
Hand down from age to age: The internal mutual influence of popular western music genres

There are many people enjoy themselves in listening to various types of music from all kinds of artists. But it is worth questioning how many of them have ever bothered to think about what genres do the artists come from, not to mention how do they affect each others' music characteristics. One of the purposes of our thesis is to find the relationships between different genres and artists, also the variation trends of music styles through time, and to give some insights into music and society change. The main feature of our thesis is that we used a huge amount of illustrations to better visualize our data, and our interpretations for the illustrations are all extremely diligent.

In **Task 1**, we built a directed graph that links from influencer to followers. By using Dijkstra algorithm, we could find the shortest path between an influencer and a follower, and go further to analyse the influence power of an artist, by processing the total lengths of the paths.

In **Task 2 and Task 3**, we utilized the datasets that divided the characteristics of music into 14 music features, regarding the features as arguments, we calculated the Euclidean distance of two artists' corresponding of data of music features to evaluate the similarity of the two. The larger the distances means the divergences are larger.

In a similar manner we also calculated the Pearson correlation coefficient of corresponding features of artists, which is a supplementary method for suggesting the similarity degree in a positive index. The results using the two methods drew very familiar conclusions about similarities.

And again harnessing the Euclidean distance formula, by calculating population average of every two artists' similarity within a genre, we calculated the single-genre similarity. And found that country genre have a rather high inner similarity, means its artists have quite similar music styles in terms of the given 13 music features. Applying the same routine, we calculated the cross-genre similarities by taking population average of similarities of artists from the two genres.

We showed all of the similarity comparison results in the similarity matrix—the most important illustration in the whole thesis. And to conclude in general, we found the similarity within one genre is more significant than cross-genres similarity. And the pop/rock genre is approximately equally influenced by other typical genres, therefore can be seen as an combination of many genres to an extent.

In **Task 4**, by calculating consistencies of music arguments between a follower and all of his alleged influencers, we found the influencers do have practical influence on followers, and the features with the smallest variance in relative size, which are tempo and loudness, are more 'contagious', meaning these two aspects exert a deeper influence on an artist than other characteristics of music.

In **Task 5**, we defined eras in which music features turned sharply, which are about 1950s and 1960s, as the time when music revolution took place. And the arguments indicators are mainly acousticalness, liveness and energy for the magnitudes of the changes are very arresting in these features.

In **Task 6**, we mainly gathered the number of artists of different genres through the time line in broken line charts using the data_by_year. We also compared the musical features of popular influencers with common music followers to determine the indicators for dynamic influencers.

In **Task 7**, analysed the influence of globalization and the wild application of Internet on music styles briefly.

keywords: Directed Graph; Similarity evaluation; Euclidean Distance; Statistical Analysis

Hand down from age to age: The internal mutual influence of popular western music genres

February 9, 2021

Contents

1	Introduction	2
1.1	Background	2
1.2	Restatement of the Problem	2
1.3	Assumptions of the Problem	3
1.4	Notations	3
1.5	Basis Analysis of the Problem	3
2	Network Model to Describe the Relationship between influencers and followers	4
2.1	Basic Preparation	4
2.2	Directed Network of Influencers and Followers	4
2.3	Methods to Measure ‘Music Influence’	8
2.4	Simple Comparison of the Three Ranking Results	10
3	Similarity between Genres, and between Artists	10
3.1	Models of Similarities	12
3.2	A Supplementary Methodology to Measure Similarity	15
3.3	The Practical Influence of the Alleged Influencers to Followers	16
3.4	Distinct Characteristics of Genres	16
4	Variation Tendency of Genres Through Time	17
4.1	The Change in Numbers of Artists in a Genre	17
4.2	The Implying Factors in the Changing of Music Styles	17
5	Indicators for Revolutions in Musical Evolution	19
5.1	Music Arguments Indicators	19
5.2	Indicators for Dynamic Influencers	19
5.3	Great Artists as revolutionists	20
6	Integral Evaluation of Strengths, Weaknesses and Conclusions	20
7	A Letter To ICM	21

1 Introduction

1.1 Background

Nowadays, it has become a more and more common phenomenon for people to listen to all kinds of music to seek pleasure and relax themselves. And we may abstract and conclude the invariable key elements in the forever-changeable styles and contents of music. That are melody, harmony, lyrics, rhythm, and timbre, etc. And in analysing the similarity and developmental portraits among some of these elements, we may have a primary clue on how the previous musicians have exerted their influence upon their same and subsequent generations of musicians and give some insights into how a new genre of music came into being.

Utilizing the interactive network model, we are able to implement the cross-tabulation of the various specimens given, and further explore the deep-rooted relationship between different musicians over time, therefore produce a elementary work on the evolutionary trajectory of music.

1.2 Restatement of the Problem

From the data we retrieved.[1] We can learn that all the artist in the world has their influences. Not only to the genre of music he specialized. But also to some specific artists from the younger generation. Our first mission is to create a network that links the influencer and the follower. Through the network, we may find the indicators of "musical influence". The weight and the form of the network can show us the relationship between different artists. They may be influencers, or followers. Or they can be both.

Similarity is also an index we focus to. Since there are experts already defined some features and factors of musical compositions. Features such as danceability, energy or liveness is hard to define by amateurs like us. However, the expert quantized these features so we can process them easily. In this case, we should find out that the similarity between artists from the same genre or from different genres.

The revolution of music through a period of time is also an significant argument we are facing. Each artists is active in different eras. Thus only the artists who became active earlier can influence the latter artists. Each era has their favorites, how does people's taste varies is also an interesting problem to solve.

In short, we can divide the whole task into the following parts.

- Build the network that can show us the relationship between artists: who is the influencer and who is the follower.
- Take the time and genre into consideration. Find out how these indicators affect our model.
- Use our model and analyze the data we processed. After all, we can have our own thoughts about it. How does music affect the society? And how much is the influence of a revolutionary artist/genre.

Our model is meant to figure out the intrinsic information from the data we already knew. By processing the data with our model, we can learn a lot of hidden information. Hence we can answer the questions like: who is the major influencer of an era and how he led the popularity, which genre of

music is the most popular one, how does the fashion changes over time. At last, we may even find out which genre will lead the major revolution of music.

1.3 Assumptions of the Problem

- Only the artists who became active earlier can influence the latter artists.
- The situation of two artists being each other's influencer and follower dually is rare.
- Most artists don't have their own genre before they are influenced by other artists.

1.4 Notations

- Typical genres: Genres with relatively large number of artists.

Notation	Definition
m_d	the direct parameter for measuring music influence of an artist
m_c	the comprehensive evaluation parameter for measuring music influence
m_u	the improved comprehensive evaluation parameter for measuring music influence
O_d	out-degree
O_{dt}	total out-degrees in a graph
I_d	in-degree
l_i	shortest path length from a node to node i
T	total number of artistS
p_{fi}	the proportion of same-genre followers to influencers ("loyal index" to a genre)
$s(a,b)$	similarity between artists a and b
$S(a,b)$	similarity between genres a and b (with capital letter S)
z_i^*	the normalized i th aspect(argument) of music in data_by_artists
$\rho_{X,Y}$	Pearson correlation coefficient

1.5 Basis Analysis of the Problem

1.5.1 The Proportion of Each Genre

As we can see very intuitively from the pie chart below that the dominant music category is pop/rock, which takes up more than half of all the genres, followed by R&B and country, occupying about 10-15% each. So we mainly focus on the pop/rock genre in the next analysis.

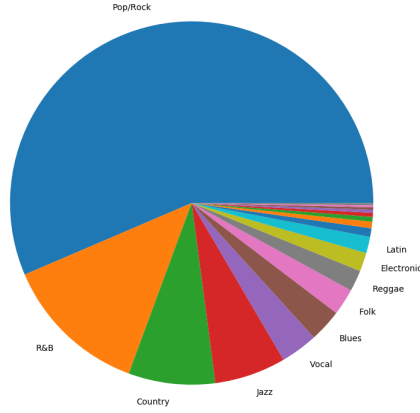


Figure 1: The proportion of each genre

2 Network Model to Describe the Relationship between influencers and followers

2.1 Basic Preparation

2.1.1 Spares Graph

From Wikipedia, we know in mathematics, a graph with only a few edges, is a sparse graph. The graph density of simple graphs is defined to be the ratio of the number of edges $|E|$ with respect to the maximum possible edges. For directed simple graphs, the maximum possible edges is twice that of undirected graphs to account for the directedness, so the density is:

$$D = \frac{|E|}{2\binom{|V|}{2}} = \frac{|E|}{|V|(|V| - 1)}$$

where E is the number of edges and V is the number of vertices in the graph. The maximum number of edges for an undirected graph is

$$\binom{|V|}{2} = |V|(|V| - 1)/2$$

So the maximal density is 1 (for complete graphs) and the minimal density is 0.

2.2 Directed Network of Influencers and Followers

We primarily concerned two ways to build the directed influential network of influencer artists and their followers:

The first way is through the adjacency matrix, where we use rownum to represent followers' ID, and column number to represent influences' ID. Using this method we can see very directly how an artist is connected to another. And the matrix for this network mainly has two characteristics: it is extremely asymmetric and sparse.

The reasons are easy to understand: there almost no notable mutual influences between two artists in the given data, and the total amount of artists is too large compared with the the number of influencers to a particular artist. The sparse matrix can be stored as dictionary of keys or list of lists or compressed sparse column to save space.

artist ID	759491	75747	84053	74	66915	81083
759491		0	0	1	0	0
75747	1		0	0	0	0
84053	1	0		0	0	0
74	0	0	0		0	0
66915	0	0	0	0		1
81083	0	0	0		0	

In this case, we chose 5 artists from 'influence_data.csv' to build an adjacency matrix. The matrix is shown above. But of course, due to the huge size of the data, we cannot write down all of them in the thesis.

The second way is to draw a directed graph to demonstrate the mutual influences of musicians from different genres. Add different nodes that representing the artists to the network, and then connect them with arrows.

Due to the content length limit, we will here further illustrate the second way to establish our first model in this paper. To begin with, in order to take better advantage of the whole influence data set from 1930 to 2010, we (choose a thousand pieces randomly from the 42771 pieces in the given influence data,) and use each node to represent an artist, we also assign different genres of artists with a distinct color to better distinguish them. The node's arc goes from the influencer to the follower. We use the coding toolkit that utilizing Adjacency list to store the sparse graph to save space, and it is also convenient for us to see the influence relationship of a particular artist in this storage format. In this case, we can observe what are the nodes that points to a particular node (the node's in-degree) to see the amount and genres of artists that have influenced this particular artist, in the same way, by observing what nodes are a particular node pointing to, (the node's out-degree), we can gain knowledge about the amount and genres of artists that this particular artist have influenced. We notice that different from the commonly seen directed graphs, which have at least some strong connectivities, our graph generated in this problem have almost no strong connectivities, that is to say we the influencers and the followers rarely change their positions, in fact according to our statistics, there are only 78 mutual-influence pairs in the 42770 edges and more than 5600 artists given.

The total number of outdegrees of a graph is equal with its indergees, and can be calculated with this formula:

$$O_{dt} = \sum_{i=1}^T O_{di}$$

The total outdegrees for this graph is 42770, and the network density of the above graph is 0.0027, suggesting that the directed network of influencers and followers is a sparse graph from a mathematical perspective, for the connections, namely the edges, are too few compared with the amount of nodes.

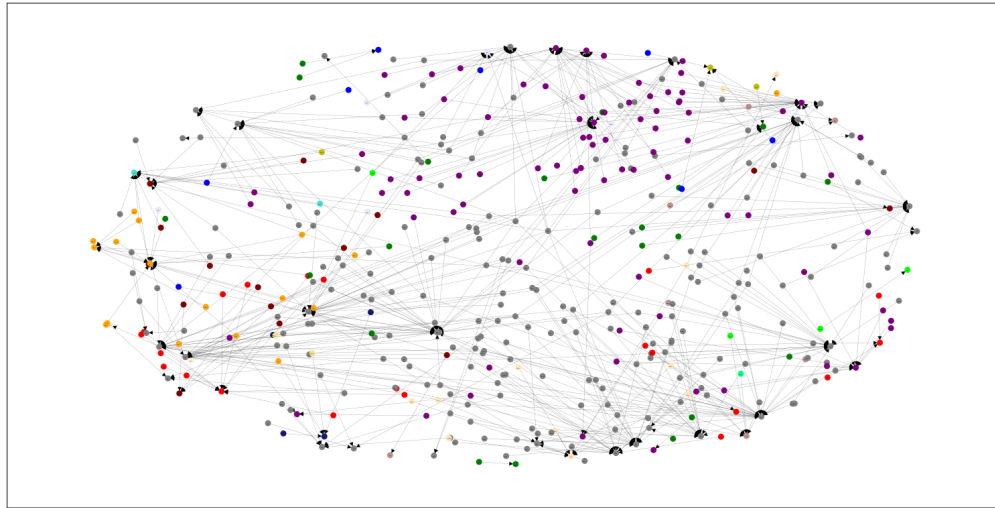


Figure 2: network of influencer-artists and their followers

2.2.1 Further Analysis of the First Model

To illustrate the first model more specifically, we can see that, generally speaking, the artist's own genre exerts more influence to him/herself than other genres.

We also drew a bar graph to qualitatively analyse the active-start time of influencers and followers. The height of the blue column represents the number of influencers, the height of the pink column represents the number of followers, and purple means the influencers and followers overlaps. It is clear that the influencer number obeys approximately normal distribution, peaks at the year of 1960, and the follower number peaks at the year of 1990. In the year of 1970 and 1980, there are almost the same number of influencers and followers.

2.2.2 A 3D Model to Further Analyse The Network

In order to have a deeper understanding of how different artists influence one another, we conducted our second model in matlab. As shown, each of the three dimensions of the broken line graph represents genres, the year from which the artist became active and the artistid respectively, and each single point is to be an artist. We connected the points that have an influence relationship with each other. Four main conclusions are drawn from the graph:

- Same with the proportion pie chart in front, we can see dots are more densely packed in the right side of the graph, meaning three categories, pop/rock, R&B and country take up most of the genres.

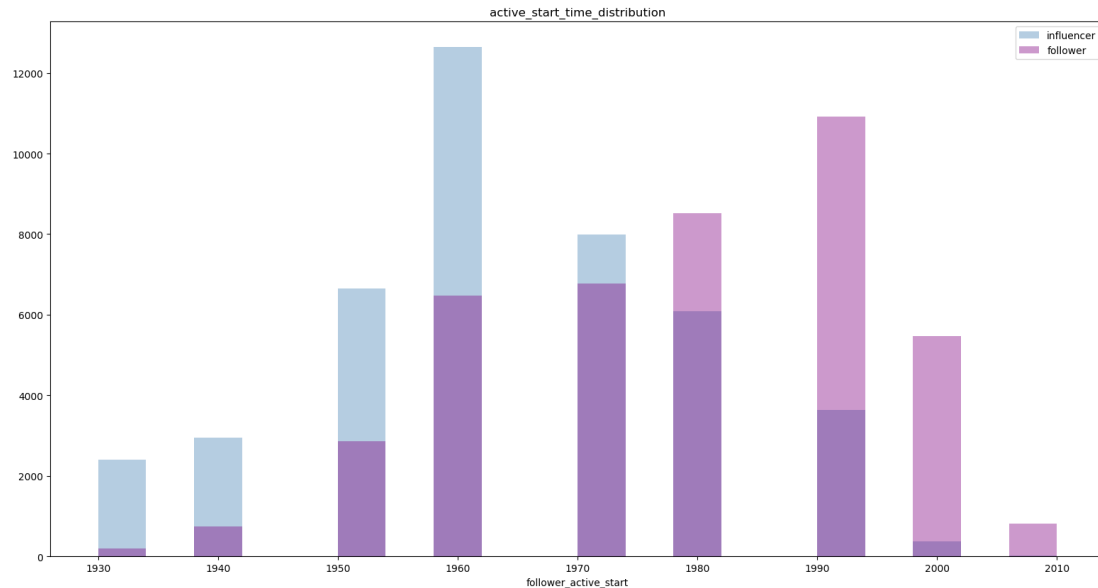


Figure 3: active-start time of influencers and followers

- The darkest lines are the vertical lines in the right, indicating artists are more influenced by the same genre they are from, and have strong connections with the artists from the same genre. The horizontal lines are comparatively few in number, implying that the influence relationships are mainly from the artists who started to become active in different years.
- We can see roughly from the graph that there is a downward trend for the number of Jazz genre artists, by truncating vertically versus the artistid axis, and upward trend for pop/rock genre artists' number as time goes.
- We can roughly see that the ABC genre have more connections with the DEF genre, meaning the two genres are more closely related, we will cross-contrast this conclusion with the following similarity analysis among genres.

2.2.3 Another Network from a Different Aspect

In a similar manner, we also depicted another directed network ignoring the genre differences of artists to demonstrate the influence relationship among artists who become be active in different years.

To do that, We assigned the artists who be active in the same year with a distinct color, no matter what genre they are from. To give an example to illustrate the graph, we made all artists who be active in 1960 as green nodes, and artists who be active in 1980 as blue nodes, we can see clearly that there is a huge amount of green nodes pointing to blue nodes, however among the different colors of the nodes that blue nodes point to, we can hardly see any green, meaning our assumption that 'Only the artists who became active earlier can influence the latter artists' is correct on the whole.

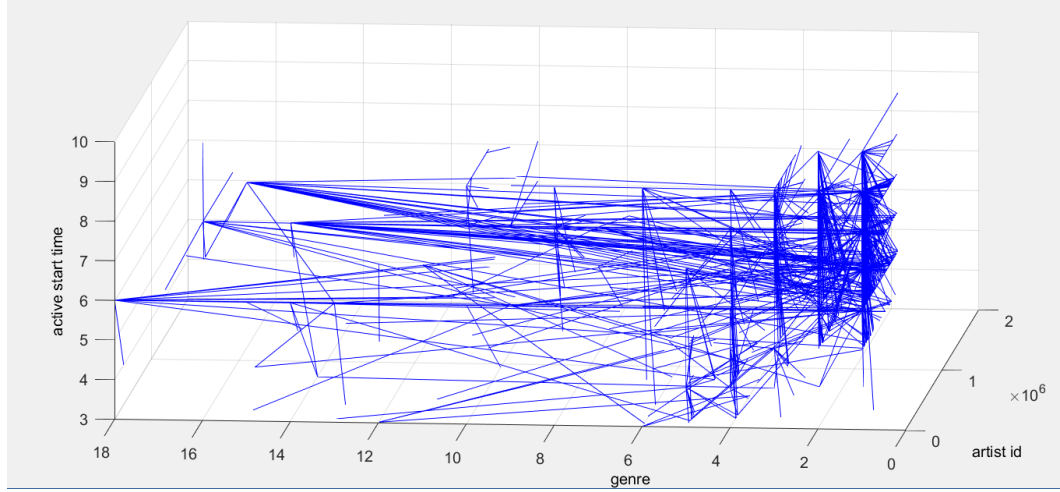


Figure 4: 3D model of the influence_data

2.3 Methods to Measure ‘Music Influence’

2.3.1 Direct Way for Measuring ‘Music Influence’

As is explained above, we may use the outdegree of a node in the first model to evaluate the power of influence of that artist. And in order to narrow the influence parameter to attain a similar effect as normalization, we use the outdegree divided by the total number of artists available.

$$m_d = \frac{O_d}{T}$$

Combining our statistics with the formula, we got the ten most influential artists, they are: **1** The Beatles, **2** Bob Dylan, **3** The Rolling Stones, **4** David Bowie, **5** Led Zeppelin, **6** Jimi Hendrix, **7** The Kinks, **8** The Beach Boys, **9** Hank Williams and **10** The Velvet Underground.

2.3.2 Comprehensive Way for Determining ‘Music Influence’

Again making use of the connectivity of our first model, and associate it with the Dijkstra algorithm in graph theory, we developed a second way for determining ‘Music Influence’ as an improved plan of the first. We divided the influence power of an artist into two aspects: his/her direct influence to others and indirect influence to others. The numerical value of direct influence is the outdegree of the node, and the indirect influence is the indirect connection of two nodes, for example, if artist A influenced artist B and artist B influenced artist C, but there is no direct connection between A and C, then we consider artist A also has influence upon artist C, in order to have a set of positive assessment criteria, we define the numerical value of indirect influence as the reciprocal of the shortest path of two nodes.

Below is an example to demonstrate the meaning of shortest path of two nodes, there are two ways that A can be connected with C, either through A→B→C or A→D→E→C, and the shortest path is 2, the indirect influence value is 1/2. We then calculated all the shortest paths of each node available of the first model, take reciprocals, and added them together as the total indirect influence value of an artist. It is worth mentioning that if there isn’t any path between two nodes, their path length is considered as 0 to simplify the calculation.

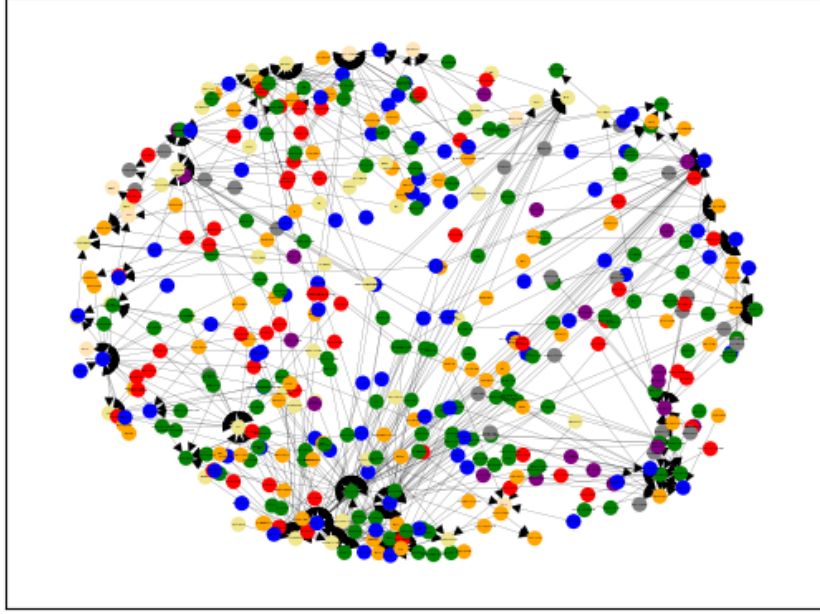
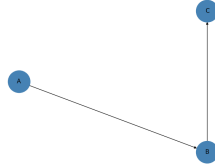


Figure 5: influence relationship among artists of different years

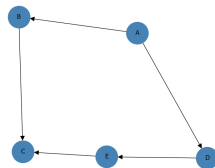


We defined our comprehensive evaluation parameter of node 'i' for 'music influence' as:

$$m_c = O_d + \sum_{i=1}^T \left(\frac{1}{d_i} \right)$$

The merit of using Dijkstra algorithm to calculate the shortest paths between nodes is its considerably low complexity, which is almost linear, and Dijkstra algorithm is just fit to be applied in the calculation of sparse graphs like ours. However, it still took more than 2 hours to run the program for the comprehensive ranking of all artists, for the extremely large number of floating point arithmetic we have to do, it will only take much longer time if we use other normal methods like the Floyd algorithm.

And the ten most influential artists calculated in this method are: **1** The Beatles, **2** Bob Dylan, **3** Chuck Berry, **4** James Brown, **5** The Rolling Stones, **6** Elvis Presley, **7** Little Richard, **8** Jimi Hendrix, **9** Muddy Waters and **10** Hank Williams.



2.3.3 Updated Version of the Comprehensive Way for Determining ‘Music Influence’

It is noticeable that the weight of indirect influence is not very convincingly fixed, namely the decision of which is rather subjective. From Wikipedia we know that in science, inverse-square law is any scientific law stating that a specified physical quantity is inversely proportional to the square of the distance from the source of that physical quantity. The fundamental cause for this can be understood as geometric dilution corresponding to point-source radiation into three-dimensional space.

And using this commonly mentioned formula format for reference, we intended to decrease the weight of indirect influence, and modified m_c as:

$$m_u = O_d + \sum_{i=1}^T \frac{1}{d_i^2}$$

And using to this index calculate the 10 most influential artists out of the data, the results are: **1** The Beatles, **2** Bob Dylan, **3** The Rolling Stones, **4** Chuck Berry, **5** Elvis Presley, **6** Jimi Hendrix, **7** James Brown, **8** Little Richard, **9** Hank Williams and **10** Miles Davis.

2.4 Simple Comparison of the Three Ranking Results

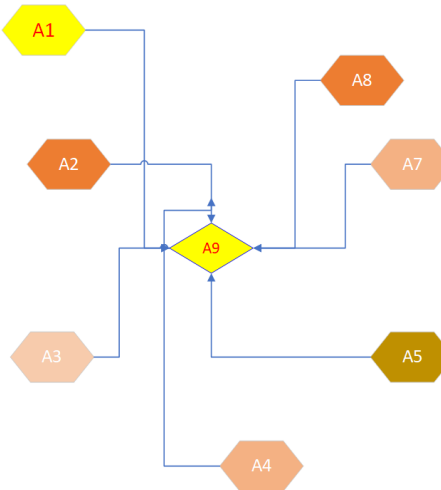
The three ranking results above are very close to each other, the first two of all rankings are consistent, they are The Beatles and Bob Dylan. In general, the artists listed from the indirect method are more “antique” compared with the direct method, for taking the secondhand influence of an artist into consideration, the earlier artist became active the more of the following generations he is likely to indirectly influence. And ranking of the comprehensive method is kind of the compromise of the direct and indirect method, for we manually minished the weight of the indirect influence.

3 Similarity between Genres, and between Artists

Below is an illustration to show the relationship between one follower and his/her influencers, we use the icon “A9” in the middle to represent the follower artist, as we can see, there are more than one influencer that had influenced the follower, either directly or through the connections with others. Our interest is to quantify the relationship of how many of these influencers are from the same genre as the follower, and how many are from different genres with him to see whether an artist is influenced more within the same genre or from other genres.

We counted the number of influencers in the data_by_artists in some typical genres, and sorted all of their followers by genres. Here is an example, we collected all the influencers from the avant genre, and their followers’ genres are shown in the next two illustrations.

And our interpretation is: the avant genre doesn’t have enough inner influences within its own genre, for more than 90% of their followers “give up” this genre and changed their musical styles into other genres like pop/rock, electronic, new age, etc.



We can also see some connections between avant genre and other genres. Since the majority of avant followers are pop/rock artists, we can say that avant genre is more closely related with pop/rock than with others.

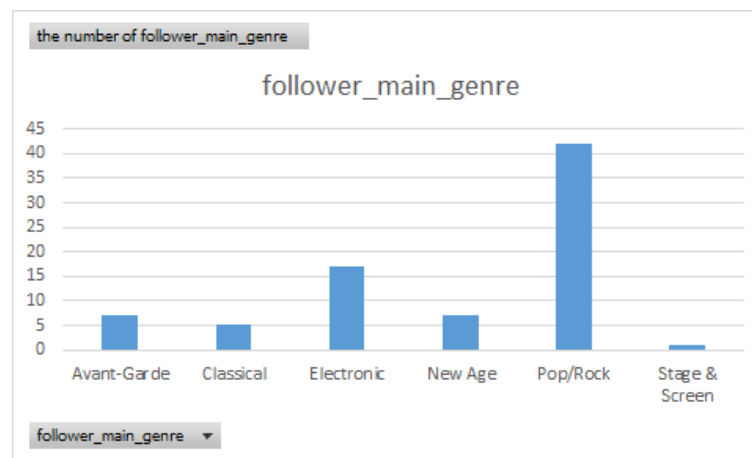


Figure 6: The categories of followers by influenced by avant artists

In the same way, we can say blue has a better inner influence than avant, about 30% followers chose to stay in the same genre. And for follower artists that set foot on different paths, the majority of them also later became pop/rock artists. We can see from the statistics that blue have a closer relationship with pop/rock and R&B than with others.

Combined these two samples, we can also draw the conclusion that the pop/rock genre is a fairly “synthetical” genre, integrated many elements from other genres to help form their own style.

The following are results of other genres we computed. If we define the proportion of same-genre followers to influencers as p_{fi} , p_{fi} can also be regarded as the “loyalty index” of a genre. we can also tell whether artists got more influenced by their own genre. The larger the p_{fi} , the more artists are influenced by their own genre. And we can draw the conclusion that the pop/rock genre is more self-influenced (has higher “genre loyalty”) compared with genres like vocal and electronic.

follower_main_genre	the number of follower_main_genre
Avant-Garde	7
Classical	5
Electronic	17
New Age	7
Pop/Rock	42
Stage & Screen	1

Figure 7: The categories of followers influenced by avant artists

follower_main_genre	the number of follower_main_genre
Blues	372
Country	6
Electronic	1
Folk	24
International	5
Jazz	17
Latin	1
New Age	1
Pop/Rock	689
R&B;	118
Reggae	3
Religious	1
Vocal	5

Figure 8: The categories of followers influenced by blue artists

3.1 Models of Similarities

3.1.1 Euclidean Distance

In mathematics, the Euclidean distance between two points in Euclidean space is the length of a line segment between the two points. It can be calculated from the Cartesian coordinates of the points using the Pythagorean theorem.[2]

In general, for points given by Cartesian coordinates in n -dimensional Euclidean space, the distance is:

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_i - q_i)^2 + \cdots + (p_n - q_n)^2}.$$

3.1.2 General Methodology

We analysed the similarity among and within the genres using the data_by_artist database. In the data set, the music attributes of an artist are divided into 14 different aspects, take out the count (the number of songs), we consider the rest 13 different arguments of music, which are danceability(da), energy(e), valence(v)... duration(du) and popularity(p). The closer the numerical value, the more similar the two artists are in musical styles.

We then regard the “music space” as a 13-dimensional Euclidean space, and used the Euclidean distances of artists to describe their similarities. Referring to the Euclidean distance definition, the similarity between artists a and b can be defined as:

$$s_0(a, b) = \sqrt{(da_a - da_b)^2 + (e_a - e_b)^2 + \cdots + (du_a - du_b)^2 + \cdots + (p_a - p_b)^2}.$$

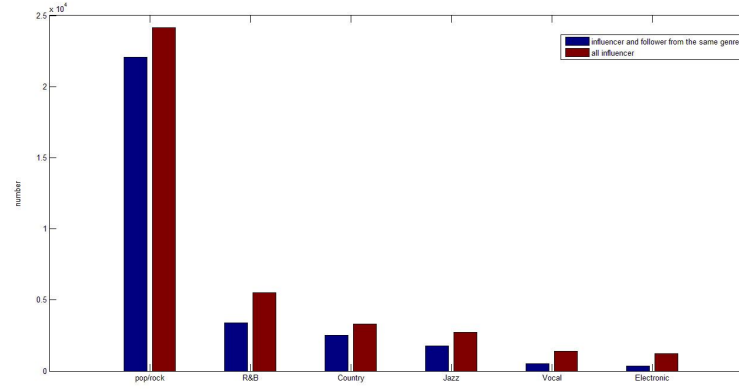


Figure 9: Variation of Music Arguments

	artist_name	artist_id	danceability	energy	valence	tempo	loudness	mode	key	acousticness	instrumentalness	liveness	speechiness	duration_ms
follower	Pat Traver	2867	0.328	0.702	0.724	180.384	-11.718	1.000	11.000	0.013	0.073	0.203	0.070	282213.500
influencer	Cream	112462	0.436	0.560	0.572	124.188	-13.135	1.000	7.000	0.267	0.116	0.288	0.044	276187.105
influencer	Led Zeppelin	139026	0.422	0.689	0.506	118.887	-10.202	1.000	2.000	0.224	0.158	0.226	0.062	327086.296
influencer	Elmore Jac	176936	0.457	0.703	0.711	129.322	-5.875	1.000	2.000	0.386	0.093	0.192	0.047	172699.500
influencer	Eric Clapton	187478	0.571	0.563	0.654	115.870	-11.202	1.000	7.000	0.358	0.151	0.345	0.052	271236.591
influencer	Deep Purple	192382	0.455	0.666	0.554	124.411	-10.464	1.000	9.000	0.085	0.180	0.184	0.048	323164.932
influencer	Jeff Beck	240865	0.480	0.646	0.598	116.257	-10.041	1.000	0.000	0.247	0.389	0.156	0.061	273635.816
influencer	Leslie West	255371	0.467	0.785	0.676	134.452	-9.172	1.000	5.000	0.100	0.028	0.110	0.054	227786.667
influencer	Jimi Hendrix	354105	0.425	0.716	0.539	120.287	-8.709	1.000	8.000	0.211	0.173	0.266	0.099	247888.965
influencer	The Yard	489303	0.481	0.638	0.648	123.367	-10.636	1.000	0.000	0.214	0.169	0.182	0.057	170852.750
influencer	Little Walter	824599	0.585	0.511	0.687	122.075	-10.781	1.000	7.000	0.742	0.116	0.166	0.059	191239.900
influencer	Robin Trower	831381	0.475	0.591	0.522	114.082	-10.838	0.000	4.000	0.149	0.075	0.159	0.051	268950.949
influencer	The Rolling Stones	894465	0.524	0.720	0.655	123.765	-7.830	1.000	0.000	0.294	0.176	0.268	0.051	229705.962
average value			0.481	0.649	0.610	122.247	-9.907	0.917	4.250	0.273	0.152	0.212	0.057	248369.569
variance			0.064	0.075	0.071	16.383	1.760	0.266	3.577	0.174	0.084	0.062	0.014	48814.798

Figure 10: Variation of Music Arguments

However, the scales of different arguments also need to be taken into account. Thus we used the maximum and minimum value to normalized all the arguments before calculating the distance:

$$z_i^* = \frac{z_i - z_{i \min}}{z_{i \max} - z_{i \min}}$$

And the modified formula:

$$s(a, b) = \sqrt{(z_{1a}^* - z_{1b}^*)^2 + (z_{2a}^* - z_{2b}^*)^2 + \dots + (z_{3a}^* - z_{3b}^*)^2 + \dots + (z_{13a}^* - z_{13b}^*)^2}.$$

Notice the magnitude of result is contrary to similarity, because if two artist's music styles are completely the same, the distance should be zero.

3.1.3 Similarity within the Same Genre

We began with sorting out the artists with the genres, then each time considered the similarity of two artists (by calculating the Euclidean distance of the two artists using the above formula), and promoted that principle to every two artist within the genre and averaging all the results to get a mean value of similarity within a genre. Assuming a genre x with n artists, the similarity within this genre is:

$$S(x, x) = \frac{\sum_{j=1}^n \sum_{i=1}^n s(i, j)}{n(n-1)} (i \neq j)$$

3.1.4 Similarities of Different Genres

Assuming we want to calculate the similarity of a genre x with n artists and a genre y with m artists, Similar with the calculation method of the prior section, the formula for calculating the similarities of different genres can be defined as:

$$S(x, y) = \frac{\sum_{j=1}^m \sum_{i=1}^n s(i, j)}{mn}$$

Also notice that the numerical pair in the calculation of one-genre similarity is order-independent, but the numerical pair in the calculation of different-genres similarity is sequential, for example the pair $s(1,1)$ here means the similarity between the first artist from genre x and the first artist from genre y , in the same manner the pair $s(1,2)$ and $s(2,1)$ have different meanings.

We calculated the similarities of 10 typical genres, the similarity matrix below is symmetric, so we only showed the upper triangle of it, and to make the data look more intuitionistic, we multiplied all of the actual results with 1000. The processed results are as below:

Similarity	pop	rb	country	jazz	vocal	blues	folk	reggae	elec	latin
pop	0.813	0.808	0.712	0.884	0.867	0.864	0.885	0.877	1.021	0.868
rb		0.810	0.703	0.846	0.892	0.842	0.885	0.858	1.024	0.816
country			0.605	0.745	0.738	0.744	0.853	0.806	1.004	0.738
jazz				0.743	0.791	0.843	0.786	0.995	1.115	0.864
vocal					0.620	0.845	0.676	1.078	1.231	0.894
blues						0.845	0.845	0.925	1.103	0.858
folk							0.729	1.026	1.189	0.885
reggae								0.855	1.030	0.891
elec									1.057	1.069
latin										0.870

The similarity matrix is the most important conclusion in the whole paper.

It contains tremendous information in the similarity among different genres and within one genre. If space allows it, we can do dozens of subconclusions from the matrix. Here we only list five examples for reference:

- The genre with the largest inner similarity is country, meaning the feature of country style music is with the brightest characteristics and easiest for people to distinguish.
- The genre with diminutive cross-genre similarities (shown with rather large values in matrix) like the electronic music, also suggests that it's a rather distinct style from others.
- Genres with larger cross-genre similarities than inner similarities, like Latin music's inner similarity is smaller than with R&B, imply that people may be hard to tell Latin music from R&B music and vice versa.
- The similarities of pop/rock with others and with itself are all quite close. As we have mentioned in the preceding section, this data better explained that pop/rock music is likely to be a combination and derivative of all other previously existing genres.

- The inner similarities are larger cross-genre similarities in general.

For purpose of better visualizing data of similarity matrix, we also drew a thermodynamic chart in below. The genres with lower similarities have a color toward red trend. Note that the graph is also symmetric Because of the attributes of similarity matrix. The color of the downward sloping diagonal is lighter (tends to be in blue series), means inner similarities are more significant. Also the row and column which represent electronic music has the darker color, in agreement with the conclusion mentioned above.

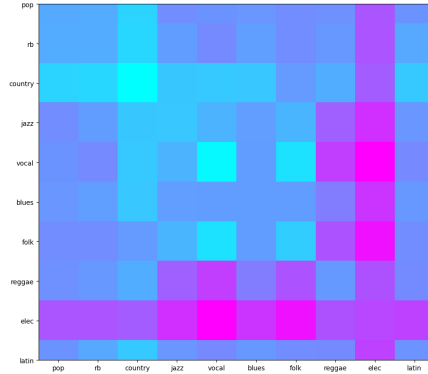


Figure 11: thermodynamic chart

3.2 A Supplementary Methodology to Measure Similarity

3.2.1 Pearson Correlation Coefficient

Pearson's correlation coefficient is the covariance of the two variables divided by the product of their standard deviations:[3]

$$\rho_{X,Y} = \frac{\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

where: σ_X and σ_Y are standard deviations of X and Y

μ_X and μ_Y are the means of X and Y

\mathbb{E} is the expectation.

In a sample, the Pearson correlation coefficient is commonly represented by r_{xy} . Given paired data $\{(x_1, y_1), \dots, (x_n, y_n)\}$, r_{xy} can be calculated as:

$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}.$$

Different from the Euclidean distance, the correlation coefficient is a positive index for measuring similarity.

We then attempted to calculate the correlation coefficient of each two artists as a supplementary to better testify the result from our first method. However, the results are not very desired, for the results came very near each other.

3.3 The Practical Influence of the Alleged Influencers to Followers

We selected ten followers randomly, and then studied the consistencies of music arguments between a follower and all of his direct influencers. First we calculated the Euclidean distance and found them rather lower than truly random artists. We found that the similarity between influencer and followers are higher the normal situations. For example, PJ Harvey's distance to his influencers average by 0.640 which is significantly lower than other than the distance between two randomly selected artists-but we should aware that most of his influencers and himself is from the same genre. And then by calculating the average value of arguments of the influencers, and compare the corresponding values of the follower, we researched to what extent a follower is practically affected by his alleged influencers.

As our example below shows, the arguments' mean value from the influencers are close with those of the follower, also by calculation (using the Euclidean distance mentioned), we found that the similarity among one follower and his/her influencers are more apparent than with similarity between the follower and random artists in his/her genre. So we can say the alleged influencers do have a practical influence upon the follower.

We also computed the variance of all influencers and the follower, and found them to be fluctuating in an acceptable range, what is also noteworthy is that the arguments of "tempo" and "loudness" are extraordinarily small in relative size. Therefore, we consider these two characteristics of music to be more "**contagious**", for these two aspects exert a much deeper and firmer influence on an artist than other characteristics.

We suggest that the mean value of these features of influencers is just fitted to the same feature of the follower. We tried to use hypothesis test. But we cannot confirm the distribution is normal. It is a pity.

3.4 Distinct Characteristics of Genres

We normalized the 13 music arguments, and calculated the mean values of which in the typical genres to determine the distinct Characteristics of genres. Here are some interesting results:

- The features of electronic music are most prominent, with high values in danceability(0.6), energy(0.6) and instrumentness.
- The features of pop/rock are high popularity and energy(0.6).
- The feature of Latin is high valence(0.64).
- The feature of folk is low energy(0.3).

4 Variation Tendency of Genres Through Time

4.1 The Change in Numbers of Artists in a Genre

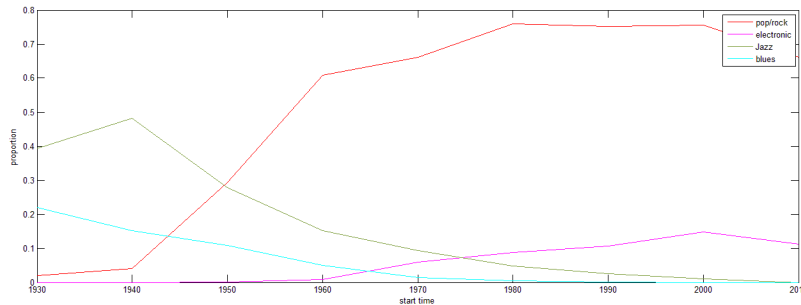


Figure 12: Variation in Number of Artists

we picked a few typical genres to study as depicted in the illustration above.

- **Pop/Rock:**

As we can see the pop/rock genre grew explosively in the last century, especially from about 1940 to about 1960, multiplied almost 14 times in its relative size with other genres.

- **Electronic:**

The electronic genre is rather a newly-developed genre, started to emerge as late as in 1960, but is only about one seventh to the size of pop/rock.

- **Jazz:**

Jazz music reached its top in reached its summit in proportion in around 1940, and has begun to fade out since then.

- **Blue:**

The blues has been seceding out of the stage of music world for almost a century, and has a tendency to become only a memory in the history.

4.2 The Implying Factors in the Changing of Music Styles

From the line chart below we can see the overall trend of different music features. And in the following sections, we will explain the variance and tendency about different features. To calculate the change, we used **difference method**

$$D_i = \frac{F_{n_i} - F_{n_{i-1}}}{1}$$

After the quantization by using the difference method, we can verify the conclusion that we seen in the chart.

- **Energy**

From the line chart, we can know that the throughout the whole 20th century, the energy in the music is going up. Furthermore, we find out that in the age of 1960s, the increase of energy in the music is bigger. We assume that it is strongly connected to the society and culture in those days.

We know that 1960s is age of rebellion and hippies. They always tend to shout out their opinions and they are never afraid to fight against the government. In the mean time, rock music, as a unique genre, is coming to the main stream of the music world. The rebellious spirit become the catalyst of music. In return, music can sensitize people and make them aware of important events and take actions.[4] Therefore, the energetic music become more and more popular. Music in this era encourages people to fight for themselves and resist the injustice. Only with energy can people be inspired.

- Acousticness

In the old time, classical music focus on melody and timbre. People go to concert to enjoy themselves. However, nowadays musicians compose songs mainly to express themselves. The situation started in 1960s, exactly at the same time, we find that the feature ‘acousticness’ is declining. We know that classical music is known for its beautiful melody and people always immerse themselves in it. But classical music is becoming less popular in modern time. In the contrary, the rising genre of music world-pop&rock is not that acoustic. To some people, it is even regarded as noise. After all, it is just about the change of people’s taste.

- Speechiness

In the data-by-year table, we also find out that compared to the age of 1950s and 1960s, the index of speechiness of 1970s is higher. We assume that it is concerned with the born of rap. A famous comedian Rudy Ray Moore, recorded his comedy in a fast-speaking tone.[5] His career come to the prime in 1960s. Then his recordings and comedies become popular and wide-accepted, it derived into rap. And Rudy Ray Moore himself became the “Godfather of Rap”. Meanwhile, in the african american community in Bronx, New York, developed rap in the strict sense. Since the rap becomes popular as a music genre. No wonder the speechiness of music is getting higher in 1970s.

- Popularity

From the line chart we can directly know that the popularity of music is increasing significantly throughout the time. It is so obvious that we don’t even need to calculate the difference. We suggest that it is due to the increase of the quality of life. With the increase of productivity, people are living a happier life. While people are having more spare time, they must find something to fulfil themselves. Also, after people get satisfied in material life, they always tend to find their pleasure in spiritual life. Without doubt, music is an indispensable part of spiritual world. Therefore, people listen to music more often. In this case, the popularity of music rises.

Besides, we hold the view that the innovation of technology can also affect the increase of popularity. From the chart we can see that in 2000s the popularity increased significantly. First, the Internet gives a huge help in spreading music. We listen to music in our app everyday, it’s hard to imagine for us that how would we enjoy music without Internet. Secondly, electronic device is a big help to music. With modern recording devices and electronic instruments, artists can compose their songs more easily. And with the aid of newest headphones, people can enjoy the best music wherever they are. Technologies definitely helped us better enjoy music and made us more like music.

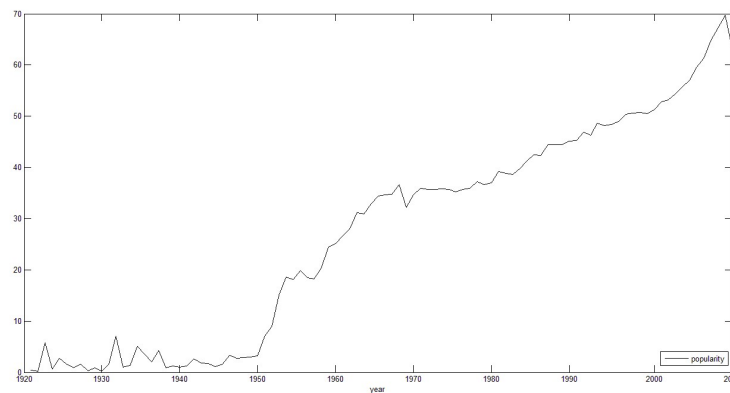


Figure 13: Variation of popularity

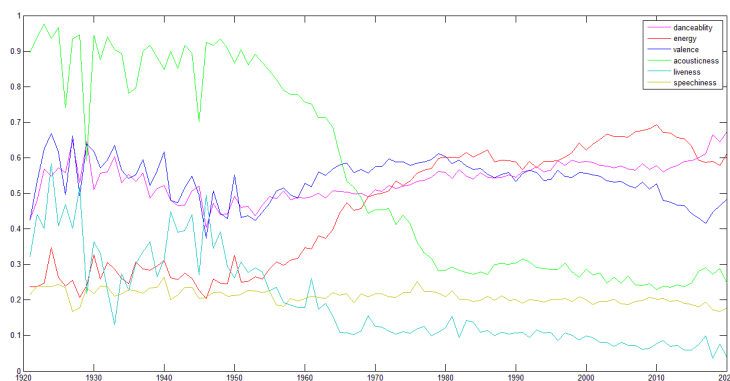


Figure 14: Variation of Music Arguments

5 Indicators for Revolutions in Musical Evolution

5.1 Music Arguments Indicators

We can also harness the iconograph above to estimate possible arguments implying musical revolutions, which are the eras that music features turned sharply, as shown, it is about the 1950s and 1960s. And the arguments indicators are mainly acousticness, liveness and energy for the magnitudes of the changes are very arresting in these features. Just as we claimed in the previous section. What we seen in the graph has also been proved by calculating the difference.

5.2 Indicators for Dynamic Influencers

In this section, in order to simplify the model, we presume the “popularity” attribute in data_ by_ artists manifests the quantitative value for “dynamical influence”.

To decide indicators for dynamic influencers, we first selected the top 100 artists with highest popularity attribute. And after kicking out those who are not influencers in the influence_data, there

are still 99 artists left. We calculated the average values of their rest 12 musical arguments (count not included). Next we picked 498 followers, and calculated the average values of musical arguments of them.

And we can conclude that in general, artists whose music have higher values of acoustiness, instrumentality, duration, lower energy, key and speechiness, etc, are more likely to become “dynamic influencers”.

5.3 Great Artists as revolutionists

As we already knew that 1950s and 1960s is the age of “music revolution”. We find some artists from 1950s and 1960s who is most influential in our network: the Rolling Stone, Chuck Berry and Elvis Presley.

We analysed their music features. Among all of the artists at the same age, their popularity is significantly higher. Specially, Chuck Berry’s music has higher tempo. And both rolling stone and Chuck Berry’s songs has more energy. Because we already knew that the rising of energy and tempo is the feature of the age and they are the most influential artists of all time. To some extent, we can say they are leading the age. They are the leader of the music.

6 Integral Evaluation of Strengths, Weaknesses and Conclusions

Strength

- We make fully use of the influence data and make it into a influence-following network. The evaluation method can precisely measure the influence of different artists. We didn’t just find the adjacent node. However, we take indirect influence into consideration. Thus the evaluation result of the network is satisfying. The highly ranked influencer is also famous and known to us all.
- We combined the data and trend with the background of the era. There are revolutionary moment of music in the history. And from the time of the change, we can look back into the history. The social event and the people’s status of life at that time is concerned to the change of music.
- Due to the large quantity of indicators of music. It’s hard to find a way to evaluate the similarity between genres and artists. The statistic tool we utilized can take all of the indicators into consideration and evaluate them in a comprehensive way.

Weakness

- Our model is totally based on the data that the problem provided. In this case, our model’s robustness is highly relied on the accuracy of the dataset. Therefore, if the data is not reliable, our conclusion cannot be assured.
- We do not know that whether the value of different features obey normal distribution or not. If the value is normal distributed, we can utilize *hypothesis test* to test if follower’s feature is fit to the influencer’s feature. Hypothesis test is a more logical tool, but it’s a pity that we can’t confirm the data fits normal distribution. So what we could do is just calculate the variance of the data and compare the feature of follower with the mean value of the feature of influencers.

7 A Letter To ICM

Dear ICM Society:

As an inalienable part of human society, music is not only for entertaining people, but also plays an important role in the process of human society development. According to your request and the data you provided us, our team has completed the model to measure the influence of music, and explored the trend of evolution and revolution of musicians and genres.

First of all, in order to determine the music influence of artists, the network reflecting the relationship between influencer artists and followers artists is constructed. By analysing the network, we find some of the greatest artist who is most influential. They are indispensable in music history. In order to show the music leader of time, we also built influence networks at different ages.

Secondly, in order to measure the influence of different artists, through correlation analysis, this paper analyzes the similarity within one and between different genres according to the music characteristics of different artists. The results show that the correlation coefficient of country within the genres is the largest, that is, the music characteristics of artists within the genres are relatively similar, and the degree of mutual influence is large. From the perspective of inter genre analysis, vocal has a high degree of correlation with other genres, which indicates that this genre has a high proportion of reference from other genres, and reflects the cooperation and learning of artists among the genres.

In the result we calculated, we found several influential features have undergone great changes in the 1950s and 1960s.

Influenced by the impact of war, cultural changes and the emergence of new technologies, American music is constantly evolving. The trend of music is changing from classical music to secular music. At the same time, new elements of music from other countries are absorbed to realize the international development of music.

Music, as the crystal of human wisdom, is a sign of the evolution of human civilization. After making the living is no longer the problem for humans, we start pursuing the peace and satisfaction of our spirit. As we all know, human's desire can never be satisfied, we all desire the better quality of life after we satisfy the older one. So in order to achieve the real happiness of life, The only way for us is to improve the level of our spirit through the graceful music. In such case, music is of vital importance for individual happiness.

The model we built is totally based on your given data. However, if we want to add some new music genres or song characteristics, our models and conclusions is going to be change. When establishing the influence model, we selected several important musical characteristics for similarity analysis. If the newly given data includes some very representative features. We can add it into the calculation of network influential. Also, our analysis of music development is based on existing data and genre popularity and characteristics. If the new data has a greater impact on the current proportion of genre popularity over the years, then our conclusions will have to be changed significantly.

Good music in a rightful place makes people fit in the atmosphere more easily.

Sincerely yours,
Team 2100545

References

- [1] Four datasets given with the problem.

[2] [wikipedia.org Euclidean_distance](#)

[3] [wikipedia.org Pearson_correlation_coefficient](#)

[4] Music as a cultural context for change : a study of Rock Music of the 1950s and 60s.Sarah Toraason

[5] [wikipedia.org Rudy_Ray_Moore](#)

Appendix

python codeto build network

```
import pandas as pd
from pandas import DataFrame
import matplotlib.pyplot as plt
data=pd.read_csv("influence_data.csv",header=None)
import networkx as nx
import numpy as np
data2=data.drop([0])
dict1=
```

```
G2 = nx.DiGraph()
for i,r in data2.iterrows():
G2.add_node(r[0])
G2.add_node(r[4])
G2.add_edge(r[0],r[4],weight=0.01)
```

python code to calculate distance and similarity

```
from sklearn.preprocessing import MinMaxScaler
mm=MinMaxScaler()
data.iloc[:, 2:14] = mm.fit_transform(data.iloc[:, 2:14])
dis=[]
import numpy as np
for i in range(80):
for j in range(80):
ja.append(np.linalg.norm(np.array(ele.iloc[i,1:11]).astype(float)-np.array(la.iloc[j,1:11]).astype(float)))
print(np.mean(ja))
```

python code to calculate the influence of artists

```
for i in G2.nodes():
a=0.0
for j in G2.nodes():
if(i!=j and nx.has_path(G2,i,j)):
a+=1/nx.shortest_path_length(G2, source=i, target=j)**2
dict1[i]=a
dict1=sorted(dict1.items(),key = lambda kv:(kv[1]),reverse=True)
print(dict1)
```

python code to sort data

```
ax=-1
w=True
for i,r in datareg.iterrows():
    ax+=1
    w=False;
    for i1,r1 in dataare.iterrows():
        if(r1[0]==r[0]):
            w=True
            break;
    if(w==False):
        datareg.drop(index=ax,inplace=True)
```