

## 2021年国际商会 问题D：音乐的影响

音乐是人类社会的一部分，是文化遗产的重要组成部分。作为理解音乐在人类集体经验中所扮演角色的努力的一部分，我们被要求开发一种方法来量化音乐进化。当艺术家创作一首新音乐时，有许多因素可以影响他们，包括他们与生俱来的创造力、当前的社会或政治事件、获得新的乐器或工具的机会或其他个人经历。我们的目标是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。

一些艺术家可以列出十几个或更多的其他艺术家，他们说他们影响了他们自己的音乐作品。还有人建议，影响可以用歌曲特征之间的相似程度来衡量，如结构、节奏或歌词。音乐有时会发生革命性的变化，提供新的声音或节奏，例如当一个新的体裁出现时，或者有一个现有的体裁的重新发明(例如。古典、流行/摇滚、爵士乐等。)。这可能是由于一系列微小的变化，艺术家的合作努力，一系列有影响力的艺术家，或社会内部的转变。

许多歌曲有着相似的声音，许多艺术家为音乐流派的重大转变做出了贡献。有时这些变化是由于一个艺术家影响另一个艺术家。有时，它是针对外部事件(如重大世界事件或技术进步)而出现的变化)。通过考虑歌曲网络及其音乐特征，我们可以开始捕捉音乐艺术家对彼此的影响。也许，我们还可以更好地了解音乐是如何随着时间的推移在社会中演变的。

您的团队已经被整合集体音乐(ICM)协会确定，以开发一个衡量音乐影响的模型。这个问题要求你研究艺术家和流派的进化和革命趋势。为了做到这一点，ICM给了您的团队几个数据集：

- 1) “*influence\_data*”<sup>1</sup> 代表音乐影响者和追随者，如艺术家自己报告的，以及行业专家的意见。这些数据包含了过去90年来5,854名艺术家的影响者和追随者。
- 2) “*full\_music\_data*”<sup>2</sup> 提供16个可变的条目，包括音乐特征，如舞蹈性，节奏，响度和键，以及98,340首歌曲的artist\_name和artist\_id。这些数据用于创建两个汇总数据集，包括：
  - a. 艺术家 “data\_by\_artist” 的平均价值”，
  - b. 意味着多年的 “data\_by\_year”。

---

<sup>1</sup> 这些数据来自All Music.com

<sup>2</sup> 这些数据是从Spotify的API中获得的

注意：这些文件中提供的数据是较大数据集的子集。 这些文件包含了你要为这个问题使用的唯一数据。

为了实施这个具有挑战性的项目，ICM协会要求您的团队通过音乐艺术家随时间的影响来探索音乐的演变，具体做法如下：

- 使用influence\_data数据集或其部分创建一个（多个）有向网络的音乐影响，其中影响者连接到追随者。 开发在这个网络中捕获“音乐影响”的参数。 通过创建你的定向影响者网络的子网来探索音乐影响的子集。 描述这个子网。 你的“音乐影响”措施在这个子网中揭示了什么？
- 使用full\_music\_data和/或两个汇总数据集（与艺术家和年份）的音乐特征，以制定音乐相似性的度量。 用你的衡量标准，流派中的艺术家是否比流派之间的艺术家更相似？
- 比较体裁之间和体裁内部的相似性和影响。 一个体裁的区别是什么，体裁是如何随着时间的推移而变化的？ 有些体裁与其他体裁有关吗？
- 指出data\_influence数据集中报告的相似性数据是否表明已识别的影响者实际上影响了各自的艺术家。“影响者”真的会影响追随者创造的音乐吗？ 有些音乐特征是否比其他音乐更具有“传染性”，或者它们在影响特定艺术家的音乐方面都有类似的作用？
- 从这些数据中确定是否有可能意味着音乐进化的革命（重大飞跃）的特征？ 什么艺术家代表革命者（重大变革的影响者）在你的网络？
- 分析音乐演变的影响过程，随着时间的推移，在一个体裁。 你的团队能找出揭示动态影响者的指标，并解释流派或艺术家是如何随着时间的推移而变化的吗？
- 你的作品如何在时间或环境中表达关于音乐文化影响的信息？ 或者，如何在网络中识别社会、政治或技术变革（如互联网）的影响？

写一份一页的文件给ICM协会，关于使用你的方法来理解音乐通过网络的影响的价值。考虑到这两个问题数据集仅限于某些类型，然后是两个数据集共同的艺术家，您的工作或解决方案将如何随着更多或更丰富的数据而变化？ 建议进一步研究音乐及其对文化的影响。

来自音乐、历史、社会科学、技术和数学领域的跨学科和多样化的ICM协会期待着您的最后报告。

您的PDF解决方案不超过25页，应包括：

- 一页汇总表。
- 目录。
- 你的解决方案。
- 提交ICM协会的一页文件。
- 参考资料清单。

**注：2021年新！** ICM竞赛现在有25页的限制。 您提交的所有方面都按25页的限制计算：摘要表、目录、解决方案主体、图像和表格、一页文档、参考列表和任何附录。

## 附件

我们为这个问题提供了以下四个数据文件。 提供的数据文件包含您应该用于此问题的唯一数据。

1. [influence\\_data.csv](#)
2. [full\\_music\\_data.csv](#)
3. [data\\_by\\_artist.csv](#)
4. [data\\_by\\_year.c](#)

## sv数据描述

### 1. [influence\\_data.csv](#)

(数据以utf-8编码，以便处理特殊字符)：

- influencer\_id: 给被列为影响者的唯一识别号码。（一串数字）
- influencer\_name: 由追随者或行业专家给出的影响艺术家的名字。（字符串）
- influencer\_main\_genre: 最能描述影响艺术家创作的大部分音乐的体裁。（如果可用）(字符串)
- influencer\_active\_start: 影响艺术家开始音乐生涯的十年。（整数）
- follower\_id: 给被列为跟随者的艺术家的唯一识别号码。（一串数字）
- follower\_name: 跟随影响艺术家的艺术家的名字。（字符串）
- follower\_main\_genre: 最能描述以下艺术家创作的大部分音乐的体裁。（如果可用）(字符串)
- follower\_active\_start: 以下艺术家开始音乐生涯的十年。（整数）

## 2. [full\\_music\\_data.csv](#)   3. [data\\_by\\_artist.csv](#)   4. [data\\_by\\_year.csv](#)

Spotify音频功能来自“full\_music\_data”、“data\_by\_artist”、“data\_by\_year”：

- artist\_name: 表演曲目的艺术家。（数组）
- artist\_id: influence\_data.csv文件中给出的相同唯一标识号。（一串数字）

### 音乐的特点：

- 舞蹈性：一种基于音乐元素的组合，包括节奏、节奏稳定性、节拍强度和整体规律性，来衡量一个曲目是否适合跳舞。 值0.0是最不可跳舞的，1.0是最可跳舞的。（浮动）
- 能量：表示对强度和活动的感知的度量。 值0.0是最不强烈/能量的，1.0是最强烈/能量的。 通常，充满活力的轨道会感觉快速、响亮和嘈杂。 例如，死亡金属有很高的能量，而巴赫的前奏在量表上得分很低。 这一属性的感知特征包括动态范围、感知响度、音色、起跳率和一般熵。（浮动）
- 价态：一种描述曲目所传达的音乐积极性的度量。 值0.0最负，1.0最正。 高价音的轨道更积极(例如。 快乐，开朗，兴高采烈)，而低价音轨听起来更消极(例如。 悲伤，沮丧，愤怒)。（浮动）
- 节奏：以每分钟节拍为单位的轨道的总体估计节奏(BPM)。 在音乐术语中，节奏是给定作品的速度或节奏，直接来源于平均节拍持续时间。（浮动）
- 响度：轨道的整体响度，单位为分贝(dB)。 值在-60到0db之间的典型范围。 响度值在整个轨道上是平均的，对于比较轨道的相对响度是有用的。 声音是声音的质量，是身体力量（振幅）的主要心理关联）。（浮动）
- 模式：一种轨迹的情态（主要或次要）的指示，它的旋律内容是从其尺度的类型。 主修用1表示，辅修用0表示。
- 关键：估计轨道的总体关键。 整数映射到点，使用标准的Pitch类表示法。 E. g. 0=C, 1=C# / D b , 2=D等等。 如果没有检测到键，则键的值为-1。（整数）

### 唱腔类型：

- 声学：衡量轨道是否声学（没有技术增强或电气放大）的置信度）。 值1.0表示高度置信，轨道是声学的。（浮动）
- 工具性：预测一个曲目是否包含没有声音。 在这种情况下，“呜”和“啊”的声音被视为工具。 说唱或口语曲目显然是“声乐”。 器乐值越接近1.0，曲目不包含声乐内容的可能性就越大。 高于0.5的值意在表示工具轨道，但随着值接近1.0，置信度更高。（浮动）
- 活力：在赛道上检测观众的存在。 较高的活性值表示轨道被实时执行的概率增加。 高于0.8的值提供了轨道运行的强烈可能性。（浮动）

- 言语：在一个轨道上检测口语的存在。更纯粹的演讲就像录音(例如。脱口秀，有声书，诗歌)，属性值越接近1.0。高于0.66的值描述了可能完全由口语构成的音轨。 值在0.33到0.66之间，描述可能包含音乐和语音的曲目，无论是在部分还是分层，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）
- 显式：检测曲目中的显式歌词（真(1)=是的；假(0)=不，它没有或未知）。（布尔值）

#### 说明：

- duration\_ms：轨道的持续时间(毫秒)。（整数）
- 流行：赛道的流行。 值将在0到100之间，其中100是最受欢迎的。 流行度是通过算法来计算的，在很大程度上是基于赛道上的总播放次数和最近的播放次数。 一般来说，现在播放频率更高的歌曲将比过去播放频率更高的歌曲更受欢迎。 重复轨道(例如。同一曲目来自单一和专辑)是独立的。 艺术家和专辑的流行在数学上来源于曲目的流行。（整数）
- 年份：轨道发布的年份。（1921年至2020年为整数）
- release\_date：轨道发布的日历日期大多采用yyyy-mm-dd格式，但日期的精度可能会有所不同，有些只是作为yyyy给出的。
- song\_title（审查）：轨道的名称。 运行软件是为了删除歌曲标题中任何潜在的显式单词。
- 计数：特定艺术家的歌曲数量表示在full\_music\_data.csv文件中。（整数）