为了量化分析音乐的影响，在这份报告中，我们针对艺术家之间复杂的关系建立了影响力网络模型，判断不同时间、不同作者、不同流派的音乐之间的相似度。并且依据网络进行分析，预测音乐流派风格的变化，尝试发现未来音乐的重大变革。

为了分析艺术家之间的影响关系，我们建立了艺术家之间的影响关系图，运用 的算法将艺术家的影响力量化，这使我们能 。我们以标准化的余弦相似度（cosine similarity）作为衡量两首乐曲各个特征之间相似度的标准，这样做的一个亮点是我们能分别分析乐曲不同特征对艺术家影响的大小，易于看出哪些特征是“易传承的”。利用对影响力图和相似度的建模，我们不仅在微观角度上比较了派系内外艺术家之间的互相影响，也从宏观上比较了不同派系之间的互相影响和相似程度。

此外，通过在我们的模型中加入时间尺度，我们寻找到了一些可能代表音乐派系的revolution出现的特征。我们还将模型与经济、政治、历史等因素共同考虑，通过使模型更贴合现实情况，分析了导致音乐派系变革的原因，并尝试预测未来音乐的发展方向。

通过模型分析，我们发现艺术家之间的互相影响在音乐的发展中起着不可或缺的作用，尤其是来自不同流派之间的艺术家之间的影响。同时，我们还发现音乐流派的变革与经济等社会因素密不可分。

我们模型一个强大的地方是它从多个角度观察了音乐家之间的联系。当需要分析的问题较为微观时，我们模拟了任何两个艺术家之间的互相影响和相似度；当问题比较宏观时，我们将流派看作整体，对流派子图整体分析。并且，通过在模型中加入更多考虑的因素，我们的模型能够很好的贴合现实情况。

重述问题：

我们拿到了有关5000多名艺术家和他们创作的98000多首乐曲的相关数据，需要根据这些数据建图，并根据图探索各种选择。我们的主要目标有：

·凭借影响力关系建立艺术家影像图，并依据图定义影响力模型，确定艺术家影响力大小并分析其代表的意义

·建立音乐相似度模型，以定量衡量不同歌曲之间的相似度，并比较流派内外的相似度

·分析流派内外的关系，并加入时间变量，分析流派随时间的变化

·分析影响者对被影响者的影响程度，以及音乐不同特征的影响能力

·找出预示音乐大变革将发生的特征，并确定引起变革的艺术家

·分析一个流派的音乐演变的过程，确定动态影响者的指标，并解释流派和艺术家如何随着时间变化

·在模型中寻找社会因素的影响

预期目标：

·建立艺术家之间互相影响的网络，在网络中分析每位艺术家的影响力

·建立相似度模型，模型能定量分析两首歌，两个艺术家，两个派别之间的相似度

·在模型中加入时间变量，研究派别各个指标随着时间的变化，寻找预示着变革的特征

·将模型与社会文化因素结合，寻找社会文化因影响音乐模型的特征

基础假设：

在建模之前，我们做出了如下基础假设：

·在这么大的数据量下，音乐是艺术家之间互相影响的唯一途径，而忽略掉人际关系，社会影响等特征

·题目中给出的歌曲数据能够很好的反应现实中所有歌曲的分布情况

建模路线

我们在建模过程中主要经历了四个阶段：

1. 将每一位艺术家看作一个结点，根据influence\_data中的艺术家影响关系，将追随者与影响者之间建立影响关系，最终形成一个复杂的影响关系图。基于此，我们根据歌曲的传播速度，来定义艺术家的影响力大小，由此建立影响力模型。
2. 我们根据题目中给出的full\_music\_data，将不同指标做差，将差值常态化，求其余弦距离，将其作为两首歌曲之间的相似度，值越小相似度越高。由此，我们可根据此模型进一步计算艺术家与艺术家之间的相似度。
3. 为了研究派系与派系之间的相似度，且在模型中加入时间变量的影响，我们简化了我们的模型，且将其与时间变量结合，研究派系在时间尺度上的变化。
4. 最后，我们在图中加入社会文化因素变量，将具体的时间结点与之前找出的音乐派系变革结合，找出预示着音乐革命可能发生的特征，将模型的变化与现实结合起来。

第一部分：艺术家间影响网络图

数据处理：

在文件influence\_data中，记录了艺术家的影响者与追随者。一个艺术家可能有十个以上的影响者，也可能有很多追随者。影响关系的数据共有42771条，面对如此大量的数据，我们选用python编写程序处理。

我们将每个艺术家视为一个结点，根据影响关系将其与其他结点链接，建立一个包含所有艺术家的有向图，依据该有向图能很方便的找到某位艺术家的追随者子图。

建立模型：

我们需要定义一个影响力的模型，量化每一位艺术家在该图中的影响力。

首先分析音乐的传播问题。我们假设：

·在不同年代音乐的传播速度是一致的，及艺术家的影响力不会因为年代不同而不同

·不同音乐家受影响的概率相同，即一个艺术家的影响力只与受他影响的人数有关

·一个艺术家受n位艺术家影响时，每位艺术家对他的影响力为单独影响他时的

·音乐的传播类似病毒，以指数速度传播

·当一个艺术家的追随者影响他人时，仍然算作该艺术家的影响力的一部分。但影响力会相应的减少

·影响力与派系无关，歌曲对不同派系的艺术家的影响力相同

拟定了这些假设之后，我们开始建立艺术家的影响力模型：

一位艺术家的影响力与受他影响的人数有关，而歌曲以指数速度传播，类似于传染病模型。但歌曲在传播之后不会像传染病一样被“治愈”，所以我们选用传染病SI模型来模拟歌曲的传播。若一位艺术家的直接追随者有n人，则他的影响力为传播人数的对数。

同时，考虑到艺术家之间可能存在类似老师和学生的一代代影响的情况，我们尝试将间接影响也纳入我们的影响力模型中：

当一位艺术家的追随者继续去影响他人时，也会为这位艺术家带来影响力。不过在多轮影响中，最初始的艺术家的影响力会逐渐降低，否则会使早期的艺术家影响力过大，不利于比较分析。所以我们选用了指数衰减的方式，即第n轮被影响者，总体影响力要除以（ζ是我们选取的一个常数）。

通过将间接影响力加入模型，我们的模型已经能较好的反应艺术家在该网络中的影响力。

影响力模型的意义

依据此模型，我们能依据给定的艺术家求出他的影响力子图，并依据此子图求出他的影响力大小。我们的影响力模型反应了他在子图中对所有受他影响的艺术家的影响力总和，这同时也是他在整个图中的影响力总和。

模型评价：

优点：

我们建立的第一个影响力模型考虑了每位追随者带来的影响，同时也将艺术家之间的间接影响纳入模型之中，能很好的反应艺术家在整个图中的影响力，而非仅仅是局部的影响力。

缺点：

这个影响力模型忽略了流派在影响力中的作用。同时，将间接影响力纳入考虑范围会使活跃年代较晚的艺术家影响力偏低，因为时间问题，他们还没有机会去间接影响更晚的艺术家。

第二部分 音乐相似性建模

数据处理：

在文件full\_music\_data中，我们能找到由上述艺术家创作的98341首歌曲的14项乐理指标，以及歌曲名字，发行时间和创作者。为了能有效处理这些数据，我们做出如下假设：

1. 给出的十二项数据能完整的描述一首歌的乐理特征，即音乐之间的相似度可以由对比这十四项指标得到
2. 这14项指标在乐曲中所占比重相同，不存在某些指标更重要的情况
3. 时间和作者不会影响两首歌的相似度，所以在对比时不会将作者和时间作为比较的参数
4. 用于比较的十四项指标全部与作者有关，不考虑多个作者创作不同部分的情况

建立模型：

基于以上假设，我们希望通过比较两首歌的这14项乐理指标来衡量它们的相似度。不过这些乐理指标大小不同，直接相减可能会使指标占的比重不同，不能很好的衡量乐曲的相似度。且有些指标的分布是离散的，需要另一种算法计算其相似度。

所以我们的团队首先将所有连续的指标进行normalize处理，使所有指标在模型中占的比重相同。然后我们选取了余弦距离来衡量两个特定乐曲之间的相似度。即若某一项指标的D-value为k，则它们的余弦距离为

而对于离散的指标，若它们不相同，则将其距离定为1（最大余弦距离的一半）；若它们相同，则将其距离定为0。

最终得到十四项指标的余弦距离，我们将其取平均值，作为两首乐曲的余弦距离。由此，我们可以定义两首乐曲的相似度：

余弦距离越接近0，两首乐曲相似度越高；余弦距离越接近2，两首乐曲相似度越低。

因为作者的创作风格与他的乐曲密不可分，所以得到乐曲的相似度后，我们可以得到作者之间相似度的定义：作者的乐理指标为他创作的所有歌曲的指标的均值（能在data\_by\_artist）中找到，两位作者的相似度为他们的乐理指标的余弦距离的均值。能够这样定义作者相似度的一个原因是：一位作者的歌曲不会太多，作者的相似性和歌曲相似性问题本质上都是针对具体对象的分析，所以可以用同一种算法。

流派内外艺术家的相似度:

依据我们建立的相似度模型，我们得以比较艺术家之间的相似度。为了研究流派内艺术家是否比流派外艺术家更相似，我们流派内部，两两艺术家之间求其相似度，再对整体取均值，作为流派内部艺术家的相似度；在流派外部，我们对两个流派的所有艺术家求相似度，取其均值作为两个流派的相似性。即

公式

如此，我们得到这样一个17\*17的矩阵，每个元素代表两个派系艺术家之间的相似度：



如图所示，元素的颜色越偏向绿色，则两个派系艺术家之间相似度越高。我们可以看出大部分派系的艺术家和本派系的艺术家更相似。这种现象在一些风格较为明显的派系上尤其明显，例如Easy Listening的派系内相似度达到了0.08，classical的拍戏内相似度达到了0.069；而对于一些派系风格不那么明显，派系内还有其他分支的派系来说，相似度一般较低，例如种类繁多的electronic派系内相似度仅有0.025。基本可以证明派系内艺术家的相似度整体上比派系外的艺术家相似度高。

模型评价：

优点：

这种建模方式能带来的优点是十分明显的，它能十分清晰的衡量两个乐曲和两个艺术家之间的相似度，数值都经过标准化，易于在其他模型种进一步加工使用。余弦距离模型的应用也使数据十分具有说服力。同时，对艺术家和乐曲使用同一个模型衡量其相似度，体现了艺术家和乐曲之间的紧密联系，使我们能用同一个标准去分析艺术家和乐曲，更加方便。

缺点：

这种建模最大的缺点在后文有所提及，即其不能直接衡量一个集合内部元素的相似度。我们只能比较不同艺术家之间的差别，而无法比较一个艺术家自己创作的乐曲的总体差别。例如在求针对这个问题，我们在第三部分给出了一种简单有效的解决方式。

第三部分 流派内外相似性的建模

在第二部分中，我们使用定义的相似度模型衡量了不同派系艺术家之间的相似度。但当我们需要从整体上衡量两个派系之间的相似度时，这种模型不再适用。因为它不能很好的反应派系整体上的相似情况，只能反应不同艺术家之间的相似情况。

而且上文中的算法需要处理每位艺术家的数据，在加入时间变量后，会使程序的复杂度变得难以接受。为了能在宏观尺度上以更简洁的算法完成计算，我们需要简化我们的算法，使我们的模型能更好的与时间结合，解决宏观上的问题。

模型建立：

首先，我们不再将每位艺术家单独分析，而将整个派系看作一个整体。遍历我们建立的艺术家图和full\_music\_data，将歌曲按派系分开，计算每个派系歌曲各项指标。

同时，我们考虑了不同歌曲popularity的影响。popularity高的歌曲在派系中所占的比重应当较高。我们对所有popularity normalize，而后将其作为这首歌曲的权值，计算所有歌曲指标的加权平均数。得到的数据如下表：

然后加入时间变量，将歌曲按年份分开分析不同年份派系的各项指标。为了能清晰的反应派系随时间的变化，我们计算了每年派系各项指标与前一年的相似度，相似度越低代表该年该派系变化越大。画出派系随时间变化的趋势图如下：

影响关系的作用

我们已经能够衡量两位艺术家之间的相似度。此时我们想分析的问题是：影响者对追随者究竟起到了多大的作用？是否影响者与追随者之间存在更加相似的情况？如果存在，某些指标是否会比其他指标更加相似？

我们根据influence\_data中的每一对影响者与追随者，计算他们每一项指标的相似度，并normalize化数据，得到的影响关系的相似度如下：

音乐革命的特征和领导者

我们相信，音乐上的革命与音乐家之间的互相影响密不可分，尤其是派系之间的互相影响。所以我们决定从派系的变化开始研究。

首先，我们从full\_music\_data中取出特定年代的数据。我们选取了1958-2020的音乐数据，将其按派系分开，研究每个派系随时间的变化。这是变化趋势：

我们可以看出，正常情况下派系与上一年的相似度是较高的。而在派系发生大的变革的时候，派系每年之间的相似度会大幅下降，这代表着在这段时间内派系每一年都在发生极大的改变。

音乐的革命是由少数领导者发起的，他们领导了音乐派系的变革。这意味着他们的音乐风格带领了派系的变革。所以，为了找到艺术的革命者，我们

第四部分 社会因素对音乐的影响

音乐与我们的社会密不可分。在前文中我们已经分析出了音乐变革的特征和发起者，但没有研究引起音乐变革的原因。所以，我们将上一部分的模型与现实世界中的因素结合，通过比较音乐的变革与社会的变化，来分析社会因素对音乐的影响。

模型评价：

优点：

·我们的模型仅使用题目中所给的数据，数据格式统一。这意味着我们的模型适用性很高，可以用于分析任意一个艺术家网络

·在不同尺度上分别建模，既减少了大范围运算的运算量，也保留了小范围内信息的完整

·模型对艺术家采用了高级抽象，使不精通乐理知识的人也能合理分析艺术家和他们的音乐

·该模型充分利用了题目中所给的数据，四个表格中的数据全部都有使用。这意味着我们的模型包含了题目中所给的全部信息，最终结果与各个方面指标都有关。

缺点：

·在比较派系相似度时，我们本准备采用聚类算法。这种方法求出的相似度能更好的反应派系的聚拢程度，但加入时间变量后，运算量太大。由于时间问题，我们在损失一部分精度的情况下减少运算量，选择了当前的模型。

·我们的模型在分析影响力和popularity参数时，没有考虑时间变量所起的作用。理论上时间越早的音乐影响会越大。

·我们定义的局部相似度与全局相似度之间因为模型不同，不能很好的统一，对比时不够直观。

敏感性分析

对于影响力模型

王稳地. (0). 传染病数学模型的稳定性和分枝. (Doctoral dissertation).

Shevy, & Mark. (2001). When is music communication? a music communication matrix based on assumption, intention, and meaning construction. Kidney International, 60(3), 969-73.

杨柳. (0). 乐曲检索与评价系统的设计与实现. (Doctoral dissertation, 电子科技大学).