or link other groups, play a mediating role and have a decisive role in the stability of the network, according to the knowledge related to the intermediary analysis.

Factions that play the role of intermediaries have an important position in the network and often determine the stability of the network. Therefore, it is reasonable to use intermediacy as part of the faction influence size index. Since the faction music influence network belongs to the edge-weighted directed graph, there are two general methods to solve for its intermediary centrality: one is to use the degree centrality representation, and then to use the nature of power-law distribution to approximate its solution. However, these two methods are less tractable, so we define the intermediary degree by using the weight matrix and adjacency matrix to find the relative outgoing and relative incoming weights, inspired by the intermediary analysis.



The intermediary propagation power can be obtained by normalizing the intermediary degree of all nodes using the maximum intermediary degree in the node



The final definition of the impact of the faction is based on a combination of negative indicators of individual original communication and intermediary communication power, and the correlation between communication and impact. 



V. Similarity metric model

5.1 Construction of Music Network

The influence between musicians is studied and analyzed by building musician influence networks and evolving them into factional influence networks. In order to explore the characteristics of music in depth, we can use the full\_music dataset to build a song scatter graph (Song Graph), and connect songs with similar musical characteristics to generate a song similarity network. Considering the one-to-one correspondence between songs and singers, we change the idea to superimpose the song scatter graph and musician influence network by using the correspondence between songs and musicians, thus generating a comprehensive network - music influence network. In the music influence network, when analyzing the similarity between songs, in addition to the musical characteristics of the song itself, there is also the influence of the relationship between the singers of that song. In this way, the similarity measure we established is closer to reality and makes the analysis of musical characteristics more comprehensive and thorough.

5.2 Music similarity measure

There are many forms of music interpretation, such as: singing, musical instruments, and computer synthesis, but almost all music can be found with its similarities and grouped into a few known categories. By analyzing similar music in chronological order, the evolution of musical characteristics over time can be discovered, making the music similarity measure very important for our study of the history of music development.

To establish a comprehensive and reliable music similarity measure, we combined the information given in the full\_music\_data dataset to quantify the music similarity in three aspects: music characteristics, vocal characteristics and song descriptions. Considering that too many referenced metrics will affect the accuracy of the similarity measure, we first analyze the 17 metrics given in the data mechanically and perform the first wave of metrics elimination.

After reviewing the literature and mechanistic analysis, we analyzed that the three indicators of year, release\_date and song\_title in song description and explicit in vocal characteristics were not strong enough to explain the musical similarity, so the above four indicators were excluded. Then, to check the correlation degree among the indicators, the correlation coefficients were solved for the remaining 13 indicators and analyzed.

(the above table with latex typed out to see the effect, not if the direct screenshot)

The table shows that the correlation coefficient between loudness and energy is the largest and only it exceeds 0.5, so we consider that the contribution of loudness and energy to similarity is duplicated. In a realistic sense, power and loudness are related by musical essence, so we eliminate loudness, leaving energy with both spiritual and loudness properties. therefore, we selected 12 indicators, and all of them are listed below according to the major categories they belong to.

1. Similarity of musical characteristics 

According to the common knowledge of music, the key of a piece of music is the biggest indicator of music recognition, mode and key values are the data of music key, for mode value, major and minor can directly determine the similarity of two songs or not, while for key value, the change range of two songs over 1 is beyond the range of transposition, so define to express the decisive influence of key value on the similarity of songs The influence of key value on song similarity is defined.

, , ,

Using , , , and represent the remaining music indicators danceability, energy, valence, and tempo, respectively, they are not as extreme as mode and key values for music recognition, so their respective weights are given to indicate the contribution of each indicator to music similarity. Therefore, it can be concluded that the similarity of music characteristics :



0.39193, 0.08823, 0.37313, 0.14671, respectively (determined by entropy method)

2. Similarity of vocal characteristics 

Among the vocal characteristics, a decisive reference like key or mode is speechiness, which represents whether the song music is interpreted in speech form, the more exclusively speechlike the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. The contribution of speechiness to similarity can be expressed as



Using ,, and represent the remaining musical indicators acousticness, instrumentalness, and liveness, respectively, all three indicators are analyzed in terms of musical tracks, and the contribution of the difference between each indicator to the musical similarity is considered to be the same. In order to obtain standardized data and a more intuitive expression, we use the cosine distance to characterize the difference between the indicators, so that we can derive the similarity of vocal characteristics as follows



3. Similarity of song descriptions 

The song description aspect does not contribute much to the similarity, so this aspect is also the one with the most eliminated indicators. For the song seconds aspect, we used the difference between the data and normalized it to obtain the song description similarity



4. Similarity measure formula

In summary, we obtainemusic similarity, vocal similarity and song description similarity, and developed the contribution coefficients for each of the three indicators to obtain the music similarity measure.



5. Determine the unknown weights using the entropy weight method

In the similarity measure formula, we assume coefficients in the formula, and reliable coefficients must be given before solving the similarity measure. The general method of giving coefficients is hierarchical analysis and other methods that use expert scoring to determine the size of the weights, but such methods are too subjective. Considering that the similarity measure method should serve the subsequent exploration of music development history, we chose the entropy weight method which gives the weights objectively through data characteristics.

According to the definition of information entropy, for a certain index, the entropy value can be used to determine the dispersion degree of a certain index, and the smaller its information entropy value, the greater the dispersion degree of the index, and the greater the influence (i.e., the weight) of the index on the comprehensive evaluation. Therefore, we use the entropy weighting method to solve for the unknown coefficients.

As an example, we show the process of the entropy weighting method by solving the contribution coefficients of the music indicators danceability, energy, valence, and tempo in The data of each indicator is standardized. The 4 indicators , where , are assumed to have a value of after standardizing the data of each indicator .



Next, the information entropy of each indicator needs to be solved for According to the definition of information entropy in information theory, the information entropy of a set of data.



where . If , then define 

With the information entropy of each index, the weights of each index can be determined. According to the formula of information entropy, the information entropy of each indicator is calculated as , and the weight of each indicator is calculated by information entropy as follows.



Through the entropy weighting method, we determined the coefficients of the unknown weights: : : : 0.24846, 0.68001, and 0.07153, respectively. The contribution coefficients of the music indicators danceability, energy, valence, and tempo in were 0.39193, 0.08823, 0.37313, and 0.14671.

5.3 Similarities within Genres

After solving for the unknown coefficients using the entropy power method, our key to unlock the world of music is finally activated. Due to the progress of society, things are changing faster and faster, and so is music. When analyzing the faction influence network, we found several songs we had heard in the POP faction, and they had retro-jazz flavored music as well as modern feeling music, so we doubted the division of the faction. With this doubt, we used the created similarity measure to solve for similarity between two two musicians within each faction in order to analyze the magnitude of similarity between musicians within the faction.

Due to the large amount of data to be displayed and the similarity result is a number between 0 and 1, it is not intuitive to display the result directly and does not play a good display effect. Therefore, we represent the number between 0 and 1 as a color change from blue to yellow, and convert the similarity matrix into a color scale diagram to show it. The color scale diagram of similarity between musicians within each faction is as follows.

The similarity of each faction can be observed very visually from the color scale chart. For example, in the chromatograms of New Age and Classical, a large part of the chart is yellow, with only a few blue horizontal or vertical lines. These few blue lines may represent alternative musicians in that faction, and from the perspective of the composition of the similarity formula, these alternative musicians may have a preference for one key or interpretation. However, the number of factions with an overall blue color in the chromatogram is high, such as the chromatograms of Pop/Rock and Unknown, which have a large or all-blue color, and Pop/Rock, which has only a few yellow spots. /Pock is a new young faction of music that emerged with the information age, and the musicians have different styles, and so do the songs in the faction, so the similarity is very low.

Based on the results, it can be seen that the similarity between factions is low, and the results are consistent with the real-life truth that music belonging to the same faction is not uniform. Moreover, the degree of similarity varies widely across factions, from those with less variation in musical style to those whose music is almost all unlike each other.

5.4 Similarities between Genres

After studying the similarity of musicians within factions, let's examine the similarity between musicians of different factions and, by the way, verify the correctness of our similarity measure. Since there are many different factions and each faction contains a large number of musicians, it is impractical to solve for them one by one, and data analysis would be very troublesome. Therefore, we designed a computer simulation experiment to compare the similarity of musicians within and between factions. The following is the algorithmic flow of the simulation process.

We conducted 500 randomized trials using a computer, and each trial was considered a success if the intra-pie musician similarity was greater than the inter-pie musician similarity, and returned a value of 1. We presented the results of the 500 trials in the form of area plots.

The graph shows that the blue area has a wider and denser line than the red area, indicating that there were more successful trials. According to the data, there were 318 successful trials out of 500 trials. We then conducted 5000 and 10,000 trials, respectively, and the success ratio was around 0.6. So, excluding chance, we have evidence that musicians within factions are more similar than musicians between factions.

5.5 Study of faction changes over time

To study the change of factions over time, starting from the data\_by\_year dataset, we calculated the year-to-year similarity

5.6 Similarity Study of Musicians' Influence Networks

Using similarity measures, we analyzed musicians within and between factions. Shifting the perspective to another known dataset, we investigated the influence relationships between musicians using the similarity measure with a view to deriving the actual role of influence relationships between musicians.

We verify whether the musical characteristics of influencers and followers are similar, and if they are not, it means that the influencers actually have little effect on the corresponding followers. Because of the large data set and the tedious process of calculating them one by one, we designed a computer simulation experiment in which a randomly selected group of interacting musicians from data\_influence was used to solve for the similarity between each group of influencers and followers separately. The experiment was cycled 500 times and the similarity of each trial was counted as a way to reveal whether the influencers actually had a large or small effect on the followers. The algorithmic steps of the computer simulation procedure are as follows.

Through computer simulations, we derived similarity data for 500 trials.

As can be seen from the figure, the similarity derived from 500 trials fluctuated around 0.4, and the number of trials where the statistical similarity exceeded 0.5 was 81 in total. Therefore, the test results can prove that the similarity between influencers and followers is not high, which can further indicate that the actual influence effect of influencers on followers is not significant.