Problem Chosen	2021	Team Control Number
D	MCM/ICM Summary Sheet	**2107959**

Heroes Make Times? Intrinsic Mechanism of Music Revolution

Summary

Studying intrinsic mechanism of the evolution of music is very important for the historical research and future development of music. The aim of our team is to model the intrinsic mechanism of music evolution by using a wide range of analyzing methods, including mathematics, statistics, and synthesis.

After a series of studies, we have constructed a logic model of music revolution with solid quantitative and qualitative support. The course can be divided into three steps following.

Build an influence topology network system. Our group develops a high-performance system based on several algorithms like **GN Algorithm**, **BFS Algorithm/DFS Algorithm** and **Branch-and-Bound**. We creatively define an indicator which is the centrality of the class feature vector to gauge the magnitude and direction of the influence of each musician and even the followers. Feeding the data to the system, our topology network system can automatically analyze the directionality and intensity of connections between any of the musicians. With the accordance to our model, T-bone Walker is one of the most influential musicians, having 48 followers with great influence. This fits the fact, since T-Bone Walker is the well-known godfather of Blues.

Explore similarities and differences of music to find intrinsic mechanism of music revolution. We first explore the similarity of songs. Using **Hierarchical Clustering Algorithm**, we measure the similarity between songs through various characteristics. Next, we use "faction" as auxiliary information to perform hierarchical clustering again. We not only prove that existing classification is statistically significant, but also clarify the closeness between factions. For example, Blues and R&B belong to one cluster, meaning they share more similarity than with other genres. It is true because R&B originates in Blues. Through the analysis of existing data and related theory, we discover that a music revolution means drastic changes in various features of music and the rise and fall of music genres. Blues prospered in 1950s, resulting changes in all characteristics of music, some of which remain still until now. To sum up, a single musician can indeed influence other musicians via the topology network, and further affect the overall music style, thereby driving the music revolution.

So far, we have established a **logic model** of individual musicians stimulating great music changes. For the briefness of understanding and expanding universality of our model, we conduct a detailed case study using T-Bone Walker as an example. Finally, by further reviewing the related materials , we also conclude that music changes cause cultural changes through music transmission, which ultimately affects all aspects of society. In turn, changes in these aspects impact on individual musicians and their creating process, and then spur music revolution.

Our model demonstrates **the superiority of the "genre" classification** commonly used in the music industry from a mathematical standpoint and explains the process by which great musicians cause and promote music revolutions. Finally, we take a firm stand on comprehensively understanding the influence of "current situation" and "hero" on the evolution of music.

Keywords: Topology Network; Eigenvectory Centrality; Hierarchical Clustering; LDM

Contents

1 Introduction

With the development of society, people's demand for cultural satisfaction is increasing day by day.

Music, as a cultural product, is deeply loved by people, developing constantly. Changes in characteristics, styles and others aspects result in drastic revolutions in music. In combination with the historical background and given questions, we focus on revealing the veil of influence of musicians, music similarity and revolution.

Our team uses the graph theory and network analysis methods to build a visualized network and analysis system of musicians' influence network. Then quantify the influence of musicians. Clustering Algorithm is used to describe the similarity and differences of music, which define the music revolution. At the same time, we have clarified the logic chain for individual musicians to initiate music revolution. That is, through influencing the network, individual musicians make full use of their influence to alter certain musical features created by their followers, thus causing music revolution. Finally, by searching relevant thesis and essays, we confirm the connection between musical change and culture, further explore to what extent individual artists are affected by cultural background, turning the logic chain into a closed loop.

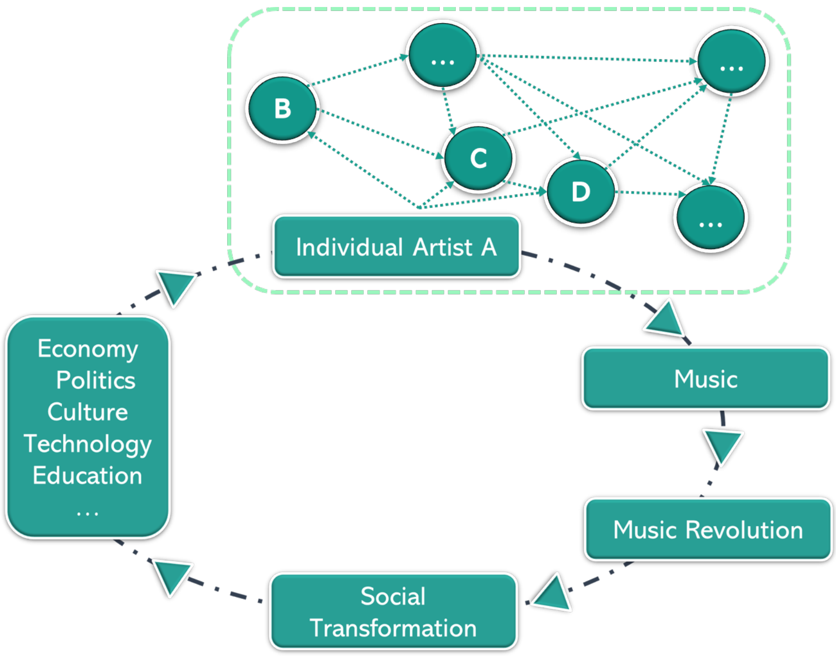


Figure 1: Logic Model

In this paper, we will elaborat the process of building the model, related concept and illustrate the operate mechanism of the model with vivid examples, extending its universality and profundity.

In the first part, firstly, a topology network of influence among musicians is established and the "influence" value of each artist between $[0, 100]$ is calculated by reflecting the relationship between artists in this network. Secondly, we first use the Hierarchical Clustering Algorithm to conduct hierarchical clustering analysis of music respectively. Then we adopt existing category as auxiliary information in further clustering analysis, proving that "faction" is indeed a good statistical classification standard, and musicians within genres are more similar than those belong to differentiated genres. After this section, we take the faction as the measurement of musical similarity, which means there are significant differences between factions. That is what makes some music prosperous, while some gradually come to a demise, forming a revolution.

In the second part, we choose a musician to perform a case study, and then explain the operation mechanism of the model. Our team selects a collection of musicians who show great influence and randomly picks T-Bone Walker. The significance level test and time series data analysis of the case clearly indicate that the musician, as a core reformer, lead to a musical revolution in the 1950s. We look at the musician's life experience and successfully found a complete agreement with our analysis. Taking this musician as an example, we have reason to believe that musicians with high "influence" value are most likely to be the core reformers of a certain musical change. Also, this prove that the "influence" value constructed by us is an indicator that can reveal the veil of dynamic influencers. This is very interesting, because such a conclusion obviously violates the materialist theory that "Time makes heroes". In the third part, we will try to give a reasonable explanation to it.

In the third part, we extend its thinking on the impact of musical changes on culture, and how these influences further affect individual musicians. Through related reference, this group believes that music revolution causes cultural revolution through the spread of music in the audience, and finally affects all aspects of politics and economy. Political, economic, and cultural changes affect individual musicians and their novelty, spurring musical change.

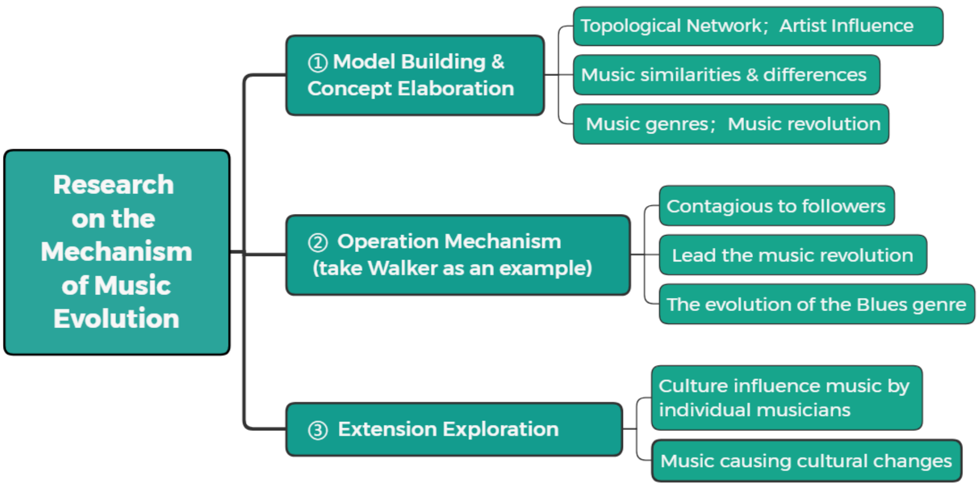


Figure 2: Framework of Paper

So far, we have established a logic model of individual musicians causing musical changes and explained the operating mechanism of the model using Walker and his genres as examples. The follow-up expounds the influence of music changes on culture and how these influences further affect individual musicians. Finally, we demonstrate the dynamics between great men and his/her time, showing our thoughts on historical materialism and idealism of music.

2 Assumptions and Symbols

2.1 Model Hypothesis

- Artists are not affected by artists with strong influence from genres not in the dataset.
- The 19 genres in the data set have complete information.
- The difference and similarity between music can be fully measured in 14 dimensions in the data set.
- There is no directed ring with more than 2 points in the topology network.

2.2 Symbols and Definitions

Table 1: Notations

Symbols	Description
$\mathbb{M}$	Music Space
$\mathbb{A}$	Artists Space
$\mathbb{N}$	Topology Network
$\mathbb{F}$	Follower Space
R	Music Revolution
S	Music Style
G	Music Genre
d	Minkowski Distance
I	Influence of Artists
B	Indicator to Reveal Revolution
$\vec{w}$	"Contagious" Ability Vector

3 Model Building and Concept Elaboration

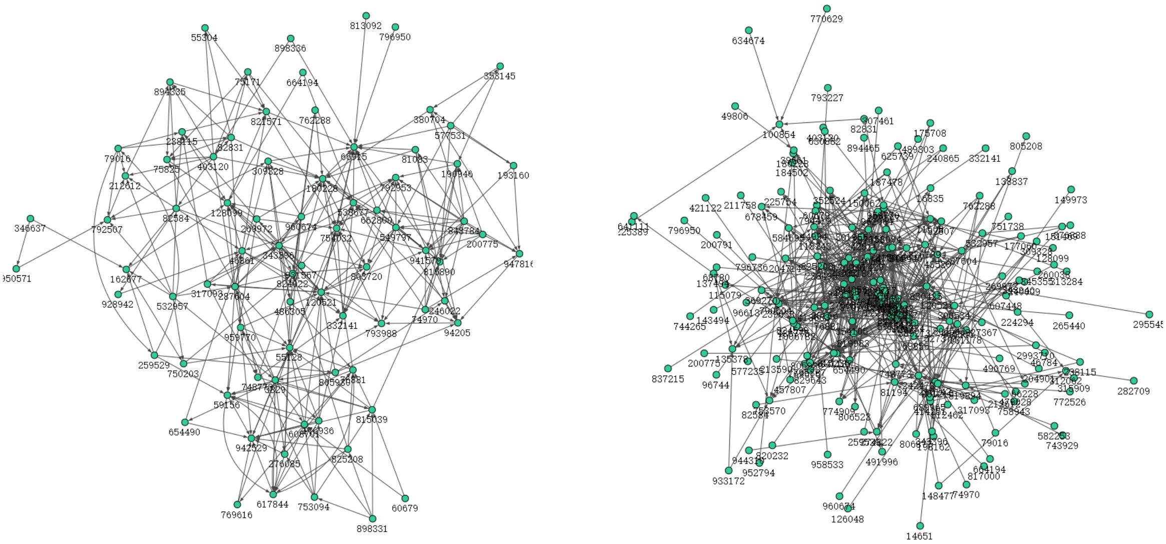
3.1 Develop Python Class of Topology Networks

In the *Influence_data* set, the influence relationship of musicians is given, on which an adjacency list of directed acyclic graph is generated. Combined with the relevant knowledge of community network analysis, we build a class that is convenient for network analysis and point importance analysis, realizing more desirable compatibility of directed graph and undirected graph in the class.

This classification implanted following functions, as visual network graph, adjacency matrix, output undirected graph point-degree centrality, density centrality, intermediate centrality, eigenvector centrality, simple community division.

In order to make subsequent analysis more practical, we realized the function of screening out the subnet from the clumsy original network by inputting some conditions, which includes the following two kinds:

- Input the influence, and the program will prune the points in the original network whose influence does not reach the standard and output the subnet after pruning.
- Input the genre, and the program will generate only the network within that genre.



(a) Network of Artists with Influence > 90

(b) Network of Artists in Blues

Figure 3: Subnet Examples

3.2 Eigenvectory Centrality as a Measure of Music Influence

Definition of degree centrality in undirected graph:

$$C_D(x) = \frac{\deg(x)}{n-1} \tag{1}$$

Definition of eigenvector centrality:

$$EC(i) = x_i = c\sum_{j=1}^{n} a_{ij}x_j \tag{2}$$

Decomposing the eigenvectory, its essence is to diffuse the degree centrality so that eigenvectory centrality not only depends on the number of points it connects, but also depends on the importance of the points it connects.

In a directed graph, the adjacency matrix is not a symmetric matrix with unideal convergence. Therefore, the definition of eigenvector centrality cannot be directly adopted. Under the circumstances and optimization considerations, we developed a new centrality based on the point-based outdegree centrality and PageRank algorithm. When measuring the importance of a point, not only the size of its out-degree is taken into account, but also the importance of the points connected by its out-degree are integrated.

Definition of outdegree centrality in undirected graph

$$C_D^*(x) = \frac{\text{outdeg}(x)}{n-1} \tag{3}$$

Definition of eigenvectory centrality:

$$EC^*(i) = x_i = c_i \sum_{j=1}^{n} a_{ij} x_j \beta \tag{4}$$

$$(i = 3 \qquad c_0 = C_D^*(x) \qquad \beta = (\frac{1}{2})^{i-1})$$

From the above formula, we calculate the centrality value of the class eigenvectors of all musicians.

For the convenience of calculation, use the following formula to convert the influence score into a percentile score:

$$I = \frac{\log(x_3)}{\max(\log(x_3))} \times 100 \tag{5}$$

we can get the final result of the influence of each musician:

Table 2: Influence Result

Artisit_id	I (Influence)
3829	100
100160	82.14
1003087	0
1003907	0
100473	17.86
100496	0
1006765	67.86
100840	0

3.3 Describe the Similarity and Difference of Music by Hierarchical Clustering Algorithm

Our team uses music characteristics to build a high-dimensional space into sample points, and then uses hierarchical clustering algorithm. We can conclude from the results that music lassified into the same cluster has a higher similarity.

First, each genre randomly selects 53 sample points, and uses Minkowski Distance as the standard to perform hierarchical clustering.

The reason why we choose 53 sample points of each genre is that:

- From the perspective of calculation speed and visualization, the original data set is inefficient and without ideal universality.

- Take the genre with the lowest number of occurrences as the floor, which eliminates scale impact and overcomes the defect of only selecting the local optimum in the greedy algorithm.

By adjusting the number of clusters, we can get some interesting results:

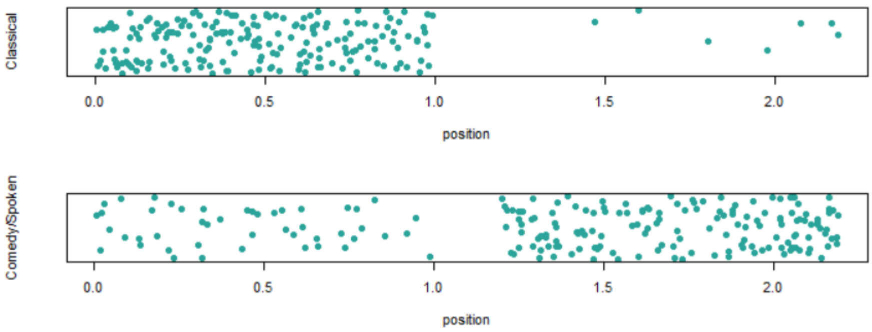


Figure 4: Result of Cluster

As shown in the above figure, the music of category A mainly concentrate in the left cluster, whereas music of category B mainly concentrate in the left cluster, which indicates that there are differences between category A music and category B music. This means that the similarity within category A is greater than the similarity between category A and category B. Similarly, for most genres, musicians within genres share more common characteristics than artists between genres.

That is to say, the difference of music according to genre is supported by statistics.

In order to make better use of music genre and distinguish the closeness and distance between different genres, we further analyzed the sample genre information as auxiliary information. The result is as follows:

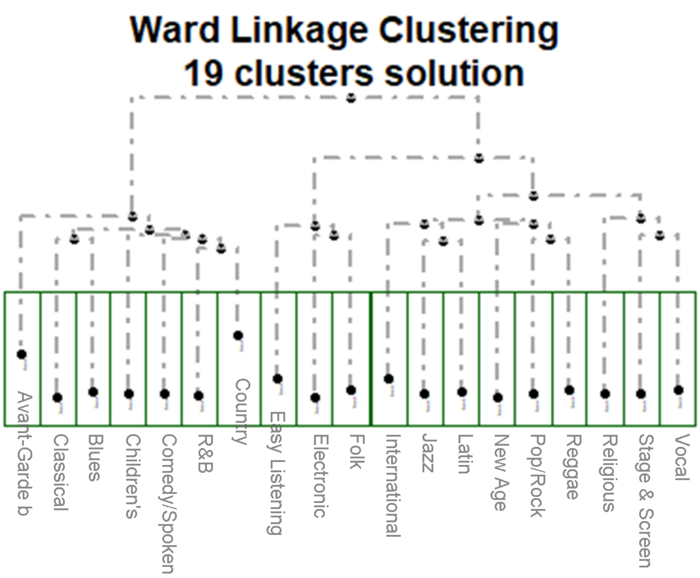


Figure 5: Relationship among Genres

3.4 Conclusion

From the perspective of influence, we use the influence measurement index based on the eigenvectory centrality to scientifically describe influence of musicians in music industry.

From the aspect of similarity and difference, we prove that "faction" is indeed a good classification standard in a statistical sense. Musicians within genres are more similar than artists between genres.

After the first part, we use factions as a measure of music similarity. There are significant differences in music between different factions, and the development and demise of different genres of music can be regarded as a musical revolution.

4 Operating Mechanism of the Model

This part selects a strong influential musician for case study and then explains the operation mechanism of the model. We use various significance level tests and time series data analysis to analyze the artists influence on his/her followers and even on the entire music industry, further verifying our logic chain, which proves that a single musician can indeed influence others through the topology network and further influence the overall music style, thereby driving musical changes.

Our group took out the musicians with high influence values in our definition, and selected T-Bone Walker who is accidently the first in the fitter result, and conducted the following analysis using him as an example.

4.1 "Contagious" Ability of Influencers

From *data_by_artist*, we can get Walker's music feature $\text{walker}_i, i \in \mathbb{N}^+, i \in [1, 14]$ and active year $acyr_w$, where i represents different genre.

Using $acyr_w$ as an index, we find the artists of the same generation as Walker and get the average music feature value of that age as mean_i

At the same time, from the *Influence_data*, using the relationship between Walker and its followers and the musician id, we find out all the songs of followers and features. We can get features of song x_{ij}, and amount of followers nf_w where i is same with previous and j indicates the writer for it.

Then for each song, we can calculate:

$$\sum_{\substack{j=0 \\ x \in \mathbb{F}}}^{nf_w} [d(x_{ij}, \text{walker}_i) - d(x_{ij}, \text{mean}_i)] = P_i \sim N(\mu, \sigma^2) \tag{6}$$

After that, we define the significance level of the music affected by Walker and establish the significance level table of each music feature affected by Walker. Result is as follows:

Table 3: Significant Level Result

Artisit_id	Danceability	Energy	Valence	Tempo	Loudness	Mode	Key
Significance	0.56	0.57	0.51	0.22	0.48	0.50	0.36
Artisit_id	Acousticness	Instrumentalness	Liveness	Speechiness	Duration_ms	Popularity	
Significance	0.95	0.81	0.55	0.82	0.52	0.98	

From this, we can conclude that the music of Walker's followers has a significantly smaller gap with Walkers than the average of era, that is, the music of followers is rather similar to that of Walker, which shows that Walker does have an influence on the music of followers. At the same time, we can draw a series of music characteristics with high similarity between followers' music and Walker's music, which are defined as more "infectious".

We've defined an indicator to reveal revolution $B = f(\Delta G)$. As we explained, change in music features ΔG is affected by influence of artist (I) and his or her "contagious" ability (w), that is $B = f(I, w)$ And obviously, an artist with low "contagious" ability, the outdegree centrality and eigenvectory centrality of him or her won't be high. Besides, Mathmatically speaking:

$$\exists k > 0, \forall I > k, |B - f(I)| < \epsilon_k \tag{7}$$

That is, when I is sufficient large, $B \approx f(I)$. Based on this, we can use influence of artists to reveal music revolution.

4.2 Leads to Music Revolution in 1950s

- Starting from Walker's musical characteristics: We select one (acousticness) from the "contagious" musical characteristics in 4.1. With that, we draw a figure based on average of year, which is extracted from *full_music_data* accordingly.

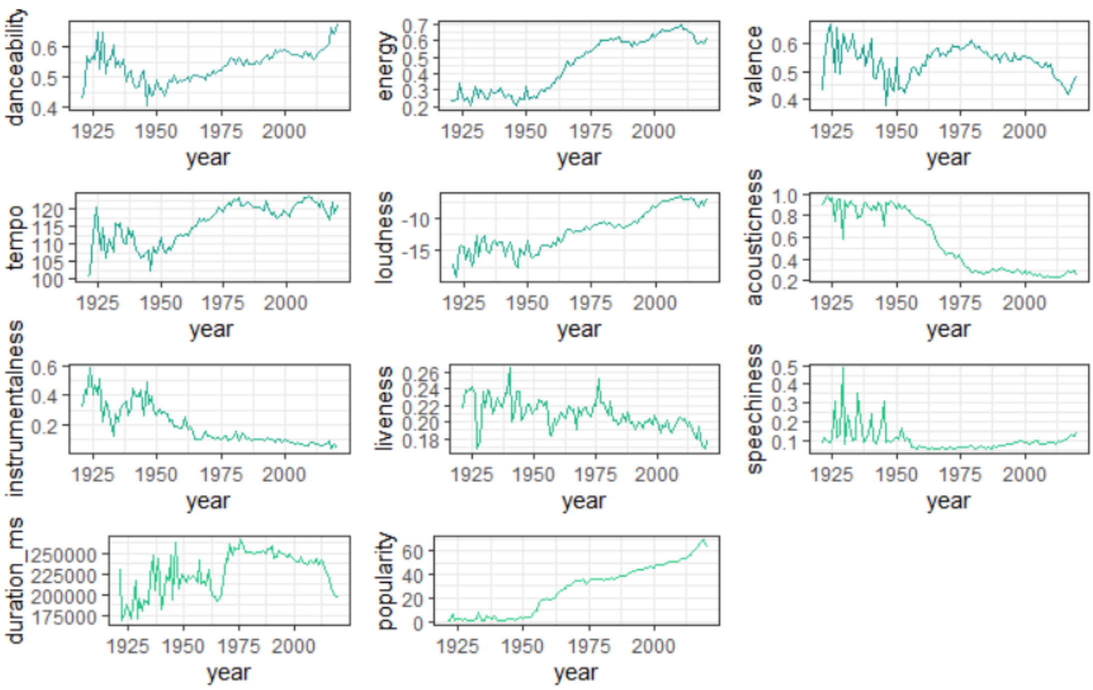


Figure 6: Trend of Features of Music

It can be seen from the figure that the acousticness of the entire music industry has dropped significantly around 1950, meaning the style of music has been affected, which can be considered as a revolution.

- Starting from the genre of Walker and its followers: We count the number of songs categorized as Blues, R&B, PopRock respectively.

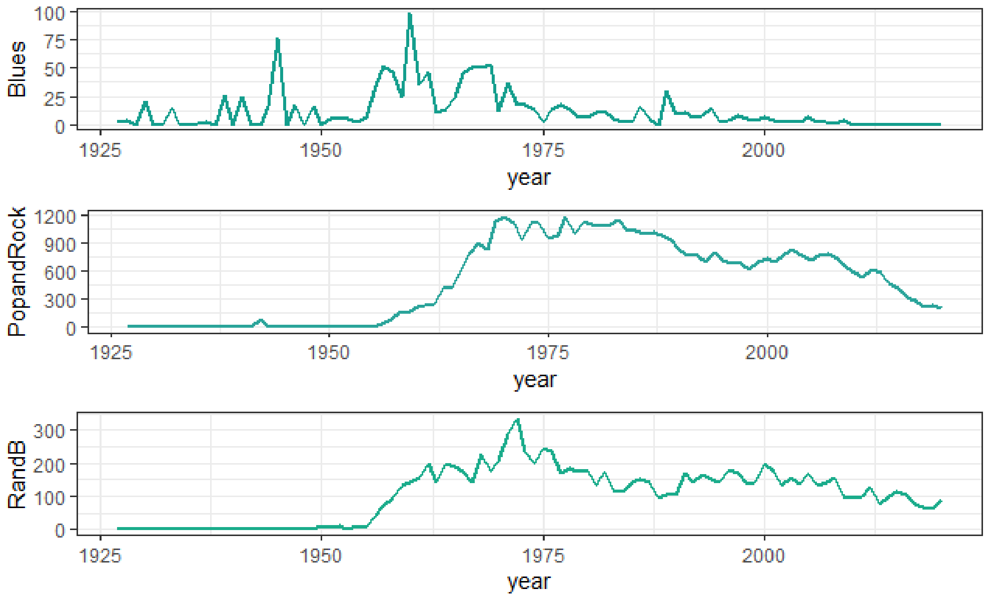


Figure 7: Trend of Genres of Music

The figure shows that the number of Blues skyrockets during 1950s, whereas R&B, PopRock increases significantly. The prosperity of the genre is a revolution in music.

4.3 Evolution in Blues

Now we use our model to explore the evolution of the Blues to which Walker belong.

From the topology network of 3.1 and the quantified influence result of 3.2, we can identify all of Walker's followers belonging to high I area with their genres and influences. We find that almost all of this influencial followers belong to the Blues, R&B, and PopRock factions.

Based on our model in 4.1, we predict that the corresponding factions will also have music revolution.

4.4 Vertify the Model by Empirical Material

Based on other martial and our previous analysis, we find that:

- Walker is the pioneer of Blues.
- Walker was very popular in the 1950s and had a great influence.

- Blues ushered in a peak from the 1950s to the 1960s, which is consistant with the Figure 8 in 4.2
- Blues is the origin of R&B and PopRock.

These data show music revolution the 1950s did exist, and Walker was one of the indispensable core changemakers, which is consistent with the conclusions of our model. Taking Walker as an example, we have reason to believe that musicians with high influence value are more likely to be the core innovators of a musical revolution, that is, the influence we construct is an indicator that can reveal dynamic influencers.

4.5 Conclusion

In this part, we choose Walker as an example to verify that our logic model is effective and solid.

Considering the connection between musicians, we build a high-dimensional space and perform hierarchical clustering to verify the statement that high-impact artists can influence music styles and genres.

Through relevant information, we corroborated our inference. We have reason to believe that musicians with high value of "influence" are more likely to be the leaders of a certain musical revolution, that is, the "influence" we have constructed is an indicator that can pinpoint dynamic influencers.

Starting from the entire music world, by observing the changes in the musical characteristics of high-impact artists and the evolution of the genres where the artists and followers belong, we can infer the occurrence of musical transform. In our example, we can infer that there was a musical revolution in the 1950s, and Walker (a powerful artist) was the core promoter of this revolution.

This is very interesting, because such a conclusion obviously violates the materialist argument of "time makes heroes". In the third part, through external data and related research, we find that such "greatness" is the result of certain historical accidents. However, the influence of "heroes" on "the current situation" cannot be ignored. This will be explained and demonstrated very well in the third part.

5 Extension Exploration

This part is mainly to extend the thinking of the model. In the previous sections, our team has established a logic chain model for musicians to influence the creation of new musicians through their own influence and to influence music and bring about musical changes. In this part, we mainly think about whether the chain can be a closed logic chain, that is, music as a cultural product closely related to social life, will music changes affect other non-musical factors, and whether these effects will be reversed impact on individual musicians.

Basically, we want to establish these two arrows in left of the figure.

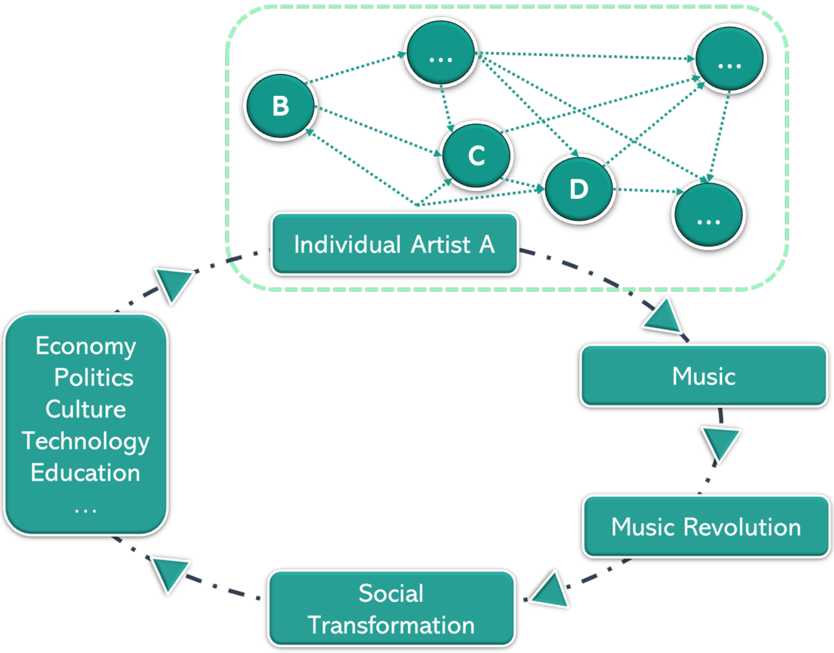


Figure 8: Logic Model

Our group believes that there are paths for music change to have other effects and counteract individual musicians. That is:

1. Musical changes cause cultural changes through the dissemination of music among the audience, and ultimately affect all aspects of politics and economy.
2. The political, economic, and cultural changes will affect individual musicians and their creative process, which in turn will promote musical changes.

Our group has read a lot of relevant music literature and pondered the nature of the influence mechanism more deeply. We believe that the old music framework is actually a kind of restriction determined by the old cultural framework, which restricts the universal musical ability and unique music outlook of human beings. The framework of the old music clamps the musicality of human beings under the music concept shaped by the old culture, forming the so-called "music consensus" and "music trend". But when the old cultural framework is affected by political, economic, scientific and other factors and develops into a new cultural framework, there will be "heroes" under the new cultural framework that will cast cultural changes into music, and will quickly influence the topology network. The new music concept was passed on to the artists who accepted the new cultural framework in that era, which caused the reform of the old music framework. From this point of view, the view that Time makes heroes is supported.

However, in the transformation of the old music framework, certain personal styles of heroes that do not depend on the times will also rapidly expand by influencing the topology network, forming a trend, and becoming an indispensable part of the new music framework, and counterproductive above culture. This trend is essentially a manifestation of the view that "heroes create the situation".

Therefore, our group believes that "heroes make times" is one-sided, and "times make heroes" is also one-sided. "Current situation" and "hero" complement each other, influence each other, and jointly create and determine the evolution of music history. One-sided historical materialism and one-sided historical idealism are biased.

6 Strengths amd Weaknesses

6.1 Strengths

1. The operation process of the model is not only supported by statistics and mathematics, but also confirmed by a large number of historical facts in the music industry. It is highly scientific and universal.

2. The concept of analyzing the history of music revealed by the model: "The process of mutual influence between "time situation" and "hero" conforms to dialectical philosophical thinking, which can enlighten us to look at the other historical developments of sociology and history from the same perspective. This shows the strong cultural and thinking of our model.

3. Musician network analysis program established by our team is multifunctional and highly compatible, built-in various conditional subnet dimensionality reduction methods, can support richer data set analysis. In the future, the program can realize more complete functions by adding various ports and methods, which means it is quite scalable.

4. Many of the mathematical and statistical inference conclusions in the model are highly theoretical, like "It is scientific to use genres to divide music in the music industry" and "Artists with high influence are often participants in core changes." can be widely accepted in the real world, not just valid in a specific data set.

6.2 Weaknesses

1. The model only uses some cases to illustrate the operation of the model, so it can only draw a conclusion of "feasibility", and it is difficult to draw a more general conclusion.

2. Due to time and data set limitations, some non-mathematical or non-statistical inferential arguments are difficult to obtain extensive data sampling support.

7 Conclusion

We have established a closed logic chain in which individual musician brings musical changes through their influence. Those musical changes have other political, economic, and cultural influences, which counteract with the musicians. Our result provides a possible/reasonable answer to the debate between "heroes create the situation." and the times make heroes" in the history of music. That is, "heroes make times" is one-sided, and "times make heroes" is also one-sided. "Current situation" and "hero" complement each other, influence each other, and jointly create and determine the evolution of music history.

8 One-Page Document to the Integrative Collective Music Society

Dear Integrative Collective Music Society,

According to your request, our team tried to answer the questions by carefully analyzing the data set provided and extra related theory. We also constructed a valid model of how past music and current politics and culture affect music novelty today. In the process of researching the problem and forming the model, we gained quite interesting discoveries and conclusions worth sharing. We are honored to share them with you.

1. **The influence network between musicians can clearly reveal the influence of musicians in the history of music.**
2. **The use of the concept of "faction" to differentiate music is supported by statistical analysis.**
3. **Musicians with great influence magnify their influence through topology network to bring about music revolution.**
4. **In the view of music history, both historical materialism and idealistic historical theory are one-sided, and we must comprehensively look at the influence of "current situation" and "hero" on the evolution of music.**

Our group believes that the view "heroes shape epoch" is one-sided, so is "time makes heroes". "Current situation" and "hero" complement each other, and jointly determine the evolution of music history.

In the process of research, we took the limitations of the current data into account. And the problem of large data volume. We have also tried our best to enhance the portability of mega data in the model, specifically in the following aspects:

1. **We have developed a function to obtain smaller subnets of large networks based on various conditions. And the function manages to reduce the amount of computation through network dimensionality reduction.**
2. **The statistical measurement of music similarity uses random sampling method, ensuring low difficulty to perform analysis and high accuracy regardless of the size of data.**
3. **When analyzing the model, we used mathematical methods to tentatively prove that our conclusions are applicable to musicians with extremely high influence. When the amount of data is abundant, the applicability of the model should remain excellent in theory.**

We are more than grateful to have this opportunity to look into the whole music industry. We sincerely hope that our findings on music influences, differences and similarities can contribute even a little to the prosperity of music industry. As we have stated, music evolution happens when a perfect guy shows up at the perfect time. Protection of intellectual property will spur novelty of music. We are looking forward to witnessing the next greatness in music.

Yours sicerely,
ICM 2021 team

References

[1] How Musical Is Man? John Blacking,1928-1990.

[2] Spizzirri L. Justification and application of eigenvector centrality[J]. Algebra in Geography: Eigenvectors of Network, 2011.

[3] P. Berkhin. A survey on pagerank computing. Internet Mathematics, 2(1):73120, 2005.

[4] W. Tecumseh Fitch. The Evolution of Music in Comparative Perspective. Ann N Y Acad Sci. 2005 Dec;1060:29-49. doi: 10.1196/annals.1360.004.

[5] WALLIN, N.L., B. MERKER & S. BROWN. 2000. The Origins of Music. The MIT Press. Cambridge, MA.

[6] Pengsheng Ji,Jiashun Jin. Coauthorship and citation networks for statisticians. Ann. Appl. Stat. Volume 10, Number 4 (2016), 1779-1812.

Appendices

Appendix A Code for Section 3.1 and 3.2

```
# This is the function to get the influence of every artists

import os, warnings, igraph
import numpy as np
import pandas as pd

class Jgraph(object):
    def __init__(self, matrix):
        '''(matrixDataFrame)'''
        self.matrix = matrix
        self.__directed = False
        self.shape = self.matrix.shape
        self.name = self.matrix.index
        self.__check()
        self.file_path = os.path.join(os.path.expanduser("~")
                            , 'Desktop') + '\\social_network.png'
        self.shortest_file_path = os.path.join(os.path.expanduser("~")
                    , 'Desktop') + '\\shortest_paths_social_network.png'
        self.g = igraph.Graph(directed=self.__directed)
        self.visual_style = {"vertex_size": 10, "bbox": (800, 800),
                            'margin': 20, 'vertex_label_size': 15,
                    'edge_arrow_size': 0.5, 'vertex_label_dist': 2}
        self.__graph()
        self.__get_centrality()

    def __check(self):
        for i in range(self.shape[0]):
            if self.matrix.iloc[i][i] != 0:
                warnings.warn("Diagonal elements don't make sense
                             , program has been setting it to 0")
                self.matrix.iloc[i][i] = 0
            for j in range(self.shape[1]):
                if self.matrix.iloc[i][j] != self.matrix.iloc[j][i]:
                    self.__directed = True

    def __nodes(self):
        self.g.add_vertices(self.shape[0])
        self.g.vs['label'] = self.name
        self.visual_style["vertex_label"] = self.g.vs["label"]

    def __edges(self):
        edges = []
        weight = []
        for index in range(self.shape[0]):
            for columns in range(self.shape[0]):
                if self.matrix.iloc[index, columns] != 0:
                    if self.g.is_directed():
                        edges.append((index, columns))
                    else:
                        if index <= columns:
                            break
                        edges.append((index, columns))
```

```
                weight.append(self.matrix.iloc[index, columns])
        self.g.add_edges(edges)
        self.g.es['weight'] = weight

    def __sort_dict(self, dist):
        ''''''
        return dict(sorted(dist.items(), key=lambda x: x[1]
                                              , reverse=True))

    def __graph(self):
        self.__nodes()
        self.__edges()
        self.base_visual_style = self.visual_style.copy()
        self.base_visual_style.update({'layout': self.g.
                                   layout("kamada_kawai")})
        p=igraph.plot(self.g, self.file_path, **self.base_visual_style)
        p.save('TEST.png'.format('title'))

    def __get_centrality(self):
        print('degree_centrality:\n', self.degree_centrality())
        print('closeness_centrality_without_weights:\n',
               self.closeness_centrality_without_weights())
        print('betweenness_centrality_without_weights:\n',
             self.betweenness_centrality_without_weights())

    def __from_name_get_id(self, *ids):
        labels = []
        for id in ids:
            labels.append(list(self.matrix.index).index(id))
        return labels

    def __get_eid_from_list(self, paths):
        id_list = []
        for path in paths:
            for i in range(len(path) - 1):
                id_list.append(self.g.get_eid(path[i], path[i + 1]))
        return id_list

    def directed(self):
        ''''''
        print('This is %sdirected graph' % (self.__directed * 'un'))

    def page_rank(self):
        '''PageRank'''
        return self.__sort_dict(dict(zip(self.g.vs['label']
                                        , self.g.pagerank())))

    def degree_centrality(data):
         ''''''
        data1=data.sum(axis=1)
        return data1

    def eigenvalue_centrality_without_weights(self):
        '''()'''
        return self.__sort_dict(dict(zip(self.g.vs['label']
                                , self.g.eigenvector_centrality())))
```

```
def link2matrix(link):
    label = np.union1d(link.iloc[:, 0], link.iloc[:, 1])
    matrix = pd.DataFrame(0, index=label, columns=label)
    for i in range(len(link.iloc[:, 0])):
        matrix.loc[link.iloc[i, 0], link.iloc[i, 1]] += link.iloc[i, 2]
    return matrix

if __name__ == '__main__':
    f=pd.read_csv('full.csv', dtype={'influencer_id':str,
                                     'follower_id':str})
    data = link2matrix(f)
    g = Jgraph.degree_centrality(data)
    a=len(data)
    g2=g
    for i in range(0,a):
        for j in range(0,a):
            if data.iloc[i,j] == 1:
                g2[i]=g[i]+g[j]

    g3=g2
    for i in range(0,a):
        for j in range(0,a):
            if data.iloc[i,j] == 1:
                g3[i]=g2[i]+1/2*g2[j]

    #
    g4=g3
    for i in range(0,a):
        for j in range(0,a):
            if data.iloc[i,j] == 1:
                g4[i]=g3[i]+1/4*g3[j]

    #0-100
    import math
    g5=g4
    for i in range (0,5602):
        if g4[i]>0:
            g5[i]=math.log(g4[i])

    g6=g5.map(lambda x : x/28*100)
    g7=g6.map(lambda x : "%.2f"%x)

    sys.setrecursionlimit(10000)

    class relationship:
        def __init__(self,influencePath,dataPath):
            self.influence = pandas.read_csv(influencePath, index_col=[0])
            self.data = pandas.read_csv(dataPath, index_col=[0])
            self.influenceNum = self.influence.values
            self.dataNum = self.data.values
            self.name = self.data.index.values
            self.foundRowList = [0] * self.data.shape[0]
            self.foundColList = [0] * self.data.shape[0]
            self.adjacencyTable = {}
            self.numSet=set()


        def findRelationshipFunc(self,num):
```

```
            if self.foundRowList[num]==0:
                self.foundRowList[num]=1
                rowOne = np.where(self.dataNum[num] == 1)[0]
                for i in rowOne:
                    if self.influenceNum[i] > self.gate:
                        if not (num in self.adjacencyTable.keys()):
                            self.adjacencyTable[num]=set()
                        self.adjacencyTable[num].add(i)
                        self.numSet.add(num)
                        self.numSet.add(i)
                        self.findRelationshipFunc(i)
            if self.foundColList[num]==0:
                self.foundColList[num]=1
                colOne=np.where(self.dataNum[:,num]==1)[0]
                for i in colOne:
                    if self.influenceNum[i] > self.gate:
                        if not (i in self.adjacencyTable.keys()):
                            self.adjacencyTable[i]=set()
                        self.adjacencyTable[i].add(num)
                        self.numSet.add(num)
                        self.numSet.add(i)
                        self.findRelationshipFunc(i)


        def findRelationship(self,key,gate):
            num=np.where(self.name==key)[0][0]
            self.gate=gate
            self.findRelationshipFunc(num)

        def saveCsv(self,path):
            newData=np.zeros(len(self.numSet)*len(self.numSet)).
                    reshape((len(self.numSet), len(self.numSet)))
            numList=list(self.numSet)
            nameList=[]
            for i in numList:
                nameList.append(self.name[i])
            for key in self.adjacencyTable.keys():
                for num in self.adjacencyTable[key]:
                    newData[numList.index(key)][numList.index(num)]=1
            panData=pandas.DataFrame(newData,columns=nameList,
                                            index=nameList)
            print(panData)
            panData.to_csv(path)




    miku=relationship('influence.csv','data.csv')
    miku.findRelationship(3829,75)
    miku.saveCsv('new.csv')
```

Appendix B Part Code for Section 3.3

```
cluster2 <- cutree(fit.average,k =19)
table(cluster2)
d2 <- aggregate(merge2byb,by = list(cluster = cluster2),median)
```

```
c2 <- aggregate(as.data.frame(merge2by.scaled),
            by=list(cluster=cluster2),median)
fit.average1 <- as.dendrogram(fit.average)
NO2 <- plot(cut(fit.average1,h = 150)$upper, cex = 0.1,
main = 'Ward Linkage Clustering \n 19 clusters solution',
edgePar = list(col = "darkgrey",lty = 4,lwd = 2.5,edge.root = TRUE),
    nodePar = list(pch = 20,lab.cex = 0.1,type = "t"),center = TRUE,
                                                   axes = FALSE)
result2 <- rect.hclust(fit.average,k=19,border = "darkgreen")
```

Appendix C Code for Section 4

```
data <- read.csv("D:/ICM//influence.csv")
j=0
InfluenceMusician<-data.frame()
for(i in 1:5603){
  if(data[i,2]>98){
    j=j+1
    InfluenceMusician[j,1]=data[i,1]
    InfluenceMusician[j,2]=data[i,2]
  }
}


data <- read.csv("D:/ICM//influence_data.csv")
j=0
follower<-data.frame()
for(i in 1: 42770){
  if(data[i,1]==InfluenceMusician[2,1]){
    j=j+1
    follower[j,1]=data[i,5]
    follower[j,2]=data[i,7]
    follower[j,3]=data[i,8]
  }
}


data <- read.csv("./data_by_artist.csv")
library(dplyr)
colnames(data)[2]<-"V1"
colnames(follower)[1]<-"V1"
follower<-left_join(follower,data,by="V1")


data <- read.csv("./data_by_artist.csv")
colnames(data)[2]<-"V1"
colnames(InfluenceMusician)[1] <- "V1"
InfluenceMusician<-left_join(InfluenceMusician,data,by="V1")


data<-read.csv("./data_by_year.csv")
library(ggplot2)
library(lubridate)
p1<-ggplot(data, aes(year, y=danceability))+
     geom_line(colour="#2AA49A")+theme_bw()
...
p11<-ggplot(data, aes(year, y=popularity))+
```

```
    geom_line(colour="#2ACE84")+theme_bw()
library(gridExtra)
grid.arrange(p1,p2,p3,p4,p5,p8,p6,p7,p9,p10,p11,ncol=3)


data<-read.csv("./data_by_year.csv")
p=array()
j=1
Walker<-InfluenceMusician[2,]
mean1930<-data[10,]#1930 in the tenth raw
compare<-data.frame()

##danceability
for(i in 1:48){
compare[i,1]=abs(follower[i,5]-Walker[4])
compare[i,2]=abs(follower[i,5]-mean1930[2])
}
mean_walker=mean(compare[,1])
var_walker=var(compare[,1])
mean_1930=mean(compare[,2])
var_1930=var(compare[,2])
mean=mean_1930-mean_walker
variance=var_walker+var_1930
p[j]=pnorm(-mean/sqrt(variance),lower.tail=F)
```