

问题1

利用邻接表建图

这里考虑到这里的数据量的大小，我们采用邻接表来建图，图的节点表示艺术家，图的有向边从影响者指向其对应的追随者，边的权重代表着影响者对某一个对应的追随者的影响大小。

此处可以普及一下邻接矩阵和邻接表的概念，顺便凑一下字数。

音乐影响力的表示

初版

通过统计每一个艺术家的出度，用出度的大小来表示艺术家影响力的大小，此时影响者对其追随者的影响均为1。一个艺术家的音乐影响力也就是其追随者数。

可以贴一下出度、入度的定义。

优化1

实际情况中对于一个特定的艺术家，他对其追随者的影响不一定是均等的，而这又与追随者自身有关。基于这样的事实，我们对我们的初版做了一些优化。

记艺术家 i 的音乐影响力为 T_i ， i 的入度为 I ，则我们现在这样定义 T_i ： $T_i = \sum t_{ij}$ 。这里的 t_{ij} 表示艺术家 i 对其追随者 j 的影响力，且 $t_{ij} = 1/I$ 。通过这样的方法，我们可以知艺术家 i 的音乐影响力由艺术家 i 影响的人数以及其影响其追随者的艺术家总数共同得出。

优化3

观察优化2，我们可以发现 t_{ij} 的表示不能很好地反应影响者们对某一个追随者 j 的影响程度，而在现实之中不同的影响者对 j 的影响往往是不同的。在此基础上，我们做出进一步的优化。

首先，对于每一位音乐家 i ，我们用一个向量表示， $\vec{x}_i = (x_1, x_2, x_3)$ ，其中 x_1, x_2, x_3 分别表示对音乐特定降维后的主成分，然后计算影响者与追随者之间的距离 $dist$ 。

$$dist_{ij} = \sqrt{\sum (x_i - x_j)^2}$$

对于 $dist$ 我们不难得出这样的结论， $dist$ 越大，则他们之间的影响越小；反之，影响越大。

作一个变换，记影响力 $t'_{ij} = C - dist_{ij}$ ，其中 C 为一个足够大的参数（可以取为 $\max(dist_{ij}) + 1$ ）

最后对其归一化，得 $t_{ij} = \frac{t'_{ij}}{\sum_u t_{uj}}$

所以一个艺术家的音乐影响力 $T_i = \sum_j t_{ij}$ 。

结果

对于初版，最后统计得出音乐影响力最大的艺术家的ID号为754032，名字为The Beatles。受其影响的艺术家数为615，由前面的定义，其音乐影响力为 $T = 615$ 。

对于优化1后的结果，其结果仍然是 The Beatles，其音乐影响力为 $T = 72.9157$