

音乐的影响

音乐是文化遗产的重要组成部分，在人类集体经验中所扮演的角色，因此在本文中为量化表达音乐演进过程，我们提出了一种衡量音乐影响力的综合评估模型，并以此模型为工具对艺术家和流派的进化和革命趋势进行考察。

首先，在音乐艺术家影响关系定向网络中，我们选取了包括影响人数、竞争力水平等在内的能够衡量艺术家音乐影响力的多个指标，通过熵权法赋权建立灰色综合评价模型，通过对多种艺术家属性的平衡评价，定量衡量影响者对追随者的影响力，根据该模型我们得到了以 Beatles、Bob Dylan、Rolling Stones 为首的极具影响力的百余名艺术家大师。

第二，为了探究音乐作品的相似程度，我们将音乐作品抽象为 12 个特征维度决定的点集，问题转化为大数据分类问题。由于相似的作品会内聚成簇，我们采用 k-means 聚类分析并由交叉验证和 PCA 降维两个手段确定了作品实际上应该分为两类。通过对内聚图的分析我们解释了决定了作品本质的并非流派，而是审美、历史和技术的演变。

第三，我们以音乐的技巧运用入手来研究流派内的联系和流派间的区别。通过对情况迥异的皮尔逊相关系数热力图的解释，我们发现了流派内对同一音乐技巧的偏好和流派间对不同音乐技巧运用的区别。但流派间同样存在着相互借鉴情况，因此我们并按照流派间的借鉴关系建立了有向图，揭示了随时间的演化各流派如何借鉴其他流派，而一些流派间的交流借鉴情况更为活跃。

第四，我们通过标称数据相似度度量法对评估模型得到的影响力及相似度模型得到的音乐整体相似度进行关联分析得到在影响力和音乐相似度之间存在很大相关性，定量表征了影响者对于追随者音乐产生了一定程度的影响，为衡量该影响对于各音乐特征指标的影响程度，我们又通过余弦相似度度量法对子网络的各音乐特征指标进行相似度度量并以其作为数据列采用灰色关联法进行关联度分析，同时为提高合理性和代表性，我们对全局网络进行特征差量分析，通过仿照神经网络的 SGD 反馈规则建立衡量追随者风格的改变量，计算发现由高影响力艺术家的追随者曲风和影响者更为类似并将结果以双层旭日图可视化呈现，从全局网络和子网络、定量分析和可视化分析多方面得出其中 valence、loudness、tempo、energy 等音乐指标主要继承了影响者的影响力，其关联度达 0.9 以上，而 key、mode 是最少被追随者模仿的属性。

第五，建立关于影响力-年份-派别的三维坐标图确定变革时间为 1957 年左右，根据变革时间确定出该时段变化较大的音乐特征为，在发生大变革的流派内部联系实际历史环境分析流派内音乐特征变化并与全音乐领域水平做关联度分析，并针对 pop 流派与全音乐领域之间各评估指标进行关联度分析，发现该流派对音乐领域进化和发展趋势起到重要影响。

第六，我们对引起音乐史上重大发展的 pop 派别进行可视化及关联度分析，对 pop 流派、全体音乐领域做出影响力、流行度、新生代艺术家人数随时间发展趋势对比图，通过计算 pop 流派与全体音乐领域发展之间的关联度（均在 0.8 以上）表征 pop 流派随时间一直影响着全局音乐领域的发展，并在第一次重大飞跃之后发展势头均有所下降，通过分析 pop 流派与全局音乐领域的各音乐指标逐年发展变化，我们得出 pop 流派发生重大变化，通过其中关键音乐特征指标***的重大变化而映射影响到全局音乐领域的音乐特征，从而导致全体音乐的飞跃发展并持续发展。

最后，我们通过评估模型的影响力，各音乐特征指标及外部因素对流派进化和革命趋势进行了进一步分析，为更好地适用于不同大小的全局网络及状态，我们对模型进行了灵敏度分析和综合优化，该模型在音乐进化及影响力衡量等方面具有着很好的迁移性和应用价值。

关键词：灰色综合评价模型、KNN，余弦相似度、灰色关联分析法

1 问题重述

1.1 Problem Background

1.2 Our Efforts

几句背景

Music is one of the best carriers of popular aesthetic culture based on the self-expression and the joy of inner life. As an important part of cultural heritage, it affects people's values and lifestyles. The series of changes it causes cannot be ignored in any society and culture. Music is also inextricably linked to new music and music artists.

Therefore, this article establishes a method to quantify musical evolution, thus understanding and measuring the influence of previously produced music on new music and musical artists. Our goals are to examine evolutionary and revolutionary trends of artists and genres.

In this article, we will:

- create multiple directed networks of musical influence, then explore a subset of musical influence by creating a subnetwork of your directed influencer network.
- develop musical similarity measurement model and use it to measure the similarity of artists within genre and artists between genres.
- compare similarities and influences between and within genres.
- indicate whether the similarity data suggest that the identified influencers in fact influence the respective artists.
- characterize signify revolutions (major leaps) in musical evolution.
- analyze the influence processes of musical evolution that occurred over time in one genre.
- express information about cultural influence of music in time or circumstances.
- Alternatively, how can the effects of social, political or technological changes (such as the internet) be identified within the network?

2 问题分析

本文旨在通过建立衡量音乐影响力的数学模型考察艺术家和流派的进化和革命趋势。为解决这一问题，我们首先根据所给条件及附件数据进行预处理，确定关键指标讨论建模。

(1) 在问题一中，问题核心旨在开发某些指标作为合理表征音乐影响力的参量，此时问题转化为选取指标对音乐影响力进行定量表征。我们使用 Influence-Data 数据集建立代表艺术家关系的一个定向网络，通过定向网络使得影响者与追随者产生多对多映射联系，开发一些潜在指标，分别通过 PR 迭代算法和对这些指标建立灰色综合评价模型以表征音乐影响力。

(2) 该问题的目标是建立音乐相似性度量模型。首先我们将音乐作品抽象为表征作品的 12 条特征信息，问题转化为大数据点集的分类问题，通过无监督学习的 k-means 算法进行聚类分析，通过对相似度相近的节点形成簇的观察，探究能够本质地反映作品

特征的分类，究竟是流派之分还是聚类之分。最后我们计划通过研究作品聚类的原因（审美、历史、技巧的演变），成功解释为何流派不能作为作品的本质分类。

（3）由我们在第二问中的发现，我们将对技巧的偏好作为入手点。音乐技巧会体现为信息特征的反常表现，在数值上则表现为较高的皮尔逊相关系数。于是我们计划按照皮尔逊相关系数作出热力图，通过解释热力图迥然不同的分布来对流派内外的联系和区别作出回答。通过建立有向图，通过演示该有向图随时间的演化揭示了流派随时间推移的变化，流派之间是在何种程度上相互借鉴的。

（4）该问题可以转化为如何衡量影响者的影响力与追随者-影响者音乐的相似性，如果影响力与相似度之间呈正相关，那么可以认为“有影响力的人”真的会影响追随者创作的音乐；相应的，问题的后半部分可以转化为比较影响者与追随者各音乐特征相似性之间，如果存在某一音乐特征相似性远大于其他音乐特征，那么可以认为该音乐特征比其他特征更具感染力，如果各音乐特征相似性差别不大，那么可以认为它们在影响某个特定艺术家的音乐方面都扮演着相似的角色。

（5）对于该问题我们逆向思维，某些艺术家变革了部分音乐特征引领了潮流从而带动整体音乐发生重大变革，故我们需要首先确定出变革时间，根据变革时间确定出该时段变化比较大的音乐特征，在发生大变革的流派内部分析流派内音乐特征变化关联度最大的艺术家确定其为所要寻找的代表变革者。

（6）该问题可转化为分析在一个音乐流派中随着时间的推移，影响者们的影响力的变化过程及关键音乐指标与音乐发展关系，计划对该流派各音乐特征随时间变化情况及影响力水平随年份变化作图分析，找出发展过程中随时间发生重大变化的音乐特征，分析其对影响力、新生代艺术家人数、流行度等等的影响，结合 model1 中评价模型指标解释动态影响者指标及其影响过程。

（7）该问题要求我们基于前几问的结果，结合实际情况进行分析。

3 模型假设

为了简化给定的问题，并修改它更适合于模拟真实情况，我们作出以下基本假设，每个假设的合理性都适当地被证明。

4 符号约定及名词定义

符号	含义	符号	含义
x_i	各数据列指标	PR	表征直接/间接影响人数水平
A_i	权重	r	相关系数
ρ	比例系数	sim	相似度

5 模型建立与求解

5.1 问题一——综合评价模型建立与求解

5.1.1 建模思路

该问题的核心旨在开发某指标作为合理表征音乐影响力的参量，此时问题转化为选取指标对音乐影响力进行定量表征。我们使用 influence_data 数据集建立表征音乐影响力的多个定向网络，通过定向网络使得影响者与追随者产生多对多映射联系。而艺术家之间的音乐影响力存在差异，这需要在我们的图中通过对节点和边计算合理权重来得以体现。

根据 PageRank 网络迭代算法依照影响人群具体分布确定网络节点权重，并以此表征艺术家自身影响力；选取能够衡量艺术家音乐影响力的多项指标，通过熵权法分配

各指标的重要程度，建立灰色综合评价模型，通过对多种音乐属性的平衡评价，定量比较影响者对追随者的影响水平，以此为网络边权重。

a. 节点权重——艺术家自身影响力

~~各艺术家与其追随者之间的映射关系反应了他本身的影响力：一方面，如果一个追随者被更多影响者所影响，那么追随者对每个影响者的重视(attention)应该就应更小。反过来，如果一个影响者能够影响更多的追随者，那么他的音乐影响力(influence)应当就更大。当一名影响力较大的追随者出现时，他的影响力必然超过该名影响者。~~

~~以上思想可以根据 PageRank 网络迭代算法体现，最终算法收敛至使得网络平稳分布的各节点权重值（称为 PR 值）。这样我们就依照影响人数确定了网络节点权重。~~

b. 边权重——两两影响关系强度

~~节点权重成功地反映了艺术家自身的影响力，但具体到影响者-追随者的两两关系，我们该如何表征这一关系的程度呢？也即寻找合适的边权重，衡量影响者对追随者的影响水平。~~

~~对此我们选出一些能够衡量艺术家音乐影响力的指标，通过熵权法客观合理地分配各指标的重要程度，然后建立灰色综合评价模型，通过对多种艺术家属性的平衡评价，定量比较影响者对追随者的影响水平，同时也确定了网络的边权重。~~

5.1.2 定向网络及其子网络的建立

根据influence_data数据集中给出的影响者-追随者关系，将音乐家个体表征为网络中各节点，将影响者（父节点）与追随者（子节点）关系表征为连接各节点的箭线，由一名影响者节点链接影响到其相邻点进而由扩散作用影响于整个网络，得到一个全局定向网络。为展示简便起见，有简化示意图如下：

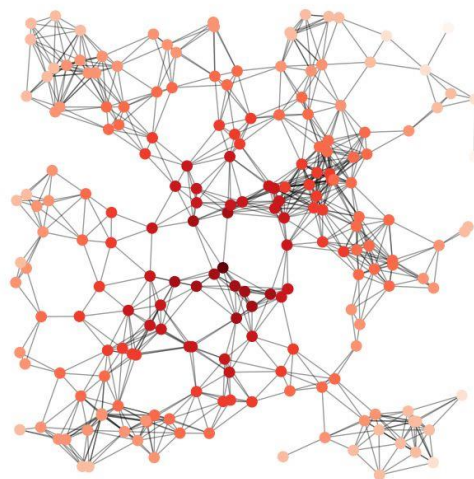


图1. 影响者-追随者定向网络

全局网络是对所有艺术家个体进行的多对多映射分析，为具体探究影响水平，选取其中任意追随者，研究其所有影响者对这一追随者的影响水平。以追随者节点为中心，影响者节点围绕中心，构成了一个任意研究对象的一对多联系子网，如图所示。

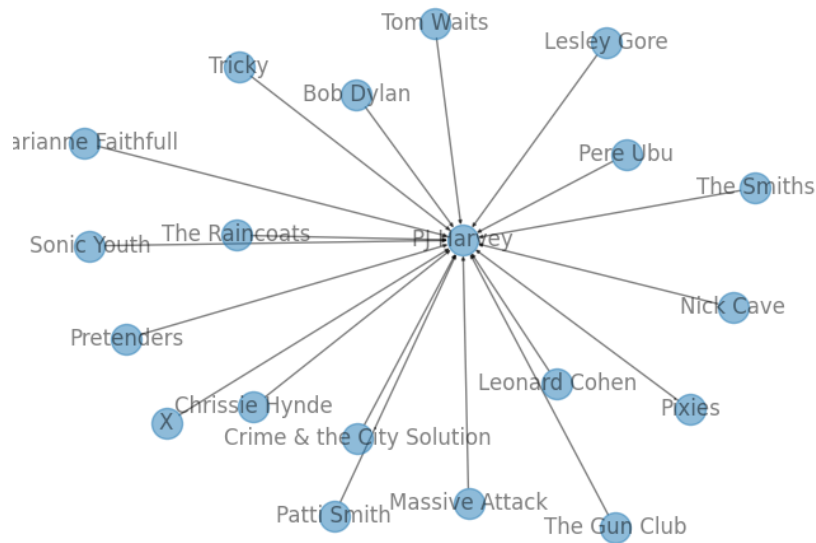


图2. 子网络

其中各节点权重为PR值，边权重由灰色评价模型得出，以下给出计算原理和方法。子网络以PJ Harvey为例进行研究，事实上，我们的模型适用于任意追随者。

5.1.2 各模型指标的确定

a. PR值 [参考文献]

在所得的各个定向网络内由于各节点都具有一定的影响力水平，当一个影响力水平较高的艺术家是被另一位艺术家影响时，那么他的影响者的影响力必然高于该名艺术家，艺术家的定向影响关系层层累积，我们将影响者和追随者之间映射关系水平赋权，首先为其随机赋权，根据PageRank算法对全局网络进行迭代可以得到使得该定向网络趋于稳定的各权值，将迭代后得到的权重称为PR值，此时，PR值表示了受影响者直接或间接影响的人数水平，将影响人数转化为PR值指标表示。

我们得到全局网络中各艺术家的PR值，取PR值较高（前100名）的影响者集合作为艺术家大师。为了可视化地展示我们的结果，我们以PR值大小作为字体大小的依据，将这些艺术家大师展现在词云中。

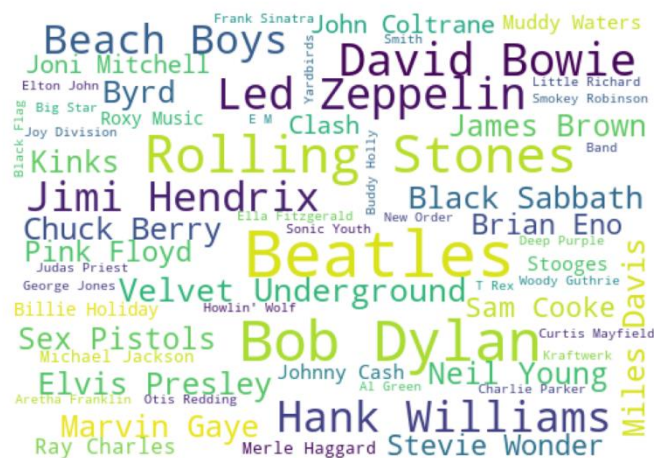


表1. 艺术家大师词云

b. 艺术家影响力属性向量

为了进行灰色关联分析需要决定哪些属性可以衡量艺术家的影响力。对此我们认为，从网络中可以发现的能够衡量音乐影响力的属性包括：

x_1 : PR 值/节点权重，表征受各影响者直接或间接影响的影响人数。

x_2 : 定向网络中与所研究的艺术家同流派人数，表征影响者所在流派的竞争激烈程度。

x_3 : 定向网络中与所研究的艺术家同年出道的同流派人数，表征影响者所在领域当年竞争激烈程度。

x_4 : 定向网络中与所研究的艺术家同年出道的总人数，表征影响者出道时的竞争激烈程度。

x_5 : 追随者是否继承了影响者的流派，相同为 1，不同为 0，是影响者对追随者影响的一个重要方面。

以上构成了艺术家影响力属性向量：

$$x' = (x_1, x_2, x_3, x_4, x_5)^T.$$

为了减少以上指标选取的主观性，我们采用熵权法，计算出指标的熵权重 w ，更为客观的艺术家影响力属性向量 x ，定义为：

$$x = wx'$$

5.1.3 综合评价模型

根据上述指标对子网络中各艺术家影响力水平建立综合评价模型进行分析。子网络的中心为追随者PJ Harvey，其影响力属性向量记为 x_0 ，子网络中围绕中心的是各影响者，他们都影响了PJ Harvey，记第 i 名影响者的属性向量为 x_i ，如表5.1.2所示。

artist-name	x1	x2	x3	x4	x5
PJ Harvey	0.00024	2808	369	1129	1
Pretenders	0.000195	2808	364	796	1
Bob Dylan	0.000252	2808	372	855	1
Sonic Youth	0.000328	2808	462	951	1
Crime & the City Solution	0.000105	2808	462	951	1
Nick Cave	0.000223	2808	364	796	1
Captain Beefheart	0.000141	2808	372	855	1
Leonard Cohen	0.000116	2808	99	540	1
Pixies	0.000165	2808	462	951	1
X	0.000156	2808	364	796	1
Lesley Gore	0.000111	2808	372	855	1
The Smiths	0.000226	2808	462	951	1
Chrissie Hynde	0.00018	2808	364	796	1
Massive Attack	0.000265	208	31	951	0
Pere Ubu	0.000146	2808	364	796	1
Tom Waits	0.000155	2808	364	796	1
Tricky	0.000165	208	61	1129	0
The Gun Club	0.000142	2808	462	951	1
Patti Smith	0.000165	2808	364	796	1

Marianne Faithfull	9.63E-05	2808	372	855	1
The Raincoats	0.000108	2808	364	796	1

根据熵权法确定的各指标权重:

权重表				
W1	W2	W3	W4	W5
0.419255	0.142749	0.167414	0.127833	0.142749

代入计算得到各艺术家影响力属性向量。

选取 x_0 作为参考列,所有影响者的属性向量 x_1, x_2, \dots, x_n 作为比较列,根据各指标数据代入灰色关联分析算法计算出对于参考列的关联度,得到各影响者与追随者的关联度作为影响关系强度的评价指标。于是我们得到,子网络中代表各艺术家影响力水平的边权重为:

艺术家	边权重	艺术家	边权重
Pretenders	0.687054	The Smiths	0.698312
Bob Dylan	0.701389	Chrissie Hynde	0.687054
Sonic Youth	0.698312	Massive Attack	0.517105
Crime & the City Solution	0.698312	Pere Ubu	0.687054
Nick Cave	0.687054	Tom Waits	0.687054
Captain Beefheart	0.701389	Tricky	0.532089
Leonard Cohen	0.611421	The Gun Club	0.698312
Pixies	0.698312	Patti Smith	0.687054
X	0.687054	Marianne Faithfull	0.701389
Lesley Gore	0.701389	The Raincoats	0.687054

以上,PR 值越高,艺术家本身影响力越强;影响者指向追随者的权重越高,他们之间的两两影响关系越强烈。至此,我们的工作从艺术家本身影响力和影响关系强度两个角度,揭示了艺术家的影响力水平。

5.2 问题二——音乐相似性度量模型建立与求解

问题二——音乐相似性度量模型建立与求解

模型思路

问题二我们使用(full_music_data and the two summary data sets(with artists and years) of music characteristics)建立音乐相似性度量模型。首先通过数值归一化、标准化手段对数据集进行预处理,使用 PCA 将多指标的数据集降维处理,使得各指标之间的特征关联度减小,通过 KNN 算法进行聚类分析,在不考虑流派影响下对各节点相似性进行度量,比较各节点相似度,相似度相近的节点形成一个个簇;根据簇点颜色区分不同流派,流派间相似度表征为各簇群中不同色分布情况,流派内相似度表征为各簇群中同色分布情况,通过分布情况定量直观表征流派内艺术家与流派间艺术家相似性

关系。

模型建立与求解

数据预处理

随附的数据文件“...”为我们提供了不同流派艺术家音乐作品的特征。多数算法对数据的相关性，可靠性，有效性都相当敏感，因此我们需要通过清理，选择和标准化数据来执行数据预处理。

数据筛选

原始数据量很大，因此我们应该首先根据信息的完整性和有用性进行数据筛选。对于多名不同流派艺术家跨界合作的作品，我们将其视为同时属于所有参与其中的流派。比如 Howlin' Wolf, Eric Clapton, Steve Winwood, Bill Wyman & Charlie Watts 共同完成的 *I Ain't Superstitious* (feat. Eric Clapton, Steve Winwood, Bill Wyman & Charlie Watts)，属于的流派为 Blues, Pop/Rock, Jazz。对于未记录流派的作品，我们将其忽略。

数据归一化、标准化

音乐的特征可以用表中从 danceability 到 duration-ms 的数值来进行量化。由于我们将指标使用不同的单位，因此我们必须对所有指标进行归一化，以利于进一步分析数据。它可以提供一种比较不同种类数据的方法，并反映不同因素的综合结果。在这里，我们使用最小-最大规格化，将所有值缩放为 [0, 1]。公式 (1) 给出了采用的归一化的一般形式。令 x' 为归一化值，公式为：

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

其中 x_{\max} 和 x_{\min} 分别是同一单位中指标的最大值和最小值。

然后，我们对各个数值特征和流行度之间的关联度采用皮尔逊相关系数 (Pearson Correlation Coefficient)，公式 (2) 进行衡量。

$$p(x, y) = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{(n-1) s_x s_y} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \quad (2)$$

得出结果如表 X 所示，然后我们删除了相关系数过小，即绝对值小于 0.15 的特征，此处用红色标出。

Characteristics	Pearson Correlation
	Coefficient
danceability	0.183133
energy	0.389819
valence	0.020748
tempo	0.094173
loudness	0.420300
mode	-0.036295
key	0.015031
acousticness	-0.477451
instrumentalness	-0.216110
liveness	-0.064326
speechiness	-0.006928
explicit	0.183029
duration_ms	0.057150
year	0.775415
popularity	1.000000

至此数据集预处理完毕。

Data Preprocessing

The attached data files *full_music_data* and the two summary data sets (with artists and years) provides us with.....

.....Therefore, we need to perform data preprocessing by cleaning, selecting and normalizing the data.

Data Screening

The amount of raw data is large, so we should first do data screening according to integrity and usefulness of the information.

Data Normalization

Since we will use indicators with various units, we have to normalize all the indicators to facilitate the further analysis of the data. It can provide an approach for comparison of different kinds of data and reflect the combined results of different factors.

Here we use Min-max normalization, scaling all the values in the range [0,1]. Formula (1) gives the general form of the adopted normalization. Let x' be the normalized value, the formula is :

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

where x_{\max} and x_{\min} are respectively the maximum and minimum value of the indicators in the same unit.

.....

模型建立与求解

要开发音乐相似性度量模型，就必须对不同音乐之间的相似性进行定量描述。一个作品包含很多指标，我们可以把每个作品看成空间上的一个点，其中 n 是每个作品所包含的指标数量。接下来，我们描述音乐之间的相似之处。

我们认为，最能够代表一个流派的作品是流行度最高的那些作品，所以我们选出每个流派中流行度最高的前 100 首作品代表各自流派来分析作品的特征信息，以此对模型进行训练可以更好地初始化聚点。

我们首先将作品信息采用 PCA 降维，然后根据流派类别给点上色，可视化地观察作品聚类情况。

PCA 降维

~~Principal component analysis is one of the most commonly used methods for data dimensionality reduction, mainly through linear transformation, to achieve the purpose of feature compression explanatory variables.~~

~~The algorithm is mainly divided into 6 steps:~~

- ~~1 Construct a variable matrix of order $p \times n$.~~
- ~~2 Standardize each row (representing an attribute field) of the variable matrix X of order $p \times n$.~~
- ~~3 Find the covariance matrix C .~~
- ~~4 Find the eigenvalues of the covariance matrix and the corresponding eigenvectors.~~
- ~~5 Arrange the eigenvectors in rows from top to bottom according to the corresponding eigenvalues into a matrix, and take the first k columns to form a matrix P .~~
- ~~6 $Y=XP$ is the data after dimension reduction to k dimension.~~

Result as follow table 2

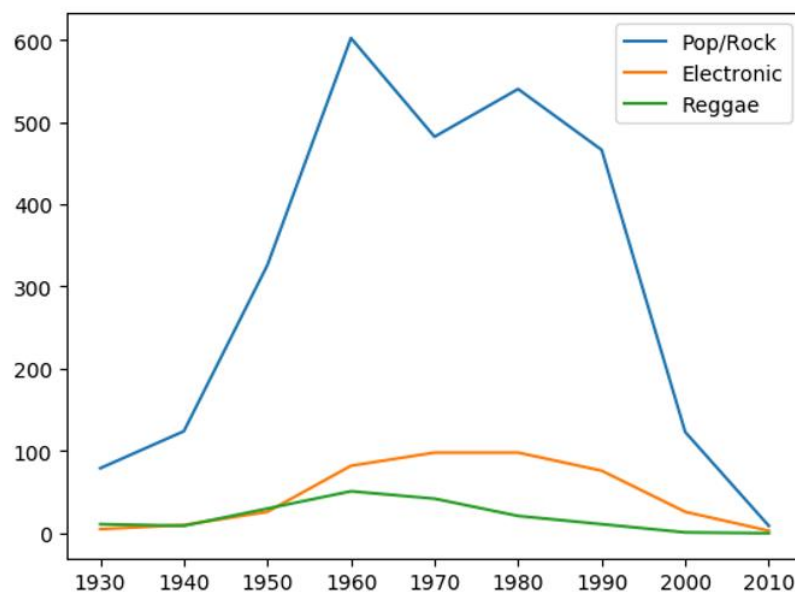
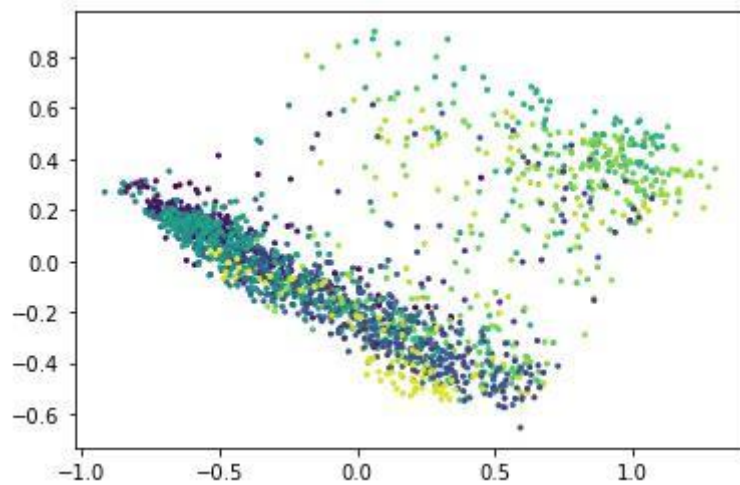


Table 2

Conclusion:

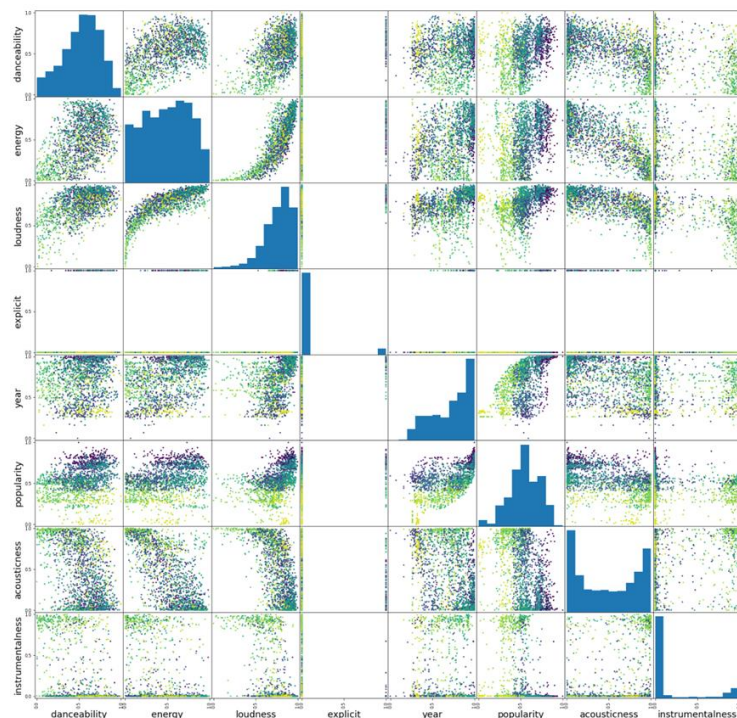
~~For society and politics, the characteristics of songs in the years of peace and stability are different from those in the years of war. The rapid development of the Cold War, World War II, Industrial Revolution and the Internet may have a great impact on music, and music also has an impact on these historical events. The emergence of the Internet has affected the form of music playing, especially the current electronic music has a great impact on the previous traditional media, such as singing machines, and the development of science and technology may also lead to differences in people's hobbies in the field of music~~

PCA 降维的二维可视化图如图 X 所示，作品的特征大致分成两簇，有些流派的内聚情况相当好（例如紫色代表的 Pop/Rock 流派和深绿代表的电子音乐流派），但并非所有流派都如此。流派的特征信息在 PCA 降维的二维可视化图中并不明显。



为了更好地观察作品的特征信息，我们接着截取每两种特征信息，将其画到二维空间中，在散点矩阵图（scatter matrix）中观察作品特征，散点矩阵图如图 X 所示。

music data divided by genres



我们以紫色代表的 Pop/Rock 为例。可以看出，在许多情况下，Pop/Rock 的内聚情况相当好，这和我们在 PCA 降维图中观察到的情况是一致的，但在部分情况中（例如在涉及到 instrumentalness 的情况下）却相当分散，这表明 Pop/Rock 在这些情况中并没有表现出相类似的特征，即，流派内部的音乐在这些方面差异较大。这是否表明其实流派并不能绝对地划分音乐之间的区别？为此，我们采用 k-means 聚类分析。值得一提的是，k-means 是无监督学习算法，这保证了流派将不作为先验知识干扰分类。

K-Means 聚类分析

在聚类算法中较为经典的算法有 K-Means, Special clustering 等。其中 K-Means 由于其高效、易于实施的特点，到目前为止仍被广泛地应用于各个领域。我们选用 K-Means 进行聚类分析，相似度近的作品代表的节点会形成一个簇，以此衡量作品的相似情况。我们在不把作品流派作为先验知识的情况下，对作品进行分类以比较各作品的相似度。输入没有标签的新数据后，将新数据的每个特征与样本集中数据对应的特征进行比较，计算相似度，然后提取样本集中特征向量最相似数据的分类标签。对于不同样本特征向量，相关系数的取值范围为 $[-1, 1]$ ，相似度在 0-1 之间变化。

具体操作步骤如下：

1、给定一组观测值

$$(x_1, x_2, x_3, \dots, x_n)$$

其中每个观测值都是 d 维实向量，k 均值聚类旨在将 n 个观测值划分为 k 个集合

$$S = \{s_1, s_2, s_3, \dots, s_k\}$$

2、分别计算数据集中每个元素与所选聚类中心的欧式距离，根据最近邻原则，将

~~元素划分到相应的类别中。~~

~~欧式距离是在欧式空间中两个样本之间的直线距离。 X_i 与 X_j 在 m 维空间中的欧式距离为~~

$$d(X_i, X_j) = \sqrt{\sum_{k=1}^m (X_{ik} - X_{jk})^2}$$

~~3、计算每个类别中元素的平均值，将其作为新的聚类中心~~

~~4、重复步骤 2 至 4，直至新的聚类中心不再变化。以便最小化群集中每个点到 K 中心的距离函数的群集内和。~~

~~换句话说，其目标是找到： $\min \sum_{i=1}^k \sum_{x \in S_i} (x - \mu_i)^2$~~

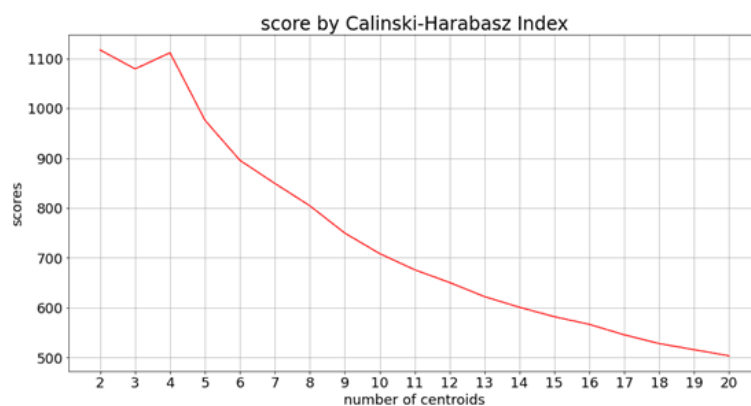
~~其中 μ_i 是 S_i 中点的平均值。~~

k-means 算法对 k 值敏感，事实上，在本题背景中， k 代表音乐按照特征信息应该划分成几类，这对我们的分析至关重要。在 PCA 降维图中，可以看出点大致内聚成两簇。为了更精确地探究这一问题，我们采用交叉验证法，用 Calinski-Harabasz 分数值 **公式(3)** 对 K 进行评估来衡量内聚问题，得到的 Calinski-Harabasz 分数值 s 越大则聚类效果越好。

$$s(k) = \frac{\text{tr}(B_k)}{\text{tr}(W_k)} \frac{m - k}{k - 1}$$

其中 m 为训练集样本数， k 为类别数。 B_k 为类别之间的协方差矩阵， W_k 为类别内部数据的协方差矩阵。 tr 为矩阵的迹。

曲线如图所示，可以看出，将音乐分为两类是最合理的做法，这也和我们在 PCA 图中观察到的情况相互印证。



~~Given a set of observations~~

$$(x_1, x_2, x_3, \dots, x_n)$$

~~where each observation is a d -dimensional real vector, k -means clustering aims to partition the n observations into k sets~~

$$S = \{S_1, S_2, S_3, \dots, S_k\}$$

~~so as to minimize the within-cluster sum of distance functions of each point in the cluster to the K center. In other words, its objective is to find:~~

$$\min \sum_{i=1}^k \sum_{x \in S_i} (x - \mu_i)^2$$

where μ_t is the mean of points in S_t .

In the K-means algorithm, K is given in advance, but selecting the K value is very difficult. According to the theory of variance, we often determine the K value based on the R^2 statistic.

$$W_t = \sum_{i=1}^{n_t} (X_{(i)}^{(t)} - \bar{X}^{(t)})'(X_{(i)}^{(t)} - \bar{X}^{(t)})$$

$$T = \sum_{t=1}^k \sum_{i=1}^{n_t} (X_{(i)}^{(t)} - \bar{X})(X_{(i)}^{(t)} - \bar{X})'$$

$$P_k = \sum_{t=1}^k W_t$$

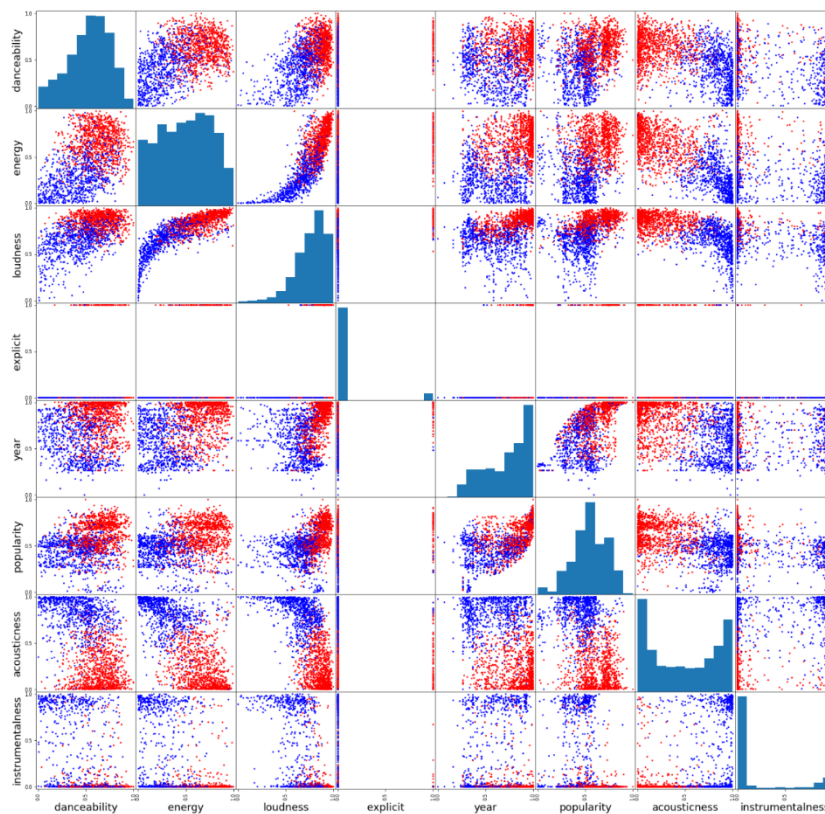
$$B_k = \sum_{t=1}^k n_t (\bar{X}^{(t)} - \bar{X})(\bar{X}^{(t)} - \bar{X})'$$

$$R^2 = \frac{B_k}{T} = 1 - \frac{P_k}{T}$$

The greater the proportion of sum of square between class is and the smaller the proportion of sum of square inner class is, the better K value is. So we choose the K value that maximize R^2 value.

接下来，我们将原先的图按照 k-means 指出的分类上色，如图 X 所示。

music data divided into 2 centroids



可以看出，现在所有点的内聚情况都相当好，这表明这些点代表的音乐作品有更为相似

的特征。让我们以其中 energy 一行为例，来解释从图中我们有什么发现。在 energy 相关的组合中，红色的点所代表的一类音乐比另一类音乐更有能量，而另一个特征信息则很自然地解释了原因：红色在 danceability、loudness、year、popularity 中数值更高，而 acousticness、instrumentalness 中数值较低。这表明此类音乐律动性强、响度更大，这无疑是红色更有能量的重要原因。从 year 和 popularity 中我们可以得知，这类音乐的年代与我们更加靠近，流行度也高，这从音乐发展历史角度告诉我们近现代的人们开始尝试创作出新的音乐，而他们的努力得到了回报，有更高的流行度，更重要的是，我们的模型表明，他们可以说与过去的音乐已经分道扬镳，形成了一种新的音乐！真是振奋人心。从 acousticness 和 instrumentalness 的角度看，这似乎表明了新音乐变化的技术原因：对人声的开发和对音轨处理的运用与新音乐的产生有相当的正相关。总的来说，红色音乐更加热情奔放，而蓝色音乐更加古典内敛，而我们的模型成功地将音乐划分为比流派之分更加本质的两类。值得一提的是，深绿色代表的电子音乐几乎在所有情况下都是坚定不移的红色音乐，这和分析也相一致，而其他的流派则两种音乐兼而有之，并没有稳定地属于其中一种，这说明大多数流派的音乐风格相当多样。因此，流派之分其实并没有那么大，不同流派的音乐常常表现出相似的特征。

5.3 问题三——综合评价模型建立与求解

模型思路

问题三需要对流派进行分析，比较流派之间和流派内部的相似之处和影响。根据 full-music-data 数据集音乐特征，我们进一步对第二问的结果进行数据挖掘，首先对流派内和流派之间作品的各项音乐特征分别画相关系数热力图，可以看出某些特征的关联度很高，这揭示了流派内部和流派之间对某些特征运用的偏好。然后在聚类图的基础上，我们将流派作为标签，单独分析各个流派的数据，通过流派之间聚点距离的比较，寻找其中相似度最高和最低的流派，定量比较流派之间的异同；做出流派相互影响矩阵，探究流派之间和内部的影响程度。我们提取出 influence-data.csv 文件中的影响者和受影响者派别，使用 Floyd 算法，画出流派两两之间演化时间和演化路径，直观地看出流派随着时间的推移的变化。

模型建立与求解

流派之间特征的关联度

根据 full-music-data 数据集音乐特征，我们进一步采用 Pearson Correlation Coefficient 定量分析了流派内作品的各项音乐特征，并画出相关系数热力图。相同流派的作品更可能对某些特征的运用有一致的偏好，这可以从皮尔逊相关系数中看出。为此我们以流派来划分音乐作品，对特征信息作热力图。首先，我们发现对所有音乐求特征相关系数的情况下，响度和能量有 0.79 的关联度，这是相当合理的，而其他特征则无明显关联。为揭示流派内的联系和他们之间的区别，我们对其中有代表性的两幅图作出解释。

图 X 为 Avant-Garde 在 danceability, energy, valence 等 13 个维度上音乐作品的相似度。对 Comedy/Spoken 流派的音乐而言，speechiness 和 liveness 的相关系数相当高，其中原因可以解释为口语单词的运用为这类音乐带来了更高的人气，这和现实情况

是一致的；而其他特征则无明显关联。这表明，Stage&Screen 音乐最明显的特征是运用口语单词与人气之间的独特关系。

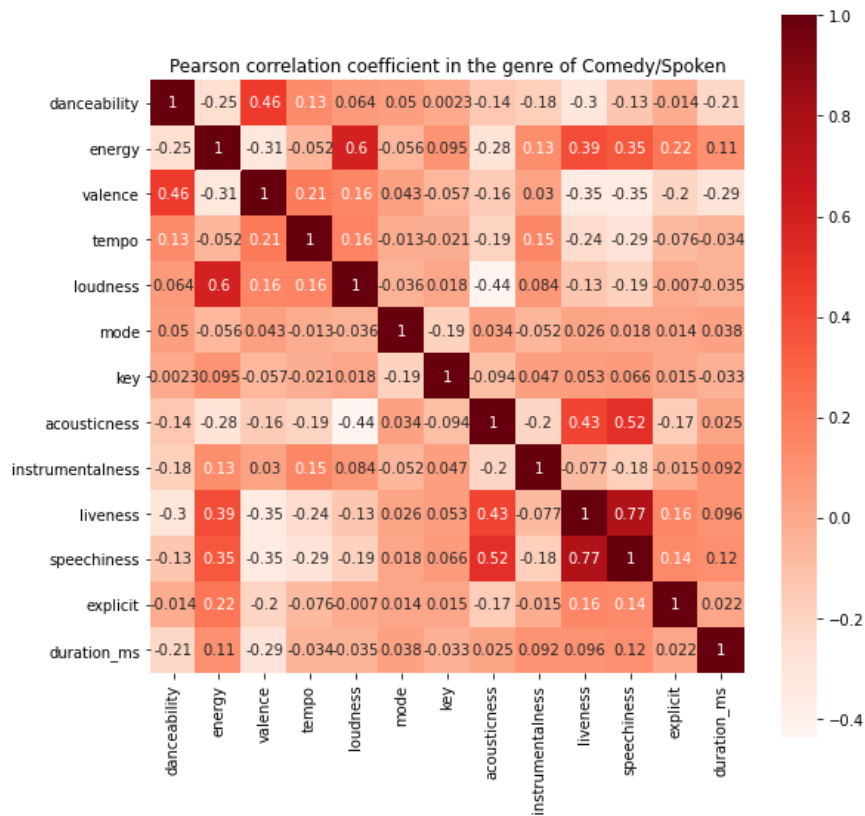
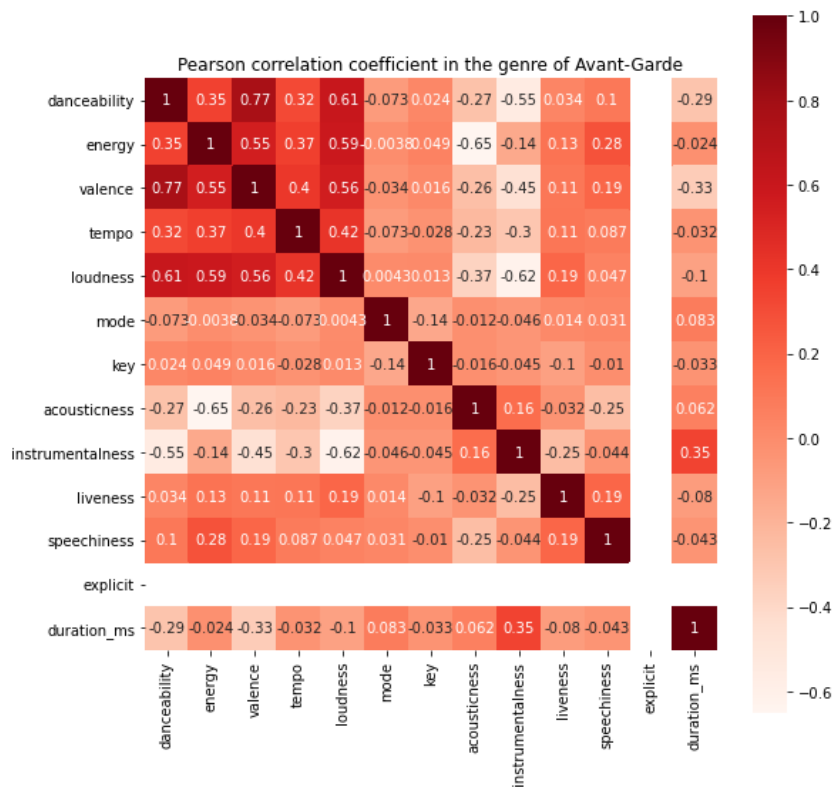


图 X 为 Avant-Garde 在 13 个维度上音乐作品的相似度。对先锋派音乐（Avant-garde）而言，有许多音乐特征的关联度并不低，这表明先锋派音乐普遍勇于尝试各种手法，音乐是他们的实验室，这也是先锋派音乐的热力图区别于其他流派最明显的地方。



以上两例表明，高的相关系数能揭示流派内的相似之处，通常这对应着流派内对某些技巧运用的共识，而流派之间热力图的差别则表明了流派之间不同的偏重点。

随着时间的推移的变化

我们提取出 influence_data.csv 文件中的影响者和受影响者派别，去重后得到二十个派别，如图所示，将二十个派别按照首字母排序后分别打上 0~19 的标签。

Music genre	Serial number	Music genre	Serial number
Avant-Garde	0	Jazz	10
Blues	1	Latin	11
Children of Bodom	2	New Age	12
Classical	3	Pop/Rock	13
Comedy/Spoken	4	R&B	14
Country	5	Reggae	15
Easy Listening	6	Religious	16
Electronic	7	Stage & Screen	17
Folk	8	Unknown	18
International	9	Vocal	19

二十个类别对应二十个节点，某一类别影响某一类别则用一条边连接两节点，于是音乐流派随着时间的推移的演变问题，转化为流派两两之间带权图从源点到终点的路径问题。

Floyd 算法 (Floyd-Warshall algorithm)

计算最短路径的算法较多，主要包括 Dijkstra 算法，Bellman-Ford 算法，Floyd 算法和 SPFA 算法等，考虑到 Floyd 算法采用三重循环紧凑结构，效率高，适用于稠密图，即复杂的多节点、多层流派演变，我们用其探究流派两两之间演化时间和演化路径。

任意流派 i 到 j 的演变有两种可能：

1. 直接从 i 到 j ;
2. 从 i 经过若干个流派演变到 j 。

~~把带权路径长度最短的那条路径称为最短路径，其路径长度（权值之和）称为最短路径长度或者最短距离。计算最短路径具体步骤如下：~~

~~设 $A = (a_{ij})_{n \times n}$ 为赋权图的权矩阵， d_{ij} 表示从 v_i 到 v_j 点的距离， r_{ij} 表示从 v_i 到 v_j 点的最短路中一个点的编号。~~

~~I. 赋初值。对所有 i, j , $d_{ij} = a_{ij}$, $r_{ij} = j$. $k = 1$. 转向 II~~

~~II. 更新 d_{ij} , r_{ij} . 对所有 i, j , 若 $d_{ik} + d_{kj} < d_{ij}$, 则令 $d_{ij} = d_{ik} + d_{kj}$, $r_{ij} = k$, 转向 III~~

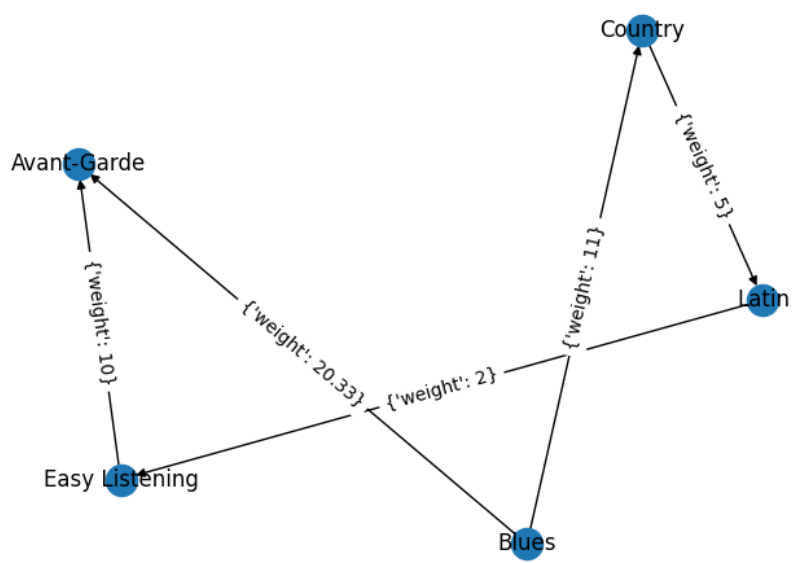
~~III. 终止判断。若 $k = n$ 终止；否则令 $k = k + 1$, 转向 II~~

~~IV. 最短路线可由 r_{ij} 得到。~~

从任意一条单边路径开始。所有两点之间的距离是边的权，如果两点之间没有边相连，

则权为无穷大。由于两节点间可以有多个边，将其权重累加，一条路径上所经边的权值之和定义为该路径的路径长度或称带权路径长度。这是由于一类影响另一类的时间用平均值较为妥当，我们可以近似认为每一个影响者影响另外一个人是在同样的条件下，其误差服从正态分布，取平均值更接近真实值，考若小于 0，实际是被影响派别在影响派别出现，从事实角度考虑是不可能发生的，就认为是同一时间点发生，传播时间为 0。算法的缺点在于：虽然在稠密图变现比 Dijkstra 好，代码也较简单，数据量较大时，由于时间复杂度时 $O(n^3)$ ， n 为数据规模，耗时较长。且如果图中有负环，则会无法计算最小值。而此题所涉及节点两两之间最短路径并无负，符合使用条件。

由此，我们得到两个流派之间演化的路径。为直观体现，我们选取 Avant-Garde、Easy Listening、Blues、Country、Latin 五个流派的演化过程，得出简化的路径图，如图 X 所示



(改)

最短演化时间见表 X，表示两节点之间最短路径，即两个流派间演化的最短时间。

	Avant-Garde	Blues	Children's	Classical	Comedy/Spoken	Country	Easy Listening	Electronic	Folk	International
Avant-Garde	0.000	25.167	30.000	6.000	26.810	28.500	23.810	19.412	18.500	28.945
Blues	20.333	0.000	38.333	20.254	27.754	3.333	10.333	36.193	20.957	16.626
Children's	46.364	53.191	0.000	52.364	53.000	56.513	50.000	54.147	51.867	55.135
Classical	14.545	19.167	24.000	0.000	30.000	22.500	19.000	28.500	12.500	27.500
Comedy/Spoken	22.000	16.000	20.000	17.500	0.000	5.000	12.000	34.318	22.000	18.293
Country	17.000	11.000	35.000	23.000	26.892	0.000	7.000	36.412	17.624	13.293
Easy Listening	10.000	29.924	30.000	16.000	27.500	32.500	0.000	29.412	27.000	18.293
Electronic	15.118	20.033	26.754	15.833	21.754	23.367	18.754	0.000	20.622	23.889
Folk	22.000	6.667	33.333	10.000	20.000	10.000	12.000	32.881	0.000	18.293
International	5.000	17.576	25.000	11.000	27.500	20.909	10.000	24.412	10.909	0.000
Jazz	10.000	12.424	20.000	14.000	10.000	15.000	8.333	29.412	17.273	15.909
Latin	12.000	24.316	30.000	18.000	21.892	26.892	2.000	31.412	19.202	8.293
New Age	19.697	11.667	31.333	15.000	25.000	15.000	17.000	21.000	5.000	20.000
Pop/Rock	6.364	13.191	18.000	12.364	13.000	16.513	10.000	14.147	11.867	15.135
R&B	15.000	9.200	29.000	5.000	12.500	12.533	16.516	21.818	17.500	20.909
Reggae	18.894	19.200	30.530	15.000	22.500	22.533	12.000	24.412	24.398	15.000
Religious	23.391	20.167	27.500	18.250	25.750	23.500	27.027	31.174	23.333	30.227
Stage & Screen	15.000	24.091	10.000	21.000	21.667	20.000	5.000	33.333	20.000	23.293
Unknown	26.364	33.191	38.000	32.364	33.000	36.513	30.000	34.147	31.867	35.135
Vocal	27.727	12.667	20.000	26.875	27.500	16.000	20.000	35.000	16.875	22.727

	Jazz	Latin	New Age	Pop/Rock	R&B	Reggae	Religious	Stage & Screen	Unknown	Vocal
Avant-Garde	26.000	24.340	13.500	13.810	25.324	30.476	27.785	20.000	31.310	24.000
Blues	20.000	8.333	27.754	22.046	15.254	23.333	18.333	28.333	26.626	21.626
Children's	53.053	50.530	50.811	40.000	51.514	56.667	53.976	48.000	57.500	56.267
Classical	20.000	20.000	7.500	20.833	23.750	34.904	30.000	14.000	37.500	18.000
Comedy/Spoken	21.892	10.000	25.000	20.909	12.500	23.654	20.000	28.909	28.293	23.293
Country	16.892	5.000	29.000	22.496	20.000	30.000	15.000	25.000	23.293	18.293
Easy Listening	17.500	10.000	22.000	23.810	31.250	39.500	20.000	20.000	28.293	17.500
Electronic	21.807	19.285	15.769	8.754	10.833	21.987	22.730	16.754	26.254	23.202
Folk	21.892	10.000	17.500	18.734	11.250	22.404	20.000	24.000	28.293	13.333
International	17.500	20.000	18.500	18.810	22.159	21.818	10.000	20.000	10.000	5.000
Jazz	0.000	12.439	21.500	23.423	21.761	25.455	15.000	18.571	25.909	16.476
Latin	11.892	0.000	24.000	24.369	21.250	30.111	10.000	20.000	18.293	13.293
New Age	16.000	15.000	0.000	13.333	16.250	27.404	25.000	21.333	30.000	18.333
Pop/Rock	13.053	10.530	10.811	0.000	11.514	16.667	13.976	8.000	17.500	16.267
R&B	16.901	14.516	12.500	16.092	0.000	11.154	24.516	19.000	30.909	12.368
Reggae	21.892	10.000	22.500	12.530	10.000	0.000	20.000	20.530	25.000	20.000
Religious	30.080	27.557	25.750	17.027	13.250	24.404	0.000	22.500	34.527	7.500
Stage & Screen	11.667	15.000	27.000	26.078	31.250	37.121	25.000	0.000	33.293	21.304
Unknown	33.053	30.530	30.811	20.000	31.514	36.667	33.976	28.000	0.000	36.267
Vocal	26.667	21.000	20.000	29.449	22.511	22.000	31.000	15.000	32.727	0.000

shortest length between genres

这里，我们列举两个派别演化。

Blues→Country→Latin→Easy Listening→Avant-Garde

Blues 到 Avant-Garde 最短路径需经过多个流派，weight 为 20.33，因而这样的演变是比较困难的，我们认为是几乎不可能的。

Electronic 到 Pop/Rock，weight 相对较低，为 14.14，这说明 Electronic 是较容易在一定的演变后变成 Pop/Rock，这与 influence_data 中 Tangerine Dream(Electronic) 受到 Pink Floyd、Jimi Hendrix (Pop/Rock) 相吻合，在 Electronic 形成的过程中，或多或少的受到 Pop/Rock 的影响，因而之后再演变为 Pop/Rock 的可能性更大。

两派别之间的关系，若是演化时间越短，则两者关联越强，繁殖越弱。经过了更多的中间流派，则两个流派关联也越弱。

5.4 问题四模型建立与求解

5.4.1 建模思路

首先我们对子网络进行分析，对于 *data_influence* data 数据集中报告的相似性数据能否表明识别出的影响者实际上影响了各自的艺术家的的问题，我们计划通过标称属性相似度量法对该问题进行分析；在 model 1 中我们提出的多个影响者对一名追随者的子网络关系中，该名追随者对所有的影响者都具有乐曲上的相似程度，在 model 2 中我们对这种相似性做出了衡量模型，从而影响者对追随者音乐是否产生影响可以转化为通过计算这些音乐的相似程度，与影响者的影响力是否有关联来表征，故我们采用相关因子-协方差分析法对影响力-相似性之间的关联度进行分析；对 *full-music-data* 和其他两个数据集分析可以得出其中某追随者音乐特征与其影响者音乐特征相似性高低，当其中某一音乐特征指标相似性远高于其他特征时，可以认为其中某些音乐特征比其他特征更具有感染力，当各音乐特征相似度差别关联不明显时，说明各音乐特征被追随者“照单全收”扮演相似角色，此时我们考虑通过余弦相似性度量方法对各音乐特征进行横向相似性的衡量，通过灰色关联分析法对关联程度进行表征，从而解决该问题。

为增加该模型的可靠性和全局性，单单对某一特定子网络分析是不够的，为此我们选取 model 1 中筛选出的前 100 名极具影响力的艺术家大师，选取 10 个样本作为影响者，各自对其相应追随者计算各音乐特征指标的特征均值改变量，为比较特征之间的区别，我们作双层旭日图对该问题从全局角度进行可视化分析。

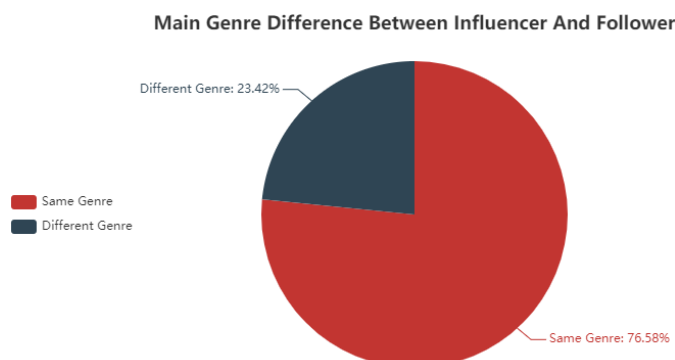
5.4.2 关于子网络的关联度模型分析

5.4.2.1 标称属性相似度量

令跟随者的主要流派为 O_1 ，影响者的主要流派为 O_2 ，数据集中共有 n 条跟随者影响者关系数据，两个对象 O_1 ， O_2 匹配的数据条数为 m ，那么相似度为匹配数占总属性数的总

数

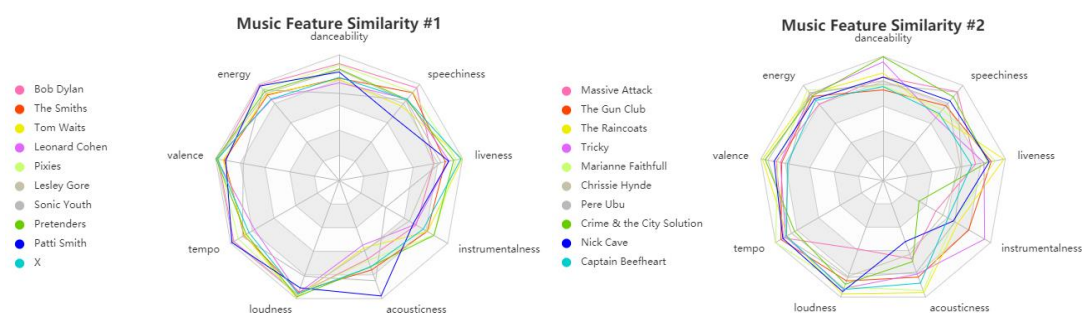
$$\text{sim}(O_1, O_2) = \frac{m}{n}$$



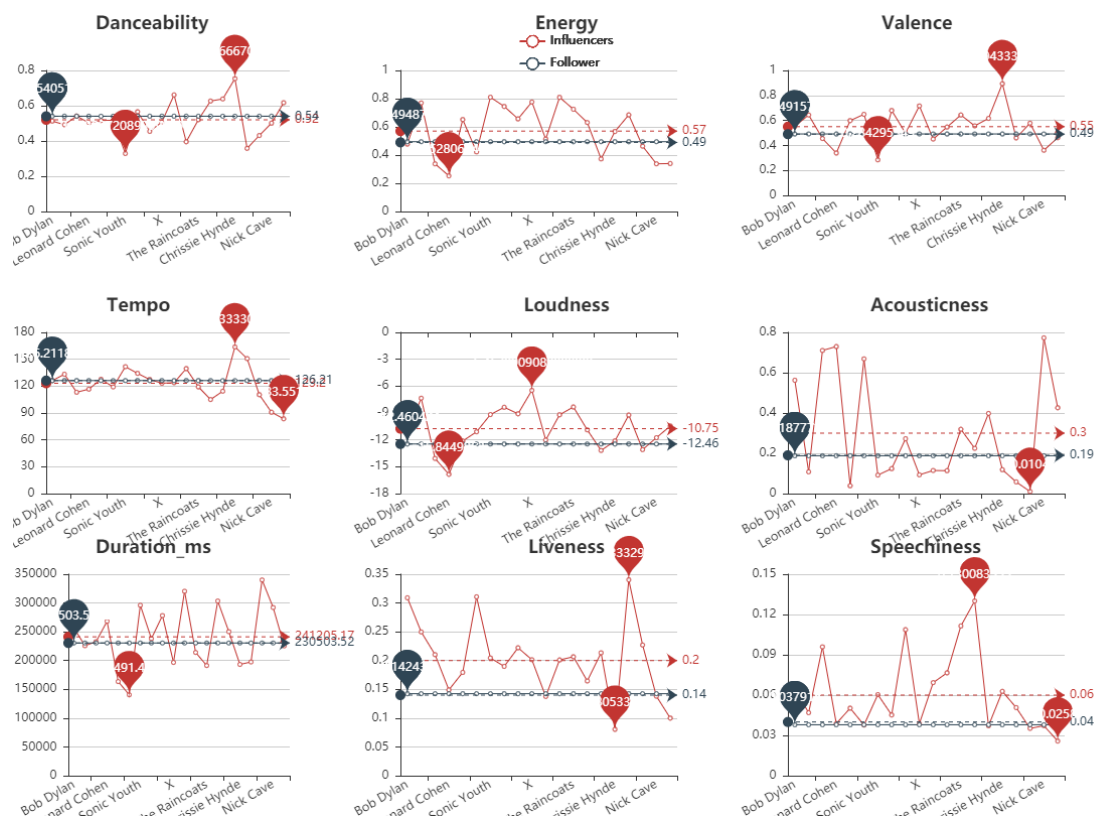
我们得到得相似度为 76.58%，可以认为数据集中报告的相似性数据可以表明识别出的影响者实际上影响了各自的艺术家。

5.4.2.2 余弦相似度度量

首先选取Danceability、Energy、Valence、Tempo、Loudness、Acousticness、Duration_ms、Liveness、Speechiness九个音乐特征指标，绘制子网络中追随者PJ-Harvey与其相应的各影响者指标关系图，由于在后续5.4.3中详细说明了key和mode对追随者的影响力小，故在此我们筛除这两个指标进行分析。

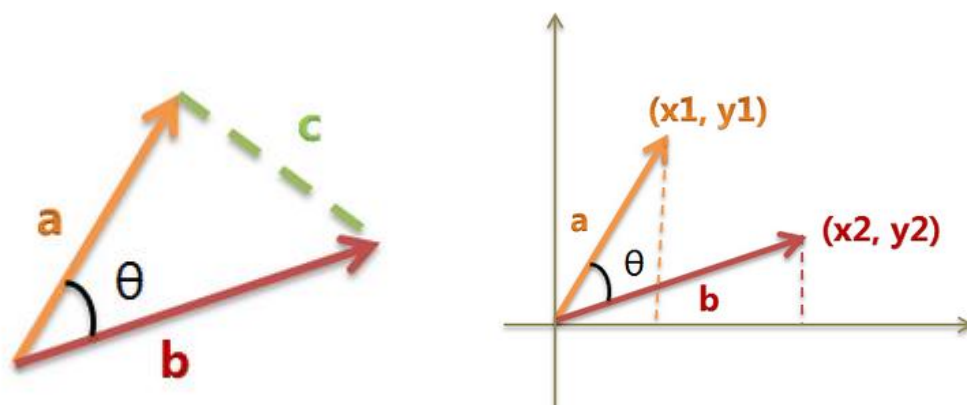


下面我们将利用余弦相似度度量方法对追随者PJ-Harvey与其影响者音乐各特征指标相似度进行衡量。



下面我们将利用余弦相似度度量方法对追随者PJ-Harvey与其影响者音乐各特征指标相似度进行衡量。

计算两组音乐特征向量的相似程度问题，可以把这它们想象成空间中的两条线段，都是从原点 $[0, 0, \dots]$ 出发，指向不同的方向。两条线段之间形成一个夹角，如果夹角为0度，意味着方向相同、线段重合；如果夹角为90度，意味着形成直角，方向完全不相似；如果夹角为180度，意味着方向正好相反。因此，我们可以通过夹角的大小，来判断音乐特征的相似程度。夹角越小，就代表越相似。



(图向量余弦相似性计算)

对于n维向量，假定A和B是两个n维向量，A是 $[A_1, A_2, \dots, A_n]$ ，B是 $[B_1, B_2, \dots, B_n]$ ，则A与B的夹角 θ 的余弦等于：

$$\cos\theta = \frac{\sum_{i=1}^n (A_i \times B_i)}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

$$= \frac{A \cdot B}{|A| \times |B|}$$

余弦值越接近1，就表明夹角越接近0度，也就是两个音乐特征向量越相似，这表示两个音乐特征向量具有"余弦相似性"。

所得子网络各音乐指标余弦相似度附录列出。

5.4.2.3 相关系数与灰色关联度分析（GRA）

- 影响者影响力-双方音乐相似性的相关系数分析表征其关联程度

我们利用model 1得到的各影响者影响力与model 2得到的各影响者-追随者音乐相似度，对子网络两指标进行相关系数-协方差分析：

首先确定数据列为 $x = \{\text{子网络中各影响者影响力}\}$ ，比较序列为 $y = \{\text{各影响者音乐与追随者音乐相似度}\}$ ，两指标数据列如下表所示。

根据下式对上述两个数据列指标进行相关系数-协方差分析：

$$r = \frac{\sum_{i=1}^{10} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{10} (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} = 0.3966$$

由于 $r=0.3966$ ，由于关联度在0.4左右，说明影响力与音乐相似度两指标之间存在关联，但关联性较弱，在该子网络中可以认为影响者影响力和双方音乐相似性之间存在关联，可以认为“有影响力的人”会影响追随者创作的音乐，并且追随者在一定程度上保持了自身音乐的创造性和独立性。

根据下式对上述两个数据列指标进行协方差分析：

$$r = \frac{\sum_{i=1}^{10} (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{10} (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}} = ? ? ?$$

由于 $r=? ? ?$ ，可以认为影响者影响力和双方音乐相似性之间存在关联，可以认为“有影响力的人”会影响追随者创作的音乐。

- 影响者&追随者各音乐特征指标相似性的灰色关联度分析

在5.4.3中我们通过余弦相似度度量方法得到了子网络中该名追随者各音乐特征和其相应的多名影响者各音乐特征指标的相似度如附录所示，针对各相似度指标进行灰色

关联度分析:

选取最优参考数据列为y，如下表所示，y为各音乐特征指标数据中的最优值。

数据列表									
k	1	2	3	4	5	6	7	8	9
Name	Danceability	Energy	Valence	Tempo	Loudness	Acousticness	Duration_ms	Liveness	Speechiness
X1									
X2									
X3									
X4									
X5									
X6									
X7									
X8									
X9									
X10									
y									

由于系统中各因素列中的数据可能因量纲不同，不便于比较或在比较时难以得到正确的结论，故要对上述各目标变量的无量纲化处理，对数据列进行均值化处理

$$x_i(k) = \frac{x_i(k)}{x_i}$$
，其中i，k含义如上表所示，无量纲化处理后按下式计算关联系数。

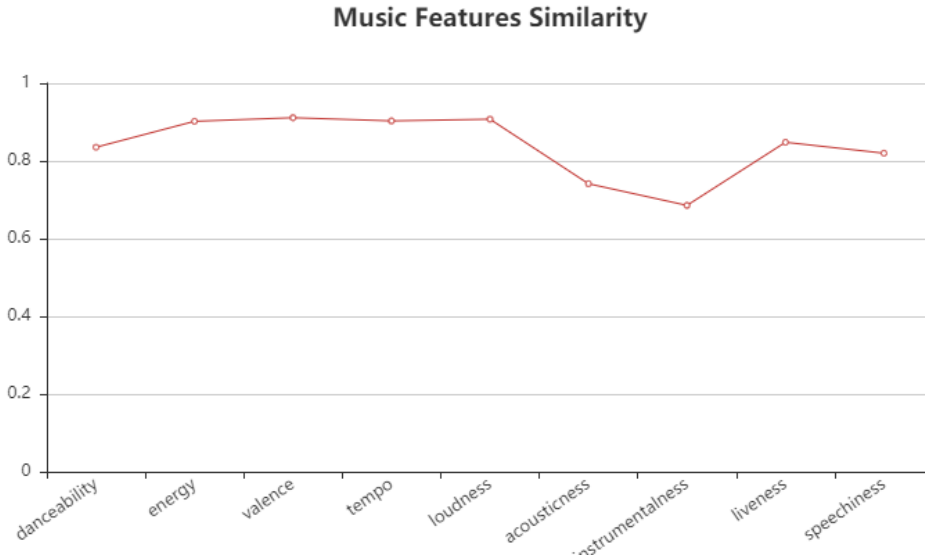
$$\zeta_i(k) = \frac{\min_i \min_k |y(k) - x_i(k)| + \rho \max_i \max_k |y(k) - x_i(k)|}{|y(k) - x_i(k)| + \rho \max_i \max_k |y(k) - x_i(k)|}$$

接下来，计算关联度 $r_k = \frac{\sum_{i=1}^{20} \zeta_i(k)}{20}$ ，i=1, 2, ..., 20, k=1, 2, ..., 9

	danceability	energy	valence	tempo	loudness	acousticness	instrumentalness	liveness	speechiness
artist_name									
Bob Dylan	0.929809	0.989175	0.978944	0.970395	0.963929	0.665416	0.709545	0.767035	0.962658
The Smiths	0.814204	0.888569	0.929871	0.875335	0.944179	0.757244	0.811725	0.860545	0.912485
Tom Waits	0.804128	0.897770	0.940211	0.869816	0.993160	0.571916	0.814740	0.998908	0.773014
Leonard Cohen	0.777680	0.840005	0.992770	0.780008	0.954374	0.547787	0.701221	0.890730	0.846677
Pixies	0.919479	0.948344	0.975647	0.929637	0.967293	0.677905	0.872867	0.949560	0.916947
Lesley Gore	0.881011	0.905848	0.988277	0.979591	0.968912	0.589701	0.647318	0.767286	0.828814
Sonic Youth	0.693402	0.943047	0.776727	0.890914	0.810887	0.849811	0.427176	0.825406	0.835686
Pretenders	0.888370	0.922956	0.998592	0.856039	0.984903	0.730113	0.873355	0.928185	0.840644
Patti Smith	0.863749	0.982460	0.918342	0.988088	0.909676	0.977968	0.684335	0.887885	0.665575
X	0.820543	0.845508	0.986224	0.830566	0.958775	0.732825	0.774542	0.987237	0.842984
Massive Attack	0.834075	0.802604	0.874713	0.925344	0.587411	0.671990	0.491168	0.755837	0.932700
The Gun Club	0.729691	0.884466	0.828712	0.950242	0.860511	0.828407	0.795864	0.884144	0.785653
The Raincoats	0.866453	0.913230	0.985302	0.891237	0.972068	0.961954	0.693261	0.987528	0.686486
Tricky	0.955032	0.907314	0.839361	0.906945	0.925810	0.802413	0.944059	0.824869	0.660752
Marianne Faithfull	0.761794	0.948942	0.860715	1.000000	0.915277	0.942203	0.656150	0.917845	0.746120
Chrissie Hynde	0.803422	0.908318	0.942162	0.851960	0.822738	0.633073	0.556037	0.633312	0.812899
Pere Ubu	0.776208	0.923754	0.770744	0.914052	0.831846	0.786634	0.668071	0.643130	0.928177
Crime & the City Solution	0.996820	0.889449	0.963852	0.822299	0.889392	0.695502	0.333999	0.858633	0.879439
Nick Cave	0.833084	0.857537	0.891540	0.932404	0.952582	0.524025	0.652816	0.868552	0.837884
Captain Beefheart	0.756432	0.842166	0.782382	0.897319	0.934723	0.879937	0.604075	0.724790	0.703846

所得关联度如图所示，根据关联度排序如图所示，由于Valence、Loudness、Tempo、Energy四个指标关联度均在0.9以上，高于其他音乐特征指标关联度，故我们可以认为Valence、Loudness、Tempo、Energy四个音乐特征比其它特征更具"感染力"。

数据列表									
排序	6	4	1	3	2	8	9	5	7
指标	Danceability	Energy	Valence	Tempo	Loudness	Acousticness	instrumentalness	Liveness	Speechiness
关联度	0.835269	0.902073	0.911254	0.903110	0.907422	0.741341	0.685616	0.848071	0.819972



5.4.2 关于全局网络的特征差量模型分析

艺术家的特征由artist_data数据集给出，而当务之急则是考虑如何衡量艺术家对追随者的影响。受神经网络中的SGD学习规则启发，影响者是其追随者的学习对象，我们仿照SGD学习规则，建立衡量影响者在各项特征上对其追随者影响力的指标。

在BP神经网络的训练学习中，网络的输出与目标输出不一致时，需要将这一差值反馈给网络以修正网络权重，而最常见的反馈方式定义为：

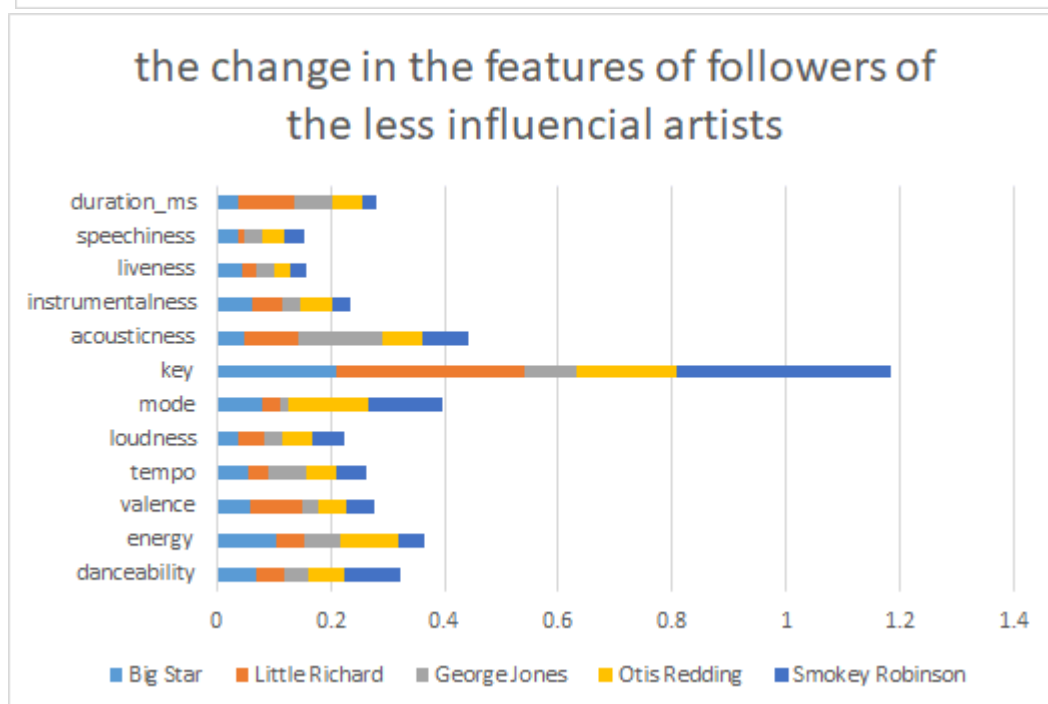
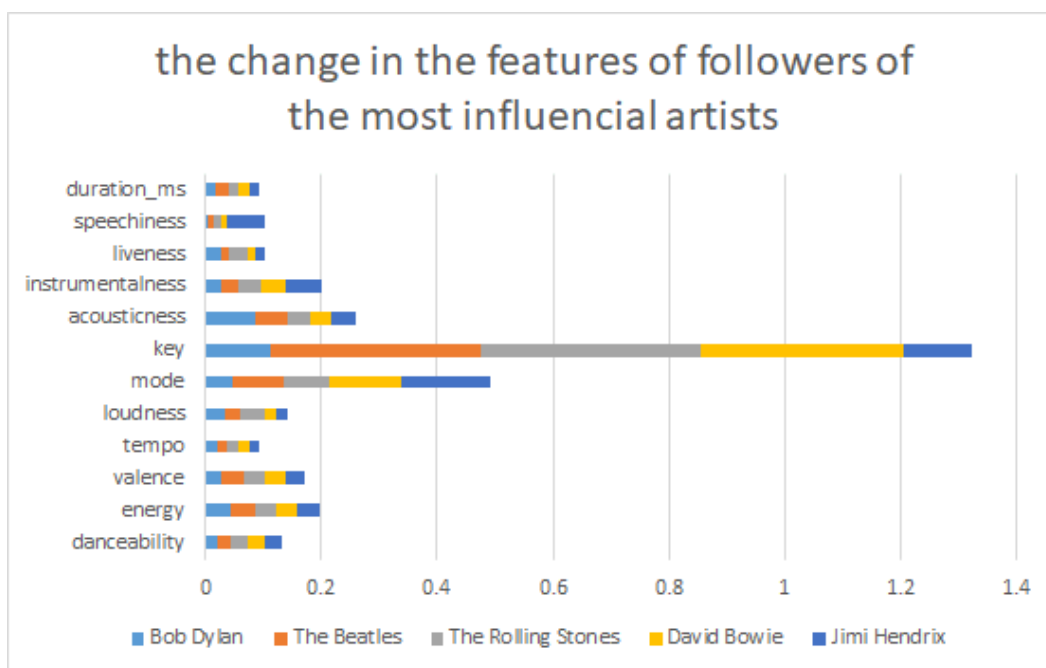
$$Loss = \frac{1}{2}(output - target)^2$$

将追随者的特征向量视作对影响者的学习结果，每个追随者在学习影响者后与影响者特征向量的不同以上述方式量化。由artist_data给出的数据中每位艺术家有12条特征数据，将影响者与追随者的特征数据一起进行归一化处理，就得到了每位艺术家的特征向量。设影响者的特征向量为 $(x_1, x_2, \dots, x_{12})$ ，某个追随者的特征向量为 $(y_1, y_2, \dots, y_{12})$ ，则第*i*个特征的改变量定义为：

$$Change_i = \frac{1}{2}(x_i - y_i)^2$$

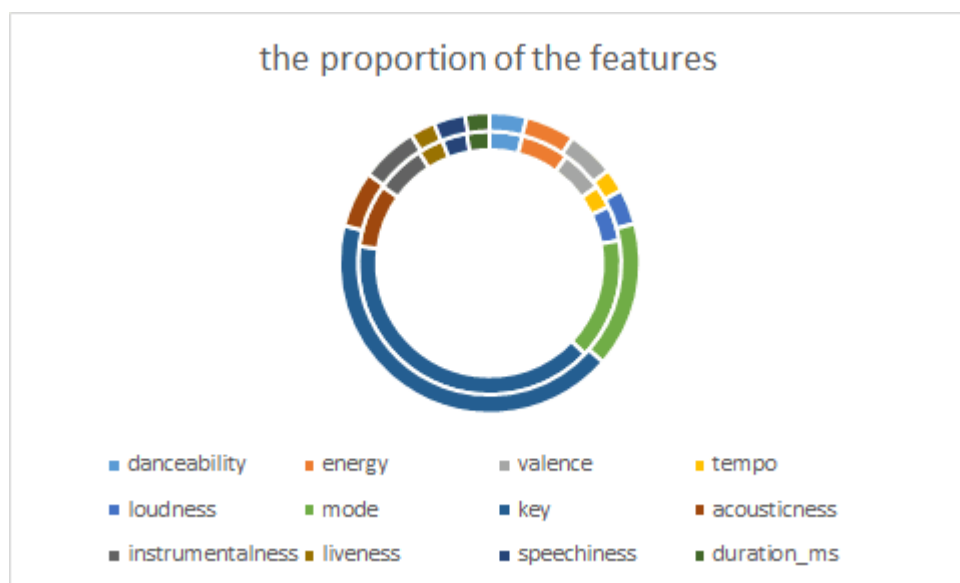
以上得到了影响者对追随者的改变量向量模型。对所有改变量向量求均值，均值改变量向量反映了影响者对追随者的影响能力，同时由于结合了每项特征的数值信息，也反映了不同特征给追随者的影响。

为了使我们的分析更典型，我们选择有一定数量的追随者的影响者来代入计算。在第一问中我们曾给出了前一百名有影响力的艺术家词云，我们选择前五名作为the most influential artists，最后五名作为the less influential artists，分别观察求值结果。



以上图中可以看出，更具影响力的艺术家，例如脍炙人口的Bob Dylan等，其追随者的改变量向量总的来说更小，可以得出结论，更具影响力的艺术家会更强烈地影响其追随者的创作，表现为追随者区别于影响者的差别更小。

从上图还可以发现，并非所有特征都会被追随者模仿，为了比较特征之间的区别，我们作双层旭日图：



上图中，不同颜色代表不同特征，比例体现特征改变量之间的大小相对关系，外环内环分别为the most influential artists和the less influential artists。图中表明，keys和mode对追随者的影响力较弱，这也符合音乐知识：多数人并不介意一首曲子是大调小调，主音是C还是E flat，调性是绝对的但受人喜爱的旋律只是相对的。或许西贝柳斯除外。

其他特征影响力的相对强弱关系也在图中被给出。

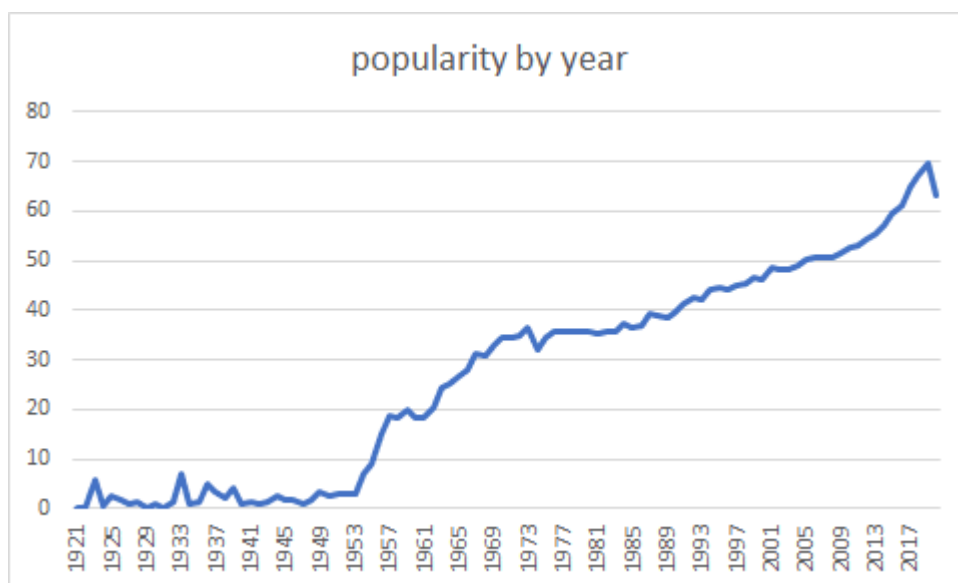
5.5 问题五模型建立与求解

5.5.1 建模思路

对于寻找音乐演进过程中可能存在的重大飞跃特征的问题，首先确定出变革时间，根据变革时间确定出该时段变化比较大的音乐特征，我们计划根据数据集绘制关于影响力-年份-派别的三维坐标图，根据其在哪些时间点出现了某种流派的突然增加或者突然减少，可以认为发生这些突发事件的时刻是发生重大历史事件的时刻，根据data-by-year 绘制流派流行程度-时间折线图，以呈现各种流派随着时间变化的程度确定出在此次音乐变革中流派的变革程度，在发生大变革的流派内部分析流派内音乐特征变化关联度最大的艺术家确定其为所要寻找的代表变革者。

5.5.2 重大音乐变革的时间和派别

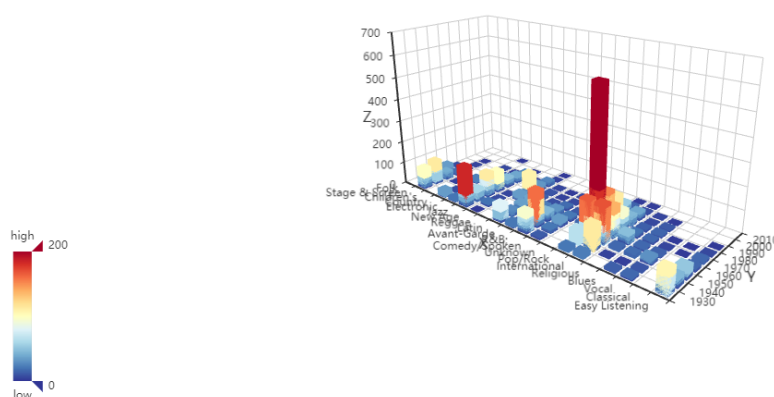
我们根据 data-by-year 绘制流派流行程度-时间折线图以反映重大历史变革时间，可视化结果如图所示。



根据图示，我们可以初步确定 1957 年为音乐演变中极具变革型的年份。

由于在音乐演变过程中，音乐发展受科技、经济、文化、环境等多方面影响，我们排除其他外部因素，只对音乐领域做研究，建立影响程度-年份-派别的三维坐标图，如图所示。

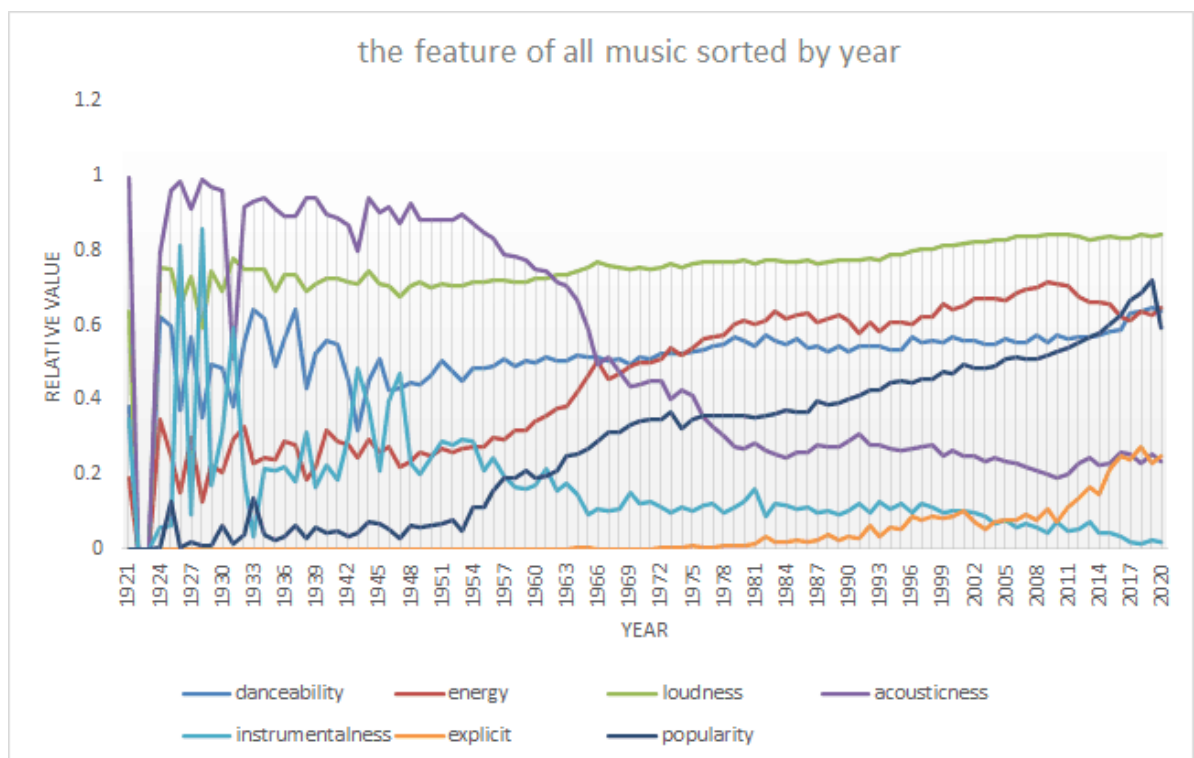
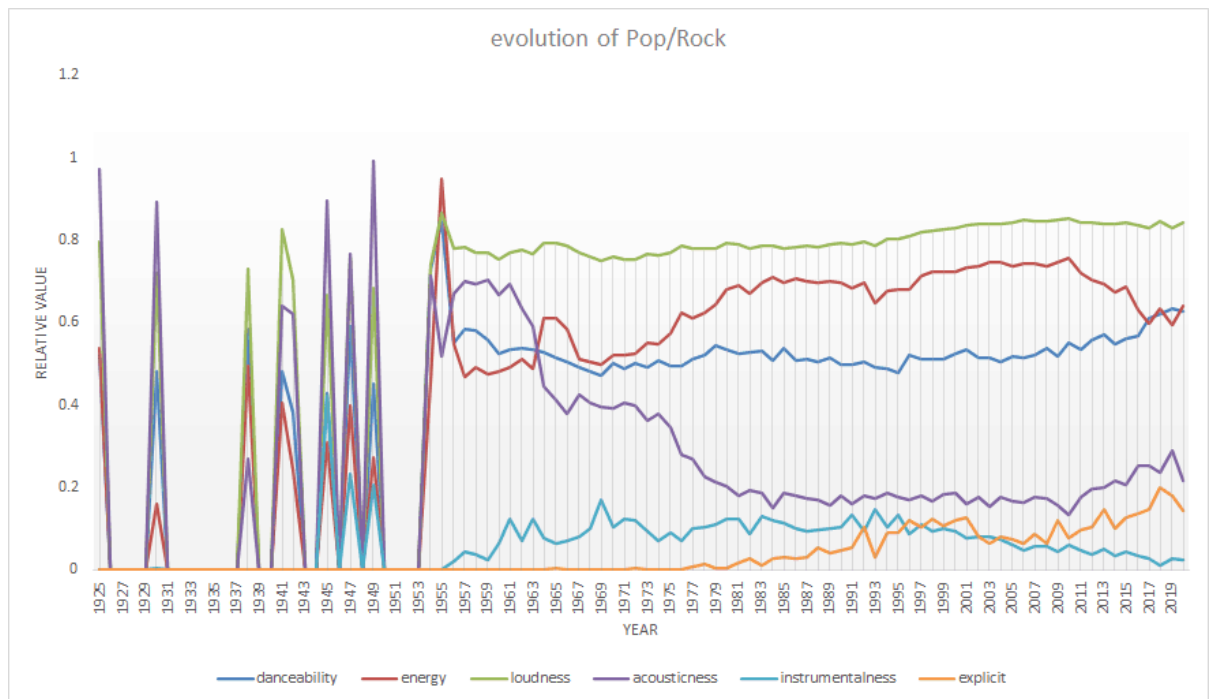
Genre Vs Year Vs Followers



观测该三维图，我们可以直观了解到1960年POP派别发生了飞跃发展，影响程度远高于各年各派，通过查找此时间点的历史事件得知，In the mid-1960s, a cultural phenomenon called British Invasion happened, when rock and pop music acts from the United Kingdom[2] and other aspects of British culture became popular in the United States and significant to the rising "counterculture" on both sides of the Atlantic Ocean.[3] Pop and rock groups such as the Beatles, the Rolling Stones, the Who, the Kinks,[4] the Dave Clark Five,[5] Herman's Hermits, the Hollies, the Swinging Blue Jeans, the Animals, Gerry and the Pacemakers, and the Searchers were at the forefront of the "invasion", 而该结果同时也验证了前述变革年份的判断，可信度更高。

5. 5. 3 1957-1960年pop派别发生重大变化的音乐特征及艺术家

我们根据数据集选取pop派别各音乐特征指标绘制逐年变化趋势图，如图。



音乐的革命性飞跃在数据上可以通过读数据折线图来理解，另一方面，历史告诉我们，音乐的革命往往与变革者横空出世密切相关。因此，我们探究音乐变革节点的杰出音乐家来理解音乐的重大进展。

在第一问中我们曾用 PageRank 算法求得每位艺术家在网络中的 PR 值，PR 值衡量了艺术家自身的影响力，而变革者更可能从高影响力的艺术家中诞生，因此我们选取 PR 值 top3 的艺术家，The Beatles, The Rolling Stones 和 Bob Dylan 这三位 Pop/Rock 艺术家来探究他们的生涯与 Pop/Rock 乃至世界音乐潮流进程的关系。

我们取出 Pop/Rock 和世界音乐的特征信息在年代中的变化，并在其中读到了有趣

的信息。首先，在现代音乐的早期（1921-约 1950）无论是 Pop/Rock 音乐还是世界音乐，都在经历日新月异的剧烈变化，早期音乐并没有形成稳定的风格，各个技术方向都在被探索、各种音乐特征都在被尝试，或许今年火爆的音乐技巧明年便无人问津，而再过一年又重新被发现其中价值。在 1950 年前后，也即二战结束不久百废待兴，Pop/Rock 音乐与世界音乐几乎同时开始趋于稳定，同时流行度也显著地升高，成熟的流派经历数十年的孕育和战火的洗礼展现出蓬勃的生命力。

目光来到 1960 年代。Pop/Rock 音乐在此节点经历了重大转变，表现为 acouticness 的热度昙花一现在此时急剧下降，而对 energy 的强调刚刚崭露头角。而三位最具个人影响力的艺术家非常巧合地几乎同时在那几年出道，他们的 energy 数值之高在同时代艺术家中独树一帜，而除了 Bob Dylan 的 acoustictness 以较高数值延续了旧时代的习惯，其他两位都摒弃了这一技巧。而世界音乐迟滞性地表达出了这一趋势，在之后三年中与 Pop/Rock 经历了类似的重大变革，受到了当时风头正盛的 Pop/Rock 的影响，而这一趋势也在之后固定下来。

artist_name	energy	acoustictness
Bob Dylan	0.477932	0.562567
The Rolling Stones	0.719915	0.293788
The Beatles	0.546827	0.360356

在数据中，我们发现了三位天才艺术家对新潮流的引领，这不仅发生在他们自身的领域，同时也让世界音乐吸收了他们的特点。世界音乐的重大变革表现位数据的急剧变化，而我们网络中的高影响力艺术家（如 Beatles、Bob Dylan、Rolling Stones）扮演了变革关头的引领者。

5.6 问题六模型建立与求解

5.6.1 建模思路

该问题可转化为分析在一个音乐流派中随着时间的推移，影响者们的影响力的变化过程及关键音乐指标与音乐发展关系。我们从 model 1 所做的多定向网络中选取某一特定音乐流派（这里选取 pop/Rock）作为子网络进行分析，首先该子网络中各艺术家的影响力水平可以通过 model 1 的综合评估模型得到，对该流派各音乐特征随时间变化情况及影响力水平、流行度随年份变化作图分析，找出发展过程中随时间发生重大变化的音乐特征，分析其对新生代艺术家人数、流行度等等的影响，结合 model1 中评价模型指标解释动态影响者指标及其影响过程。

5.6.2 Pop/Rock流派发展分析

首先根据全局定向网络找出Pop/Rock流派子网络。由5.5.3中我们获悉，pop流派在1950年之前pop流派处于起步阶段，发展不稳定不明显，而随着1957的飞跃发展，pop流派影响度、流行度逐年升高，这与其历史发展环境也有着很大的关系[参考文献]。

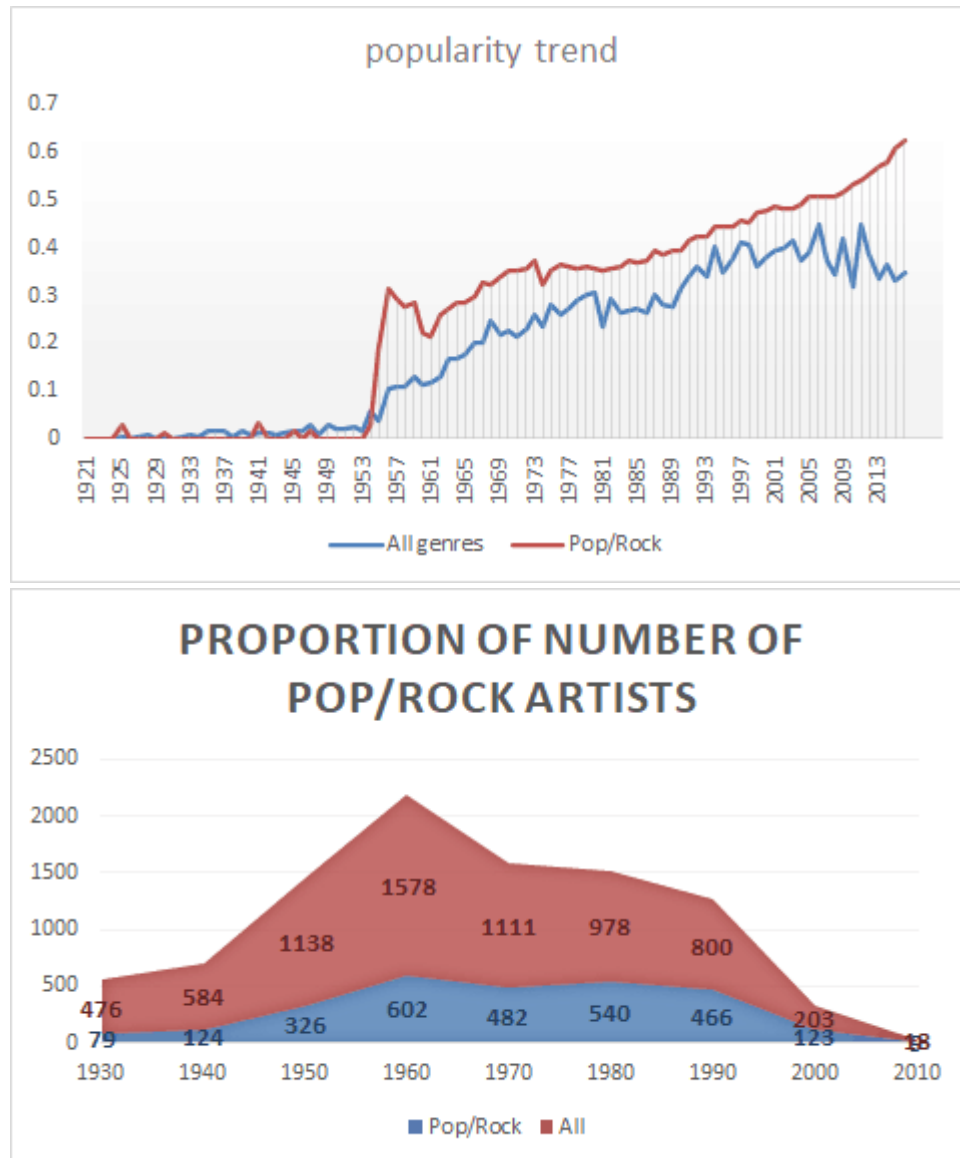
Pop/Rock流派中各音乐指标随时间变化折线图（图 ）分析我们可以知晓 音乐特征指标对该流派发展起到了至关重要的作用，故我们以此作为演变发展过程中变化较大的音乐特征的动态影响指标。

根据各流派音乐指标参数均值对时间作图（图 ）我们可以得到，在发生重大变革的时间段中，代表全体领域的音乐特征走势受动态影响指标影响，即pop流派中重大音乐指标变化已经影响到全体音乐艺术领域音乐指标风格，进而影响整个社会音乐发展。

5.6.2 音乐发展及艺术家发展影响过程

首先绘制流派全部领域与pop领域新生代艺术家人数、流行度关于时间的变化的对

比趋势图，根据不同时间序列对新生代人数、流行度关于各音乐特征的关联度分析，我们可以得到pop流派流行度与全体音乐领域流行度变化趋势相关 ($r=0.96181$)，并且pop流派新生代艺术家人数与全体音乐领域新生代艺术家人数变化趋势相关 ($r=0.884123$)，并且两者相对比，pop流派变化幅度稍大于全体音乐领域变化，关联度分析定量说明了pop流派发展带动了整体音乐领域的发展。



结合前述内容，我们可以得到pop流派发生重大变化，通过其中关键音乐特征指标***的重大变化而映射影响到全局音乐领域的音乐特征，从而导致全体音乐的飞跃发展。并且无论是pop流派还是全局音乐领域中新生代艺术家在1960年左右达到峰值，说明1960年pop流派因代表性人物及音乐变革在各流派乃至全音乐领域掀起了一场规模浩大的发展，涌现出大量新生代艺术家，而对各年份影响力均值做判断，结果依旧可以佐证该结论，而在之后一段时间里，新生代艺术家人数逐年下降，流行度增长幅度也不断降低（由于流行技术和艺术家总人数的增加，流行度依旧不断增加），这也说明了由于pop流派的影响，音乐领域发展不断繁荣但发展势头逐年下降。

(7) 你的工作如何表达音乐在时间或环境中对文化的影响？或者，如何在网络中识别社会、政治或技术变化(如互联网)的影响？

How does your work express information about cultural influence of music in time or circumstances? Alternatively, how can the effects of social, political or technological changes (such as the internet) be identified within the network?

思路

问题七要求我们基于前几问的结果，结合实际情况进行分析。我们寻找问题五音乐演进过程有代表性的变化时间段，结合这个位置所处的时间点的历史事件进行分析，识别社会、政治或技术变化对其的影响。并探究因音乐发展而产生的新文化现象以及受音乐影响的相关文化发展，分析它反作用于时间与环境的结果。在问题六的基础上，分析流派各个指标在各个时间点的变化，找到突变位置，观察音乐整体的风格的变化，结合现在社会的潮流趋势得到结论

背景部分补充：流派因共性而形成，而每个作曲家又因个性而成为永恒，了解流派形成中的一系列复杂现象，目的在于从宏观上把握音乐发展的规律。（可以提前）

问题五

流派的形成是各种艺术相互支撑、相互影响、思潮共融的结果，其产生是音乐发展过程中的客观事实，处于一定时期音乐家的创作风格必然受到那个时期的哲学观念、艺术思潮、社会历史现象的影响，并在作品中显现出历史独特的风貌。

问题五中，我们寻找到了音乐演进过程中可能存在的重大飞跃特征。1953-1957 年间为总体音乐演变中极具变革型的年份。1977-1983 年间出现平台期。

1953-1957 年二战结束，经济开始复苏。结合第五问的分析

1977-1983 年间的平台期

战后第二次世界经济危机后，各国经济进入滞胀阶段，经济回升、增长乏力，失业率和通货膨胀率却居高不下。1979—1982 年出现资本主义世界经济危机，也被称为战后第五次资本主义世界经济危机。主要资本主义国家的企业倒闭家数创战后最高纪录，失业率则达到 30 年代大危机以来的最高水平。

危机期间工业生产下降幅度美国 11.8%，英国 14.8%，法国 7.4%，联邦德国 12.2%，日本 41%。最高失业率美国 10.8%，英国 12.5%，法国 8.2%，联邦德国 8.5%，日本 2.5%。以美国、英国为例，美国失业人口从 1979 年 611 万增至 1200 万，失业率高达 10.8%，刷新上一次的记录。英国于 1979 年 7 月陷入危机，于 1981 年 5 月达到最低点。工矿生产指数降幅为 12.1%，刷新上一次危机的记录。其中纺织、冶金、和建筑业下降最大，1981 年第二季度与 1979 年同期相比，分别下降了 29.3%，29%和 18.6%。

危机期间，各国物价持续上涨，财政状况恶化，利率居高不下，世界贸易严重萎缩。1980 年美国通货膨胀率高达 13.4%。对外贸易逆差也创下 364 亿美元的新记录，出口值从 1981 年第二季度到 1983 年第二季度下降 19.8%。

艺术领域中出现的思潮往往很快会波及到其他的领域，也会反作用于时间与环境。我们很高兴地看到，我们的工作可以表达音乐在时间或环境中对文化的影响。

问题五中，我们同样发现 90 年代后则又再次迅速平滑上升。

90 年代后的平滑上升期

在文化、经济开始步入活跃期的 80 年代末、90 年代初，也是流行音乐领域日益兴起的

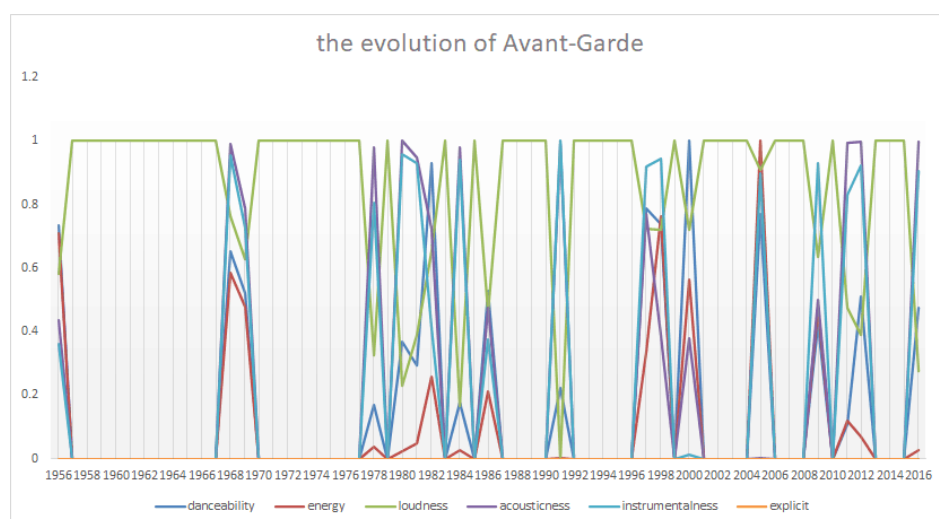
时期。伴随着网络媒介的发展，饭圈经济和饭圈文化（表达不很确切）也不断发展。Fans 意指对某一事物沉迷的爱好者。从一开始买专辑，看演唱会，催生是为偶像买周边（衍生产品）、租广告位做宣传、投票以及做慈善公益活动等多种方式。

Fans 具有群体性的概念，饭圈文化是一种有意识地寻找一种集体的认同感。费斯克在《理解大众文化》一书中就注意到了“迷”这一文化群体，认为他们是大众文化文本的过度读者，他们不同于文化精英对文本所持的批判性和欣赏性的态度，他们对文本的投入是主动参与式的，并带有热烈、狂热等强烈的感情色彩。

正是由于 Fans 的集体认同感和主动参与性才使得他们具有无以伦比的创造性和生产力。问题一与问题二中，我们探究的追随者对所有的影响者都具有乐曲上的相似程度，以及问题四中我们探究的音乐特征的感染力与继承，都印证了这一点。新音乐家作为旧音乐、旧音乐家的 Fans，使我们有理由相信，饭圈文化会给现有流派再创造，亦或是新流派出现奠定新的基础。

问题六

我们将音乐称作艺术，不仅取决于其艺术类型的主体特征，还取决于其社会功能。先锋派音乐对传统法则的颠覆，甚至打破人们的审美习惯，进而促使人们对社会现状进行重新思考。



由此图，我们看到 Avant-garde 流派的演变，是剧烈而狂野的。

Avant-garde 特指那些推动音乐整体向前发展或至少能够独辟蹊径的具有独创性的艺人。先锋派也不是单指一个艺术流派，而是泛指时间上有先后、形式上有差别、甚至观点上有对立的多种艺术派别。包括先锋派摇滚（Avant-garde Rock）、先锋派摇滚（Avant-garde Rock）指在技术上和表现上采用了与传统形式根本对立的手法的艺术探索。

先锋派音乐观念的形成可以溯源至文艺复兴时期的新艺术（Ars Nova），先锋派作曲家们普遍勇于尝试各种手法，而音乐是他们的实验室。他们的艺术，他们的创新精神在当时很长一段时间里并没有受到欢迎，但它却为日后的艺术的发展做出了持久的贡献，推进艺术进化历程并在未来人类文化的发展进程中被接受和采纳。

结合第六问的分析，Pop/Rock

六、结论

- 根据数据集创建的定向网络为图 所示，我们选择一追随者对多影响者子网络（图 ）进行分析，通过求出PR值成功表征了其自身影响力水平，并为了表征影响者对某一具体追随者的影响强度，我们考虑艺术竞争情况、流派传承关系等多个指标建立灰色综合评估模型对影响力进行衡量，所得结果如表 所示其中对研究对象最具影响力的艺术家前三名为Beatles、Bob Dylan、Rolling Stones。这里的“音乐影响力特征”是用以下指标衡量：影响者直接/间接影响人数水平、流派传承情况、各定向网络领域人数、所在领域按年人数分布、所在年份竞争激烈程度各项指标的综合评估的水平，通过熵权法提升客观性，很大程度上全面反应了音乐影响力，具有现实意义。该子网络以任意追随者为研究对象，因此我们的模型也具有良好的拓展性。
- 通过比较按流派分类的散点矩阵图和按聚类分类的散点矩阵图，我们否定了流派是音乐作品的本质分类，接着在对聚类情况的分析解释中，我们揭示了音乐作品以其在特征空间的内聚情况为更本质的分类，而对音乐分类起到决定性影响的因素是人类社会在审美、潮流、技巧方面的演化。相似度量模型已于5.2中进行了详述，流派内的艺术家并不比流派间的艺术家更相似，只有电子音乐相似性较好。
- 通过对热力图的计算和分析，我们发现流派内往往对技巧的运用有一致的偏好，流派间则不然。我们解释了在Comedy/Spoken流派的热力图中为何Speechiness和liveness相关度这么高，并在其中发现了因果关系。我们解释了先锋派的热力图为何明显区别于其他流派，并感叹于先锋派对各种技巧的大胆尝试。为了理解流派随着时间推移的变化，我们建立了流派之间的有向图并让其在时间中演化，由此我们了解到流派之间的相互借鉴、转变情况，并发现某些流派之间的借鉴关系比其他流派更为紧密。
- 根据标称属性相似度量法，其相似度达到76.58%，可以认为报告的相似性数据可以表明识别出的影响者实际上影响了各自的艺术家；根据影响力-音乐相似性的相关度分析，两个指标之间关联度达到 ，可以认为“有影响力的人”真的会影响追随者创作的音乐；根据对影响者-追随者的各个音乐特征指标余弦相似度做灰色关联分析可以得到其中Valence、Loudness、Tempo、Energy四个音乐特征与最优参数指标关联度达0.9以上，高于其他音乐特征-最优指标间关联度，而Key和mode关联度很小，故可以认为Valence、Loudness、Tempo、Energy比其它特征更具“感染力”。
- 在音乐演进过程中发生了重大飞跃，在1957年左右的pop流派出现了飞跃式发展，分析得出影响其快速发展的关键音乐特征指标为****，在定向网络中，****艺术家是此次重大发展的代表变革者。
- 详细结论见5.6 & 5.7

Sensitivity test

1、In previous section, we assume that the distinguishing coefficient $\rho = 0.2$ to reflect the ranking scores of the musicians more differently. Therefore, we need to change the value of ρ to see if the value of ρ will exert significant influence on our model.

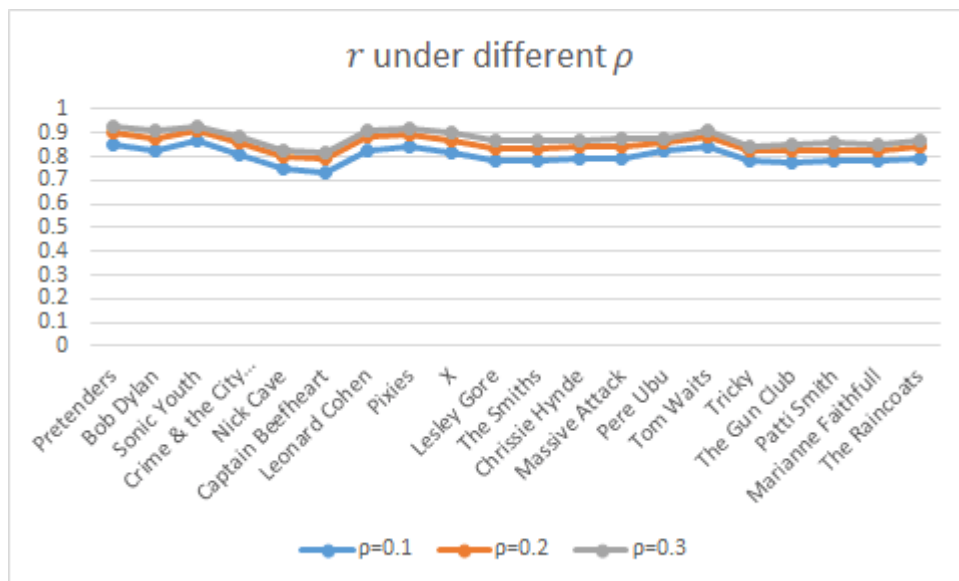
Table XX shows the rank of music influence when $\rho = 0.2$, while Table XX and Table XX show the rank of music influence when ρ is equal to 0.1 and 0.3 respectively

artists	$\rho=0.1$	$\rho=0.2$	$\rho=0.3$
Pretenders	0.85329	0.905058	0.92927
Bob Dylan	0.824561	0.8798	0.907631
Sonic Youth	0.868413	0.907122	0.927516
Crime & the City Solution	0.810047	0.860749	0.888415
Nick Cave	0.747849	0.797855	0.826917
Captain Beefheart	0.73265	0.788279	0.81858
Leonard Cohen	0.827649	0.886102	0.914215
Pixies	0.842758	0.895868	0.921529
X	0.814759	0.871768	0.901046
Lesley Gore	0.780628	0.837453	0.868956
The Smiths	0.783642	0.837902	0.867494
Chrissie Hynde	0.791441	0.842033	0.870652
Massive Attack	0.791212	0.846409	0.876518
Pere Ubu	0.82383	0.856245	0.876262
Tom Waits	0.844884	0.887602	0.910759
Tricky	0.780245	0.822202	0.845178
The Gun Club	0.778939	0.821863	0.846944
Patti Smith	0.780318	0.828872	0.855792
Marianne Faithfull	0.782329	0.829339	0.855248
The Raincoats	0.795682	0.841548	0.868191

Table XX. Rank of music influence ($\rho =0.2$)

Table XX. Rank of music influence ($\rho =0.1$)

Table XX. Rank of music influence ($\rho =0.3$)



By comparing the above three tables, we find that the rank of music influence are identical, which means that our model is insensitive to the change of ρ .

2、【在先前基于 Kmeans 聚类的音乐相似性研究中，我们根据 Calinski-Harabasz 准则确定了 K=2 为最优值，考虑到 K=4 时为较优选择，we try to change the value of K to see if the value of K will exert significant influence on our model.】

In the previous music similarity research based on Kmeans clustering, we determined K=2 as the optimal value according to the Calinski-Harabasz criterion, and considering that K=4 is the sub-optimal choice, we try to change the value of K to see if the value of K will exert significant influence on our model.

FigureXX shows the clustering of music data when K = 2, while FigureXX show the clustering of music data when K is equal to 4

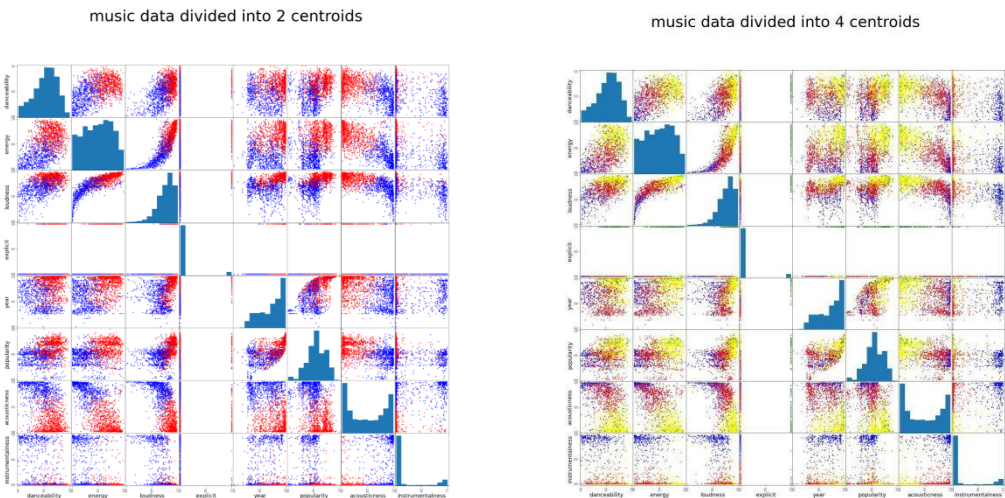


图 XX XX(K =2) 图 XX XX(K =4)

对比两图，说明如果聚成四类，其中有一类数量相当稀少，其他三类占主流；而新出现的一类处于原两类的中间地带，这表明了变革存在过渡期，过渡期前后同样分成了不同的类别，这与之前的分析相一致，并进一步补充了音乐变革中的细节。Comparing the two figures, it shows that if they are clustered into four categories, one of them is quite rare, while the other three are the mainstream; and the newly emerging category is in the middle of the original two categories, which shows that there is a transitional period for change, and the same before and after the transitional period It is divided into different categories, which is consistent with the previous analysis and further supplements the details of the music revolution.

致 ICM 协会的信

亲爱的 ICM 协会：

我们的团队开发了一种量化音乐演进的方法，了解和衡量以前制作的音乐对新音乐和音乐艺术家的影响，通过建立衡量音乐影响力的灰色综合评估模型，从而对艺术家和流派的进化和革命趋势进行考察，根据关联度来对各音乐指标受影响程度进行衡量，通过对时间序列的量化探索出音乐发展进化及革命趋势，在模型基础上，我们又考虑了诸如环境、政治等外部因素对音乐进化过程的影响，对艺术家子网络和全局网络都进行了分析，以适应更多音乐发展情况，满足了模型的迁移性和应用价值。

首先，向您做出一些关于基本模型的阐述。在音乐艺术家影响关系定向网络中，我们选取了包括影响人数、竞争力水平等在内的能够衡量艺术家音乐影响力的多个指标，通过熵权法赋权建立灰色综合评价模型，通过对多种艺术家属性的平衡评价，定量衡量影响者对追随者的影响力，根据该模型我们得到了以 Beatles、Bob Dylan、Rolling Stones 为首的极具影响力的百余名艺术家大师。

在我们所创建的艺术影响者-追随者定向网络中，各映射联系中以影响力度量，同时音乐的传承是从影响者到追随者的过程，在此过程中影响力大追随者相应增多，其传承能力相应增加，同时以音乐为载体的传承势必造成部分音乐特征指标一定程度的相似导致艺术家音乐的相似性，此时相似性通过相似度模型度量，同时影响程度的显著性以关联度模型来度量，音乐通过网络在不同派别之间传播，当某一派别发生飞跃发展时，相应整个网络产生连锁反应从而带动全局音乐领域不断发展。

由于所解决问题中提供的数据集的内容仅限于某些流派，且存在共有的艺术家的限制，由于在模型中对局部网络 and 整体网络、定量表征和可视化分析等多方面完善，随着更多或更丰富的数据引入，我们的模型亦可以起到影响力衡量作用，此时我们根据引入数据类型对综合评价模型进行一定的完善，添加数据相关指标到模型的衡量指标中，采用熵权法赋权并建模得到相应影响力，主要的改进体现在模型评价体系中新指标的引入和加权方面。

在文章中我们不仅从数学角度对问题定量表征解决，还对相关结果结合历史环境、经济政治因素、社会背景等方面均进行了分析和联系，具体论述见文章 5.6&5.7 部分详述，随着音乐的发展，社会文化不断发展丰富。我们欢迎大家对我们的模型和分析进行建议。

六、模型评价、推广与优化

七、一封信