问题说明：

1. 使用influence\_data这个数据来创造一个或多个有向图influencer连接到followers ，利用这个图设计一个指标“音乐影响力” ，如果分割一个子图出来，重新计算的子“影响力”在子图里面代表什么？

**建立有向图和计算影响力**

2.使用full\_music\_data或者3. data\_by\_artist.csv 4. data\_by\_year.csv来弄一个“音乐的相似度”，根据你的相似度，同流派作家的相似度比不同流派作家的相似度更高吗？

**计算相似度**

3.在流派比较相似度和影响，是什么导致流派的不同？流派是怎样随时间变化的？有流派和其他流派有联系吗？

**探究流派之间关系**

4.这个“受影响” 真的有在相似度上体现吗？某些音乐特征是不是更容易“传染”（指被影响者影响）？还是说所有的都一样？

**探索影响与相似度的关系**

5.找出容易引起“音乐革命“的音乐特征，哪些音乐家是导致音乐革命的”革命者“？

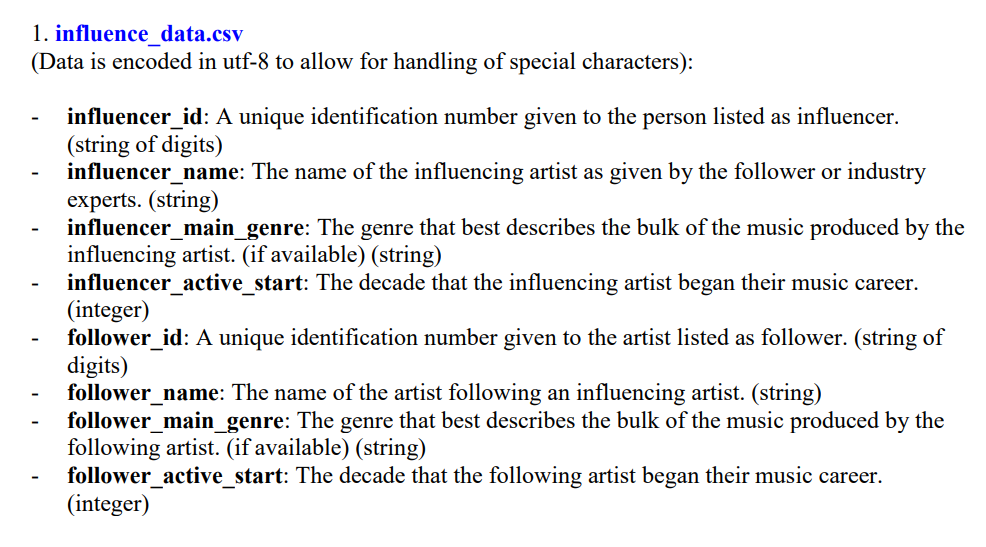
**音乐革命**

6.分析一个流派中音乐的变化历史，确定一个指标来反映动态影响者（这啥 dynamic influencers） ，解释流派和艺术家是怎样随时间变化的。

**与时间的关系**

7. 你的工作如何表达音乐在时间或环境中的文化影响？或者说，如何在这个图中辨别出社会、政治或技术变化（如互联网）对音乐的影响？

**从图中反映社会变化**



下面是解题思路（还没写）

再下一页是樊要干什么

***樊要做的事情*** *yi bai a ru wa ne，go ku ro sa ma ，orz*

推荐使用python（matlab主要是不熟

**Influence data是描述音乐家（节点）和影响关系（边）的**

**Data by year是按照年份中位数各音乐特征的 反映时代特征**

**Data by artist是按照音乐家中位数的每个人的音乐特征 要记录到节点里面去**

**注意id不是连续的，可能要先用字典整个映射把id搞成连续的省内存**

**可以再整个字典把流派给编个号，流派没有给编号的**

**第一问**

1.建立有向图 把音乐家作为一个节点 influence\_data数据中每行表示一条边，节点包含 id 名字 流派 开始活动年份 ，并且followers表示这个节点所指向的节点，如果所指节点没信息是空的话就填上这行给出的和follower相关的信息

2.检验这个图有没有环，需要保证是无环图才能下一步 如果有环跟我说我重新想想，如果没有环说一声然后继续下面步骤，dfs 请

**有环啊 那没事了**

/\*

~~3.计算每个节点的子节点数量，就每个节点再加个信息叫影响力，底下有多少点就有多少影响力，注意可能这个图不是树，所以记得遍历过的节点不要重复遍历了（~~

~~这个有向图后面还要用很多次 别删了（~~

\*/

3.算了 还是按照原来的方法整，就是不能重复就行了 （开摆qwq）

计算每个节点作为起点进行不重复的dfs所遍历的节点个数作为影响力

**第二问**

1.结构体（类） music{

各种音乐特征（danceabilty,energy,valence,tempo,loudness,mode,key,acousticness,instrument,liveness,speechines,explicit）

}

2.整一个函数 sim(music a, music b) 返回两个对象的相似度

Sim(a,b)=上面那12个音乐特征的皮尔逊相关系数Pearson Correlation Coefficient

实现方法 网上一堆 百度 请

[推荐算法入门（1）相似度计算方法大全 - 知乎 (zhihu.com)](https://zhuanlan.zhihu.com/p/33164335)

[如何理解皮尔逊相关系数（Pearson Correlation Coefficient）？ - 知乎 (zhihu.com)](https://www.zhihu.com/question/19734616)

3.整波大的 估计很花时间

对艺术家两两比较算sim 数据来源是data by artist

**O(n^2) 起飞！**这个二维数组记得存下来 后面还要用

4.然后把所有同流派的音乐家比较出来的相似度取个平均数，不同流派音乐家的比较出来的相似度取个平均数，这两个值拿来比较大小而且要写进论文

*不多不多 也就几亿次计算（*

**第三问**

1.对每个流派 算整个流派的所有音乐家的各种音乐特征的中位数，作为整个流派的音乐特征，

2.计算流派内艺术家影响其他流派的“影响数” 比如a流派有3个艺术家影响了b流派的艺术家，b流派有1个艺术家影响了a流派的艺术家，3>1,那就说 b流派受a流派的影响

3.这一问想不动了 待会继续想（

**第四问**

1.遍历每条边，对每条边来说，找到前后两个节点之间的相似度，

给chaos写进论文里

2.12个音乐特征计数，对每条边，对每个特征 某特征的前后相似度为influencer的某特征+1除以follower的某种特征+1 。。。然后对每条边都找到一个特征的前后相似度是最接近1的，排个序，然后把第几位直接加到数组里面，最后计算哪几个特征数字最小，最好12个特征的被选中次数都给出来

**第五问**

1.

**解题思路：给chaos看的东西**

**第一问**

1.对于影响力网络（即有向图）来说，影响了多少人就是最直观的表现，

所谓“宗师” ，影响力越高，影响的人越多，就越接近于“行业泰斗“

~~所以直接算某节点的子节点数量就行了~~

~~但是问题是如果有环就不能这样 所以还要检验是否存在环~~

~~（我影响了 影响了我的人）~~

**寄 居然真有环**

算了 还是按照原来的方法整，就是不能重复就行了 （开摆qwq）

意思就是 计算每个节点作为起点进行不重复的dfs所遍历的节点个数作为影响力

自己翻译翻译 我说的是人话 要翻译成鬼话给评委看（

**第二问**

1.相似度用12种音乐特征（种类）决定，问题就转化为了求两个12维向量的相似度。鉴于12个变量的取值范围各有不同，俺懒得去搞标准化，所以就采用了Pearson Correlation Coefficient作为相似度。

2.枚举所有艺术家之间的相似度可以得出是否同流派艺术家更加相似。

**第三问**

R.I.P.

**第四问**

1.直接计算影响关系中前后的相似度的平均值（中位数），是否平均地普遍地高于正常值，即可判断是否真的有影响

2.寻找“影响前后”的特征较为相似的几个特征总的统计一下就行了。