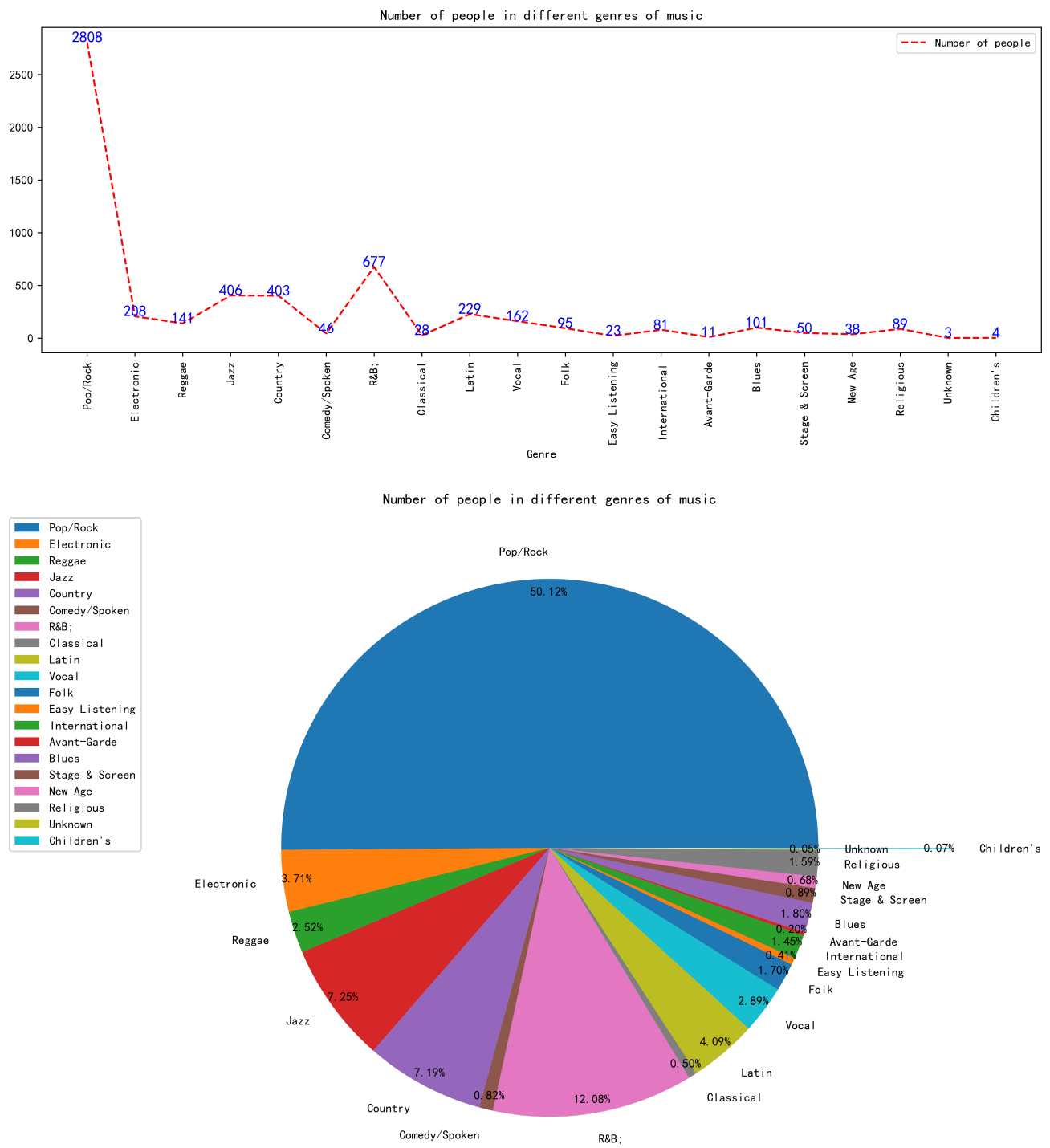


1.建图

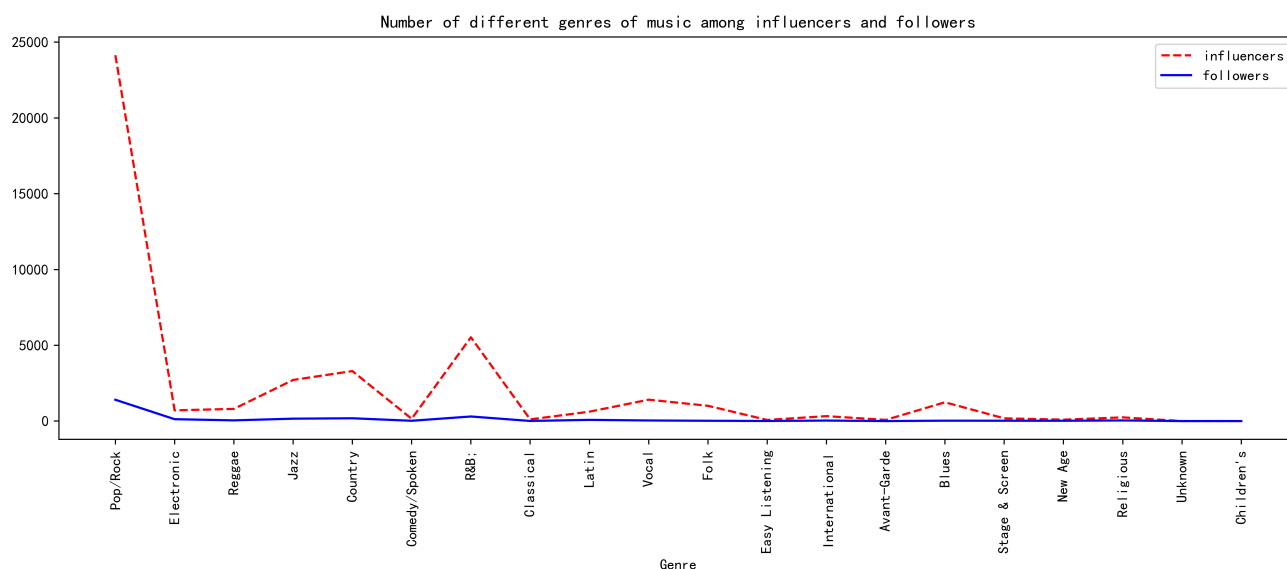
通过使用Influence\_Data数据集创建了音乐影响力的定向网络，其中影响者和追随者为点，由影响者连向追随者构成有向图。通过对网络中不同领域的艺术家进行统计，不同领域艺术家人数如图所示：



通过折线图和饼状图两种形式展现各流派中艺术家的数量可以达到更好的效果。为探究出影响者与追随者之间的联系以及影响者的影响大小，本团队在此提出三种假设以求得到相应的结论。

### 假设一：在各流派间追随者的数量会受影响者的数量的影响

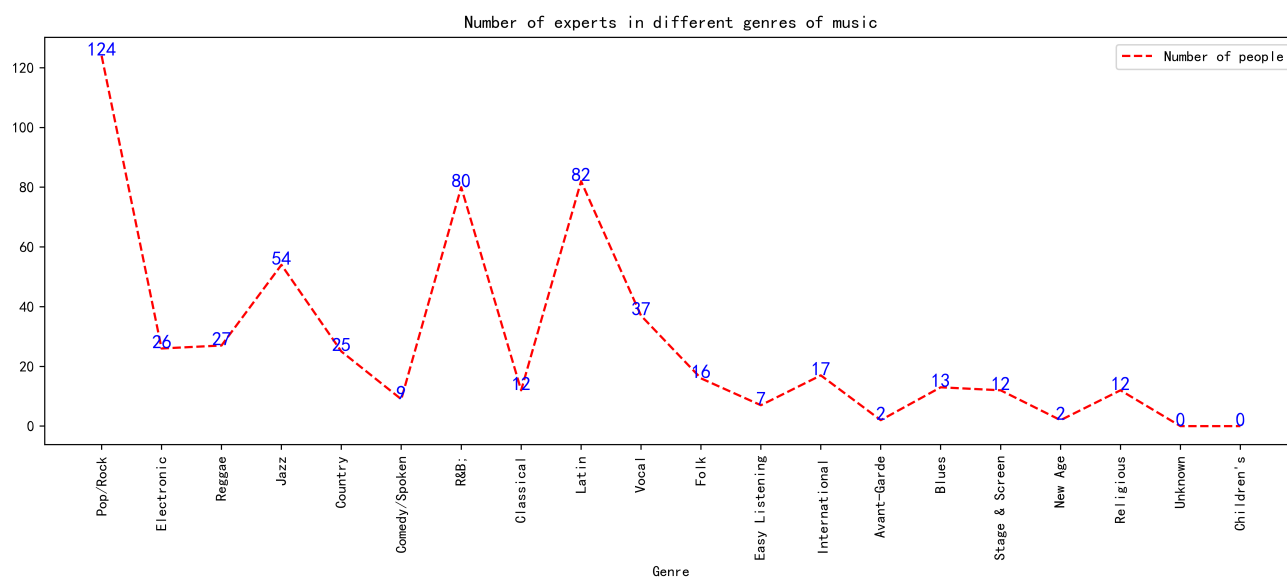
我们对各流派中影响者和追随者的数量进行统计，主要数据如图所示：

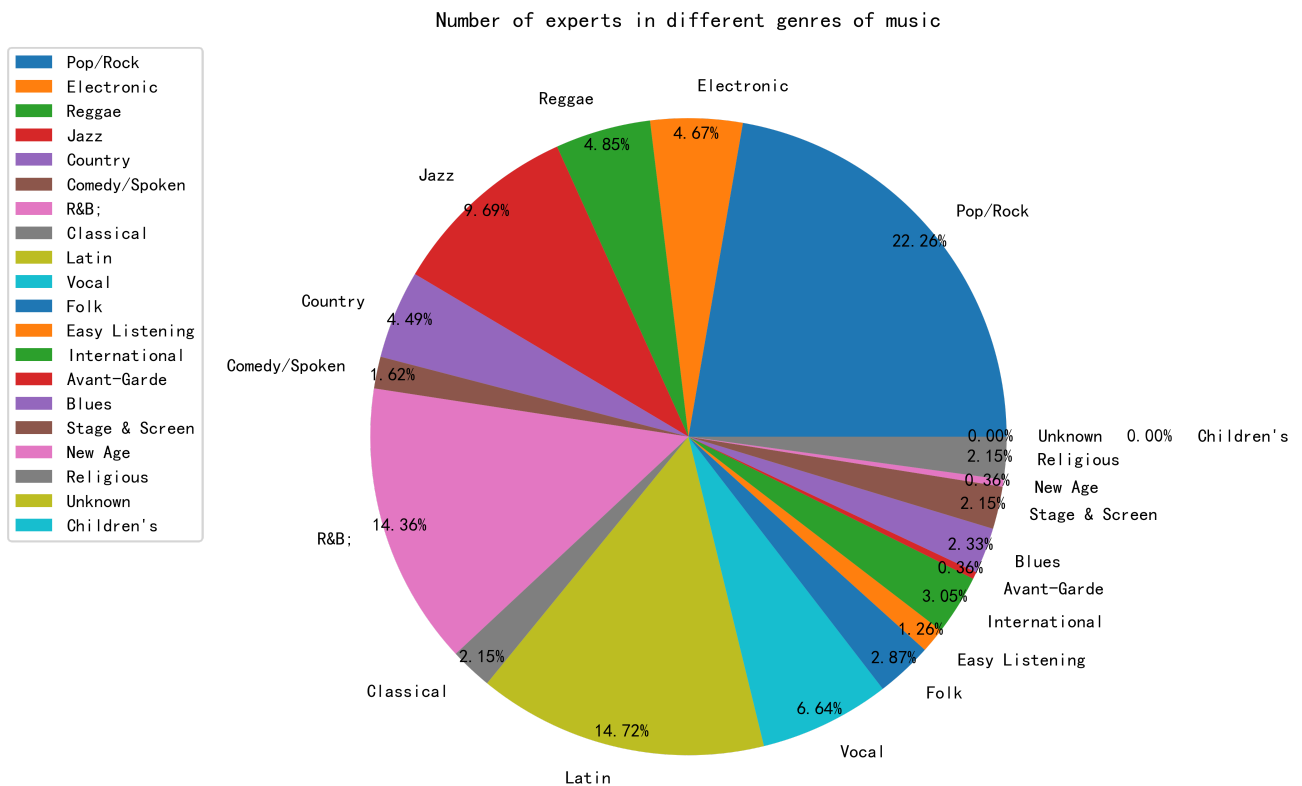


可以发现的是，随者流派的不同追随者会随者影响者数量的增加而增加，如果将图中同一流派中影响者和追随者数量相加会大于该领域真实的艺术家数量，证明同一艺术家有可能不仅是影响者还是追随者，这就促成了各个流派之间有着相互的影响。

### 假设二：各流派专家级别的艺术家人数和流派中影响者的数量有联系

为证实这一假设，我们通过对建立的网络中的每个点（影响者/追随者）进行出度和入度的统计，当一个点入度为零时，说明该艺术家不受任何人的影响，反而会影响别人。那么我们认为该艺术家可以称之为该流派的专家，具体数据如图所示：

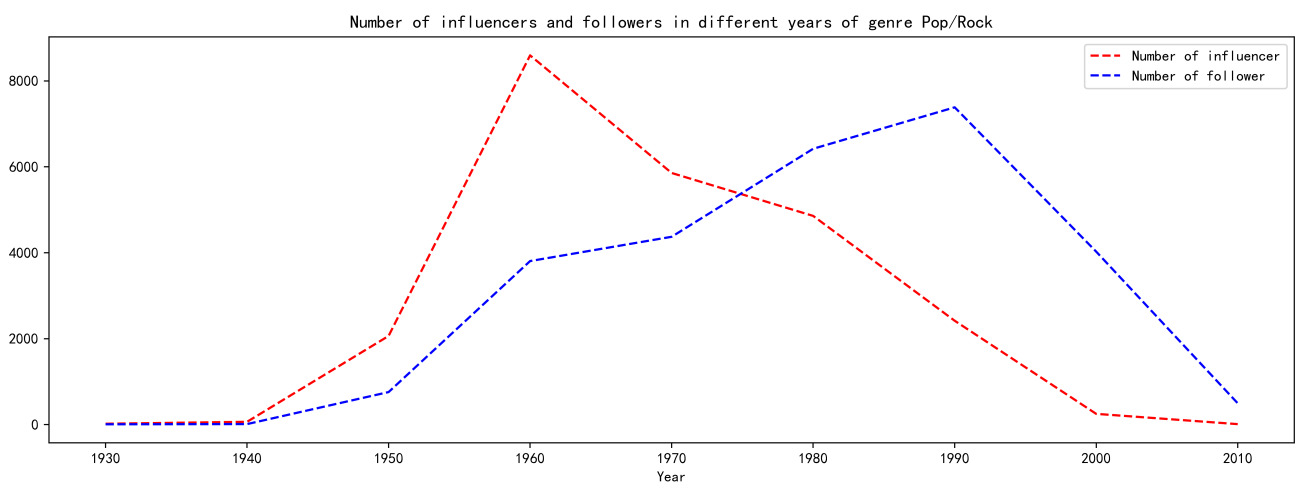




我们同样通过折线图和饼状图两种形式更直观的观察各流派中专家级别的艺术家的数量。有趣的是，随者流派中专家级艺术家数量的增加，该流派的影响者数量也在增加，可见两者之间确实存在一定的联系。

**假设三：**影响者对追随者的影响大小也会受到不同的年份因素的影响

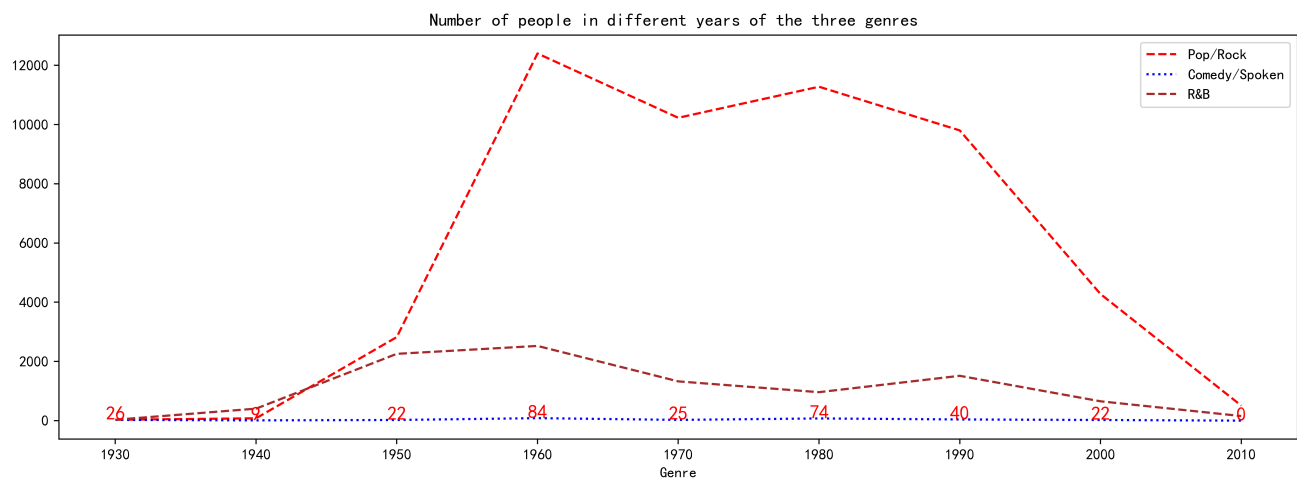
为证实这一假设，我们首先随机挑选了 Pop/Rock 流派，统计它不同年份中影响者和跟随者的人数，具体数据如图所示：



有趣的是，我们从图中可以看出随着年份的不同，影响者和跟随者大多数情况下都呈相同的上升或下降趋势。说明随着年份的不同，同意流派的影响者和跟随着之间存在着相互影响的作用。

## 2. 建立层次分析模型

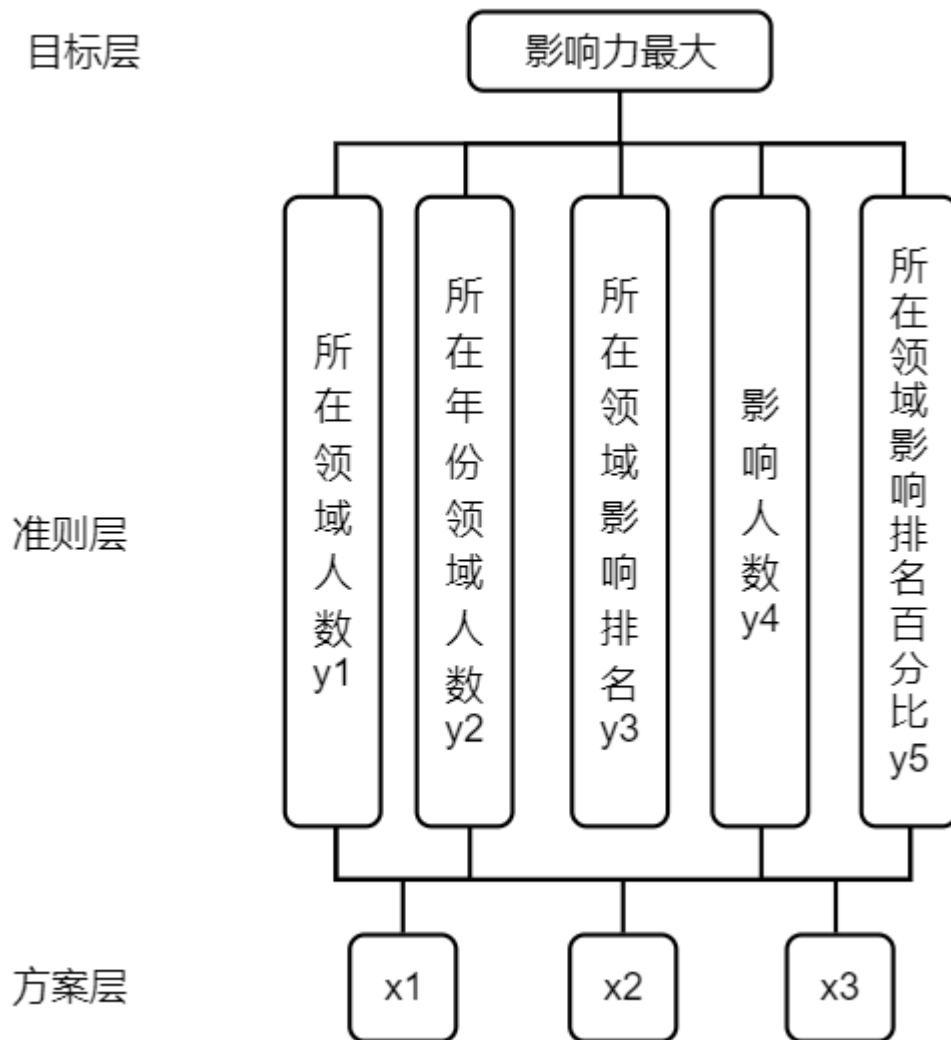
为了探究什么样的艺术家影响力最大，本团队从所有流派中挑选了 Pop/Rock、Comedy/Spoken 和 R&B; 三个流派作为研究对象，首先根据influence\_data数据集进行分析后得到三种流派随者在不同年份的影响者和跟随者的人数，具体情况如图所示：



我们通过建立层次分析模型对三种流派的影响力进行深入分析

在 Pop/Rock、Comedy/Spoken 和 R&B 三个派系中挑选三位歌手(x1, x2, x3)通过层次分析法建立层次分析模型以求 music\_fluence 指标的含义

通过对三位艺术家 x1, x2, x3, 比较 所在领域人数、所在年份领域的人数、所在领域影响力排名、影响人数和所在领域排名百分比 五个标准来挑选出影响力最大的一位歌手，构成如下层次分析模型



### 构造成对比较矩阵

在比较第*i*个元素与第*j*个元素相对上一层某个因素的重要性时，使用数量化的相对权重 $a_{ij}$ 来描述

设共有*n*个元素参与比较，则  $A = (a_{ij})_{n \times n}$  称为成对比较矩阵

成对比较矩阵中  $a_{ij}$  的取值可参考 Satty 的提议，按下述标度进行赋值。 $a_{ij}$  在 1-9 及其倒数中间取值。

- $a_{ij} = 1$ , 元素 *i* 与元素 *j* 对上一层次因素的重要性相同；
- $a_{ij} = 3$ , 元素 *i* 比元素 *j* 略重要；
- $a_{ij} = 5$ , 元素 *i* 比元素 *j* 重要；
- $a_{ij} = 7$ , 元素 *i* 比元素 *j* 重要得多；
- $a_{ij} = 9$ , 元素 *i* 比元素 *j* 的极其重要；
- $a_{ij} = 2n$ ,  $n = 1, 2, 3, 4$ , 元素 *i* 与 *j* 的重要性介于  $a_{ij} = 2n - 1$  与  $a_{ij} = 2n + 1$  之间；
- $a_{ij} = \frac{1}{n}$ ,  $n = 1, 2, \dots, 9$ , 当且仅当  $a_{ji} = n$

成对比较矩阵的特点：  $a_{ij} > 0$ ,  $a_{ij} = 1$ ,  $a_{ij} = \frac{1}{a_{ji}}$ , 当  $i = j$  时,  $a_{ij} = 1$

挑选影响力最大的五个标准：  $y_1$ ,  $y_2$ ,  $y_3$ ,  $y_4$ ,  $y_5$ . 本人使用成对比较法，得到成对比较阵如下：

$$A = \begin{pmatrix} 1 & 2 & 7 & 5 & 5 \\ \frac{1}{2} & 1 & 4 & 3 & 3 \\ \frac{1}{7} & \frac{1}{4} & 1 & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{5} & \frac{1}{3} & 2 & 1 & 1 \\ \frac{1}{5} & \frac{1}{3} & 3 & 1 & 1 \end{pmatrix}$$

$a_{14} = 1$ ，即 所在领域人数和影响人数之比为1，代表本人认为两个标准同样重要

通过计算该矩阵的特征向量  $\lambda = [0.47439499 \ 0.26228108 \ 0.0544921 \ 0.09853357 \ 0.1102982]$ ( $\lambda$  的各项相加和为 1)

### 做一致性检验

从理论上分析得到：如果A是完全一致的成对比较矩阵，应该有  $a_{ij}a_{jk} = a_{ik}, i \leq j, k \leq n$

但实际上在构造成对比较矩阵时要求满足上述众多等式是不可能的。因此退而要求成对比较矩阵有一定的一致性，即可以允许成对比较矩阵存在一定程度的不一致性。

由分析可知，对完全一致的成对比较矩阵，其绝对值最大的特征值等于该矩阵的维数。对成对比较矩阵的一致性要求，转化为要求：的绝对值最大的特征值和该矩阵的维数相差不大。

检验成对比较矩阵A一致性的步骤如下：

- 计算衡量一个成对比较矩阵A ( $n > 1$  阶方阵) 不一致程度的指标CI：

$$CI = \frac{\lambda_{\max}(A) - n}{n - 1}$$

RI 是这样得到的:对于固定的 n,随机构造成对比较阵A,其中  $a_{ij}$  是从  $1, 2, \dots, 9, 1/2, 1/3, \dots, 1/9$  中随机抽取的.这样的A是不一致的,取充分大的子样得到A的最大特征值的平均值

n	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

其他参数计算方法：

$$\lambda_{\max} = \frac{\sum (AW)_i}{nW_i}$$

$$AW = \lambda * A$$

三位艺术家的信息如下：

ID	姓名	所在领域人数	所在年份领域的人数	所在领域影响力排名	影响人数	所在领域排名百分比
295415 (Comedy/Spoken)	Richard Pryor	46	84	3	10	6.5217%
622805(R&B;)	Stevie Wonder	677	2523	3	138	0.4431%
66915 (Pop/Rock)	Bob Dylan	2808	12401	2	389	0.0712%

先对所在领域人数进行比较，得到对比较阵

$$B_1 = \begin{pmatrix} 1 & \frac{1}{3} & \frac{1}{8} \\ 3 & 1 & \frac{1}{3} \\ 8 & 3 & 1 \end{pmatrix}$$

经计算后  $B_1$  的特征向量为  $[0.08199023 \ 0.23644689 \ 0.68156288]$

$$\lambda_{\max}(B_1) = 3.002, CI = 0.001, \frac{CI}{RI} = \frac{0.001}{0.58} < 0.1$$

故  $B_1$  的不一致程度可接受。 $\omega \times 1(Y)$  可以直观地视为各候选人在 **所在领域人数** 方面的得分。

分别比较三位的 **所在年份领域的人数**、**所在领域影响力排名**、**影响人数** 和 **所在领域排名百分比** 得成对比较阵

$$B_2 = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{5} \\ 2 & 1 & \frac{1}{2} \\ 5 & 2 & 1 \end{pmatrix}$$

$$B_3 = \begin{pmatrix} 1 & 1 & \frac{1}{3} \\ 1 & 1 & \frac{1}{3} \\ 3 & 3 & 1 \end{pmatrix}$$

$$B_4 = \begin{pmatrix} 1 & \frac{1}{3} & \frac{1}{4} \\ 3 & 1 & 1 \\ 4 & 1 & 1 \end{pmatrix}$$

$$B_5 = \begin{pmatrix} 1 & \frac{1}{3} & \frac{1}{4} \\ 3 & 1 & 1 \\ 4 & 1 & 1 \end{pmatrix}$$

可得5个特征向量

$$\lambda_1 = [0.081990230.236446890.68156288]$$

$$\lambda_2 = [0.594887960.276610640.1285014]$$

$$\lambda_3 = [0.428571430.428571430.14285714]$$

$$\lambda_4 = [0.632748540.192397660.1748538]$$

$$\lambda_5 = [0.632748540.192397660.1748538]$$

$$\text{得分 } S_i = \lambda_i * \lambda$$

经计算后三人得分分别为

$$S_1 = 0.35041541005496113$$

$$S_2 = 0.2482514719859697$$

$$S_3 = 0.4013331279590692$$

所以 Bob Dylan 的影响力更大