

Towards Data Analytics for Music Influence: A Network Science Perspective

Summary

Nowadays, music seeps into our daily life and plays a dominant role in human society. On the other hand, social transition, political revolution and technological progress make a great difference to the genre and style of music in return. To study music influence and recognize revolutions in music timeline, we model a series of networks and subnetworks to demonstrate influence among artists, extend an indicator to quantize similarity between music, artists and genres, and then conduct work on the change of artists, the evolution of genres, the leaps of music and the influence of external factors.

We first build a network to visualize the influence among artists, where nodes correspond to artists and directed edges correspond to the influence relationships. To polish this network, we introduce year gap to weigh different ages, distance between nodes to simulate the influence score on a certain path. Thus popularity of every node can be acquired by summing scores of all directly or indirectly influenced nodes together.

As for the measurement of music similarity, we extract ten key music characteristics to form a feature factor, and use Euclidean distance to evaluate the distinction between music. Moreover, we extend the similarity measurement method to artists and genres, finding that music within genre shows higher similarity than between genres. Consistently, artists behave more similarly with their influencer, whose music shows small differences in speechiness and instrumentality, mainly due to the high probability that artists tend to choose the same genre as their influencers.

To illustrate the evolution of artist, music and genre, we combine influence and similarity analysis together to figure out milestones in music history. It is found that 1956-1960 and 2016-2020 saw a major transition credited to several celebrities. Certain genre and artist are taken for example to justify the model.

The model established in this paper makes full use of the given data to describe and quantify music influence and similarities. It is validated by statistical tests and verified by KNN (k -nearest neighbors) algorithm. The development process of music and genre coincides with historical facts and reflects on social, cultural, political and technological movements.

Keywords: influence network; music similarity; dynamic influence

Contents

1	Introduction	2
1.1	Problem Background	2
1.2	Restatement of the Problem	2
2	General Assumptions	2
3	Symbols	3
4	Model Methodology	3
4.1	Network of musical influence	3
4.2	Measure of music similarity	5
5	Model Analysis	7
5.1	Analysis of genre features	7
5.1.1	Similarity	7
5.1.2	Influence	9
5.1.3	Statistical characteristics	10
5.1.4	Conclusion	11
5.2	Analysis of influencer's impact	11
5.2.1	Similarity	11
5.2.2	Role of different characteristics	11
5.2.3	Conclusion	13
5.3	Analysis of music revolution	14
5.3.1	Music revolution	14
5.3.2	Representative revolutionaries	14
5.4	Analysis of the influence processes of musical evolution	15
5.4.1	Dynamic network of musical influence	15
5.4.2	Case study: Jazz	16
5.4.3	Case study: Michael Jackson	18
6	Conclusion	19

6.1	Summary of Results	19
6.2	Strengths	19
6.3	Possible Improvements	20
7	Reference List	20
	Appendices	21

1 Introduction

1.1 Problem Background

Music is the art of arranging sounds in time to product a composition through the elements of melody, harmony, rhythm, and timbre[1]. Throughout history, music has been always evolving and new forms or styles of music are still emerging today. During the history of music, some significant artists influenced others' musical work. For example, Elvis Presley, also known as the "King of Rock and Roll", was a central figure in the development of rockabilly, and he had a great influence on later artists, including John Lennon, a member of the Beatles, who once proudly said that "Nothing really affected me until Elvis." [2]

Our team has been employed by Integrative Collective Music (ICM) Society to develop a method to identify the influence of previously produced music and external factors on new music and musical artists.

1.2 Restatement of the Problem

- Create a directed network to show the relationship between influencers and followers, raise measurement to quantify *music influence* in this network and instantiate it in a subnetwork.
- Propose an indicator to measure music similarity. Using this indicator to analyze music similarity and relationship between and within genres.
- Examine the impact of identified influencers on followers' music. Figure out whether there is a certain characteristic where influencer and follower show a high similarity.
- Portrait the development of artist, genre and music over time. Identify characteristics that can signify the advent of revolutions. Combine the analysis with influence network to seek for artists leading and contributing to these revolutions. Figure out whether these revolutions can integrate with cultural influence, social change, political transition or technological breakthrough.

2 General Assumptions

- Every artist produces music of only one genre, and he/she will not switch to another genre over time.
- Music of the same genre has a high similarity, while music of different genres has a low similarity.

- Music of artists and their followers has a high similarity.
- Revolutions in music will lead to significant rise in music difference.

3 Symbols

Symbols	Definition
$L_i(A)$	The set of nodes with the distance i from A
$S(A)$	The influence score of node A
\mathbf{x}	Feature vector
s	Similarity between music
dif_a	Difference vector between artists
S_a	Similarity between artists
S_g	Similarity between genres
I_a	Influence between artists
I_g	Influence between genres
$A^{(i)}$	Average features vector from year i to year $i + 4$
$\mathbf{x}^{(j)}$	Features vector in year j
$C^{(i)}$	Contribution of artist i
$M^{(i)}$	All music works from year i to year $i + 4$
p_s	Popularity of music work s

4 Model Methodology

4.1 Network of musical influence

In order to illustrate the influence that artists have on one another, we construct a network of musical influence (known as **influence network**) where influencers are connected to their followers.

In this network (or directed unweighted graph), each node is assigned to an artist, and each artist has his unique artist ID. We check *influence_data* for relationships and if B is influenced by A, we draw an edge from node A to B.

After constructing the network, we develop a criterion to assess the influence of artists by assigning a **score** to each node. The score is divided into two parts, of which the former is determined by the number of artists influenced by him, while the latter is determined by the active year "gap" between other artists influenced and him.

To better demonstrate this, we take an example. In Figure 1, we detach a subnetwork from the original network by spanning a forest rooted in A, in order to calculate the score of A.

Definition 4.1. $L_i(A)$ is the set of nodes with the distance i from A .

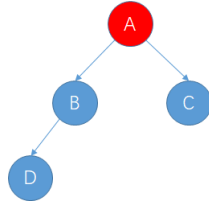


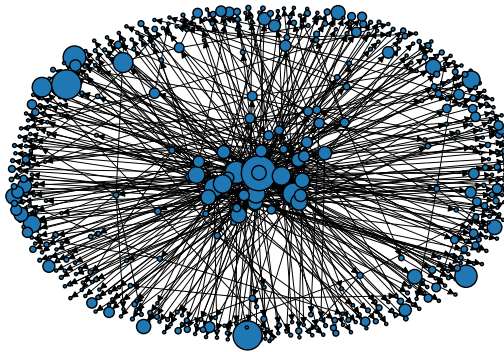
Figure 1: A simple graph

In this simple graph (shown in Figure (1)), there are 1 node(A) with distance 0, 2 nodes(B and C) with distance 1, and 1 node(D) with distance 2. So the score of A can be calculated as ¹

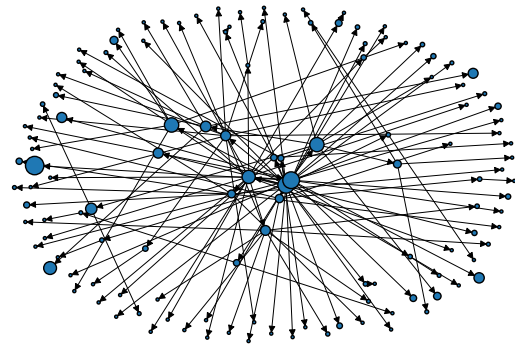
$$S(A) = \sum_{i=0}^{\infty} \alpha^i [|L_i(A)| + \beta \sum_{K \in L_i(A)} (\text{year}(K) - \text{year}(A))] \quad (1)$$

We can complete the calculation by improved **Breadth First Search** algorithm. After the calculation, we can obtain a list of scores of influence of all artists.

Now we take the forest rooted in artist 678339 and 792370 as example. In Figure (2), each circle represents a node and its radius is in proportion to its score of influence. Figure (a) has much more nodes than figure (b) does, indicating that artist 678339 has a higher influence than artist 792370.



(a) Forest rooted in 678339



(b) Forest rooted in 792370

Figure 2: Comparison of two forests rooted in different artists

After calculation, we list top five artists ranked by scores of influence:

1. **The Beatles (Pop/Rock)** were an English rock band and they are regarded as the most influential band of all time. They sparked the "Beatlemania" phenomenon, the fanaticism surrounding the band in the 1960s. ²

¹year(.) denote the year (or decade) when the artist began their music career.

²The Beatles, https://en.wikipedia.org/wiki/The_Beatles

2. **Bob Dylan (Pop/Rock)** is an American singer-songwriter, author and visual artist, and is regarded as one of the greatest songwriters all of the time. He was awarded the Nobel Prize in Literature in 2016. ¹
3. **The Rolling Stones (Pop/Rock)** is an English rock band whose members are of the best-selling music artists all of the time. The band has won three Grammy Awards and a Grammy Lifetime Achievement Award, and they were inducted into the Rock and Roll Hall of Fame in 1989. ²
4. **Elvis Presley (Pop/Rock)** was an American singer, musician and actor, and he is often referred to as the "King of Rock and Roll". He is the best-selling solo music artist of all time, and he won three Grammy Awards, received the Grammy Lifetime Achievement Award at age 36. ³
5. **Chuck Berry (Pop/Rock)** was an American singer, songwriter and guitarist. He was also one of the pioneers of rock and roll music by the name of "Father of Rock and Roll". He was among the first musicians to be inducted into the Rock and Roll Hall of Fame. ⁴

4.2 Measure of music similarity

According to previous assumptions, we ought to develop such a measure of music similarity that music of the same genre has a high similarity, while music of different genres has a low similarity. After some attempts, we determine to introduce **Euclidean Distance** to measure music similarity.

Features of a piece of music can be concluded as a **feature vector** \mathbf{x} , with dimensions associated with danceability, energy, valence, tempo, loudness and other features. We select ten features and the vector can be expressed as

$$\mathbf{x} = (x_1, x_2, \dots, x_{10}) \quad (2)$$

where x_1 for tempo, x_2 for loudness, x_3 for duration, x_4 for danceability, x_5 for energy, x_6 for valence, x_7 for acousticness, x_8 for instrumentalness, x_9 for liveness, x_{10} for speechiness.

The distance of two vectors \mathbf{x} and \mathbf{y} can be measured as :

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{10} (x_i - y_i)^2} \quad (3)$$

For two pieces of music with feature vectors \mathbf{x} and \mathbf{y} , their **similarity** can be defined as

$$s(\mathbf{x}, \mathbf{y}) = \frac{1}{d^2(\mathbf{x}, \mathbf{y})} = \frac{1}{\sum_{i=1}^{10} (x_i - y_i)^2} \quad (4)$$

¹Bob Dylan, https://en.wikipedia.org/wiki/Bob_Dylan

²The Rolling Stones, https://en.wikipedia.org/wiki/The_Rolling_Stones

³Elvis Presley, https://en.wikipedia.org/wiki/Elvis_Presley

⁴Chuck Berry, https://en.wikipedia.org/wiki/Chuck_Berry

We have defined similarity between different pieces of music, and now we are going to extend the concept of similarity to artists and define similarity between artists.

Definition 4.2. Let p and q be two artists and S_p, S_q are sets of music created by corresponding artists. We define difference vector between artist p and q as

$$\mathbf{dif}_a(p, q) = \left| \sum_{\mathbf{x} \in S_p} \frac{\mathbf{x}}{|S_p|} - \sum_{\mathbf{y} \in S_q} \frac{\mathbf{y}}{|S_q|} \right| \quad (5)$$

Thus the similarity between artist p and q can be defined as

$$S_a(p, q) = s\left(\sum_{\mathbf{x} \in S_p} \frac{\mathbf{x}}{|S_p|}, \sum_{\mathbf{y} \in S_q} \frac{\mathbf{y}}{|S_q|}\right) = \frac{1}{d^2(\mathbf{dif}_a(p, q), \mathbf{0})} \quad (6)$$

In other words, we take the center of all feature vectors of music created by a artist as his feature vector, and calculate similarities between feature vectors of artists as we do above.

To verify the effectiveness of the measure above, we execute KNN algorithm[3] on the artists dataset with Euclidean distance as the measure to classify all artists into 19 genres. We utilize k-fold cross validation to calculate the accuracy of KNN and achieve a maximum accuracy of 55% (shown in Figure 3).

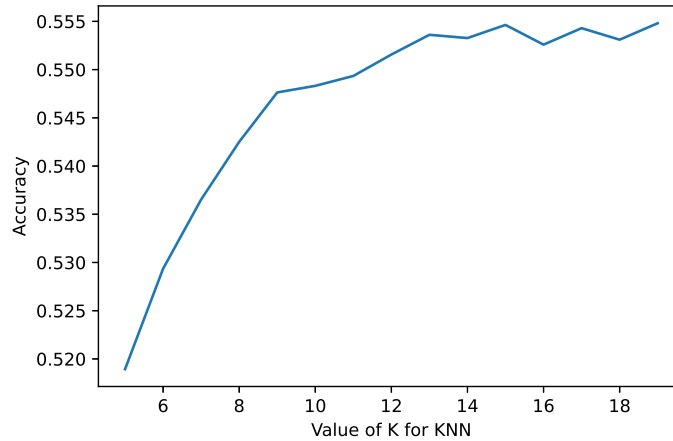


Figure 3: Accuracy of KNN with different values of K (neighbors)

5 Model Analysis

5.1 Analysis of genre features

5.1.1 Similarity

In 4.2 we have defined similarity between music pieces and artists, and now we are going to extend the concept of similarity to artists and define similarity between genres.

Definition 5.1. Let A and B be two genres and S_A, S_B are sets of artists from corresponding genres. We define similarity between genre A and B as

$$S_g(A, B) = \frac{\sum_{p \in S_A} \sum_{q \in S_B} S_a(p, q)}{|S_A||S_B|} \quad (7)$$

In addition, we sort out ten artists with the highest influences from a genre as representatives of that genre. Let A and B be two genres and S'_A, S'_B are sets of representative artists from corresponding genres. The similarity between genres A and B can be calculated as

$$S_g(A, B) = \frac{\sum_{p \in S'_A} \sum_{q \in S'_B} S_a(p, q)}{|S'_A||S'_B|} \quad (8)$$

To verify the effectiveness of the measure mentioned above, we check whether it is consistent with the assumption by calculating music similarities between and within genres. We utilize *data_by_artist* and **normalize** it. After that, from each genre we sort out ten artists with the highest influences as representatives of that genre. We leave out the genre Children's because there are only four artists in total. Then we evaluate similarities between artists from the same genre and use the average values as music similarities within that genre, and evaluate similarities between artists from different genres and use the average values as music similarities between those genres. Finally we get a similarity matrix of 10×10 size and plot a heatmap to visualize it (Figure 4).

From the figure, we can easily find that on-diagonal blocks have deeper colors, indicating that music of the same genre has relatively high similarities; off-diagonal blocks have lighter colors, indicating that music of different genres has relatively low similarities. The result corresponds to our assumption, so this measure of genre similarity is reasonable.

For example, this heatmap shows a strong correlation between R&B and Pop or Rock. In history, R&B music contributed to the development of rock music, and there are artists who mix rock and roll and R&B as "British rhythm and blues". Contemporary R&B also combines rhythm and blues with elements of pop, soul, funk, hip hop and electronic music.¹

¹Rhythm and blues, https://en.wikipedia.org/wiki/Rhythm_and_blues

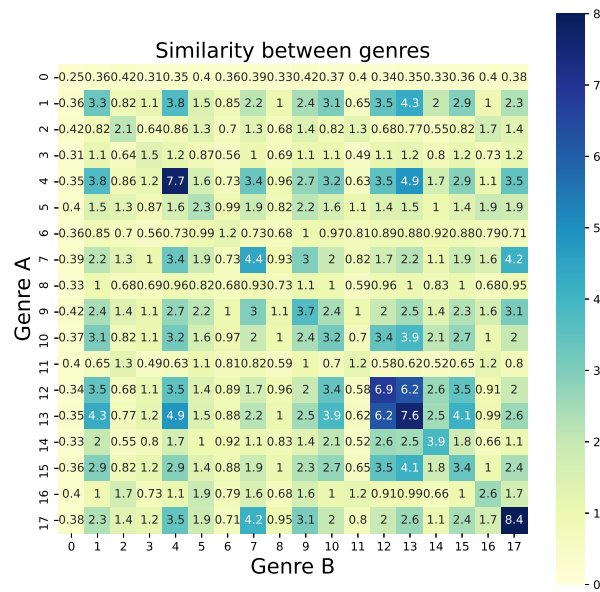


Figure 4: Heatmap of similarity matrix between genres (represented by IDs in Table 1)

Table 1: Genres and corresponding IDs

Genre	Genre ID
Avant-Garde	0
Blues	1
Classical	2
Comedy/Spoken	3
Country	4
Easy Listening	5
Electronic	6
Folk	7
International	8
Jazz	9
Latin	10
New Age	11
Pop/Rock	12
R&B;	13
Reggae	14
Religious	15
Stage & Screen	16
Vocal	17

We also conduct a more sensible analytical method named **ANOVA** (Analysis of Variance). We randomly sample 100 pairs of artists from the same genre and 100 pairs of artists from different genres, and calculate their similarities between and within genres. The result of ANOVA shows that artists from the same genre have high similarities than those who come from different genres, with a confidence level of 95%.

5.1.2 Influence

We have built an influence network and set up a method to determine influences of artists and we are going to extend the influence of similarity to genres. In the network, an artist can diffuse his influence to another specified artist by different paths and for an artist with a high influence, the number of paths is extremely large. So we apply a simplification to the influence from artist p to q as

$$I_a(p, q) = \gamma^{\text{dis}(p, q)} \quad (9)$$

$\text{dis}(p, q)$ is the distance from artist p to q (Noting that the network is a directed unweighted graph), and we can easily calculate distances by **Dijkstra** algorithm. Now we introduce the definition of influences between genres

Definition 5.2. Let A and B be two genres and S_A, S_B are sets of artists from corresponding genres. We define influence from genre A to B as

$$I_g(A, B) = \frac{\sum_{p \in S_A} \sum_{q \in S_B} I_a(p, q)}{|S_A||S_B|} \quad (10)$$

In addition, we sort out ten artists with the highest influences from a genre as representatives of that genre. Let A and B be two genres and S'_A, S'_B are sets of representative artists from corresponding genres. The influence from genre A to B can be calculated as

$$I_g(A, B) = \frac{\sum_{p \in S'_A} \sum_{q \in S'_B} I_a(p, q)}{|S'_A||S'_B|} \quad (11)$$

By calculation we get a influence matrix of 10×10 size and plot a heatmap to visualize it (Figure 5).

From the figure, we can easily find that on-diagonal blocks have deeper colors, indicating that the same genre has relatively high influences; off-diagonal blocks have lighter colors, indicating that different genres has relatively low influences. The result corresponds to our assumption, so this measure of genre influence is reasonable.

In this heatmap, R&B and jazz music have a strong influence on Reggae. Reggae is a music style that was strongly influenced by traditional mento as well as American jazz and rhythm and blues, which corresponds to the evidence in the heatmap.¹

¹Reggae, <https://en.wikipedia.org/wiki/Reggae>

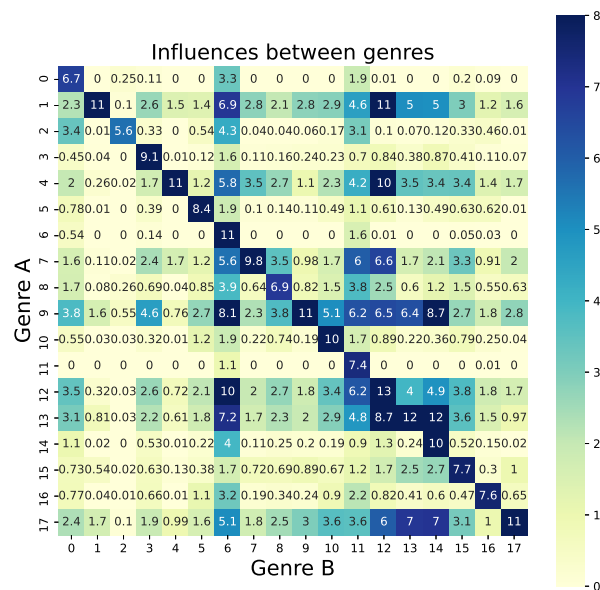


Figure 5: Heatmap of influence matrix between genres (represented by IDs in Table 1)

5.1.3 Statistical characteristics

Each genre has its own feature, and we try to analyze it by investigating genres' statistical characteristics. We compute average values of genre feature vectors and draw average charts in Figure 6.

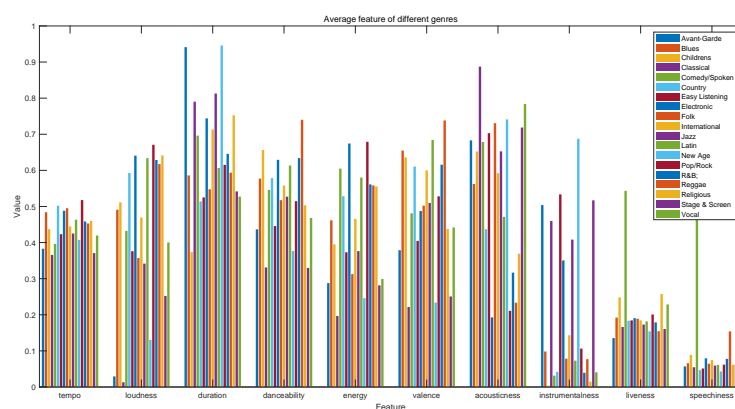


Figure 6: Average feature of different genres

Case study: In charts we can see that Children's genre has relatively high danceability, valence and acousticness, as well as low duration, indicating that

children songs are generally brief, positive, without technology enhancements or electrical amplification and suitable for dancing.

Pop/Rock genre has relatively high energy and valance, with low acousticness, because this kind of music is characterized by a strong rhythmic element and a reliance on electronic amplification for their performance.

Comedy/Spoken genre has an especially high level of liveness and speechiness, for this kind of music is often performed live and uses more words to express feelings.

5.1.4 Conclusion

From the analysis of genre features above, we can conclude that

- Artists from the same genre tend to have more similarities, while artists from different genres have less.
- Artists from the same genre tend to influence each other more, while there appears to be less influence between artists from different genres.
- Statistical characteristics vary from genres to genres, and we can also utilize them to distinguish genres.

5.2 Analysis of influencer's impact

5.2.1 Similarity

The most direct method to study influencer's impact on their followers is to measure the similarity between them. To run this method, we randomly select 100 influencer-follower pairs in *influencer_data*, and for each follower, we randomly select an artist who is not one of his influencers as control group. Using their characteristics data in *data_by_artist*, we can calculate similarities of influencer-follower pairs and control group. Then Welch's t test is used to examine whether there is a significant difference of similarities.

From Welch's t test in figure 7, it can be concluded that followers tend to have higher similarities with influencers than others.

5.2.2 Role of different characteristics

From 5.2.1 we can see that prestigious musicians make their followers' music more similar with theirs. What we are more concerned of is the role different characteristics play in the influencing process. Based on the study in 5.2.1, we not only calculate the similarity of influencer-follower pair, but also focus on

Robust Tests of Equality of Means

dis

	Statistic ^a	df1	df2	Sig.
Welch	203.930	1	1450.311	.000
Brown-Forsythe	203.930	1	1450.311	.000

a. Asymptotically F distributed.

Figure 7: Welch's t test result, where p is much less than 0.05

the difference of each characteristic between influencers and followers, which can be quantized as difference vectors (see in Definition 4.2). The average of each characteristic in the difference vector is shown in Figure 8:

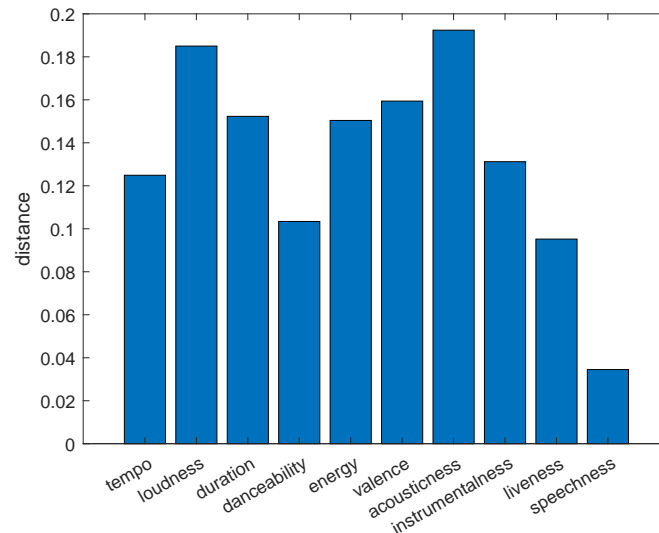


Figure 8: Average characteristic difference

Note that different characteristics may be subject to different distributions, which will affect their values in difference vectors. As a compensation, we standardize the data in *data_by_artist*, and use standardized data to calculate the variance-corrected difference vector shown in Figure 9.

From Figure 9, we can see that the music of influencers and followers shows an obvious similarity in instrumentalness and speechiness, while other characteristics of them seem to differ from each other. In fact, we go through all 42770 influencer-follower pairs, unsurprisingly to find that there are 32753 pairs in which follower holds the same genre as influencer, which accounts for 76.6% of all. So we can draw the conclusion that the idols of artists will guide them to choose the same genre as their leaders, which results in the similarity between them and their influencers.

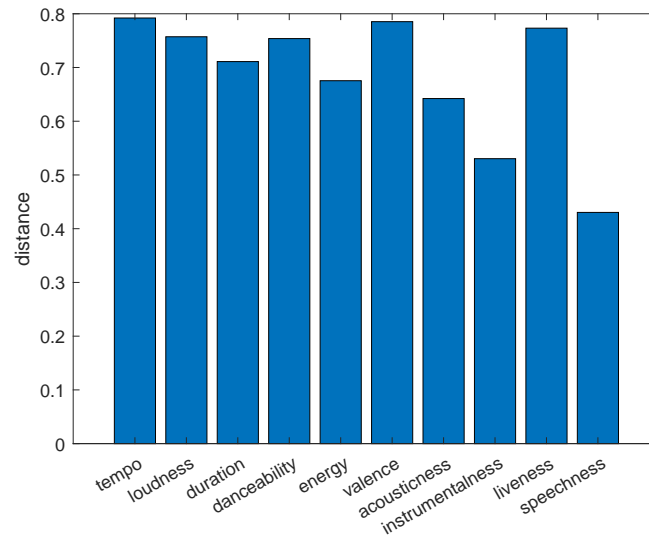


Figure 9: Average characteristic difference after variance correction

To prove opinions above, we select some artists with numerous followers, such as Hank Williams and James Brown, and divide their followers into two groups, one belonged to the same genre as their influencer, while the other doesn't. By calculating the two groups' difference vector with their influencer, we find that followers of the same genre with influencer hold higher similarity. The calculation results of James Brown and Hank Williams are shown in Figure 10.

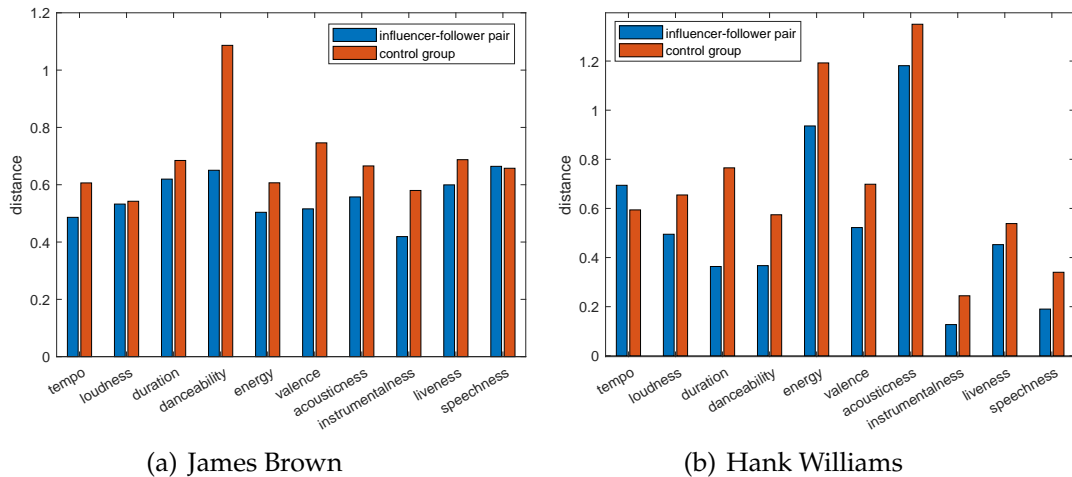


Figure 10: Comparison of variance-corrected difference vector between influencer-follower pairs and control group

5.2.3 Conclusion

From the analysis of influencer's above, we can conclude that

- Artists will influence their followers, making their followers' music more similar with theirs.
- Music created by artists and their followers shows an obvious similarity in speechiness and instrumentality.
- It is more likely that artists and their followers are from the same genre.
- If artists and their followers belong to the same genre, they tend to have more similarities.

5.3 Analysis of music revolution

5.3.1 Music revolution

We consider music revolutions as great changes of music features between different periods. Thus we introduce the average feature vector to describe the overall features of a period(5 years). It can be expressed as

$$A^{(i)} = \frac{\sum_{j=i}^{i+4} \mathbf{x}^{(j)}}{5} \quad (12)$$

where i indicates the year and \mathbf{x} indicates the normalized features vector in *data_by_year*. We calculate Euclidean Distance between the average feature vectors of adjacent periods and get Figure 11. We can notice that there are 3 extreme points(1926-1930, 1956-1960 and 2016-2020) in the figure. Due to lack of data in 1921-1930, we discard this period and consider other periods as music revolutions. The conclusion is consistent with the facts because during the 1956-1960 period Jazz and Pop/Rock were rising rapidly and before 2016-2020 modern pop music had made great progress on various media.

5.3.2 Representative revolutionaries

From 5.3.1 we capture two periods and regard them as music revolutions. In this part, we try to find representative revolutionaries who played important roles in music revolutions. It is assume that representative revolutionary should have both innovation and popularity. Specifically, what measures innovation is how different the artist's music works are from those works released in last period.

To solve the problem, we define the contribution of an artist in the music revolution as

$$C^{(i)} = \sum_{s \in M^{(i)}} d(\mathbf{x}_s, A) p_s \quad (13)$$

where i indicates artist's id, M indicates all music works during the specific music revolution, \mathbf{x}_s indicates the features vector of s , A indicates the average features vector of last period and p_s indicates the popularity of s . For both music

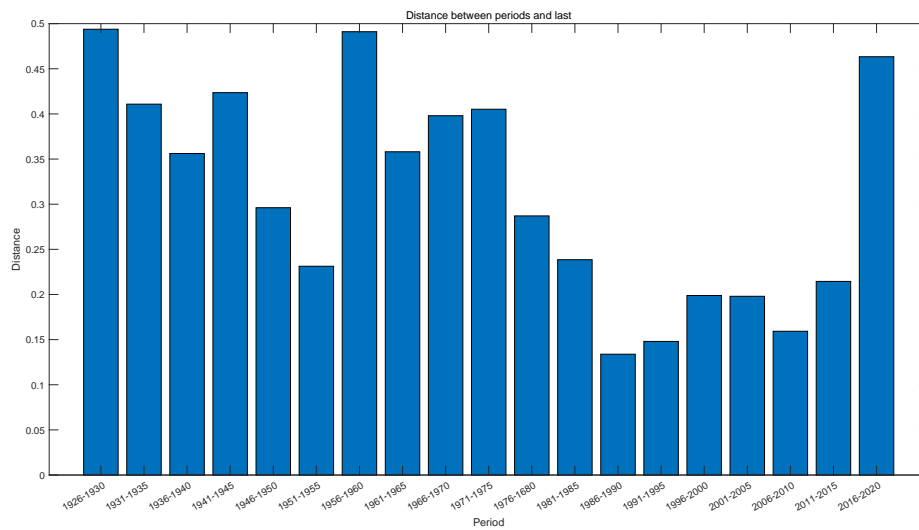


Figure 11: Distance between periods and last

revolutions, we calculate every artist's contribution using the formula above to sort them. After that, we select the top 5 artists in both music revolutions to check the result with facts, which is shown in the Table 2.

period	rank	artist
1956-1960	1st	Glenn Gould
1956-1960	2nd	Ella Fitzgerald
1956-1960	3rd	Miles Davis
1956-1960	4th	Elvis Presley
1956-1960	5th	Frank Sinatra
2016-2020	1st	Billie Eilish
2016-2020	2nd	The Weeknd
2016-2020	3rd	J Balvin
2016-2020	4th	Ariana Grande
2016-2020	5th	Frank Ocean

Table 2: Top 5 artists in 1956-1960 and 2016-2020 music revolutions

5.4 Analysis of the influence processes of musical evolution

5.4.1 Dynamic network of musical influence

Now we are going to analyze the influence processes of musical evolution that occurred over time, so we extend network of musical influence first, in order to take the evolution of the network over time into consideration:

- In a specified period, if an artist did not create any music works, we regard him inactive and remove it from the original network constructed in 4.1.
- Contribution of an artist in the music revolution ($C^{(\cdot)}$) is considered, so the **dynamic influence score** of an artist A now can be evaluated by

$$S(A) = \sum_{i=0}^{\infty} \alpha^i \sum_{K \in L_i(A)} [C^{(K)} + \beta(\text{year}(K) - \text{year}(A))] \quad (14)$$

The extended network is named **dynamic network of musical influence**, where "dynamic" means it evolves over time. After calculation, we obtain scores of artists in 1956-1960 and 2016-2020. We sort out five artists with the highest scores in 1956-1960 and 2016-2020 respectively.

Top five artists ranked by scores of dynamic influence in 1956-1960 are listed as follows:

1. **Billie Holiday (Vocal)**: Her vocal style formed a new way of manipulating phrasing and tempo and she was known for her vocal delivery and improvisational skills.¹
2. **Lester Young (Jazz)** was an American jazz tenor saxophonist, and he was one of the most influential players on his instrument.²
3. **Billy Eckstine (Jazz)** was an American jazz and pop singer. He was noted for his rich, almost operatic bass-baritone voice.³
4. **Nat King Cole (Jazz)** was an American singer and jazz pianist. He was awarded the Grammy Lifetime Achievement Award in 1990.⁴
5. **Ella Fitzgerald (Jazz)** was an American jazz singer, sometimes referred to as the First Lady of Song, Queen of Jazz, and Lady Ella. She won fourteen Grammy Awards in total.⁵

5.4.2 Case study: Jazz

Jazz, a centuried music genre continuing its popularity until now, originated in the late 19th and early 20th centuries. Since the 1920s Jazz Age when jazz music and dance gained popularity in America rapidly, it has been a major form of musical expression in traditional and popular music.⁶ We perform analysis on jazz

¹Billie Holiday, https://en.wikipedia.org/wiki/Billie_Holiday

²Lester Young, https://en.wikipedia.org/wiki/Lester_Young

³Billy Eckstine, https://en.wikipedia.org/wiki/Billy_Eckstine

⁴Nat King Cole, https://en.wikipedia.org/wiki/Nat_King_Cole

⁵Ella Fitzgerald, https://en.wikipedia.org/wiki/Ella_Fitzgerald

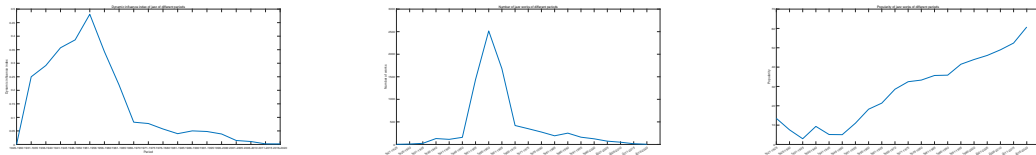
⁶Jazz, <https://en.wikipedia.org/wiki/Jazz>

music by investigating changes over time of its popularity, productivity and dynamic influence, and we attempt to introduce social, political and technological changes as a perspective of our explanations.

Firstly, we divide years from 1921 to 2020 into 20 periods : 1921-1925, 1926-1930, \dots , 2016-2020. Then we calculate the total number of jazz music works, average popularity of works, and **dynamic influence index** of jazz artists during periods.

To calculate dynamic influence index, we rank all artists by their dynamic influence during one period, and define dynamic influence index as

$$\text{Dynamic influence index} = \frac{\text{Number of jazz artists rank top 50\%}}{\text{Total number of artists rank top 50\%}} \quad (15)$$



(a) Dynamic influence index of jazz of different periods (b) Number of jazz works of different periods (c) Popularity of jazz music of different periods

Figure 12: Characteristics of jazz music of different periods

In figures we observe that since the 1920s Jazz Age, dynamic influence of jazz continued increasing together with the number of works. The 1930s belonged to popular swing big bands, but a turning point in jazz occurred as the outbreak of World War II. Although big bands of jazz had made a great success previously, the war presented difficulties for the big-band format. Instead, small groups of young musicians developed a different, highly experimental style of jazz and this musical development became known as bebop.

In the post-war years, various subgenres emerged including Hard Bop, Cool Jazz and Free Jazz, and 1959 is called "The Year that Changed Jazz" by some artists. In the 1960s, there was growing unrest in the world, and in America there existed domestic problems and events such as the Vietnam War, racial issues and the civil rights movement. Jazz music achieved its highest dynamic influence during 1956-1965, and 1956-1965 is also the time when most jazz music works were created.

After 1965, it shows the beginning of a decline in the dynamic influence and productivity of jazz music as other genres of music such as Pop/Rock and soul music emerged and gained popularity among the youth and artists experimented with Jazz fusion including Jazz-rock as an innovation.¹ However, jazz music continues its popularity until now as it still holds a major form of musical

¹Jazz, <https://en.wikipedia.org/wiki/Jazz>

expression in traditional and popular music. With the development of Internet and online music platforms, people have more opportunities to appreciate jazz music.

5.4.3 Case study: Michael Jackson

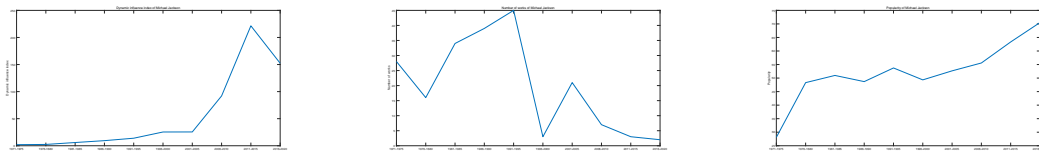
Michael Jackson was an American singer, songwriter and dancer. He is regarded as one of the most significant cultural figures of the 20th century due to his contributions to music, dance and fashion. For instance, he popularized the dance "moonwalk"; his album *Thriller* is the best-selling album all of the time; he was awarded 13 Grammy Awards, the Grammy Legend and Grammy Lifetime Achievement awards.¹ We are going to conduct an investigation on Michael Jackson through changes over time of its popularity, productivity and dynamic influence.

Firstly, we divide years from 1921 to 2020 into 20 periods : 1921-1925, 1926-1930, \dots , 2016-2020. Then we calculate the total number of his works, average popularity of works, and **dynamic influence index** of him during periods.

To calculate dynamic influence index, we rank all artists by their dynamic influence during one period, and define dynamic influence index as

$$\text{Dynamic influence index} = \frac{\text{Total number of artists}}{\text{Rank of Michael Jackson}} \quad (16)$$

In figures we observe that after Jackson began his solo career in 1971, he created



(a) Dynamic influence index of Michael Jackson of different periods
(b) Number of Michael Jackson's works of different periods
(c) Popularity of Michael Jackson's music of different periods

Figure 13: Characteristics of jazz music of different periods

numerous works and soon obtained a high dynamic influence index after 1980. Most of his influential works were released during 1980-1995. In 1982 Jackson's sixth album, *Thriller*, was released, and won eight awards at the 26th Annual Grammy Awards. *Thriller* is also credited with breaking racial barriers due to the contribution to promote racial equality made by music videos.

In 1984, a widespread famine raged in Africa and Michael Jackson created a song *We are the World* together with Lionel Richie to support the foundation United Support of Artists for Africa (USA for Africa), which would then provide

¹Michael Jackson, https://en.wikipedia.org/wiki/Michael_Jackson

humanitarian aid including food and relief aid to starving people in Africa. *We are the World* won a great success and this also helped Jackson gain more dynamic influence.

Michael Jackson passed away accidentally in 2009. Although he has passed away, still a lot of people remember him all the time, and his dynamic influence remains today.

6 Conclusion

6.1 Summary of Results

- We construct a influencer-follower network to illustrate musical influence. We sort out artists with the highest musical influence and confirm that they have a great influence in the history of music indeed.
- We develop a measure of music similarity and verify that music of the same genre has a high similarity, while music of different genres has a low similarity.
- Based on basic models, we fuse music similarity into the network of musical influence to analyze influencers' impact on followers' music works and find out characteristics with most similarities.
- By analyzing similarities of musical works in different periods, we find out periods when musical revolutions occurred and identify artists who represent revolutionaries.
- We extend the network of musical influence to a network which evolves over time, describing dynamic influence of artists. With the help of the dynamic network, we sort out artists with the highest dynamic influence in different periods.
- Through statistical characteristics of dynamic influence, popularity and productivity, we conduct case studies of a genre and an artist as an application of our models. We also take into account factors beyond music including social and political changes, and cultural influence.

6.2 Strengths

- Analysis results of our influence network correspond to the truth. We sort out some artists with the highest music influence from the network and we look up the encyclopedia to find that they are all great figures in the history of music.

- Models are of high flexibility. By utilizing various methods to evaluate node values or edge weights, we can introduce more factors to describe an artist's influence such as the time. By extend the dimension of feature vectors, we can assess music similarity more precisely and reasonably.

6.3 Possible Improvements

- We suppose all artists only create music belonging to the same genre, which is contrary to the fact. Moreover, some music are created in cooperation by several artists, making it more inaccurate to deduct its genre by its authors' genre. If the genre of each piece of music can be acquired, a better analysis of music characteristics and comparison between and within genres can be obtained.
- We use Euclidean Distance to assess music similarities and treat each dimension of feature vectors equally. However, features may weigh differently as different features may subject to different distributions. If weights of different dimensions of vectors are determined in a more proper way, we can obtain a more precise assessment of music similarities.

7 Reference List

References

- [1] Company, Houghton Mifflin Harcourt Publishing. *"The American Heritage Dictionary entry: Music"*. ahdictionary.com.
- [2] Davies, Hunter. *The Beatles*. 2nd rev. ed. W. W. Norton; 1996. ISBN 978-0-393-31571-4.
- [3] Fix, Evelyn; Hodges, Joseph L. (1951). Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties (Report). USAF School of Aviation Medicine, Randolph Field, Texas.

Appendices

We computed influence scores of all artists and list their ranks in Table 3 below.

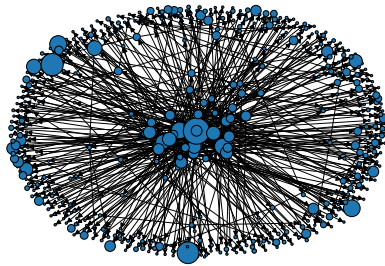
Table 3: List of artists ranked by influence (Top 50)

Artist name	Genre	Artist name	Genre
The Beatles	Pop/Rock	The Isley Brothers	R&B;
Bob Dylan	Pop/Rock	The Band	Pop/Rock
The Rolling Stones	Pop/Rock	John Coltrane	Jazz
Elvis Presley	Pop/Rock	The Stooges	Pop/Rock
Chuck Berry	Pop/Rock	Jerry Lee Lewis	Pop/Rock
Little Richard	Pop/Rock	Howlin' Wolf	Blues
Jimi Hendrix	Pop/Rock	Pink Floyd	Pop/Rock
James Brown	R&B;	Fats Domino	R&B;
Buddy Holly	Pop/Rock	The Yardbirds	Pop/Rock
The Kinks	Pop/Rock	Woody Guthrie	Folk
The Beach Boys	Pop/Rock	Jimmy Reed	Blues
David Bowie	Pop/Rock	The Everly Brothers	Pop/Rock
Hank Williams	Country	Little Willie John	R&B;
Smokey Robinson	R&B;	Roy Orbison	Pop/Rock
The Who	Pop/Rock	Billie Holiday	Vocal
Bo Diddley	Pop/Rock	Eddie Cochran	Pop/Rock
The Velvet Underground	Pop/Rock	Otis Redding	R&B;
Ray Charles	R&B;	Carl Perkins	Pop/Rock
Sam Cooke	R&B;	Nat King Cole	Jazz
Muddy Waters	Blues	Carole King	Pop/Rock
Miles Davis	Jazz	Neil Young	Pop/Rock
Led Zeppelin	Pop/Rock	The Clash	Pop/Rock
Marvin Gaye	R&B;	Johnny Cash	Country
The Byrds	Pop/Rock	Ravi Shankar	International
Sex Pistols	Pop/Rock	Brian Eno	Pop/Rock

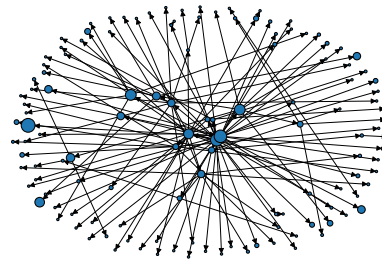
Influence network —State-of-art method to understand music

Music plays an important role in human society and is indispensable for social science and culture studies. To examine interaction between music and external environments, the understanding of influence in music over time must be prioritized.

However, the study of music consists of a variety of genres and numerous artists, making it difficult to clarify the connections between genres and artists. To get over it, our team propose a **directed network** to visualize the influence among artist and introduce year gap and distance between nodes to quantize the strength of influence. By extracting subnetworks, the relationship and influence among artists can be clearly presented. The influence network of XTC and Clint Black is exemplified as follows:



(a) XTC



(b) Clint Black

To study the transition of music and genre, measurement of music similarity is fused into the influence network. Taking the amount of composition into consideration, the influence network can be extended to **dynamic network of music influence**, with the assistance of which the main revolutions in history and its major contributor are successfully recognized. 1956-1960 and 2016-2020 prove to be the milestones in music development after calculation, and the top five artists with the highest dynamic influence are: Glenn Gould, Ella Fitzgerald, Miles Dvis, Elvis Presley and Frank Sinatra; Billie Eilish, The Weekend, J. Balvin, Ariana Grande and Frank Ocean.

Moreover, our study builds the connection between social transition and leaps in music. Take jazz as an example. Jazz originated in the late 19th and early 20th centuries, and still continues its popularity nowadays. Jazz's dynamic influence reached its peak after WWII during 1956-1965, but after that a decline in the influence and productivity as other genres such as Pop/Rock and soul music began gaining popularity among the youth.

In summary, our study is of great value to promoting research in music influence. We can't illustrate the characteristics and transition of all genres due to data limitation, but our model is flexible and can be applied to richer data with fewer modification.