2021 ICMProblem D: The Influence of Music

Music has been part of human societies since the beginning of timeas an essential component of cultural heritage.As part of an effort to understand the role music has played in thecollective humanexperience, wehave been asked to develop a method to quantify musical evolution.There are many factorsthat can influence artistswhen they create a new piece of music, includingtheirinnate ingenuity,current social or political events, access to new instrumentsor tools, or other personal experiences.Our goal is to understand and measure the influence ofpreviously producedmusicon new music and musical artists.

自古以来，音乐一直是人类社会的一部分，是文化遗产的重要组成部分。作为努力理解音乐在人类集体经验中所扮演的角色的一部分，我们被要求开发一种量化音乐进化的方法。 创作新音乐时可能会影响艺术家的因素，包括他们的天才，当前的社会或政治事件，使用新的乐器或工具或其他个人经历。我们的目标是理解和衡量先前制作的音乐新音乐和音乐艺术家的影响。

Some artists can list a dozen or more other artists who they say influenced their own musical work. It has also been suggested that influence can be measured by the degree of similarity between songcharacteristics, such as structure, rhythm, or lyrics. There are sometimes revolutionaryshifts in music, offering new sounds or tempos, such as when a new genre emerges, or there is a reinvention of an existing genre(e.g. classical, pop/rock, jazz, etc.).This can be due to a sequence of small changes, a cooperative effort of artists, a seriesof influential artists, or a shift within society.

一些艺术家可以列出十几个或更多他们认为对自己的音乐作品有影响的艺术家。还建议可以通过歌曲特征之间的相似程度（例如结构，节奏或歌词）来衡量影响力。有时音乐会发生革命性的变化，提供新的声音或节奏，例如何时出现新的流派，或者对现有流派（例如古典，流行/摇滚，爵士等）进行重新发明。小变化的顺序，艺术家的合作努力，一系列有影响力的艺术家或社会内部的转移。

Manysongshavesimilar sounds, and many artists have contributed to major shifts ina musical genre.Sometimes these shifts are dueto one artist influencing another. Sometimesit is a change that emergesin response to external events (such as major world events or technological advances). By considering networks ofsongsand theirmusicalcharacteristics, wecanbegin to capture the influence that musical artistshave on each other.And, perhaps, we can also gain abetterunderstanding ofhow music evolvesthrough societiesover time

许多歌曲具有相似的声音，许多艺术家为音乐流派的重大转变做出了贡献。有时，这些转变是由于一位艺术家影响了另一位艺术家。有时，它是对外部事件（例如重大世界事件或技术进步）的响应而出现的变化。通过考虑歌曲的网络及其音乐特性，我们可以开始捕捉音乐艺术家之间的相互影响。也许，我们还可以更好地了解音乐随着时间的流逝在整个社会中的发展。

Your team has been identified by the Integrative Collective Music (ICM) Society to develop a model that measures musical influence. This problem asks you to examine evolutionary and revolutionary trends of artists and genres. To do this, your team has been given several data sets by the ICM:

集成集体音乐（ICM）协会确定了您的团队，以开发一种衡量音乐影响力的模型。这个问题要求您检查艺术家和流派的进化和革命趋势。为此，ICM为您的团队提供了一些数据集：

1)“influence\_data”1represents musical influencers and followers, as reported by the artists themselves, as well as the opinions of industry experts.These data contains influencers and followers for 5,854artists in the last 90years.

2)“full\_music\_data”2provides 16variableentries, including musical features such as danceability, tempo, loudness, and key, along with artist\_nameand artist\_idfor each of98,340songs.These data are used to create two summary data sets, including: a.mean values by artist “data\_by\_artist”, b.means across years“data\_by\_year”

1）“ influence\_data” 代表艺术家自己报告的音乐影响者和追随者以及行业专家的意见。这些数据包含过去90年中5,854位艺术家的影响者和追随者。

2）“ full\_music\_data” 提供了16个变量项，包括音乐性，例如可跳舞性，节奏，响度和音调，以及98340首歌曲中的每首歌曲的artist\_name和artist\_id。这些数据用于创建两个摘要数据集，包括：a。艺术家的平均值“ data\_by\_artist”，b。表示跨年“ data\_by\_year”

Note:DATAprovided in these filesareasubset of larger data sets.These files CONTAIN THE ONLY DATA YOU SHOULD USE FOR THIS PROBLEM.

注意：这些文件中提供的数据是较大数据集的子集。这些文件包含您应为该问题使用的唯一数据。

To carry out this challenging project, the ICM Society asks your teams toexplore the evolution of music through the influence across musical artists over time,by doing the following

为了执行这个具有挑战性的项目，ICM协会要求您的团队通过以下方法，通过音乐艺术家随时间的影响来探索音乐的演变：

•Use the influence\_data data set or portions of it to create a (multiple) directed network(s)of musical influence, where influencers are connected to followers. Develop parameters that capture ‘music influence’ in this network. Explore a subset of musical influence by creating a subnetwork of your directed influencer network. Describe this subnetwork. What do your ‘music influence’ measures reveal in this subnetwork?

•Use full\_music\_data and/or the two summary data sets (with artists and years) of music characteristics, to develop measures of music similarity. Using your measure, are artists within genre more similar than artists between genres?

•Compare similarities and influences between and within genres. What distinguishes a genre and how do genres change over time? Are some genres related to others?

•Indicate whether the similarity data, as reported in the data\_influence data set, suggest that the identified influencers in fact influence the respective artists. Do the ‘influencers’ actually affect the music created by the followers? Are some music characteristics more ‘contagious’ than others, or do they all have similarroles in influencing aparticular artist’s music?

•Identify if there are characteristics that might signify revolutions (major leaps) in musical evolution from these data? What artistsre present revolutionaries (influencers of major change) in your network?

•Analyze the influence processes of musical evolution that occurred over time in one genre .Can your team identify indicators that reveal the dynamic influencers, and explain how the genre(s)or artist(s) changed over time?

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

•使用influence\_data数据集或其中的一部分来创建一个（多个）音乐影响定向网络，将影响者连接到跟随者。开发可捕获此网络中“音乐影响力”的参数。通过创建定向影响者网络的子网来探索音乐影响力的子集。描述此子网。您的“音乐影响力”措施在此子网络中体现了什么？

•使用full\_music\_data和/或音乐特征的两个摘要数据集（包括艺术家和年份）来制定音乐相似度的度量。使用您的度量，流派的艺术家是否比流派的艺术家更相似？

•比较类型之间和内部的相似性和影响。什么是流派的区别，流派如何随时间变化？有些类型与其他类型有关吗？

•指出data\_influence数据集中报告的相似性数据是否表明所确定的影响者实际上影响了各自的艺术家。 “影响者”实际上会影响追随者创作的音乐吗？是某些音乐特征比其他音乐特征更具“感染力”，或者它们在影响特定艺术家的音乐方面都有相似的角色？

•根据这些数据，确定是否存在某些特征可以指示音乐发展中的革命（重大飞跃）？哪些艺人是您网络中的革命者（重大变革的影响者）？

•分析一种类型音乐随时间变化的影响过程。您的团队能否确定能揭示动态影响者的指标，并解释不同类型或艺术家随时间的变化？

•您的作品如何表达有关音乐在时间或环境方面的文化影响的信息？或者，如何在网络中识别社会，政治或技术变化（例如互联网）的影响？

Write a one-page document to the ICM Society about the value of using your approach to understanding the influence of music through networks .Considering the two problem data sets were limited to only some genres ,and subsequently to those artists common to both data sets, how would your work or solutions change with more or richer data? Recommend further study of music and its effect on culture.

向ICM协会写一份单页的文件，说明使用您的方法通过网络理解音乐的影响的价值。考虑到两个问题数据集仅限于某些流派，然后仅限于这两个数据集共有的艺术家，随着更多或更丰富的数据，您的工作或解决方案将如何变化？建议进一步研究音乐及其对文化的影响。

The ICM Society, an interdisciplinary and diverse group from the fields of music, history, social science, technology, and mathematics, looksforward to your final report.

Your PDF solution of no more than 25 total pages should include: •One-page Summary Sheet. •Table of Contents. •Your complete solution. •One-page document to ICM society.•References list.

从音乐，历史，社会科学，技术和数学领域来看，ICM协会是一个跨学科且多元化的团体，它期待着您的最终报告。

您的PDF解决方案的总页数不超过25个，应包括：一页摘要表。目录。 您的完整解决方案。 面向ICM协会的一页文档。引用列表。

Attachments

We provide the following four data files for this problem. THE DATA FILES PROVIDED CONTAIN THE ONLY DATA YOU SHOULD USE FOR THIS PROBLEM.1.influence\_data.csv2.full\_music\_data.csv3.data\_by\_artist.csv4.data\_by\_year.csv

附件

针对此问题，我们提供了以下四个数据文件。提供的数据文件包含您应用于此问题的唯一数据1.influence\_data.csv2.full\_music\_data.csv3.data\_by\_artist.csv4.data\_by\_year.csv

Data Descriptions

1. influence\_data.csv(Data is encoded in utf-8to allow for handling of special characters):

-influencer\_id: A unique identification number given to the person listed as influencer.(string of digits)

-influencer\_name: The name of the influencing artist as given by the follower or industry experts. (string)

-influencer\_main\_genre: The genre that best describes the bulk of the music produced by the influencing artist. (if available) (string)

-influencer\_active\_start: The decade that the influencing artist began their music career.(integer)

-follower\_id: A unique identification number given to the artist listed as follower.(string of digits)

-follower\_name: The name of the artist following an influencing artist. (string)

-follower\_main\_genre: The genre that best describes the bulk of the music produced by the following artist.(if available)(string)

-follower\_active\_start: The decade that the following artist began their music career.(integer)

资料说明

1. impact\_data.csv（数据以utf-8编码，以允许处理特殊字符）：

-influencer\_id：给列出为影响者的人的唯一标识号。（数字串）

-influencer\_name：影响者的名称，由关注者或行业专家给出。 （串）

-influencer\_main\_genre：最能描述影响艺术家创作的音乐的流派。 （如果有）（字符串）

-influencer\_active\_start：有影响力的艺术家开始其音乐生涯的十年。（整数）

-follower\_id：给予跟随者的艺术家的唯一标识号。（数字串）

-follower\_name：跟随有影响力的艺术家的艺术家的名字。 （串）

-follower\_main\_genre：最能描述以下艺术家创作的大部分音乐的流派。（如果有的话）（字符串）

-follower\_active\_start：以下艺术家开始其音乐生涯的十年。（整数）

2.full\_music\_data.csv 3. data\_by\_artist.csv 4.data\_by\_year.csv

Spotify audio features from the “full\_music\_data”, “data\_by\_artist”, “data\_by\_year”:

-artist\_name: The artist who performed the track.(array)

-artist\_id: The same unique identification number given in the influence\_data.csv file.(string of digits)

2.full\_music\_data.csv 3. data\_by\_artist.csv 4.data\_by\_year.csv

从“ full\_music\_data”，“ data\_by\_artist”，“ data\_by\_year”中发现Spotify音频功能：

-artist\_name：执行曲目的艺术家。（数组）

-artist\_id：influence\_data.csv文件中给出的相同唯一标识号。（数字字符串）

Characteristics of the music:

-danceability: A measure of how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.(float)

-energy: A measure representing a perception of intensity and activity. A value of 0.0 is least intense/energetic and 1.0 is most intense/energetic. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.(float)

-valence: A measure describing the musical positiveness conveyed by a track. A value of 0.0 is most negative and 1.0 is most positive. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).(float)

-tempo: The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.(float)

-loudness: The overall loudness of a track in decibels (dB). Values typical range between -60 and 0 db. Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude).(float)

-mode: An indication of modality (major or minor), the type of scale from which its melodic content is derived, of a track. Major is represented by 1 and minor is 0.

-key: The estimated overall key of the track. Integers map to pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C♯/D♭, 2 = D, and so on. If no key was detected, the value for key is-1. (integer)

音乐的特点：

-舞蹈性：一种基于节奏，节奏稳定性，拍子强度和整体规律性的音乐元素的组合来衡量轨道适合跳舞的方式。值0.0最低可跳舞，而1.0最高可跳舞。（float）

-能量：表示对强度和活动的感知的量度。值0.0最小强度/能量，而1.0强度最大/能量。通常，充满活力的曲目会感觉快速，响亮且嘈杂。例如，死亡金属具有较高的能量，而巴赫前奏的得分则较低。促成该属性的感知特征包括动态范围，感知响度，音色，开始频率和一般熵。

-价：一种衡量音轨传达音乐积极性的量度。值0.0最负，1.0最正。价态高的音轨听起来更积极（例如，快乐，开朗，欣快），而价态低的音轨听起来更负面（例如，悲伤，沮丧，愤怒）。

-tempo：曲目的总体估计速度，以每分钟的拍数（BPM）为单位。在音乐术语中，速度是指给定乐曲的速度或节奏，直接从平均节拍持续时间得出。

-响度：轨道的整体响度，以分贝（dB）为单位。值的典型范围是-60至0 db。响度值是整个轨道的平均值，可用于比较轨道的相对响度。响度是声音的质量，它是身体力量（振幅）的主要心理关联。

-mode：音轨的形式（主要或次要）的指示，是其旋律内容所源自的音阶类型。 Major用1表示，minor用0表示。

-key：轨道的估计总体密钥。整数使用标准音高类别符号映射到音高。例如。 0 = C，1 =C♯/ D♭，2 = D，依此类推。如果未检测到密钥，则密钥的值为-1。 （整数）

Type of vocals:

-acousticness: A confidence measure of whether the track is acoustic(without technology enhancements or electrical amplification). A value of 1.0 represents high confidence the track is acoustic.(float)

-instrumentalness: Predicts whether a track contains no vocals. “Ooh” and “aah” sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly “vocal”. The closer the instrument alness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.(float)

-liveness: Detects the presence of an audience in a track. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live. (float)

-speechiness: Detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.(float)

-explicit: Detects explicit lyrics in a track(true (1) = yes it does; false (0) = no it does not OR unknown). (Boolean)

人声类型：

-声学：轨道是否是声学的置信度（没有技术增强或电放大）。值1.0表示音轨具有良好的置信度（浮动）

-instrumentalness：预测音轨是否不包含人声。在这种情况下，“哦”和“啊”的声音被视为乐器。说唱或口语单词轨迹显然是“声音”。乐器的响度值越接近1.0，轨道中没有人声内容的可能性就越大。高于0.5的值旨在表示乐器轨迹，但是当值接近1.0时，置信度更高。

-liveness：检测曲目中观众的存在。较高的活跃度值表示增加了实时执行轨道的可能性。高于0.8的值很可能会显示该轨道处于活动状态。 （浮动）

-语音：检测曲目中是否存在口语。录音越像语音（例如脱口秀，有声读物，诗歌），属性值越接近1.0。大于0.66的值描述的曲目可能完全由口语组成。介于0.33到0.66之间的值描述了可能同时包含音乐和语音的曲目，无论是分段还是分层的（包括说唱音乐）。低于0.33的值最有可能代表音乐和其他非语音类曲目。（浮动）

-explicit：检测到曲目中的显式歌词（true（1）=是； false（0）= no，它不是OR未知）。 （布尔值）

Description:

-duration\_ms: The duration of the track in milliseconds. (integer)

-popularity: The popularity of the track. The value will be between 0 and 100, with 100 being the most popular. The popularity is calculated by algorithm and is based, in the most part, on the total number of plays the track has had and how recent those plays are. Generally speaking, songs that are being played more frequently now will have a higher popularity than songs that were played more frequently in the past. Duplicate tracks (e.g. the same track from a single and an album) are rated independently. Artist and album popularity are derived mathematically from track popularity. (integer)

-year: The year of release of a track. (integer from 1921 to 2020)

-release\_date: The calendar date of release of a track mostly in yyyy-mm-dd format, however precision of date may vary and some just given as yyyy.

-song\_title (censored):The name of the track. (string) Software was run to remove any potential explicit words in the song title.

-count:The number of songs a particular artist is represented in the full\_music\_data.csv file.(integer)

描述：

-duration\_ms：轨道的持续时间（以毫秒为单位）。 （整数）

-popularity：这首歌的受欢迎程度。该值将在0到100之间，其中100是最受欢迎的值。受欢迎程度是通过算法计算的，并且在很大程度上取决于音轨的总播放次数以及这些播放的最近时间。一般而言，现在播放频率更高的歌曲将比过去播放频率更高的歌曲具有更高的知名度。重复曲目（例如，同一首曲目和一张专辑中的同一曲目）将被独立评估。艺术家和专辑的流行度是从曲目流行度中数学得出的。 （整数）

-year：曲目发行的年份。 （1921年至2020年的整数）

-release\_date：曲目发布的日历日期，大多数采用yyyy-mm-dd格式，但是日期的精度可能会有所不同，有些只是以yyyy给出。

-song\_title（经过审查）：曲目的名称。 （字符串）已运行软件以删除歌曲标题中的任何潜在显式单词。

-count：full\_music\_data.csv文件中代表特定歌手的歌曲数。（整数）