Problem Chosen	2021 MCM/ICM Summary Sheet	Team Control Number
D		2103274

Music evolution

Summary

We construct 4 models to study musical influence, musical similarity, the evolution of music and relationship between music and society.

Firstly, we build Model 1 to study the musical influence. Before we start, we do the data processing to eliminate invalid data and narrow down the data range. To get the parameters which measure the influence of music, we study this problem in both "local" and "global" perspectives. By using the knowledge of graph theory, we produce 4 directed networks of each major genre and find interesting patterns: the existence of a few very influential artists and the huge relationship network centered around them. As a consequence, in "local" perspective, the out degree of node in the networks is selected to measure the musical influence of each artist within genres. In "global" perspective, the average out degree of the nodes in one genre towards another is used as the parameter, and in order to give it a more meaningful and quantitative description, we further develop another parameter on the strength of PageRank algorithm. From Model 1, we totally form 3 parameters for music influence and found specific artists and genres that are more influential.

In addition to musical influence, we also want to study the similarity of music. In model 2, we use data standardization and Principal Component Analysis (PCA) to pre-process all the data, forming 8 principal components to describe musical similarity. Then K-means Clustering Analysis is adopted to see whether songs of same genre and distinct genres have different similarity. Furthermore, to jump out of the local optimal solution, Simulated Annealing (SA) is used, and we obtained a clear clustergram of randomly selected samples. At last, we conduct a sensitivity analysis for SA and the final conclusion is that songs of same genre are more similar than others, while it might be true songs from some specific genres are quite similar actually.

The next model we built is to figure out what characteristics can be used to signify major leaps in music history. This model is based on Time Series Analysis and we use an algorithm called FTTO to eliminate vibration and noise in the data set and mark turning points. We choose popularity and number of songs as 2 valid characteristics and explored the similar pattern. Based on that, we can signify those turning points as major leaps. Processing other characteristics and comparing their turning point distributions respectively with popularity and number of songs as we mentioned, we can know whether they can be used to mark out major leaps.

After studying the musical influence, similarity and evolution of music, Model 4 was made to check how external factors affect music and how music affect the society. Ideas of Time Series Analysis are also used in this model. We pick 2 parameters each which are strongly associated with influence. Additionally, we search for and choose some important events to better explain the influences we conclude.

Then we reevaluate our 4 models and summarize major strengths and weaknesses of our models, which is meaningful for us to improve them. At last, we form a report which reviews what we have got from 4 models and what are worth using for reference and recommendation for further study.

Keywords: Quantification of musical evolution, social network analysis, simulated annealing

Contents

1 Introduction

1.1 Problem background

Music is an integral part of human society as Andrews reported in 2018 that the history of music is as old as human itself. It is valuable to study and analyze the changing process and its influence on society over centuries of this subject, which may be helpful when concluding the past and forecasting the future trend. Although it is universally acknowledged that music is part of the literal arts that cannot be described by number, quantifying music still reveals its great value especially during the era of data. Focusing on artists, the innate ingenuity, acquired experiences, and instruments acquired of artists can all be quantified to make an evaluation of artists.

Quantifying the music statistics may contribute to understanding the evolution of music. To be specific, comparing the data of music feature over the time may help us find the major shift of music trend, current most popular music genres, and the generation of some types of music. Moreover, complex influential relationship between genres can be discovered.

With regards to practical effect, the evolution of music not only exerts impact on music itself. Latent social, economic, and political influence may be found during the research, which is also conducive for us to understand the society.

1.2 Project breakdown

To understand the evolution of music by evaluating the influence relation, features comparison and time series analysis, several research and analysis is done according to ICM suggestions.

1. Create the directed networks to illustrate the influence between and within genres and find reasonable indicators.
2. Find the similarity between artists' work and further compare it with original genres. Set up standards to distinguish different genres.
3. Analyze the change of music over time and find the major leaps. Further research the mutual influence between society and music.
4. Write an report to the ICM Society to present what we have found on this project and our edge. Demonstrate the potential more indicating factors of music change and give advise on further research.

2 Task I: Exploration of Musical Influence

2.1 Assumptions

- The single influence of each influencer on its follower is the same regardless of who the artists are. In other words, the weight of single influence between artists is the same.
- Time factor does not change the musical influence between artists. That is, the musical influence between artists under different ages is the same.

2.2 Notations

The primary notations used in this paper are listed in Table 2.

Table 1: Notations

Symbol	Definition
λ_{ij}	The out degree from genre i to genre j
n_{ij}	the number of artists in genre i influencing the artists in genre j
N	The total artists number of this genre
w	The weight parameter
R	The comprehensive score of music influence
k	PageRank parameter
d	PageRank parameter

2.3 Preliminary work

Faced with the trouble fitting the large amount of statistics into the networks models, we think it is reasonable to divide the influence into that within genres and that between genres. Before specific operation, we exclude faulty data via comparing the same influencer name and follower name.

2.4 Local analysis

To analyze the network within genres, we first picked out the main 4 influential genres by calculating the total influence number. (See in Figure 1)

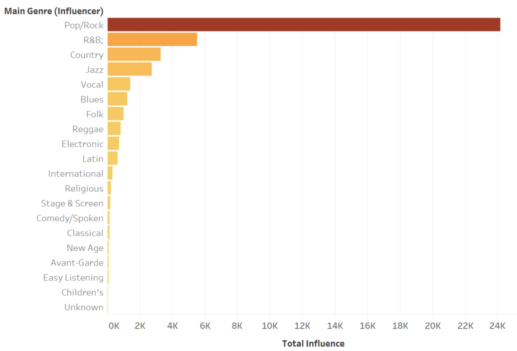
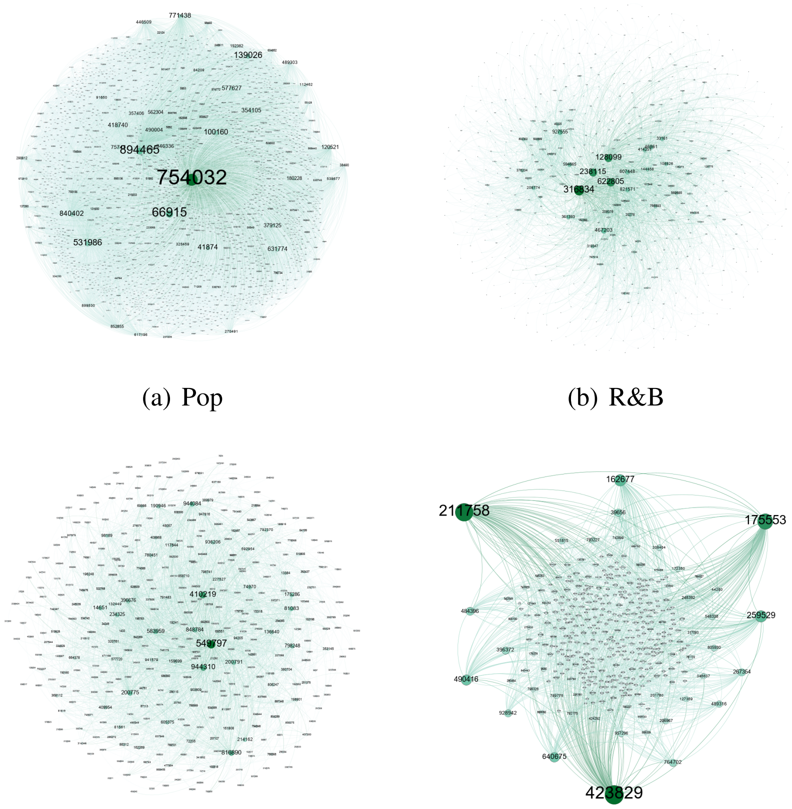


Figure 1: Genre Influence

It can thus be seen that Pop /Rock, R&B, Country and Jazz are the main genres with regards to influencer number. We may assume that the networks of these genres can be representative and others may have the similar directed networks.

Applying Gephi to visualize the data, the respective networks of Pop/Rock, R&B, Country, and Jazz are shown as below (Figure 2).

The density of line round a certain node reveals the out degree, namely the number of followers of this influencer, which can denote the musical influence we want to explore. It can be



(a) Pop (b) R&B

(c) Country (d) Jazz

Figure 2: Comparisons

seen that among the 4 networks, there all exist several “Big nodes” with relatively higher out degree. We denote these artists “Big wheels” who may be representative of the genre in terms of influence.

It is therefore safe to conclude that the pattern of musical influence is similar to the networks above. We list top 3 influencers with statistics, see in Figure 3.

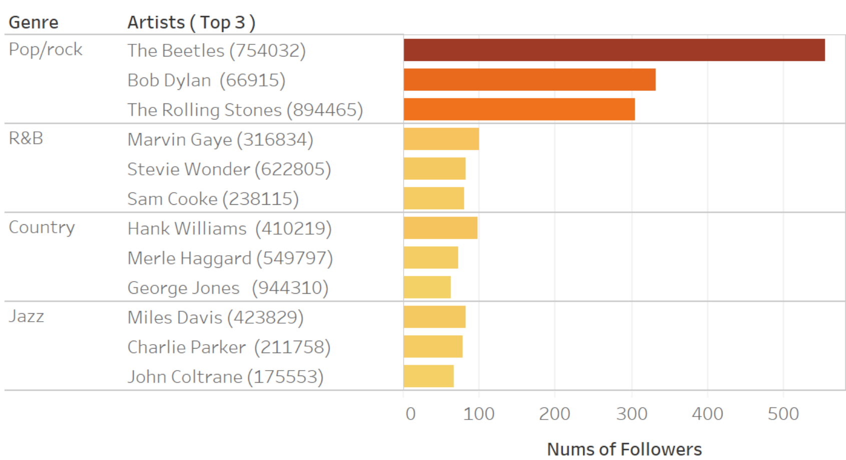


Figure 3: Top 3 influencers

2.5 Global analysis

2.5.1 Data choosing

Analysing Figure 1, it can be found that the influence number of the last 10 genres concentrate to a relative low volume. Hence, we choose the statistics of first 10 genres and a total statics of the last 10 genres to research to form a data set with 11 elements.

No.	1	2	3	4	5	6	7	8	9	10	11
Genre	Pop	R&B	Country	Jazz	Vocal	Blues	Folk	Reggae	Electronic	Latin	Others

2.5.2 Model establishment by PageRank

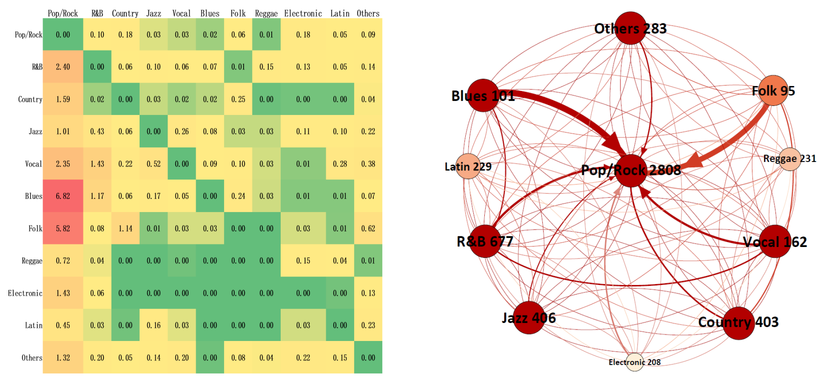
To simplify and quantify the influence number between different genres, we denote λ_{ij} as the out degree from genre i to genre j, which can be seen equivalent to the vote for i from j. ($\lambda = 0$ when $i = j$)

$$\lambda_{ij} = \frac{\sum_{p \in i} \sum_{q \in j} n_{pq}}{N_i}$$

Where n_{pq} is the number of artists in genre i influencing the artists in genre j, and N_i is the total artists number of genre i. We have

$$\lambda_j = \sum_{i=1}^{11} \lambda_{ij}, \quad j = 1, 2, 3, \ldots, 11$$

The adjacent matrix and directed network shown in the Figure 4 demonstrate the influence relation between genres to some extent. The deeper the color is, the more effect the node exerts on others. However, due to the difference in the sample quantity, error is unavoidable. We hence adopt PageRank algorithm to make a better comment system.



(a) Adjacent Matrix (b) Directed network

Figure 4: General influence

PageRank Processing: We denote w_{ij} as the weight of genre j to genre i, which can be seen equivalent to the proportion genre j contributes to genre i.

$$w_{ij} = \frac{\lambda_{ij}}{\lambda_j}$$

Therefore, we have matrix W containing all weight from each genre.

$$W = \begin{bmatrix} w_{11} & w_{12} & w_{13} & \cdots & w_{111} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{211} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{111} & w_{112} & w_{113} & \cdots & w_{1111} \end{bmatrix}$$

We define R as the comprehensive score of a certain genre to evaluate the musical influence.

$$R_i = \sum_{j=1}^{11} w_{ij} R_j, \quad i = 1, 2, \ldots, 11$$

Solving the equation $R = kWR + d$, where

$$d = [\frac{1-k}{N}, \frac{1-k}{N}, \frac{1-k}{N}, \ldots, \frac{1-k}{N}]^T, d \in \mathbf{R}_{11},$$

$$k = .85.$$

We have $R = [0.1915, 0.1406, 0.1210, 0.1150, 0.1143, 0.0926, 0.0669, 0.0569, 0.0525, 0.0249, 0.0238]^T$, and the dot plot is shown in Figure 5.

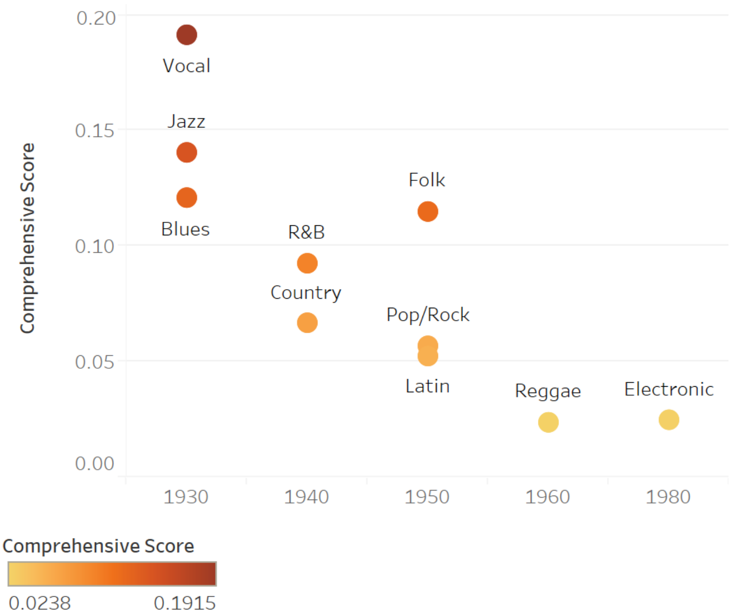


Figure 5: Comprehensive scores

According to the plot which is similar to negative exponential trend, we can conclude that the music genres generated earlier may have a greater influence among all genres.

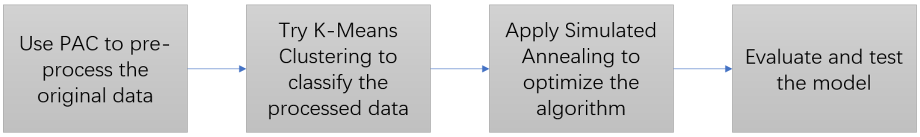
Summary After drawing the directed networks the music influence within the genre, it is safe to conclude that the “Big wheel” plays an important role in influencing others when we denote the out degree as the musical influence parameter. When it comes to the networks between genres, the results turn out that the influence degree is strongly positively related to the years the influencer has been generated.

3 Task II: Measures of Music Similarity

3.1 Overview

In order to find out the similarity between music. The basic process of this section is to classify the original feature statistics into different groups. We therefore focus on data pre-processing, finding clustering algorithms and optimizing it if necessary.

Therefore, we can compare the outcome of the model and the initial genre classification to evaluate the accuracy of the model.



3.2 Assumptions

- Assume that Year factor does not affect characteristic of different music genres. To be more specific, Year factor will not be regarded as a parameter in PCA process.
- In PCA process, if the cumulative contribution rate is higher or equal to 70%, then the principal components are valid to measure and describe the characteristic of a certain song.

3.3 Notations

The primary notations used in this paper are listed in Table 2.

Table 2: Notations

Symbol	Definition
x	The corresponding coefficient of PCA result
C	The cluster result set
C_e	The set of centroid of clustering set
T_0	The initial temperature in SA
T_f	The final temperature in SA
L	The iteration times
d_{pq}	The euclidean distance between x_p and x_q

3.4 Data pre-process

3.4.1 Standardization

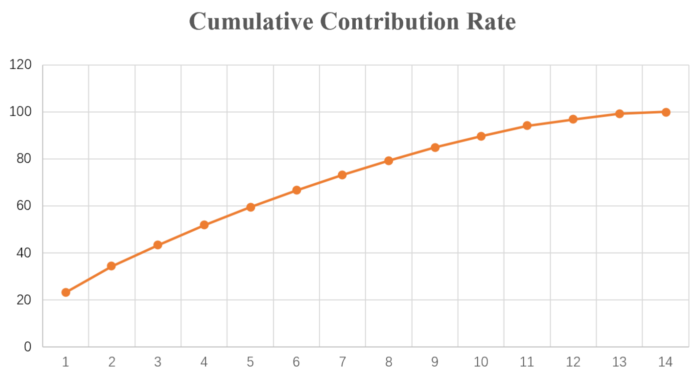
Due to type difference of the data provided, we need to first process the data set to further operate. Since all the data is positive, we use 0-1 Standardization which is:

$$b_{ij} = \frac{a_{ij} - a_j^{min}}{a_j^{max} - a_j^{min}}$$

Where a_{ij} is the ith data of x_j in its corresponding data sample S_i and let $a_j^{max} = max\{S_i\}$; $a_j^{min} = min\{S_i\}$; b_{ij} is the value of a_{ij} after normalization.

3.4.2 Multivariate analysis by PCA

To abate the dimension of original data set, we first remove the data of year for a separate study of time series, and then we apply PCA (Principal Component Analysis). Via MATLAB, the cumulative contribution rate is shown as below (Figure 6):



Numbers of Principal Components	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Cumulative Contribution Rate	23.23	34.31	43.29	51.75	59.47	66.54	73.23	79.16	84.84	89.73	94.13	96.79	99.14	100.00

Figure 6: Cumulative Contribution Rate

The Figure 6 reveals that when the number of principal reach 7, the cumulative contribution rate exceeds 70%. Hence we denote $x \in S$(sample space)$, x = [z_1, z_2, \ldots, z_8]^T$ and the corresponding coefficient matrix is shown as below (Table 3).

Table 3: PCA result

		z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8
danceability	x_{01}	0.22	0.54	0.05	-0.30	0.05	0.28	0.04	0.00
energy	x_{02}	0.48	-0.15	0.01	0.10	-0.14	0.08	0.02	0.00
valence	x_{03}	0.27	0.53	0.14	0.03	-0.27	0.20	0.15	0.00
tempo	x_{04}	0.17	-0.11	0.04	0.37	-0.34	-0.36	0.54	0.00
loudness	x_{05}	0.47	-0.08	-0.06	0.09	0.01	-0.03	-0.11	0.00
mode	x_{06}	-0.02	0.12	-0.01	0.59	0.36	-0.02	-0.21	0.00
key	x_{07}	0.03	-0.02	0.06	-0.39	-0.45	-0.52	-0.43	0.00
acousticness	x_{08}	-0.45	0.18	0.17	-0.01	0.05	-0.08	0.00	0.00
instrumentalness	x_{09}	-0.25	-0.19	-0.07	-0.10	-0.24	0.20	0.45	0.00
liveness	x_{10}	0.05	-0.27	0.54	0.23	-0.17	0.26	-0.36	0.00
speechiness	x_{11}	0.08	-0.10	0.68	-0.15	0.11	0.03	0.15	0.00
explicit	x_{12}	0.14	-0.20	0.24	-0.33	0.50	-0.21	0.30	0.00
duration_ms	x_{13}	-0.01	-0.40	-0.19	-0.18	-0.14	0.56	-0.05	0.00
popularity	x_{14}	0.33	-0.15	-0.29	-0.16	0.28	-0.08	-0.06	0.00
year	t	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

3.5 Model establishment by K-Means and SA

3.5.1 Clustering analysis

To find similarity, namely to classify the processed statistics, we want to group the initial statistics into several groups based on the euclidean distance d between each point. Such group should be relatively distant from each other to make the classification clear.

We thus adopt K-means clustering to primary grouping. Based on the research in Task I, setting the number of groups k to be 4 is reasonable when we assume that there should be 4 mainstream music genres.

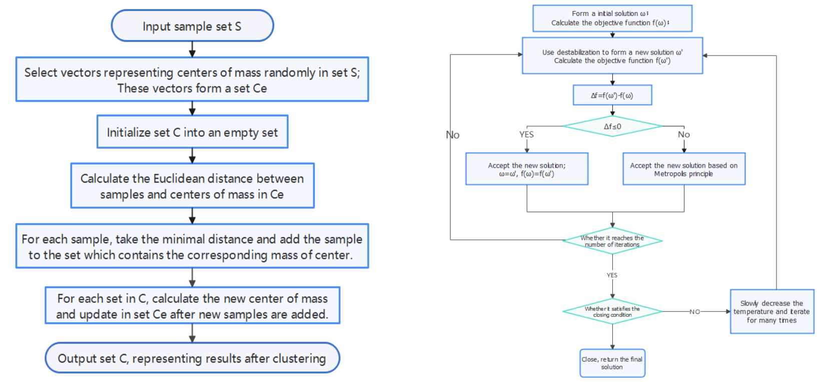
Hence we denote set $C = \{C_1, C_2, C_3, C_4\}$ to be the cluster result set and $C_e = \{\mu_1, \mu_2, \mu_3, \mu_4\}$ to be the set of centroid of clustering set, where

$$\mu_i = \frac{\sum_{x \in C_i} x}{|C_i|}, \quad i = 1, 2, 3, 4$$

This algorithm has it drawback in that it only operate single time clustering, which is likely to be restricted to local optimization instead of the global.

3.5.2 Model optimization

We thus use SA (Simulated Annealing) to perform multiple inner and outer loops in order to obtain the global optimization. K-means is treated as a disturbance process to continually update the results. The flow diagram is shown as below (Figure 7):



(a) K-means as disturbance process (b) SA process

Figure 7: Overall optimization process

Simulated Annealing Processing: The annealing object function is

$$f(C, C_e) = MSE = \sum_{k=1}^{k} \sum_{x \in C_i} |x - \mu_i|^2$$

We set the initial temperature T_0 to be 1000 and final temperature $T_f = 1$. According to , to ensure bigger searching space, the attenuation function and parameter are

$$T_{i+1} = \alpha T_i, \quad \alpha = 0.7$$

For the inner loop Markov chains, we set iteration $L_k = 1000$.

Disturbance Processing: The object of this process is to create a new solution and determine whether to update it, where

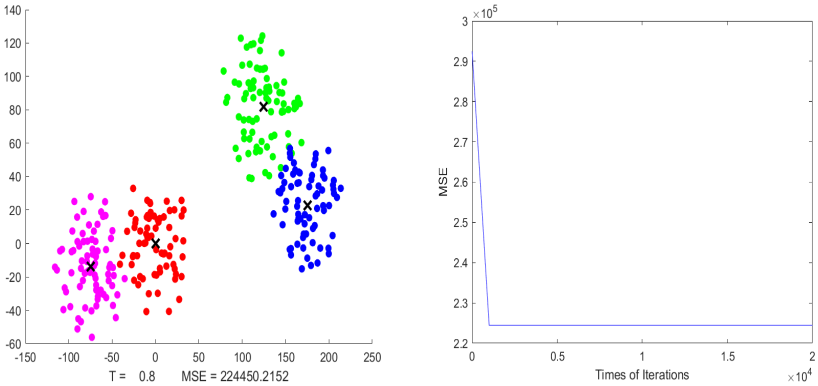
$$C_{new} := C_{old}, \qquad C_{enew} := C_{eold}$$

And the flow is shown in Figure 7(a).

3.6 Model application

3.6.1 SA result

To test and evaluate the accuracy of model of genres classification, we should compare the model result and original result. Simply use simple random sampling method to pick 300 work as a sample.



(a) Clustering diagram (b) MSE convergence

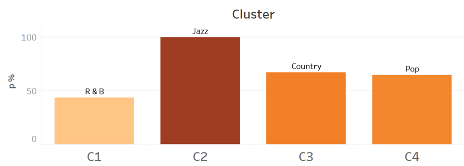
Figure 8: SA result

From Figure 8 (a), it is clear that 4 clusters C are produced by the model. It is valid because the MSE is quickly convergent to a relatively low point when times of iterations are big enough.

3.6.2 Classification standard

- For a certain C_i, if there is a obtained genre G_k whose proportion is far larger than those of any others', we denote this cluster to be similar to genre G_k.
- For more than one cluster $C_i, C_j, \ldots$, if they have the same genre G_k to be the largest proportion, compare the number of their genres in different clusters and denote the cluster containing the most to be similar to genre G_k. Operate the others in same procedure.
- **Expectation:** 4 clusters obtained have a valid one-to-one relation to the genres manually chose.

Cluster	Pop %	R&B %	Country %	Jazz %	Nums of Sample
C1	60.87	39.13	0.00	0.00	69
C2	0.00	0.00	25.00	75.00	80
C3	0.00	45.21	54.79	0.00	73
C4	100.00	0.00	0.00	0.00	78



(a) Result diagram (b) Accuracy bar-chart

Figure 9: Model classification

The Figure 9 shows the result of comparison of the model and original genre where p_k in (b) represents the proportion of sample in clusters that can match the original genre. The result shows that 120 samples can match their original genres so that the total accuracy $p = 40\%$.

3.6.3 Definition of measure of similarity

We define the similarity between x_p and x_q in S to be the euclidean distance $d(x_p, x_q) = d_{pq}$.

We aim to find a range or an interval I, such that if $d_{pq} \in I$, then x_p and x_q are similar. If $d_{pq} \notin I$, then x_p and x_q are not similar. Furthermore, if $d(x_p, \mu_k) \in I$, then $x_p \in C_k$.

1. Use the standard in 3.6.1 to find the one-to-one map.

2. Calculate the average distance $\overline{d}$ between samples in the same cluster

$$\overline{d} = \frac{\sum_{i=1}^{4}\sum_{d\in L_i} d}{\sum_{i=1}^{4}|L_i|}$$

where $L_i = \{d(x_p,x_q)|x_p,x_q \in C_i\}$ is the set of all distance between samples in C_i and $L = \bigcup_{i=1}^{4} L_i$.

3. Since $|L| > 50$, we may treat the sample distribution of $\overline{d}$ to be normal. Hence the 99% confidence (judge) interval and relative parameters are as below (Table 4):

MeanCI (99%)	34.04	34.97
SDCI (99%)	18.09	18.75
Judge Interval (I)	0.00	53.71
Other parameters	Mean = 34.5	d_{max} = 53.71

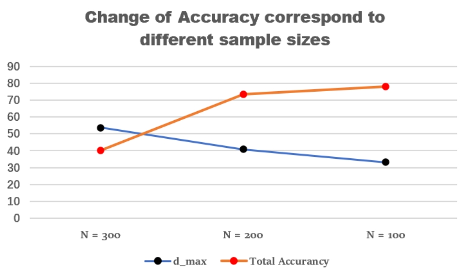
Table 4

3.7 Sensitivity analysis

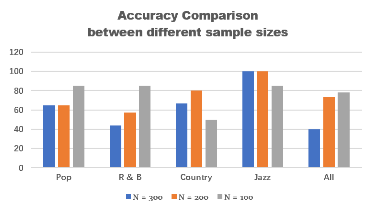
Due to the randomness when we choose the sample size of music work, it is necessary to test the change of result by adjusting the sample size. We have retested the model with sample size 100 and 200,and the results are shown as below:

Sample Size	d_max	Total Accuracy
300	53.71	40
200	40.9	73.5
100	33.19	78

(a) Change of d_{max}



(b) Change of accuracy



(c) Accuracy comparison

Figure 10: Sensitivity analysis

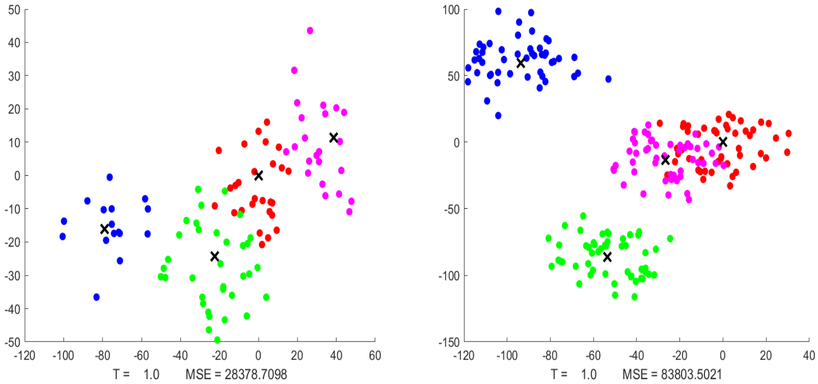
According to Figure 10 we have 3 conclusions about the sensitivity analysis:

- The change of d_{max} is sensitively related with the change of sample size. 26% rise is observed when we increase the sample size by 100.

- The accuracy tends to decrease with the increment of sample. Larger sample size to analyze the sensitivity is necessary for the discussion on a larger scale.

- The sharp decrease of accuracy from sample size 200 to 300 may be related with the initial centroid μ_k in K-means. Further research can focus the initial centroid of "Big wheels" chosen in Task I.

3.8 Relation and difference between genres



(a) Clustering with size 100 (b) Clusering with size 200

Figure 11: Clustering with different sizes

3.8.1 Relation among artists within genres and between genres

The intersection in Figure 11 reveals that similarity may appear between artists from different genres. However, since the probability for intersection is relatively low, we may conclude that the similarity within genre is relatively higher than that between genres.

$$d(x_p, x_q) > d(x_p, x_s), \quad x_p, x_q \in C_i(G_{k_1}), \quad x_s \in C_j(G_{k_2})$$

3.8.2 Similarities and influences between and within genres

The results of clustering show the likelihood that multiple clusters may have intersection. That is to say, the distance between centroids of cluster is small.

This to some extent reveals the similarity between genres, which can be shown in Figure 11(b) that the genres Country and R&B have high similarity.

From above, certain different genres can be speculated to have some mutual effect.

Remark Cluster cannot completely represent the genre, thus the intersection we analyze can only give a fuzzy relation between genres.

4 Task III: The Revolution of Music over time and Mutual Influence

4.1 Assumptions

- The data provided includes the reasonable numbers (the sample size is big enough) of the musical work for processing.

4.2 Revolutions identification by Time Series Analysis

4.2.1 Indicators setting

In this section, we consider the music as a whole to analyze the change of music over time. In order to establish the evaluation indicator, we consider to analyze the total number of music publication and popularity. Because the number indicates the vitality of music creation of a certain year and the popularity indicates the general impact of music on the public. Two sides parameters give a complete description of the music circumstance.

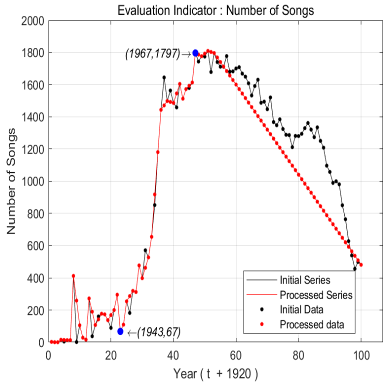
4.2.2 Turning point evaluation

Since the data set is distributed along with the years, we consider to analyze the statistics in terms of time series.

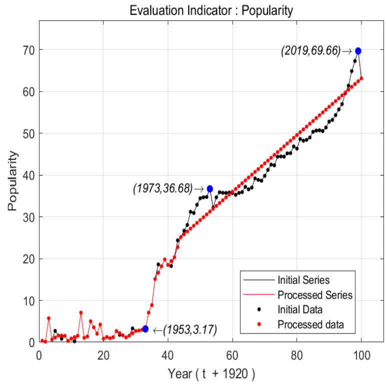
We define the turning point (leap) of time series data with two characteristics:

- For a certain point, there exists a notable difference of slope between the point and the points of few years nearby.
- The point has a lasting influence for the following years, i.e., the slope of the data will stay stable for a relatively long time.

To search and identify the revolutions of music, we only need to find the turning point of the time series analysis. We further apply the interpolation method and FTTO algorithm by Xing, et al. in 2018 to find the turning points. The plots are drawn as below (Figure 12), where processed series lines are denoise processed to show the potential trend and the initial series lines are the plot of original data.



(a) Number of songs



(b) Popularity

Figure 12: Indicators evaluation

After compare and analyze two lines respectively, we find that:
With regards to number of songs, 1943 and 1967 turn out to be the turning points.
With regards to popularity, 1953, 1973 and 2019 turn out to be the turning points.

Since the revolution should last for a period of time and impact on every indicating characteristics, it is safe to assume that the major leaps we want to find are just around the interval between the turning points we have found.
i.e. $(1943, 1953)$; $(1967, 1973)$ and 2019.

In terms of other characteristics, operate the same procedure. We can determine whether it can signify the major leap by comparing the composition location of turning points from every characteristic. The expected composition is shown as below:

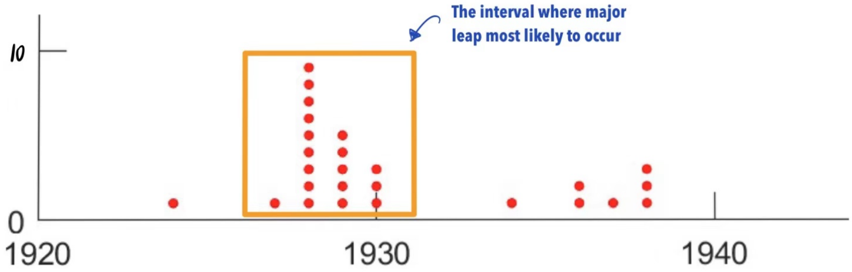


Figure 13: Composition location

4.3 Various influence on music

Based on the revolution time we assessed in 4.3, we search and collect background information at that time to find the reason behind. Analysis in 4.2.2 reveals 3 major leaps in music history, which may result from the changes in factors like society, politics or technology. As a result, in this model, we try to find the influence on music by searching for the related global events or issues during the period of major leaps.

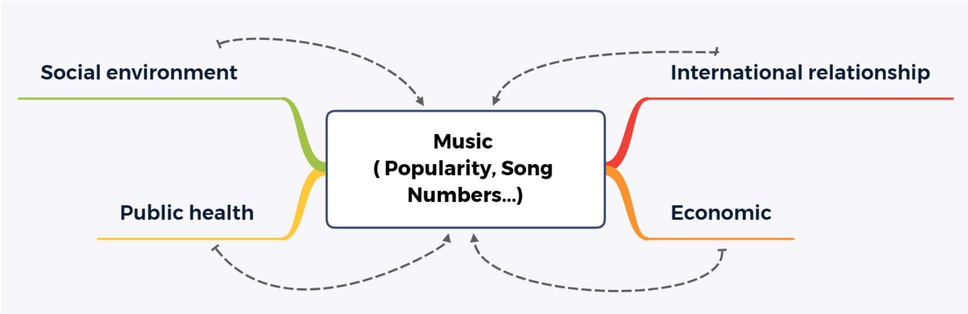


Figure 14: Influence on music

- **During 1943-1953 (major leap 1):** In 1945, The World War 2 ended; in 1945, the United Nations was set up; In 1947 and 1948, the US proposed The Truman Doctrine and the Marshall Plan respectively.
- **During 1950-1970 (between major leap 1 and 2):** This period of time is called the "The golden age of capitalist economic".

These events show that the peaceful social environment, cooperative international relationship as well as good economic environment have positive impact on music, promoting the popularity and increasing number of songs.

- **During 1967-1973 (major leap 2):** Around 1973, Eastern Europe was in upheaval.
- **After 1970s:** The "Information Revolution" began, information technology developed a lot.

After major leap 2, the number of songs had a decreasing trend. The turbulent international situation was supposed to have negative in music with this regard; However, the popularity of music is still increasing, though the speed is slowing down. This shows that the quality of music is better, which might be affected by Information Revolution.

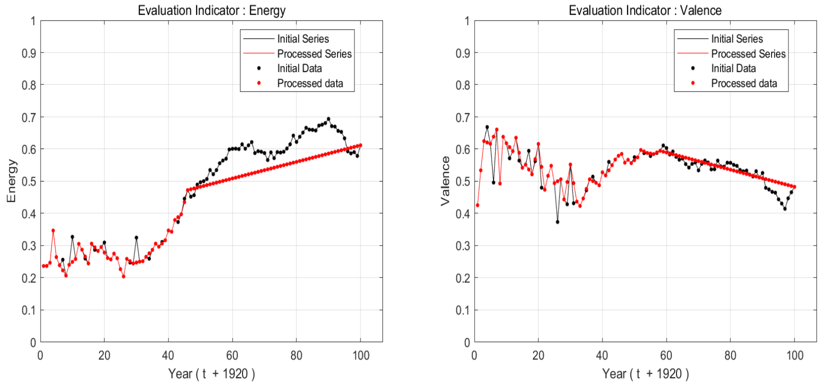
- **In 2019 (major leap 3):** The existence and spread of Covid-19 greatly influenced the whole world.

The poor public health environment harms the music industry. Both popularity and number of songs have a fair number of decline.

4.4 Musical influence on the society

After analyzing the social influence on music, it is natural and reasonable to try to find the alternative influence of music on the society.

Based on limited data we have about the music, we choose characteristics of music "Energy" and "Valence" to describe the perception of intensity and activity from the public and the positiveness public hold at the certain time.



(a) Energy (b) Valence

Figure 15: Musical influence time series evaluation

From Figure 15, we can describe the pattern as follows: Over the past 100 years, the overall trend of energy is on the rise. To be more specific, from 1920 to about 1963, the rate of growth is increasing and from 1963, the rate become stable and energy of music still increases with relatively low speed. This trend means songs become more and more energetic over the time, which indicates that music might have the impact of making people and the whole society energetic and innovative.

Similarly, in the first 45-50 years, the valence shows a similar pattern comparing with energy of music. However, from about 1970, valence of music reveals a slow and uniform descent, which means people have the preference of emotionally negative music. The impact of that might make the whole society less happy and motivated. This kind of influence needs our introspection of musical preference.

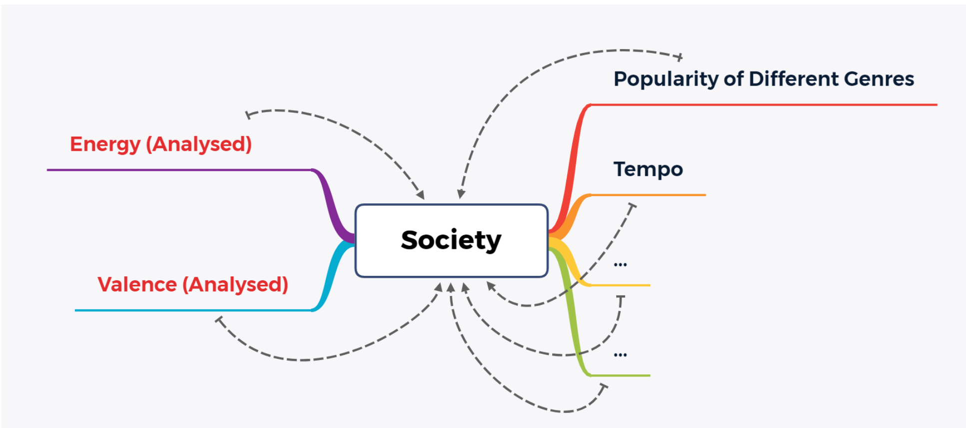


Figure 16: Musical influence

5 Strengths and Weaknesses

5.1 Strengths

- The dimension reduction of data help to build a more visualized network. The classification of directed network is reasonable and describe the relation both within and between genres with clear algorithm.

- We avoid the drawback of traditional clustering method when finding the similarity of music and give a global analysis.
- Mutual influence between music and society is analyzed and concluded in the research.

5.2 Weaknesses

- The research targeted at individual musical genre needs further research.
- The dimension reduction by PCA makes the model unable to specifically demonstrates the influence by detailed factors.
- The sample choice will make a great impact on the clustering result accuracy of the model. Use stratified sampling method to choose representative people may help to reduce the error.
- The provided data may not include the all of the musical work, which may lead to the error in terms of publication number analysis.

One-page document

To: ICM Society
From: Team 2103274
Date: Feb 8th, 2021
Subject: Report on music evolution

The classification of networks enables the researcher to analyze the respective networks within genres and between genres. From a local or an individual view, this model could be helpful to find out the vital roles among the genres. From a global view, the model provides quantified relationship between different genres which may be conducive to analyze the origin of various genres of music. Moreover, the result of the model reveals a likelihood that the influence is positively related to the years the influencer has been generated instead of the quantity of artists in the genre.

The sample data provided may not cover the total work during the research period which may lead to error in networks setting and time series analysis. The limited items of characteristics may hinder the assessment about the musical influence on the society. Richer data could bring more kinds of impact the music exerts on the society such as economic and politics.

The data by year is relatively complete. However, there is still more space to research based on more detailed data. For example, the basic information of listener may contribute to the analysis of the music audience.

When it comes to the recommendation for further study, our suggestions are that:

- Operate more turning point evaluation to conclude a summary on major leap of music.
- Take the authority of Big Wheel into consideration to make a more through research on the artists' network.
- Focus the initial centroid of "Big wheels" chosen in Task I to operate the SA model.

Based on the further study of music, we may be able to forecast the music trend based on current society circumstance. Relative workers and artists can devote more to the analyzed music trend to cater for the society need.

References

[1] Brusco, Michael J., and Hans-Friedrich Köhn. “Exemplar-Based Clustering via Simulated Annealing.” Psychometrika, vol. 74, no. 3, 17 Mar. 2009, pp. 457–475, 10.1007/s11336-009-9115-2. Accessed 7 Feb. 2021.

[2] Humenberger, H. “How Does Google Come to a Ranked List?–Making Visible the Mathematics of Modern Society.” Teaching Mathematics and Its Applications, vol. 30, no. 3, 12 May 2011, pp. 107–119, 10.1093/teamat/hrr007.

[3] Andrews, E. (2018). What is the oldest known piece of music? [online] HISTORY. Available at: https://www.history.com/news/what-is-the-oldest-known-piece-of-music.

[4] Pankavich, Stephen, and Rebecca Swanson. “Principal Component Analysis: Resources for an Essential Application of Linear Algebra.” PRIMUS, vol. 25, no. 5, 20 Dec. 2014, pp. 400–420, 10.1080/10511970.2014.993446. Accessed 2 Dec. 2020.

[5] XING Han, SHI Xiaoda, SUN Lianying, et a1. Trend Turning Point Extraction Algorithm for Time Series Data[J]. Computer Engineering, 2018, 44(1): 56-61, 68.

Appendix: Program Codes

Here are the program codes we used in our research.

PCA.m

```
% PCA_main
clc,clear;
data = xlsread('full_data.csv');
[m,n] = size(data);
r = corrcoef(data);
[x,lamda,p] = pcacov(r);
totalctb = sum(lamda);
cumctb = cumsum((lamda/totalctb)*100);
k = 7;
z = zeros(m,k);
for i = 1:k
coe = ones(m,n);
for j = 1:14
coe(:,j) = x(j,i);
end
z(:,i) = sum(coe.*data,2);
end
```

SA.m

```
% SA_main
clc; clear;
% load Data
Data = xlsread("DATA3.xlsx");
Data = Data';
global sampleSpace; global n; global centers; global C;
[m,n] = size(Data);
sampleSpace = cell(1,n);
for i = 1:n
sampleSpace{i} = Data(:,i);
end
% plotGraph,sethandles
plotGraph();

% initialize centers and C
k = 4; centers = cell(1,k); C = zeros(1,n);
choose = ceil(rand(1,4)*n);
for i = 1:k
centers{i} = sampleSpace{choose(i)};
end
[C, centers] = change(C,centers);

% initialize parameter
T = 1000; % Starting T
Tf = 1; % End TF
a = 0.7; % Decline ratio a
Lk = 1000; % Length of Markov c h a i n n is the scale of the problem
```

```
% initialize iteration times
iteration = 0;
iterationplot = 0;
final_MSE = MSE(C,centers);
iterationTimes = [iteration];
MSEChange = [final_MSE];

% set initial plot
[~,~,~,~] = update(centers,C,T);

% Simulated annealing start
while T>Tf
for i = 1:Lk
[newC, newcenters] = change(C,centers); % change the solution
new_MSE = MSE(C,centers); % calculate the new MSE
delta = new_MSE - final_MSE; % calculate the difference
% Metropolis
if(delta<0)||(rand < exp(-delta/(T)))
C = newC;
centers = newcenters;
final_MSE = new_MSE;
end
iteration = iteration + 1;
iterationplot = iterationplot + 1;
if iterationplot == 20
[~,~,~,~] = update(centers,C,T);
iterationplot = 0;
end
end
iterationTimes = [iterationTimes;iteration];
MSEChange = [MSEChange;final_MSE];
T = a*T; % cooling
end

% set final plot
[X,Y,Xp,Yp]= update(centers,C,T);
hr = plot(iterationTimes,MSEChange,"b-");
xlabel("Times of Iterations");
ylabel("MSE");
\begin{lstlisting}[language=MATLAB, name={FTTO.m}]
% FFTO_main
Data = xlsread(" d a t a _ b y _ y e a r valence .csv");
global X; global Xc; global se; global fi;
Q = 9;
X = Data(:,2);
n = length(X);
Xc = zeros(n,1);
Xc(1) = X(1);
fi = 1;
se = 1;
for i = 1:(n-4)
P = (X(i+2)-X(i))*(X(i+3)-X(i+1));
FTTO(P,X(i),X(i+1),X(i+2),X(i+3),"global");
```

```
if (se-fi)>Q
FTTO_G(fi,se);
fi = se;
end
end
Xc(n) = 1;
```

```
% FTTO_function

function FTTO(P,x0,x1,x2,x3,type)
global Xc; global se; global Gc; global sl;
switch type
case "global"
if P<0
if x3-x1<0
if x1>=x2
se = se+1; Xc(se) = 1;
elseif x1<x2
se = se+2; Xc(se) = 1;
end
elseif x3-x1>0
if x1>x2
se = se+2; Xc(se) = 1;
elseif x1<= x2
se = se+1; Xc(se) = 1;
end
end
elseif P==0
if ((x3-x1)~=0)&((x2-x0)==0)
se = se+2; Xc(se) = 1;
elseif ((x3-x0)==0)&((x2-x0)~=0)
se = se+1; Xc(se) = 1;
end
end
case "local"
if P<0
if x3-x1<0
if x1>=x2
sl = sl+1; Gc(sl) = 1;
elseif x1<x2
sl = sl+2; Gc(sl) = 1;
end
elseif x3-x1>0
if x1>x2
sl = sl+2; Gc(sl) = 1;
elseif x1<= x2
sl = sl+1; Gc(sl) = 1;
end
end
```

```
elseif P==0
if ((x3-x1)~=0)&((x2-x0)==0)
sl = sl+2; Gc(sl) = 1;
elseif ((x3-x0)==0)&((x2-x0)~=0)
sl = sl+1; Gc(sl) = 1;
end
end
end
end

% FTTO_G
function FTTO_G(fi,se)
global X; global sl; global Gc; global Xc;
G = [];
x = [fi;se]; y = [X(fi);X(se)];
for w = fi:se
pw = [w,X(w)];
G = [G;distance3(x,y,pw)];
end
n = length(G);
Gc = zeros(n,1);
Gc(1) = G(1);
sl = 1;
for j = 1:(length(G)-4)
P = (G(j+2)-G(j))*(G(j+3)-G(j+1));
FTTO(P,G(j),G(j+1),G(j+2),G(j+3),"local");
end
Tc = Gc + fi;
Xc(Tc) = 1;
end
```