Problem Chosen	2021	Team Control Number
D	MCM/ICM	2100640
	Summary Sheet	

Dynamic Directed Network Modelling the Evolution of Influence Between Music Genres and Relevant Data Analysis

Abstract

The similarity and influence between and within music genres has been an interesting topic all the time, which actually implies relationships between music genres, the evolution of existing genres, and the origin of emerging genres. Our team identified by **Integrative Collective Music (ICM) Society** aims to design appropriate indicators and models measuring the similarity and influence, and study the evolution of music genres. We conduct this research through the given tasks step by step.

As for task 1, we establish a dynamic directed influencer network based on the start year of artists and give a weighted score by **Entropy Weight Method (EWM)** to evaluate the influence of an artist from popularity, the number of followers and the number of his/her works.

As to task 2, we introduce two metrics to measure the similarity between artists and genres: **Cosine Similarity** and **Euclidean Distance** with the aide of **Principal Components Analysis (PCA)**. We compare the results of influence between and within genres and explain the limitations of using Euclidean distance here according to its performance.

In task 3, we compare the most similar genre measured by cosine similarity and the largest source genre of influence, for the genres with the sufficient number of listed artists, in order to find the relationships between genres. We then give a criterion for distinguishing the genres and do an accuracy test for the classification of the genre of artists.

With regard to task 4, we justify the association between influence and similarity for certain artists through proving the difference in the distributions of similarity of followers and non-followers of an artist within the same genre, by **Kolmogorov–Smirnov Test**. Regarding task 5, we find out the characteristics signifying the music revolution by analyzing the change of each feature from a global view. We then introduce two indicators to measure the revolutionary degree of an artist and list those representative artists by ranking.

In task 6 and 7, we design a dynamic influence factor to measure the influence of an artist at a specific period and give its correlation with the influence score calculated in Task 1. We explain the evolution in our sub-netowrks of Pop/Rock, Jazz and Blues by associating with historical events and social, economic, technological, and cultural changes at that time.

We give strengths and weaknesses of our model and summarize the literature of dynamic musical influence network with detailed answers for questoons raised in each task and a one-page letter to **Integrative Collective Music (ICM) Society**.

Keywords: Dynamic directed network, Evolution of music genre, Entropy weight method, Principal components analysis, Cosine similarity, Kolmogorov–Smirnov test, Euclidean distance.

Contents

1 Introduction

1.1 Restatement of the problem

Music is a valuable cultural asset in the development of human history. In recent decades, its development and changes have been particularly rapid. Different music genres influence each other and inspire each other, but also have their own characteristics. Studying the evolution of music helps to understand the role of music in human culture and its correlation with society trends. **Integrative Collective Music (ICM) Society** is interested in modelling musical influence and has identified our team. Therefore, this literature aims to study the self and mutual influence of different music genres, and the evolutionary and revolutionary trend of music genres and artists in the last 90 years.

In order to study this evolutionary trend step by step, the problem is divided into several sub tasks. A further simplified clarification of each specified task specified is listed as follows, and the steps for each task are shown in Figure 1.

1. Create a directed network of musical influence for artists and design an indicator to measure this influence between artists. Extract a sub-network and analyze its behaviour and music influence.
2. Measure the music similarity of artists between and within genres.
3. Compare the similarity and influences between and within genres, determine a criterion that can distinguish genres, and describe the evolution of genres.
4. Justify the association between influence and similarity and identify "contagious" characteristics.
5. Find characteristics signifying music revolution, and revolutionary artists.
6. Design indicators revealing dynamic influencers and analyze the evolution of influence.
7. Link the network's evolution with society changes from different aspects.

2 Assumptions and Nomenclature

2.1 Assumptions

To simplify the problem, we introduce some assumptions. Each assumption will be re-mentioned once it is used in this literature.

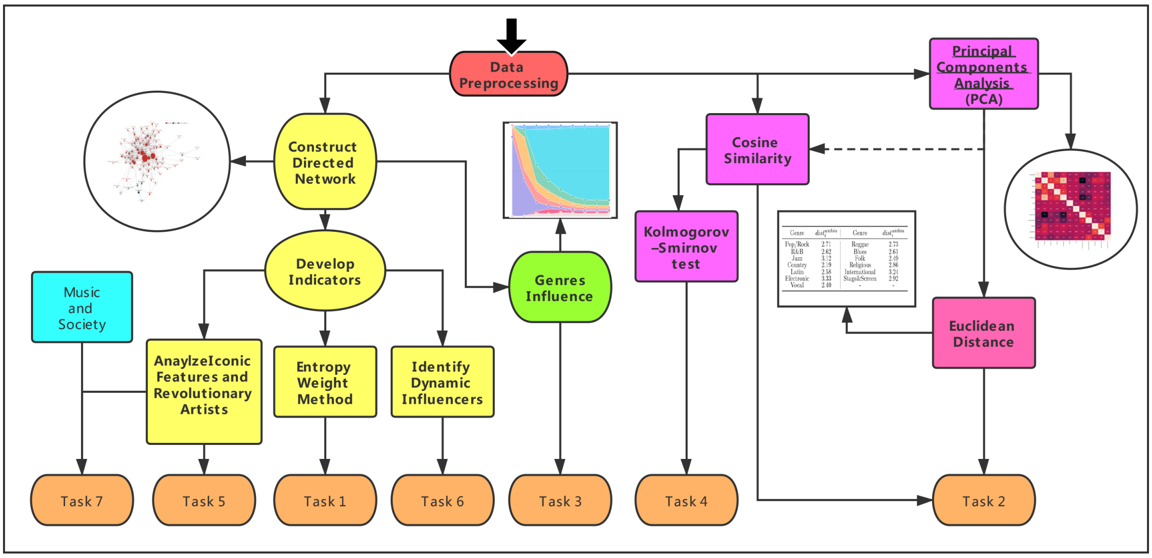


Figure 1: Work flow chart

1. **Artist and genre**: One artist performs one genre of music, there is no artist performing more than one genre, and therefore we do not consider the situation where an artist changes his/her genre. A genre can be viewed as the aggregate of its artists and the characteristics of a genre can thus be taken as the average or added-up characteristics of its artists. A revolutionary artist must have followers to spread his/her influence of music revolution and therefore we do not consider the listed artists without followers.

2. **Characteristics**: We assume that popularity and the number of works have no connection with an artist's music style but with the influence of the artist. When it comes to characteristics of genre or artists, we focus on the 12 features: danceability, energy, valence, tempo, loudness, mode, key, acousticness, instrumentalness, liveness, speechiness, explicit, and duration. We link the rest 2 features, popularity and the number of works, to the influence of an artist. The distribution of each of these features of the same genre obeys a normal distribution respectively.

3. **Influence**: We assume that the influence on an artist comes only from his/her direct influencers in order to simplify our model.

2.2 Nomenclature

The notations used in this literature are listed in Table 1.

3 Data Preprocessing

Before starting to analyze the data, we clean and normalize the data. Since the data in file *"data_by_artist.csv"* and file *"data_by_year.csv"* is the mean value extracted from file *"full_music_data.csv"*, we concentrate on these 2 files of mean values and the file *"influence_data.csv"* that contains the connections between influencers and followers. We find out that there is no explicit missing value but some artists with their main genres of value "Unknown" after joining the data in file *"data_by_artist.csv"*

Table 1: Nomenclature

Symbol	Definition	Unit
$S_i^{influence}$	The score of music in influence of artist i in task 1	-
PC_i	The i^{th} principal component in PCA	-
N	The number of artists in total	-
Sim_{cos}	The cosine similarity between two artists or two genres	-
Sim_{L2}	The similarity derived from Euclidean distance	-
$\mathbb{D}_{revolutionary}$	Indicator of revolutionary degree of an artist	-
L_f	The number of followers of an artist	-
L_i	The number of influencers of an artist	-
$C_{i,t}$	The temporal mean number of works of artist i at period t	-
$M_{i,t}$	The temporal mean popularity of artist i at period t	-
$S_{i,t}^{dynamimc}$	The dynamic influence factor of artist i at period t	-

with the data in *"influence_data.csv"*. There are in total 5,854 artists mentioned in these files of whom 255 have no information of their genres, with a missing rate 4.36%. Therefore, we remove the records of these artists and do not consider them in the following analysis.

There are 14 types of quantitative feature describing the artist and his/her works, which are respectively danceability, energy, valence, tempo, loudness, mode, key, acousticness, instrumentalness, liveness, speechiness, explicit, duration, popularity, and the number of songs. Mode and explicit are Boolean variables, key and the number of songs are integers, and the rest are continuous variables. We apply the **min-max normalization** on each feature respectively, as shown in Equation (1) with x the column vector of one feature for different artists, x_i its i^{th} element, and $x_i^{normalized}$ the corresponding normalized element.

$$x_i^{normalized} = \frac{x_i - max(x)}{max(x) - min(x)} \tag{1}$$

All the feature variables appeared in following discussions are normalized.

4 Task 1: Directed Network of Influencers and Followers

With the relationship of follower and influencer in file *"influence_data.csv"*, we can create a dynamic directed network $G = (V, E)$ with each vertex V representing an artist and each edge E representing an influence relationship from an influencer to a follower. The dynamic property comes from that we add vertices based on the artist start year of their career and that we add edges based on the later year between the start years of follower and influencer. We can use concepts from social network to analyze our network. The influencer of an artist can be partly measured by the **density** of a vertex, i.e. the ratio between the number of existing connections and that of possible connections, and **centrality** in terms of out-degree of a vertex, i.e. the number of out-going interactions [2].

4.1 Parameters to Capture Music Influence of an Artist

We propose to estimate the influence magnitude of an artist with 3 factors: popularity, the number of songs and the number of the direct followers that is exactly the out-degree of a vertex. We do not discuss about the density of a vertex because we focus on the out-going interactions of a vertex. **Entropy weight method** [1] is one of the most popular approaches for the determination of criterion weight according to the dispersion of values in different criteria. We define N the total number of artists and $\boldsymbol{Q^{influence}}$ the matrix of normalized factors (the number of followers is also normalized) related to influence, with each line corresponding to the three factors (popularity, the number of songs and the number of the direct followers) of an artist.

$$\boldsymbol{Q^{influence}} = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ \cdots & \cdots & \cdots \\ q_{N,1} & q_{N,2} & q_{N,3} \end{bmatrix}_{N\times 3} \tag{2}$$

We denote p_{ij} the relative probability of factor j of artist i, and E_j the entropy of factor j. The weight attributed to factor w_j is given by Equation (5). We define the weighted sum $S_i^{influence}$ of these 3 factors as a score to evaluate the music influence of artist i.

$$p_{ij} = \frac{q_{ij}}{\sum_{i=1}^{N} q_{ij}} \tag{3}$$

$$E_j = -\sum_{i=1}^{N} p_{ij} \log_N p_{ij} \tag{4}$$

$$w_j = \frac{1 - E_j}{\sum_{j=1}^{N}(1 - E_j)} \tag{5}$$

$$S_i^{influence} = \sum_{j=1}^{3} w_j q_{ij} \tag{6}$$

We rank the final influence score of every artist and list the top 10 most influential artists with their parameters and scores in Table 2. It consists with the reality.

Table 2: Top 10 most influential artists with their parameters and scores

Rank	Artist name	Popularity	Number of followers	Number of works	Score ($\times 10^{-3}$)
1	The Beatles	48.06	615	823	9.451
2	Bob Dylan	30.86	389	1092	7.292
3	The Rolling Stones	34.57	319	1035	6.302
4	The Beach Boys	27.96	186	994	4.573
5	Elvis Presley	33.39	166	990	4.319
6	Led Zeppelin	41.42	221	676	4.274
7	Frank Sinatra	26.00	71	1396	4.022
8	Miles Davis	22.70	160	864	3.954
9	David Bowie	41.73	238	446	3.952
10	Johnny Cash	26.61	112	1104	3.917

4.2 Analysis of Sub Influencer Networks

Analyzing such a great directed network with more than 50,000 edges is not easy thus we select some typical sub-networks from it by different standards to analyze. Figure 2 shows the sub-network of the top 3 artists, namely The Beatles, Bob Dylan, and The Rolling Stones and their direct and indirect followers. The indirect followers here refers to the followers of the followers of the influencers.

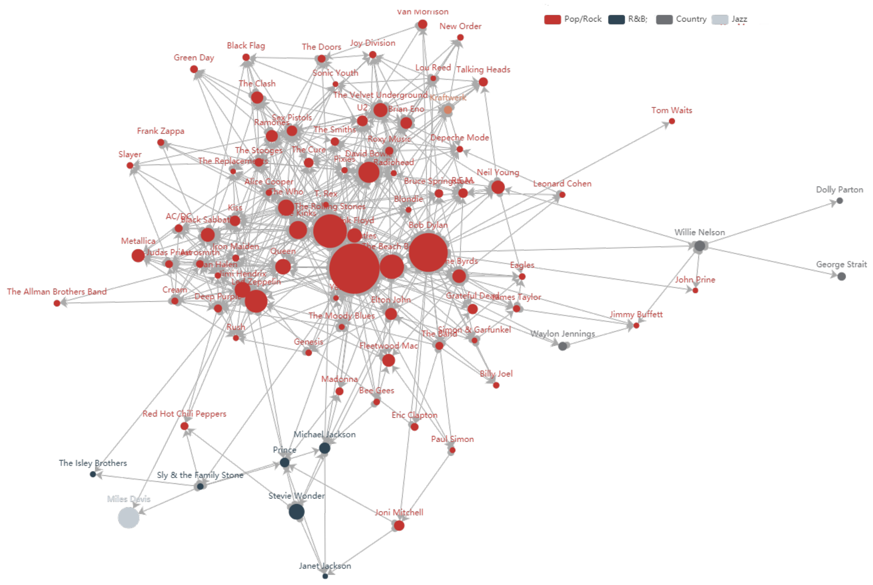


Figure 2: Directed sub-network of top 3 artists and their followers

In Figure 2, the size of a vertex represents the influence score of an artist and the direction of an edge is from an influencer to a follower. From the perspective of artist, it can be seen that the influence of top 3 artists situated in the center of the network is significant and outweighs far more than the others, which consists with the reality that The Beatles is regarded as the most influential band of all time [4], Bob Dylan is one of the greatest songwriters, and The Rolling Stone is one of the most famous rock bands. It signifies that their vertices have a high centrality in terms of out-degree. Moreover, an influencer can also be a follower, even The Beatles is influenced directly by the artist named "The Band" and Bob Dylan. This explains the mutual influence of artists on each other. From the perspective of genre, the genre of Pop/Rock has influenced the genres including Jazz, Country, and R&B at current.

5 Task 2: Similarity of Artists Between and Within Genres

We are interested in the similarity in **music style** of artists between and within genres hence we introduce two metrics, cosine similarity and Euclidean distance, to measure it. To measure such similarity, we consider 12 features, danceability, energy, valence, tempo, loudness, mode, key, acousticness, instrumentalness, liveness, speechiness, explicit, and duration but remove popularity and the number of works with the assumption that these 2 features have nothing to do with the music style but with the artist's influence.

5.1 Analysis of Correlation and Dimension Reduction with PCA

Before measuring the similarity, we try to figure out whether some of the 12 features are correlated with each other. We calculate the **correlation matrix** and Equation (7) gives the correlation coefficient between feature X and Y with μ_X, μ_Y their expected values and σ_X and σ_Y their standard deviations.

$$\rho_{X,Y} = \operatorname{corr}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathrm{E}\left[(X - \mu_X)(Y - \mu_Y)\right]}{\sigma_X \sigma_Y} \tag{7}$$

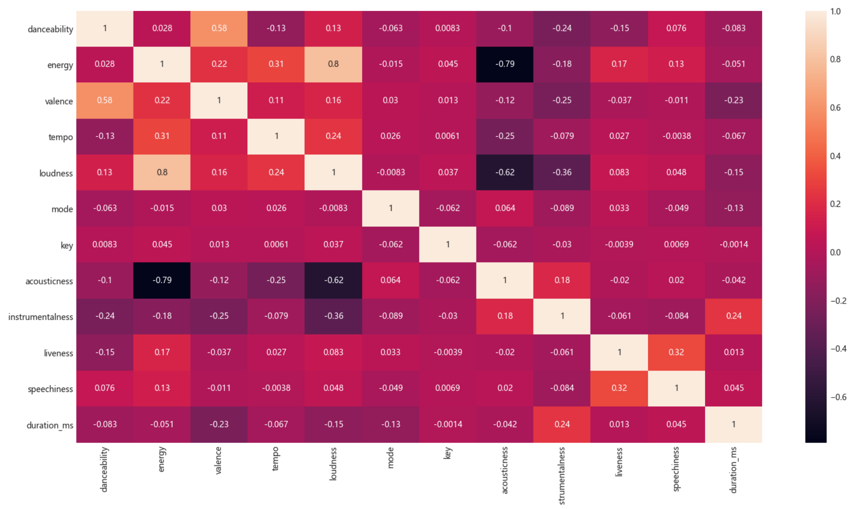


Figure 3: Correlation matrix of 12 features

We observe in Figure 3 that there is a strong positive correlation between loudness and energy, and between and that there is a strong negative correlation between acousticness and energy, and between acousticness and loudness, considering a coefficient greater than 0.6 as a strong degree of correlation. Hence, we may say that there are 9 features strongly linearly independent, and we standardize the features as Equation (8) shows, use **principal components analysis**, and conserve the 9 components related to the 9 largest eigenvalues of correlation matrix. The results of PCA are shown in Table 3 with PC_i the i^{th} principal component and γ the **cumulative explained variance ratio**.

$$X = \frac{X - \mu_X}{\sigma_X} \tag{8}$$

Table 3: Results of principal components analysis

Feature	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	PC_7	PC_8	PC_9
Danceability	-0.168	-0.577	0.045	0.353	0.132	-0.031	0.17	0.008	0.034
Energy	-0.522	0.232	-0.045	0.045	0.068	-0.074	0.056	-0.199	0.102
Valence	-0.250	-0.526	0.036	0.095	0.085	0.288	0.287	-0.231	-0.184
Tempo	-0.233	0.181	-0.136	-0.269	-0.013	0.783	0.193	0.336	-0.032
Loudness	-0.507	0.108	-0.051	-0.030	0.009	-0.203	-0.155	-0.007	0.062
Mode	0.001	-0.076	0.101	-0.628	0.056	-0.368	0.650	0.035	0.151
Key	-0.046	0.018	-0.047	0.202	-0.939	-0.033	0.264	-0.021	0.016
Acousticness	0.469	-0.206	0.216	-0.131	-0.068	0.164	-0.058	0.068	-0.033
Instrumentalness	0.277	0.273	-0.217	0.189	0.149	0.222	0.306	-0.663	0.319
Liveness	-0.077	0.255	0.638	-0.042	-0.018	0.057	0.057	-0.341	-0.588
Speechiness	-0.070	0.095	0.672	0.236	0.035	0.121	0.053	0.206	0.622
Duration	0.122	0.309	-0.091	0.496	0.239	-0.160	0.478	0.435	-0.303
γ	24.07%	38.28%	49.24%	59.46%	67.78%	75.12%	82.08%	88.11%	93.57%

5.2 Cosine Similarity of Artists' Music Style

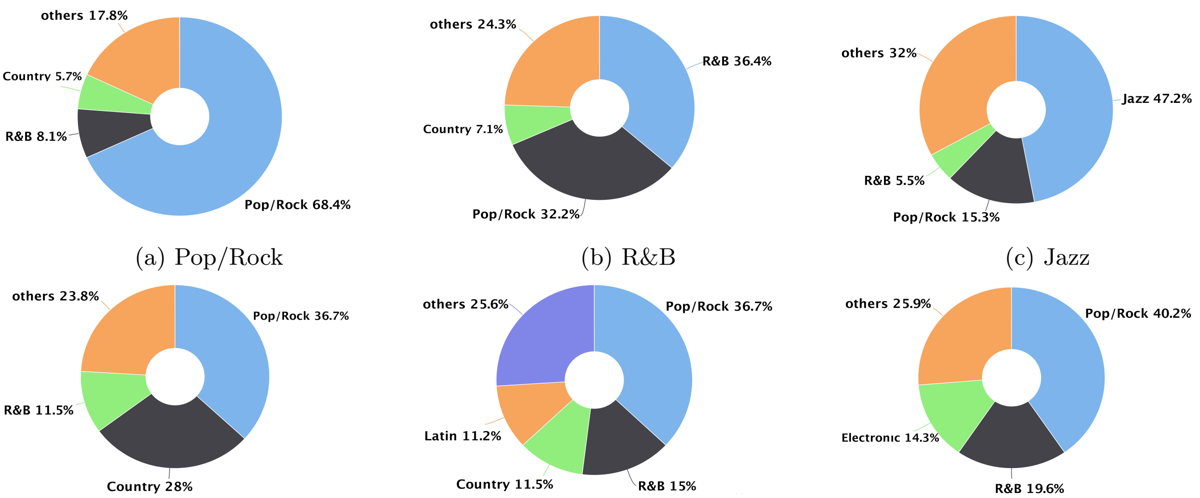
We replace the 12 features of an artist with the 9 principal components to calculate the **cosine similarity** between artist i and artist j with Equation (9). To compare the similarity of artists between and within genres, for each artist **A** in the specified genre, we find another artist **B** with the highest cosine similarity between them from all the 5,599 artists. We note the genre of artist **B** as the most similar genre of artist **A**. We then count respectively the number of artists Num_i in this specified genre grouped by their most similar genre i. We take the ratio of Num_i to the total number of artists in this specified genre as a metric to measure the similarity in **music style** of artists.

$$Sim_{cos}^{i,j} = \frac{\sum_{k=1}^{9} PC_k^i PC_k^j}{\sqrt{\sum_{k=1}^{9}(PC_k^i)^2}\sqrt{\sum_{k=1}^{9}(PC_k^j)^2}} \tag{9}$$

Table 4: Number of artists in each genre

Genre	Number of artists	Genre	Number of artists
Pop/Rock	2807	Religious	89
R&B	677	International	81
Jazz	406	Stage&Screen	50
Country	403	Comedy/Spoken	46
Latin	299	New Age	38
Electronic	208	Classical	28
Vocal	162	Easy Listening	23
Reggae	141	Avant-Garde	11
Blues	101	Children's	4
Folk	95	-	-

The number of artists in each genre is shown in Table 4. If the number of artists in a genre is small, then the corresponding analysis result could be accidental. We then study the similarity of artists between and within the top 6 genres: Pop/Rock, R&B, Jazz, Country, Latin, and Electronic. In the calculation related to similarity, we do not consider the genres with a small number of artists, namely Children's, Avant-Garde, Easy Listening, Classical, New Age, and Comedy/Spoken.



(a) Pop/Rock (b) R&B (c) Jazz

(d) Country (e) Latin (f) Electronic

Figure 4: Similarity between and within genres, calculated by the ratio of Num_i to the total number of artists in different specified genre, with Num_i the number of artists in this specified genre grouped by their most similar genre i.

From Figure 4, we observe that the similarity of artists between and within genres depends on the genre itself. The similarity within Pop/Rock far exceeds that between Pop/Rock and any other genre. The similarity within Jazz also exceeds that between Jazz and any other genre. The similarity within Country is a little less than that between itself and Pop/Rock. The similarity within R&B is almost the same as that between itself and Pop/Rock. For the genres Latin and Electronic, the similarity between themselves and Pop/Rock ranks first, which means that artists of these 2 genres are more similar to the artists of Pop/Rock.

The similarity within the genre is higher than that between genres can be easily explained by that a music genre is an ensemble of artists classified by their music styles therefore the similarity is relatively high. The case of Latin, Country and Electronic is mainly because of the impact of sub-genres: Latin Pop a fusion sub-genre of Latin and Pop/Rock, Country Pop a fusion of Country and Pop/Rock, and Electropop a genre merging Electronic and Pop/Rock.

5.3 Measuring Similarity with Euclidean Distance and the Limitations

We try to measure the similarity between and within genre by Euclidean distance. More precisely, we use a function of the average Euclidean distance between every artist and the **centroid** of the genre to estimate the similarity within genre, and use a function of Euclidean distance between two centroids of the genres to measure the similarity between these two genres, in the orthogonal vector space of PCA with a dimension of 9.

We denote N_i the number of artists in genre i. The distance $dist_{i,j}$ between genre i and j is then given by Equation (12) and the distance $dist_i^{within}$ within genre i is given by Equation (13).

$$dist_{i,j} = \sqrt{\sum_{k=1}^{9}(PC_k^{i,mean} - PC_k^{j,mean})^2} \tag{10}$$

$$dist_i^{within} = \frac{1}{N_i}\sum_{q=1}^{N_i}\sqrt{\sum_{k=1}^{9}(PC_k^{i,mean} - PC_k^{q})^2} \tag{11}$$

We suppose $Sim_{L2}^{i,j}$ the similarity of artists between genres i and j and $Sim_{L2}^{i,within}$ the similarity of artists within genre i by using Euclidean distance.

$$Sim_{L2}^{i,j} = \frac{1}{dist_{i,j} + 1} \tag{12}$$

$$Sim_{L2}^{i,within} = \frac{1}{dist_i^{within} + 1} \tag{13}$$

We normalize the similarity for each genre with its sum equal to one. The obtained results are quite **questionable** considering the same 13 genres as Section 5.2 in the process of calculation. We obtain that artists of Pop/Rock are more similar to those of Electronic while the similarity within Pop/Rock ranks merely the 10^{th} place and that the similarity within R&B is the smallest compared with that between R&B and any other genre.

We explain this disastrous results by the “**curse of dimensionality**”. The previous work of Michel Verleysen and Damien François [6] shows that the ratio of the volume of a unit-radius sphere to that of a cube with same edge length (the sphere is the tangent to the cube) decreases rapidly to a negligible value as the dimension increases to 10, and therefore for the points uniformly distributed in this high-dimension space, their norm is not randomly distributed but mostly concentrated on the maximum value, i.e. the square root of the dimension [6]. The precedent calculation shown in Table 5 proves that the average Euclidean distance between the point representing an artist and the centroid of his/her genre, no matter which genre is, is almost the same, ranging from 2.19 to 3.36 due to the shape of the cuboid but near the square root of dimension, which is 3. The centralized distribution of Euclidean distance shows that it is not a good measure of similarity.

Table 5: The distance $dist_i^{within}$ within genre for considered genres

Genre	$dist_i^{within}$	Genre	$dist_i^{within}$
Pop/Rock	2.71	Reggae	2.73
R&B	2.62	Blues	2.61
Jazz	3.12	Folk	2.49
Country	2.19	Religious	2.86
Latin	2.58	International	3.24
Electronic	3.33	Stage&Screen	2.92
Vocal	2.40	-	-

6 Task 3: Characteristics of Genres

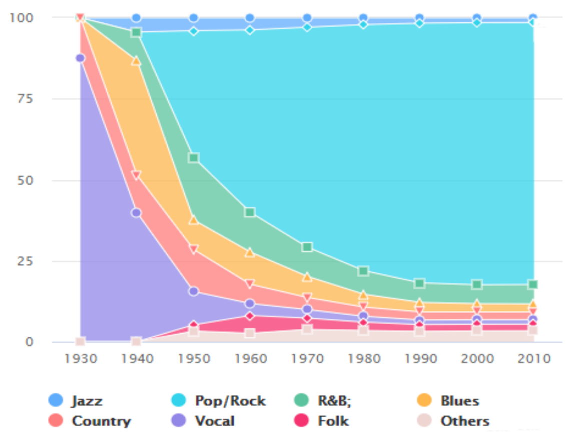
We describe the influence on a genre by the evolution of influencer genres for this genre, analyze whether the influence is associated with the similarity between and within genre, and figure out what characterizes a genre and then try to classify a genre by its characteristics.

6.1 Evolution of Influencer Genres from the Perspective of a Genre

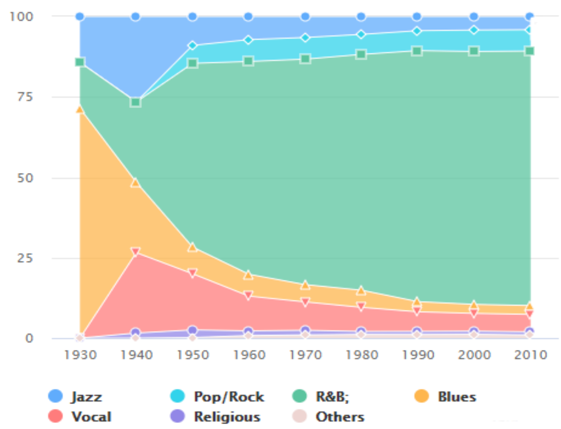
To discuss the evolution trend of influencer genres for a genre, we place an emphasis on the cumulative influence from the genre itself and other genres. We define a **record** of influence as one claim from an artist of his/her influencer artist (data in *"influence_data.csv"*) and an artist can have multiple claims because he could be influenced by several artists. We use the statistic that the ratio between the number of **cumulative** records of influence of genre **A** on genre **B** and the total **cumulative** number of records of influence of any genre including genre **B** on itself before a given year, e.g. 1970, to measure the cumulative influence on genre B before the given year. When counting the number of records of influence, we consider all the genres including those with a small number of artists but we focus on the top 6 genres: Pop/Rock, R&B, Jazz, Country, Latin, and Electronic.

Figure 5 shows the evolution trend of influencer genres of these 6 genres. Pop/Rock music emerged in the 1940s and 1950s. It was most inspired by Vocal and also influenced by Blues, Jazz, R&B and Country in the early period. After the 1950s, it began to flourish, and the degree of mutual influence of artists within the genre increased greatly. Today, around 80% of the influence on its artists comes from within the genre, and the remaining influence mainly comes from Blues, Jazz, R&B, Folk and Country. This consists well with the similarity analysis of Pop/Rock where today 68.4% of its artists' music styles have the highest similarity to that of the artists within in genre, 8.1% is most similar to R&B and 5.7% is most similar to Country, as is shown in Figure 4(a).

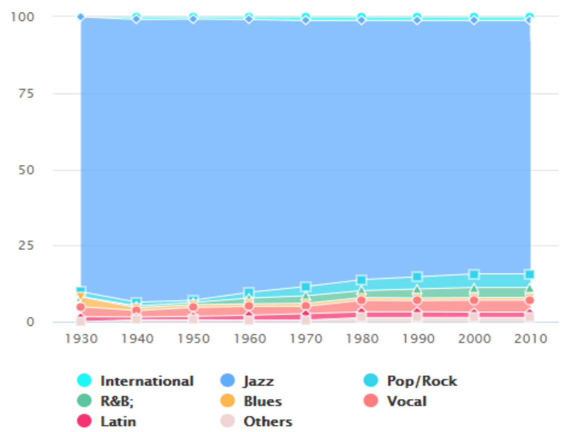
R&B originated from Blues, initially developed in the 40s and 50s, and then flourished. Today, the



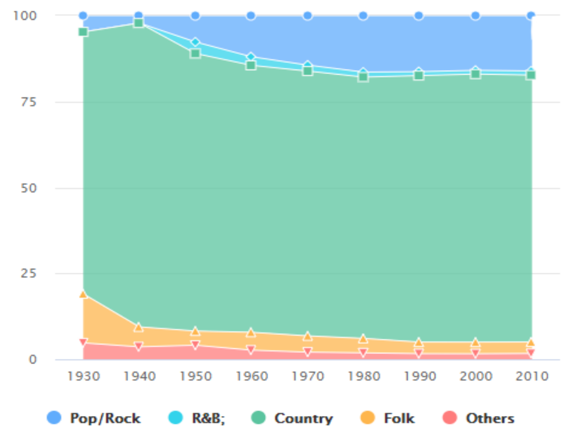
(a) Pop/Rock



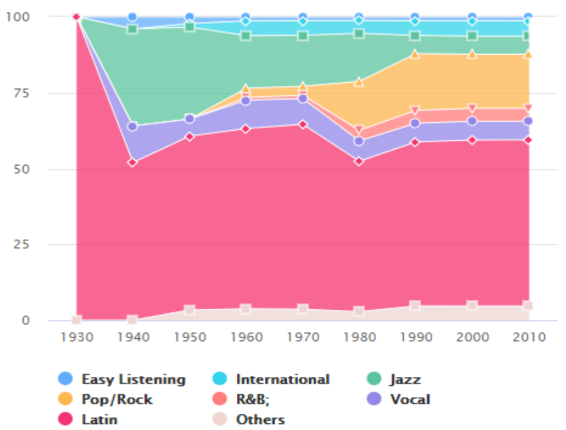
(b) R&B



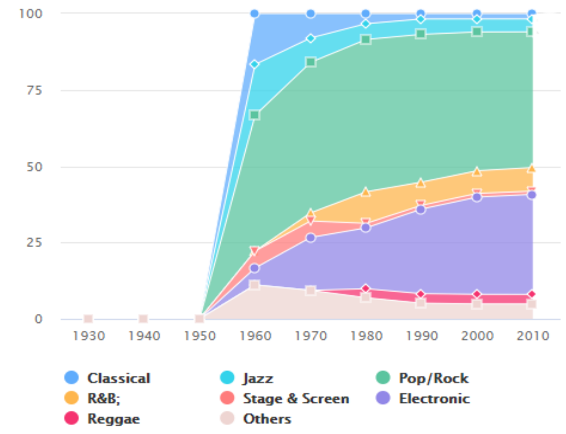
(c) Jazz



(d) Country



(e) Latin



(f) Electronic

Figure 5: Evolution for top 6 genre **B** of the ratio between the number of cumulative records of influence of genre **A** on genre **B** and the total cumulative number of records of influence of any genre including genre **B** on itself in the past 90 years

influence of artists within the genre is mainly from the artists of the same genre. This complies with the known fact and consists with the results of its similarity analysis in Figure 4(b). Jazz originated earlier than 1930 and the results of influencer genres and its similarity are also coherent by comparing Figure 5(c) and Figure 4(c).

There is a little incoherence in the analysis of influence and similarity for genre Country and Latin, where their ranks of the first and second place are exchanged. The results of Electronic also resemble to that of similarity analysis by comparing Figure 5(f) and Figure 4(f), where the artists of Pop/Rock have a larger impact than those within this genre on the genre Electronic itself. From the results of the analysis of similarity and influence between and within genres, we can conclude that at current Electronic is highly related to Pop/Rock, Country and Latin are related to Pop/Rock to some degree.

6.2 Distinguishing Characteristics of Genres

To distinguish a genre from the others, we propose to take the mean normalized feature vector of artists within the genre as the genre's characteristics of distinction. The normalized feature vector of an artist has a length of 12, containing danceability, energy, valence, tempo, loudness, mode, key, acousticness, instrumentalness, liveness, speechiness, explicit, and duration. We assume that the distribution of each feature of artists within the same genre obeys a **normal distribution**. We can then use the vector of mean value of the joint normal distribution of dimension 12 to differentiate each genre. This vector of mean value is actually the mean normalized feature vector of artists within the genre.

There are 19 genres in total but we do not consider genre Children's because there are only 4 artists listed within this genre. To know the accuracy of this characteristics, we design an accuracy test, where we randomly choose 1000 artists from these 18 genres, classify them based on the cosine similarity and calculate the accuracy of classification. We measure the probability of artist j belonging to genre i by Equation (14). We denote $\mathbb{A}_i$ the set of artists within genre i and $\vec{q}$ the characteristics vector of a genre or an artist. Table 6 shows the test results. We obtain an accuracy of 31.2% if the true genre has the highest probability of prediction and an accuracy of 56.7% if the true genre exists in the top-4 predictions. This is quite reasonable considering that there are 18 genres.

$$\text{Prob}_{j \in \mathbb{A}_i} = Sim_{cos}^{i,j} = \frac{\vec{q_i} \cdot \vec{q_j}}{\|\vec{q_i}\| \|\vec{q_j}\|} \tag{14}$$

Table 6: Test results of genre prediction using our characterizing method

Evaluation criterion	True genre = top-1	True genre ∈ top-4
Accuracy	31.2%	56.7%

Among all the 12 features, some are far more distinguishing, namely energy, instrumentalness, and speechiness, while some do not have distinctive variance in different genres, such as danceability, tempo, loudness, liveness, duration, and mode. Figure 6 shows the normalized values of these 3 distinguishing features. A large value of speechiness distinguishes Comedy/Spoken from other genres. Similarly, a large value of instrumentalness can separate New Age, Easy Listening, and Avant-Garde from the rest. Most of the features that we do not plot behave just like Energy but with less variance. Therefore, it is necessary to distinguish a genre by all the 12 features.

To develop how music genres have changed in the past years, we study the evolution of the mean feature of music from a global view. Features such as tempo, mode, instrumentalness, liveness, speechiness, and duration are quite stable. We focus on the features with dramatic change, shown in Figure

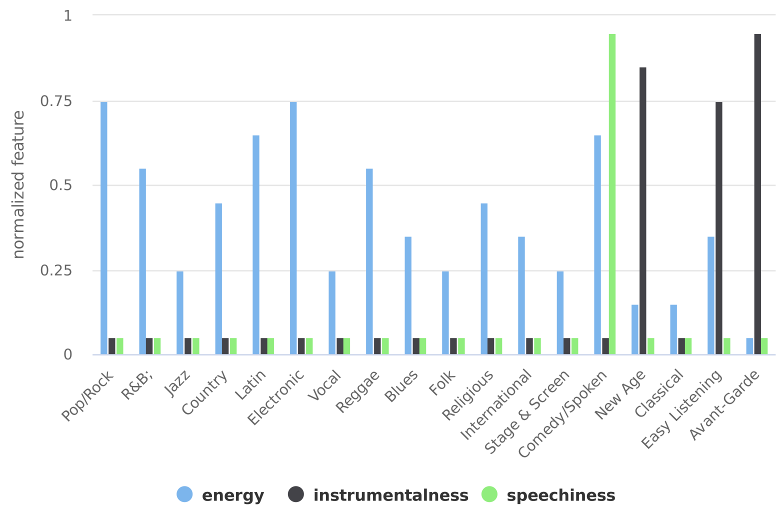


Figure 6: Energy, instrumentalness, and speechiness in different genre

7 and try to explain it. The steady increase in energy and loudness results from the popularization of Pop/Rock music when drum, electric guitar and bass became increasingly important [3]. The significant drop of acousticness is due to the development of technology enhancements. More precisely, the invention of the transistor promoted the transition from magnetic tape to electronic equipment, reducing the acousticness.

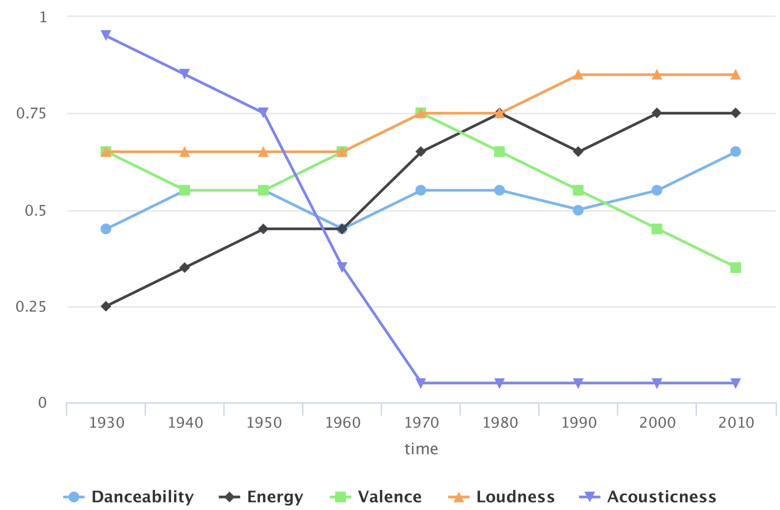
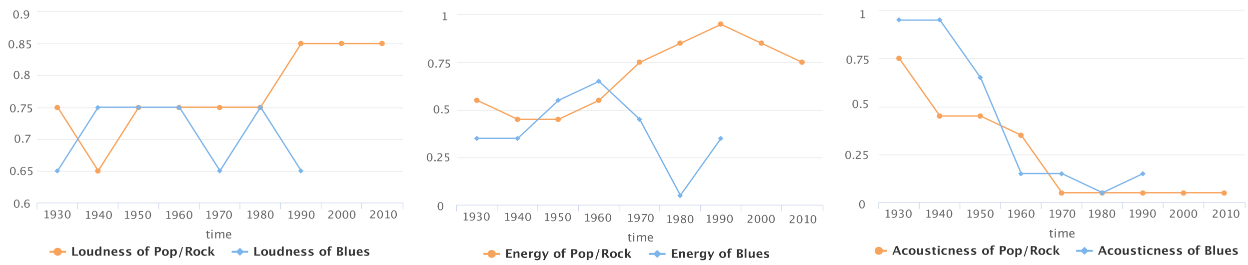


Figure 7: Evolution of danceability, energy, valence, loudness, and acousticness from a global view

We select two genres Pop/Rock and Blues to illustrate the change of loudness, energy, and acousticness within the genres, as shown in Figure 8. The acousticness within the 2 genres consists with the global trend. The loudness and energy of Pop/Rock gets larger increasingly but Blues keeps its loudness and even reduces its energy in a period, which indicates the evolution of their music characteristics: Pop/Rock prefers to make people more and more energetic but Blues maintains its melancholic nature.



(a) Loudness (b) Energy (c) Acousticness

Figure 8: Evolution of loudness, energy, and acousticness in Pop/Rock (orange) and Blues (blue)

7 Task 4: Correlation Between Similarity and Influence

7.1 Justification of the Correlation Between Similarity and Influence

To prove the existence of the correlation between similarity of music style and the influence on the follower, we try to prove that the **expectation of cosine similarity** between an influencer and his/her followers is greater than that between this influencer and those non-followers, and that the **distributions of similarity** of the two groups of artists are different. To rule out the impact of different genre on the similarity, we select the influencers, followers and non-followers from the same genre. To avoid accidental error, the genre that we choose must have a large number of artists. We decide to choose top-3 most influential artists, The Beatles, Bob Dylan, and The Rolling Stones, as the identified individual influencers. They all come from genre Pop/Rock that has 2807 listed artists in total.

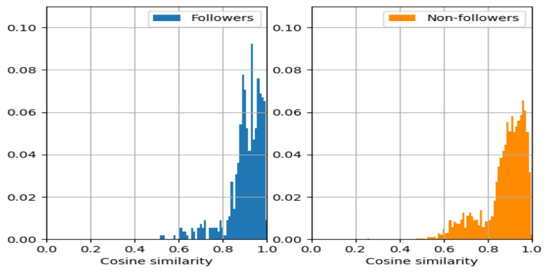
For each identified influencer, we divide the rest artists in Pop/Rock into followers and non-followers, considering only the direct influence. We calculate the **cosine similarity** between two artists using all the 12 min-max normalized features (we do not use principal components here because we want to obtain a **non-negative** similarity). The 12 features are danceability, energy, valence, tempo, loudness, mode, key, acousticness, instrumentalness, liveness, speechiness, explicit, and duration. We plot the empirical probability density function of cosine similarity for the two groups of artists, followers and non-followers, and for each influencer respectively, as is shown in Figure 9. We round the value of similarity to 2 decimals so as to divide intervals.

The probability density function of cosine similarity shows obviously that the related distribution is not normal. We use **two-sided Kolmogorov–Smirnov test** with notations $F(x)$ the cumulative distribution function of follower's similarity, $G(x)$ the cumulative distribution function of non-follower's similarity, and $D_{max} = \sup_x |F(x) - G(x)|$, and then set up the null hypothesis and alternative hypothesis as follows:

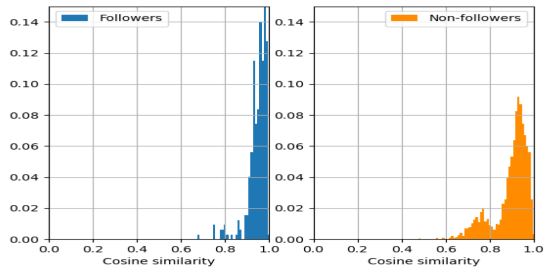
$$\text{Null Hypothesis } \mathbf{H0}: \; F(x) = G(x)$$

$$\text{Alternative Hypothesis } \mathbf{H1}: \; F(x) \neq G(x)$$

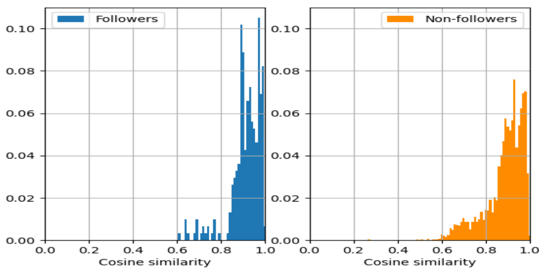
We apply on the two-sided KS test on their emperical cumulative distribution functions (CDF), plotted in Figure 10. The KS test results are shown in Table 7. With a significant level below 0.01%, we can reject null hypothesis and argue that the distributions of cosine similarity between influencers (The Beatles, Bob Dylan, and The Rolling Stones) and their followers within the genre, and between



(a) The Beatles



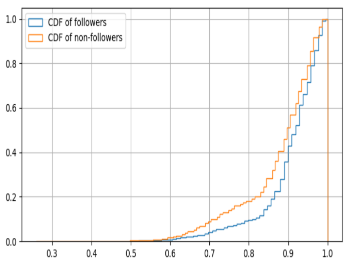
(b) Bob Dylan



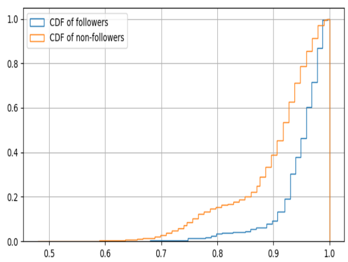
(c) The Rolling Stones

Figure 9: Empirical probability density function of cosine similarity between influencers (The Beatles, Bob Dylan, and The Rolling Stones) and their followers (blue), and non-followers (orange)

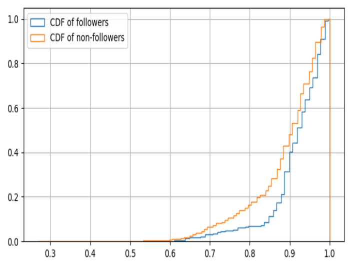
these influencers and non-followers within the genre are different.



(a) The Beatles



(b) Bob Dylan



(c) The Rolling Stones

Figure 10: Empirical cumulative distribution function of cosine similarity between influencers (The Beatles, Bob Dylan, and The Rolling Stones) and their followers (blue), and non-followers (orange)

Table 7: Results of two-sided Kolmogorov–Smirnov test on the similarity distributions

Influencer	Number of followers	Number of non-followers	D_{max}	p-value
The Beatles	553	2254	0.135	1.407×10^{-7}
Bob Dylan	322	2485	0.345	6.661×10^{-16}
The Rolling Stones	304	2503	0.160	1.617×10^{-6}

For these 3 influencers, we assume that the empirical CDF reflects the real CDF. We can observe from Figure 10 that the area covered by $F(x)$ with x-axis is larger than that covered by $G(x)$ with x-axis. Since we perform cosine similarity on mix-max normalized features, cosine similarity x here is non-negative, then $\int_0^{+\infty} F(x)\mathrm{d}x > \int_0^{+\infty} G(x)\mathrm{d}x$. The expectation of a non-negative variable Y can be given by $E[Y] = \int_0^{+\infty}(1 - CDF(y))\mathrm{d}y$. Denote X_f and X_g two random variables of cosine similarity for followers and non-followers respectively, then we can obtain that the expectation of cosine similarity $E[X_f] > E[X_g]$ for each of the 3 identified influencers. Therefore, for identified influencers The Beatles, Bob Dylan, and The Rolling Stones, they actually have influence on their followers.

7.2 Contagious Music Characteristics

We pay more attention on the music characteristics of an artist that imply music styles. In this sense, the influence of influencer on follower is passed by these music style characteristics which do not include acousticness, duration and liveness that imply more the technology in the music. To find those "contagious" features, we propose a simple metric $d_{i,j}$ measuring this single-feature difference in Equation (15), where q_i, q_j, q_{max}, and q_{min} are the original value of a given feature for artist i and artist j, and its maximum value and minimum value respectively. The smaller $\epsilon_{i,j}$ is, the more similar the given features between artist i and artist j are, and then the more contagious this feature is.

$$\epsilon_{i,j} = \frac{|q_i - q_j|}{q_{max} - q_{min}} \tag{15}$$

The contagious feature can vary from different influencers in different genre. We try to analyze the contagious features of artist The Rolling Stones of genre Pop/Rock. There are 304 followers of The Rolling Stones within the genre Pop/Rock, and we choose another 304 non-followers of this artist within Pop/Rock and 304 non-followers of this artist from Country, to compare their average difference in different single feature. From the results shown in Table 8, we can argue that the average difference in energy for followers is less than that for energy and energy is the contagious feature for The Rolling Stones. Since we do not observe an obvious reduction of difference of other features, considering the limit of sample size, we argue that the rest features have a similar role in the influence.

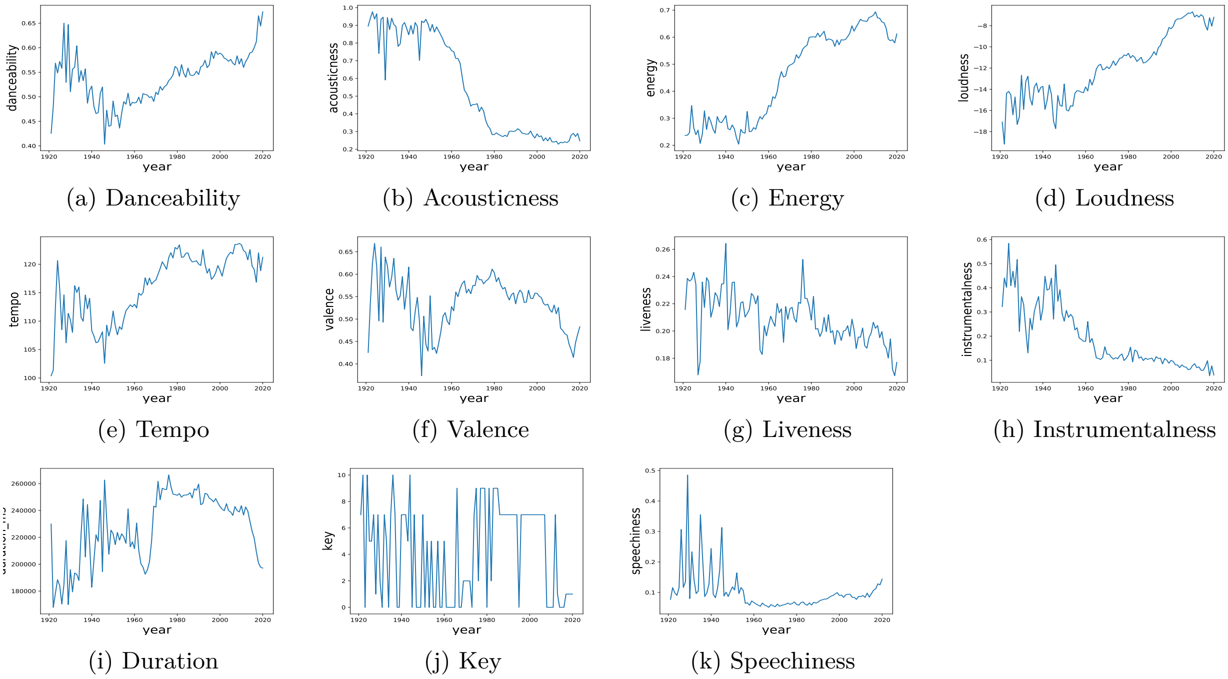
Table 8: Average difference $\epsilon_{i,j}$ in different single feature of different groups of artists

Artist group	Follower $\in$ Pop/Rock	Non-followers $\in$ Pop/Rock	Non-followers $\in$ Country
danceability	9%	8%	8%
energy	12%	15%	21%
valence	14%	15%	14%
tempo	7%	3%	8%
loudness	7%	8%	9%
key	53%	46%	54%
acousticness	18%	17%	24%
instrumentalness	15%	13%	20%
liveness	9%	7%	13%
speechiness	3%	3%	4%
duration	3%	3%	5%

8 Task 5: Characteristics and Artists of Music Revolution

8.1 Iconic Features of Music Revolution

To identify the iconic characteristics (feature) of the music revolution, we should find the features with a large temporal change in the whole field of music. We are interested in the 12 features describing music style as mentioned before, and use the data in file *"data_by_year.csv"*. Figure 11 shows the



(a) Danceability (b) Acousticness (c) Energy (d) Loudness

(e) Tempo (f) Valence (g) Liveness (h) Instrumentalness

(i) Duration (j) Key (k) Speechiness

Figure 11: Global evolution of music characteristics (normalized) in the past 90 years

evolution of all the 12 features except mode because it always equals to 1 in this file. We observe that there has been a significant increase in danceability, energy, loudness, tempo and valence since the 1950s, and that there has been a dramatic decrease in acousticness and instrumentalness since the 1950s. All of these changes signified that there was a music revolution in the 1950s. The increase of danceability, energy, loudness, tempo, and valence signified the revolution in music style, which is mainly due to the rise of Pop/Rock. The decrease of acousticness and instrumentalness indicates the revolution in music equipment, which is mainly because in the 1950s the invention of the electronic transistor by Bell Laboratories contributed to the invention of electric guitar, bass, and other electronic instruments and equipment. Moreover, Valence and tempo peak in the 1970s which means that people at that time preferred high-pace positive music, and that is exactly the golden age of Pop/Rock.

8.2 Representative Artists of Music Revolution

The great change in these typical features are leaded by some creative artists. These revolutionary artists generally have the characteristics of strong creativity and great influence, and their music styles are broadly spread. We assume that a revolutionary artist must have followers. To quantify the revolution degree of an artist, we introduce a measuring indicator in Equation (16), where function h means normalizing the input by its maximum, L_f is the number of followers of this artist i.e. the number of out-going edge in our directed network, L_i the number of the artist's influencers i.e. the number of in-going edge in the network, Sim_i the average cosine similarity between the artists and his/her influencers, Sim_f the average cosine similarity between the artists and his/her followers, which can be calculated by Algorithm 1, and $S_{influence}$ the score of influence that we have obtained in task 1.

$$\mathbb{D}^{add}_{revolutionary} = h(\frac{L_f}{L_i}) + h(\frac{Sim^f_{cos}}{Sim^i_{cos}}) + h(S_{influence}) \tag{16}$$

We interpret our indicator of revolutionary degree with an aggregate of three parts: **spreading ability**, **creativity** and the **actual influence**. Spreading ability is interpreted by $\frac{L_f}{L_i}$ the ratio between the number of his/her followers and that of his/her influencers: the larger the ratio $\frac{L_f}{L_i}$ is, the broader his/her music style propagates. Creativity is measured by the ratio $\frac{Sim_f}{Sim_i}$ between the relevant followers' similarity and relevant influencers' similarity: a larger ratio $\frac{Sim_f}{Sim_i}$ indicates a higher creativity of the artist. The influence score $S_{influence}$ calculated in task 1 gives an estimation of influence not only on other artists but also on the public: a larger score $S_{influence}$ indicates his/her works attract more attention of the public and then promote its propagation more widely beyond artists.

Apart from the above additive model of revolutionary indicator, we also propose a multiplication model of this indicator, as given in Equation (17), where all the variables and functions are the same.

$$\mathbb{D}^{mul}_{revolutionary} = h(\frac{L_f}{L_i}) \times h(\frac{Sim^f_{cos}}{Sim^i_{cos}}) \times h(S_{influence}) \tag{17}$$

Table 9: Revolution degree of top-10 artists by additive indicator $\mathbb{D}^{add}_{revolutionary}$

Rank by $\mathbb{D}^{add}_{revolutionary}$	Influencer rank	Artist Name	$\mathbb{D}^{add}_{revolutionary}$
1	1	The Beatles	2.104
2	20	Hank Williams	2.095
3	11	Billie Holiday	2.063
4	58	Nat King Cole	1.798
5	2	Bob Dylan	1.797
6	78	Howlin' Wolf	1.706
7	55	Thelonious Monk	1.649
8	12	The Kinks	1.591
9	3	The Rolling Stones	1.546
10	4	The Beach Boys	1.505

Table 10: Revolution degree of top-10 artists by multiplication indicator $\mathbb{D}^{mul}_{revolutionary}$

Rank by $\mathbb{D}^{mul}_{revolutionary}$	Influencer rank	Artist Name	$\mathbb{D}^{mul}_{revolutionary}$
1	1	The Beatles	0.287
2	11	Billie Holiday	0.277
3	20	Hank Williams	0.230
4	2	Bob Dylan	0.162
5	12	The Kinks	0.128
6	58	Nat King Cole	0.114
7	4	The Beach Boys	0.102
8	55	Thelonious Monk	0.094
9	28	Ella Fitzgerald	0.093
10	78	Howlin' Wolf	0.092

We calculate the revolution degree using our proposed indicators for the top 100 influential artists that we have selected in task 1 and list the top-10 results in Table 9 and Table 10. We can see that an influential artist is not necessarily a revolutionary artist. We select the intersection of the top-10 most revolutionary artists between the two indicators as representative revolutionary artists: The Beatles, Billie Holiday, Hank Williams, Bob Dylan, The Kinks, Nat King Cole, The Beach Boys, Thelonious

Monk, and Howlin' Wolf.

Algorithm 1: Calculate parameters of revolutionary indicator $\mathbb{D}^{mul}_{revolutionary}$ and $\mathbb{D}^{add}_{revolutionary}$

Input: Directed artist influencer network $G = (V, E)$

Output: L_f, L_i, Sim^f_{cos}, Sim^i_{cos}

$Sim^f_{cos} = 0$;
$Sim^i_{cos} = 0$;
for *artist* $v_k \in V$ **do**
    $L_i = \deg^{in}(v_k)$;
    $L_f = \deg^{out}(v_k)$;
    for $(v_k, v_q) \in E$ **do** $Sim^f_{cos} \leftarrow Sim^f_{cos} + Sim^{(k,q)}_{cos}/L_f$;
    for $(v_p, v_k) \in E$ **do** $Sim^f_{cos} \leftarrow Sim^f_{cos} + Sim^{(p,k)}_{cos}/L_i$;
end
return L_f, L_i, Sim^f_{cos}, Sim^i_{cos}

9 Task 6: Influence Process of music evolution in Pop/Rock

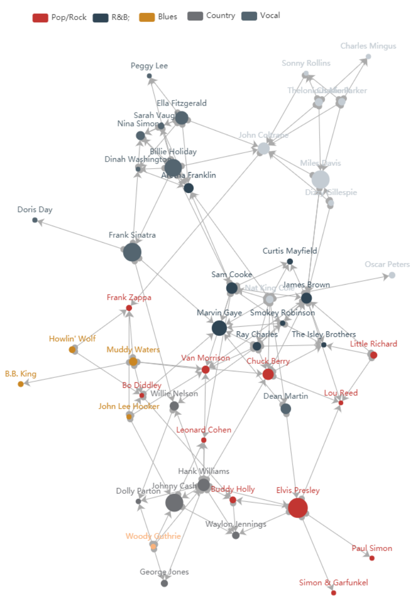
9.1 Influence Process in Pop/Rock

The temporal music influence propagation within the genre can described by the temporal evolution of the directed sub-network where all the vertices represent artists belonging to this genre. We visualize this influence process of Pop/Rock in Figure 12 by plotting the expansion of the selected sub-network (We select the vertices of some influential artists in Pop/Rock for better visual effects) every 10 years from the 1950s to the 1970s. Pop/Rock originated in the 1950s when Elvis Presley and some early Pop/Rock artist were influenced by Country music. Pop/Rock was flourishing in the 1960s with the appearance of famous bands and musicians, such as The Beatles, The Rolling Stones, and Bob Dylan. Since the 1970s, it has been developing with more and more Pop/Rock artists as in the network the number of links between artists keeps increasing.

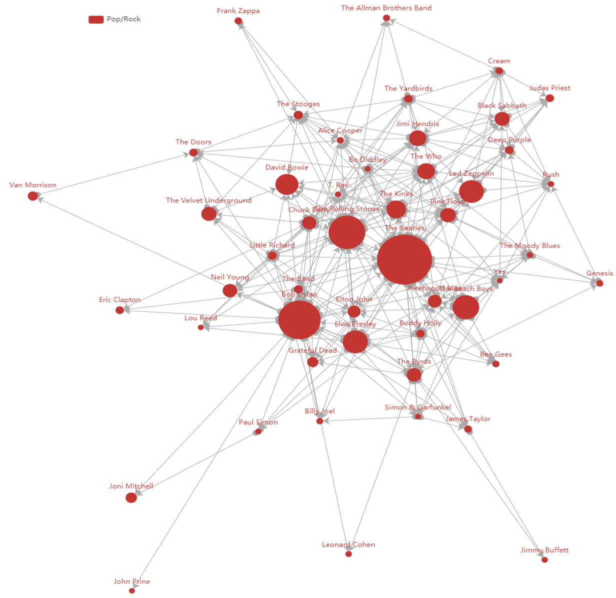
9.2 Dynamic influencers in Pop/Rock

We choose Pop/Rock to study its dynamic influencers for its large number of artists. We argue that generally, the **dynamic influence factor** during a period t of a genre i of an artist i can be measured basically from 3 aspects: the follower number $L^f_{i,t}$ within this period t (classified by the follower's start year), the temporal mean popularity $M_{i,t}$ during this period, and the temporal mean number $C_{i,t}$ of works during this period. The **dynamic influence factor** $S^{dynamic}_{i,t}$ can be measured by the weighted sum in Equation (18), where the weights w_1, w_2, w_3 are recalculated within in the genre using the same **entropy weight method** as in task 1. The absolute influence score in task 1 can be viewed approximately as a sum of the series of dynamic influence factors multiplied by period length T, i.e. $S^{influence}_i = \sum_t TS^{dynamic}_{i,t}$ where $t \in \{1930\text{s}, 1940\text{s}, \cdots, 2010\text{s}\}$. We can use the rank in $S^{influence}_i$ to identify the dynamic influencers in different period.

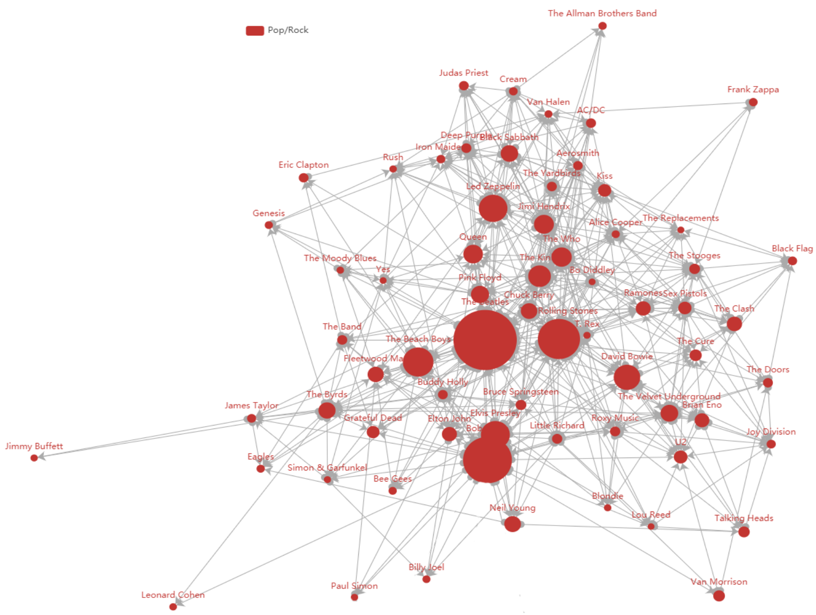
$$S^{dynamic}_{i,t} = w_1 L^f_{i,t} + w_2 M_{i,t} + w_3 C_{i,t} \tag{18}$$



(a) Pop/Rock with other genres in the 1950s



(b) Pop/Rock in the 1960s



(c) Pop/Rock in the 1970s

Figure 12: Temporal influence process of music evolution in Pop/Rock

We select some influential artists, The Beatles, Elvis Presley, David Bowie, and Buddy, to analyze their dynamic influence factors and their absolute influence score. It can be clearly seen from Figure 13 that although Elvis has no outstanding dynamic influence under this measurement system, but it has a large time span, from the 1950s to the present. The absolute influence is approximately the area between the curve of an artist or genre and the time axis, therefore Elvis has a large absolute influence. For the whole genre Pop/Rock, with the substantial development in the 1960s, its dynamic influence has also shown an upward trend. Buddy Holly, who passed away in 1959, still gets a dynamic influence in the 1990s. The possible reason may be that with the development of the Internet, people can get in touch easily with the songs created several decades ago. Artists like Buddy Holly had decreasing dynamic influence in the 1970s and 1980s, are re-found by the Internet and paid more attention for their qualified works.

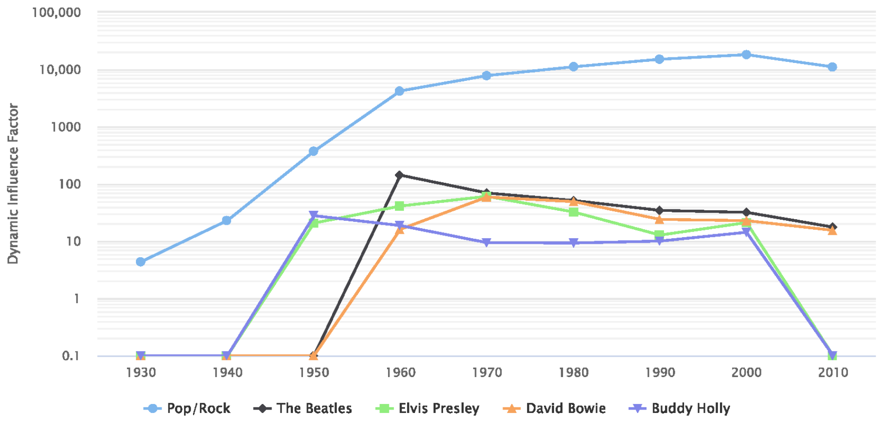


Figure 13: Dynamic influence factor of several artists and genre

Further improvements in our dynamic influence factor may consist on multiplying the follower's number by their average similarity $Sim_{cos}^{mean}{}_{i,t}$, which writes in Equation (19), and add a term to describe the influence inherited from the previous period with a damping factor β ($0 < \beta < 1$), which writes in Equation (20).

$$S_{i,t}^{dynamic} = w_1(Sim_{cos}^{mean}{}_{i,t}L_{i,t}^{f}) + w_2M_{i,t} + w_3C_{i,t} \tag{19}$$

$$S_{i,t}^{dynamic} = \beta S_{i,t-1}^{dynamic} + w_1(Sim_{cos}^{mean}{}_{i,t}L_{i,t}^{f}) + w_2M_{i,t} + w_3C_{i,t} \tag{20}$$

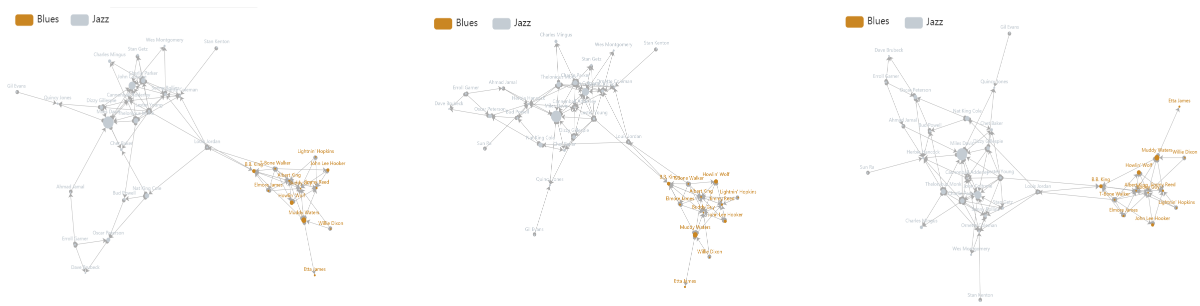
10 Task 7: Mutual Influence Between Music and Society

We try to explain the rapid expansion of Pop/Rock in the network since the 1960s from its cultural influence and social changes at that time.

- **Political change**: On November 22, 1963, U.S. President John F. Kennedy was assassinated during a live TV broadcast, which shocked the country. Then on December 17, 1963, the WWDC radio station in Washington, D.C. aired The Beatles' "I Want to Hold Your Hand". This song was a comfort to the people who were immersed in the gloomy atmosphere. Pop/Rock and The Beatles walked into the middle of the stage.
- **Technological change**: Since the appearance of the radio in the 1920s, this novel method of information dissemination has been rapidly popularized. By the end of the 1950s, the popularity of radio in the United States was already very high, and combined with the WWDC radio station' s promotion of Pop/Rock, the genre quickly became popular.
- **Social change**: The baby boom in the 1950s leads to a sharp increase in the number of young people, and the awakening of the young generation' s sense of autonomy has also prompted

them to seek new and more exciting forms of music, and the rise of rock music also heralds young people' s rebellion towards "traditional order, mainstream culture and authority".

- **Cultural influence of Pop/Rock on the society**: Rock music broke racial barriers. With the emergence of rock music, blues, R&B, Jazz, Country and other music forms were integrated and reconstructed, forming their own style, and they were welcomed by all races [5].



(a) Jazz and Blues in the 1950s (b) Jazz and Blues in the 1960s (c) Jazz and Blues in the 1970s

Figure 14: Sub-network of influential artists of Jazz and Blues

We observe that the sub-network of influential artists of Blues and Jazz keeps almost a constant size, shown in Figure 14, when Pop/Rock was flourishing from the 1950s to the 1970s. We explain as follows:

- **Characteristics of Jazz**: Jazz requires listeners to have a high level of appreciation, and at the same time requires the performing artists to have a high level of music performance, so it is classified as "elite music" and it is difficult to achieve popularization.

- **Economic and social change**: Some Jazz musicians gave up jazz during the Great Depression. Jazz music in the 1940s was influenced by the United States' participation in World War II. The new generation of young people who grew up after the Second World War prefers the bright and vibrant rock. Jazz and Blues receive the impact from Pop/Rock.

11 Strengths and Weaknesses

Strengths: We combine the out-going edge number of an artist vertex in influencer graph with popularity and the number of released works to objectively measure the artist's influence. We build a dynamic network to reflect the evolution of different genres, making it easier to find the changes in each decade. We use cosine similarity to measure the similarity between artists and genres meanwhile pointing out the limitations of using Euclidean distance for this measurement. We prove the association between influence and similarity by Kolmogorov–Smirnov test. We introduce some indicators to measure the revolutionary degree of an artist and identify the dynamic influencers, respectively. We also exemplify some changes in our sub-networks with social, economic, technological, and cultural changes to prove the effectiveness of our network.

Weaknesses: In the influencer network analysis, we do not take into consideration the propagation of the influence from an influencer to its indirect followers when we measure the magnitude of influence of artist. This is also the direction of further optimization of our model, for which we may apply an exponential decay series of weights on the two or three following artist vertices in the chain of influence propagation.

12 Conclusion

We can use cosine similarity to measure similarity between artists and genres and measure influence by a weighted sum of popularity, the number of followers, and the number of works. Through analysis of similarity and the source of influence, we find that artists of some music genres, like Electronic, are more similar to those of other genres, but that artists of most genres are similar to those within the same genre. The mean vector of 12 characteristics (danceability, energy, valence, tempo, loudness, mode, key, acousticness, instrumentalness, liveness, speechiness, explicit, and duration) can be used to distinguish genres from each other. We confirm the association between similarity and influence between artists , and find that energy is a "contagious" characteristics for artist The Rolling Stones within genre Pop/Rock. Characteristics, such as danceability, energy, loudness, tempo, valence, acousticness, and instrumentalness, can signify music revolutions by their drastic change. The most revolutionary artists in Pop/Rock by our measurement are The Beatles, Billie Holiday, Hank Williams, Bob Dylan, The Kinks, Nat King Cole, The Beach Boys, Thelonious Monk, and Howlin' Wolf. The dynamic directed influencer network can fully interpret the evolution of influence on both artists and genres and we do find some actual reasons for certain identified change in the sub-networks.

References

[1] H. Garg, N. Agarwal, and A. Choubey. Entropy Based Multi-criteria Decision Making Method under Fuzzy Environment and Unknown Attribute Weights. *Global Journal of Technology and Optimization*, 6:182, 2015. doi: doi.org/10.4172/2229-8711.1000182.

[2] P. Huang, H. Liu, C. Chen, and P. Cheng. The Impact of Social Diversity and Dynamic Influence Propagation for Identifying Influencers in Social Networks. *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, 1:410–416, Nov 2013. doi: 10.1109/WI-IAT.2013.58.

[3] M. Mauch, R. M. MacCallum, M. Levy, and A. M. Leroi. The Evolution of Popular Music: Usa 1960–2010. *Royal Society Open Science*, 2, 2015. doi: 10.1098/rsos.150081.

[4] H. Nick. *You Really Got Me: The Story of The Kinks*. Omnibus Press, 2017. ISBN 978-1-7855-8851-8.

[5] M. Regev. Pop-Rock Music as Expressive Isomorphism: Blurring the National, the Exotic, and the Cosmopolitan in Popular Music. *American Behavioral Scientist*, 55:558–573, 2011. doi: 10.1177/0002764211398079.

[6] M. Verleysen and D. François. The Curse of Dimensionality in Data Mining and Time Series Prediction. *Lecture Notes in Computer Science*, 2, 2005.

Dear **Integrative Collective Music (ICM) Society**,

We have completed the model with indicators measuring dynamic musical influence of artists and music genres. Through our modeling, the influence and creativity of each artist and genre have been quantitatively analyzed, and the process of the spread and evolution of music genres can be better understood. We use force-oriented algorithms to create and visualize multiple directed graphs similar to a "social network" and show the artist's musical influence as the size of the node. Each of these graphs represents the relationship between influencers and followers in a particular era and all of these graphs constitute a dynamic influencer network to show the temporal evolution.

When calculating the artist's personal musical influence, we comprehensively consider popularity, the number of works, and the influence by time. When measuring the changes of the entire genre, the scale of the genre is taken into consideration. When studying the interaction between genres, we focus on the change makers in each genre, and use the artist's influence and creativity to evaluate its role in the entire music genre.

Through observation and analysis at different time and space scales, we can better understand the process of music propagation and evolution. Any music genre has its source of development. The creation of a genre was not born out of thin air, but drew nutrients from the soil of other genres and transformed it into its own characteristics. In the early 1950s, Pop/Rock began to emerge. Its original founders were directly influenced by Jazz, Country and Blues. Pop/Rock learned from not just a single genre but multiple genres. Afterwards, different pop styles based on the "British style" began to collide and blend with the original style, and a large number of artists with cross-generation influence, such as The Beatles and The Rolling Stones, emerged.

From the perspective of time scale, some music genres have developed much more slowly than Pop/Rock. This is very intuitive from the network we built. In our network, Jazz and Blues are just such examples. The styles of popular music have been changing over the past 90 years. The style of music such as Blues and Jazz has basically not changed. It may be due to the many instruments required for its performance, but this may make people form a stereotype of these genres.

If more data is added to our model, it may mitigate the current genre size gap to a certain extent. At the same time, our model will provide a more accurate description of the development and evolution of some new genres such as children' s and Avant-garde. We could also find out who has played the role of "influencer" and "revolutionist" in the development of small genres.

In future music research, we can add indicators to measure innovation and the continuous active time span of an artist. The more innovative an artist is, the easier the artist is to influence or even create a new genre, and the longer the artist keeps active, the more people may be influenced. We also propose that the musical instruments used by these genres and their development should be included in the statistics.

Yours sincerely,

Team #2100640