Q4

1. Indicate whether the similarity data, as reported in the data\_influence data set, suggest

that the identified influencers in fact influence the respective artists. Do the ‘influencers’ actually affect the music created by the followers?

To some extent，the influencers identified by influence\_data.csv in fact influence the respective artists’s music. We assume that influence can be measured by the similarity of song characteristics, like structure, rhythm or lyrics. Therefore, we compare the similarity of song features among artists with the influence of artists in Model 1(我不知道你在论文里把第一个模型称作什么，这里需要改一下). If the similarity between artist A and B is higher than it between A and C, A’s impact on B will be stronger than it on C. Or it does not conform to such a relationship. Then it shows that the influencers does not actually affect the music created by the followers.

According to the data set, the influence and song similarity between artists in the same genre is not highly related. While the influence and song similarity between artists from different genres has closer relationship. It should be because the similarity of music in the same genre is higher than is between genres, the influence between artists in the same genre is not obvious. However, influence from other genre would lead to obvious changes in the characteristics of songs of follower. In return, the follower’s characteristics of music tends to influencer’s.

主要思路是认为 影响可以通过歌曲特征(如结构、节奏或歌词)之间的相似程度来衡量，所以我们通过前面所提取出来的艺术家间歌曲特征相似度的表，再结合Influence model 里各个艺术间之间的影响力大小来对比，我们假设，若两个艺术家之间歌曲特征相似度高，那么相应的他们之间的影响或被影响关系就相对更强，若不符合这样的关系，即由influence model得到的影响力大的两个artist间歌曲特征相似度低，那么说明do not actually affect the music created by the followers。

根据所得到的数据(inf\_cor.csv,inf\_cor\_differe.csv，10\_inf\_cor.csv)发现，同一个流派内部，影响力与歌曲相似性关系不高；而不同流派间影响力与歌曲相似性关系更加密切，大体符合同向增长的特点。这应该是因为同一流派间歌曲本身相似度会更高一些，所以互相之间的影响没那么明显。而不同流派间的影响会导致follower歌曲特征有较为明显的改变，明显会趋向于influencer的歌曲特征。

We select artists, who has great influence(value>100，此处看模型中的记号进行修改) and are influenced by more artists(indegree>30), as representative to analyze.

我们选取了影响力较大(value>100)并且入度较高(indegree>30)的结点来作为代表分析。

↓这边应该是表头

对66915 id的artist（”Bob Dylan”）：其影响者如下（对同一genre内的数据按influence升序排列）

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| id | **influence** | **correlation** | **genre** | |
| **Sonny Terry** | 0.419970162 | 0.564698856 | Blues | |
| **Mississippi Sheiks** | 0.766864316 | 0.619248512 | Blues | |
| **Blind Boy Fuller** | 1.082601149 | 0.809011495 | Blues | |
| **Hank Snow** | 1.75092982 | 0.555799846 | Country | |
| **Johnny Cash** | 2.836068086 | 0.546448631 | Country | |
| **Bill Monroe** | 3.032812353 | 0.630900184 | Country | |
| **Hank Williams** | 4.99948725 | 0.78083013 | Country | |
| **Cisco Houston** | 0.207653621 | 0.699211838 | Folk | |
| **John Jacob Niles** | 0.268171062 | 0.7069571 | Folk | |
| **Doc Watson** | 0.270404483 | 0.762647247 | Folk | |
| **Odetta** | 0.275298517 | 0.818197602 | Folk | |
| **Alan Lomax** | 0.275743036 | 0.786005169 | Folk | |
| **Ewan MacColl** | 0.285640008 | 0.880937936 | Folk | |
| **Billy Lee Riley** | 1.007297408 | 0.137111484 | Pop/Rock | | |
| **Buddy Holly** | 7.079100426 | 0.238907941 | Pop/Rock | | |
| **Little Richard** | 8.94174187 | 0.323498559 | Pop/Rock |
| **Elvis Presley** | 9.740188266 | 0.677291639 | Pop/Rock |

我们可以发现，在固定genre这一变量时，随着A对B的influence增大，A与B歌曲间的相关系数变高，这也证实了确实影响到了相关的follower的音乐。

We find out that when the variable gene is fixed, as A’s impact on B increases, the correlation between A and B becomes higher. Therefore, it confirms that Bob Dylan actually influence his follower’s music.

However, data became more complex, we also found some anomalies. For example, there is no obvious relationship between influence and correlation of artist in Pop/Rock. The reason might be that Pop/Rock is a big genre itself. From the perspective of music system, Pop music is developed on the basis of American popular music such as Blues, Jazz, Rock and R & B. For instance, a large proportion of Michael Jackson’s(id=467203,main\_genre= R&B;) songs can be classified as Pop/Rock.

但是，在数据分析变多时，我们也发现一些反常，比如在所给数据里流派占比最多的PoP/Rock流派内（follower id=884686，对同一genre内的数据按influence升序排列，仅列出部分数据）

|  |  |  |  |
| --- | --- | --- | --- |
|  | **influence** | **correlation** | **genre** |
| **Phoebe Snow** | 0.599773244 | 0.858490347 | Pop/Rock |
| **Annie Lennox** | 0.681619165 | -0.23855874 | Pop/Rock |
| **Sarah McLachlan** | 0.75203365 | 0.76736194 | Pop/Rock |
| **Rickie Lee Jones** | 0.843534107 | 0.820600617 | Pop/Rock |
| **Bonnie Raitt** | 0.870107477 | 0.934304358 | Pop/Rock |
| **Carole King** | 2.980210576 | 0.916031613 | Pop/Rock |
| **Elton John** | 5.561115162 | -0.114147621 | Pop/Rock |
| **Joni Mitchell** | 6.480264695 | 0.943932196 | Pop/Rock |
| **Johnny Mercer** | 0.589335522 | 0.919215283 | Vocal |
| **Nina Simone** | 0.771161653 | 0.834338186 | Vocal |
| **Dean Martin** | 1.116535633 | 0.918034292 | Vocal |
| **Sarah Vaughan** | 1.388508823 | 0.84865684 | Vocal |
| **Billie Holiday** | 7.066089922 | 0.990610378 | Vocal |

我们可以发现PoP/Rock流派内的作家相互之间的影响大小与歌曲特征相似度间没有明显的关系了，波动较大。比如influence model里显示id=796734对id=884686影响力比较大，但两者间歌曲的相似程度较小，correlation=- 0.114147621。

分析原因可能是因为PoP/Rock流派本身是一个比较大的流派，从音乐体系看，流行音乐是在Blues、jazz、Rock、R&B等美国大众音乐架构基础上发展起来的音乐，比如Michael Jackson(id=467203,main\_genre= R&B;),但其很多歌曲可以被归为PoP/Rock。所以，可能同一流派内的歌手歌曲相似程度反而不大，也就是说尽管influence model显示A对B影响力大，但可能事实上A并未真的影响到B，两者创作的音乐上还是存在很大差距。

In addition, there is no obvious relationship between the influence of Vocal artists on Pop/Rock artists and the correlation. However, the similarity between them are generally high. This may be related to the vague positioning of Vocal singers, refers to the singer who takes his own singing talent and singing skills as the selling point. So in fact, it is more appropriate to Pop/Rock genre, and the overall similarity between is very high.

Vocal流派作家对Pop/Rock流派作家的影响力与他们之间歌曲特征的相似程度也没有明显的关系，但歌曲相似程度都普遍偏高。这可能与Vocal流派歌手的定位模糊有关，Vocal系歌手就是指以自身的歌唱天赋及歌唱技巧为卖点的歌手，所以其实本质上和大众化的PoP/Rock流派比较贴切，整体相似度都很高。

2. Are some music characteristics more ‘contagious’ than others, or do they all have similar roles in influencing a particular artist’s music?

主要思路：对一个有名的艺术家进行探讨，在不同流派里选取对其影响力较大的几个influencer，针对不同的指标，查看影响者与被影响者之间指标的相似程度，做出归一化数据的指标的雷达图，如果间隔远，说明这个指标影响小，如果间隔大，说明这个指标more ‘contagious’

We select Bob Dylan to analyze. Bill Monroe(id=81083 , genre=Country), Hank Williams(id=549797,genre=Country),Woody Guthrie(id=577531,genre=Folk) are Bob Dylan’s influencers. From the radar map, if attribute A is closer than B, it shows A is more ‘contagious’ than others. If we set 0.1 as threshold, we can find that loudness, tempo, and danceability are more ‘contagious’ .

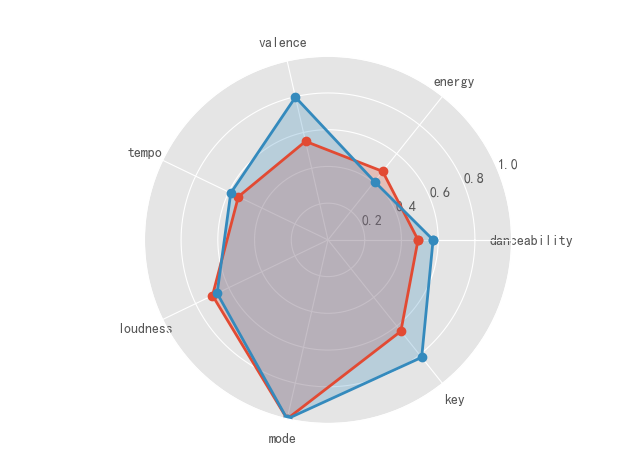
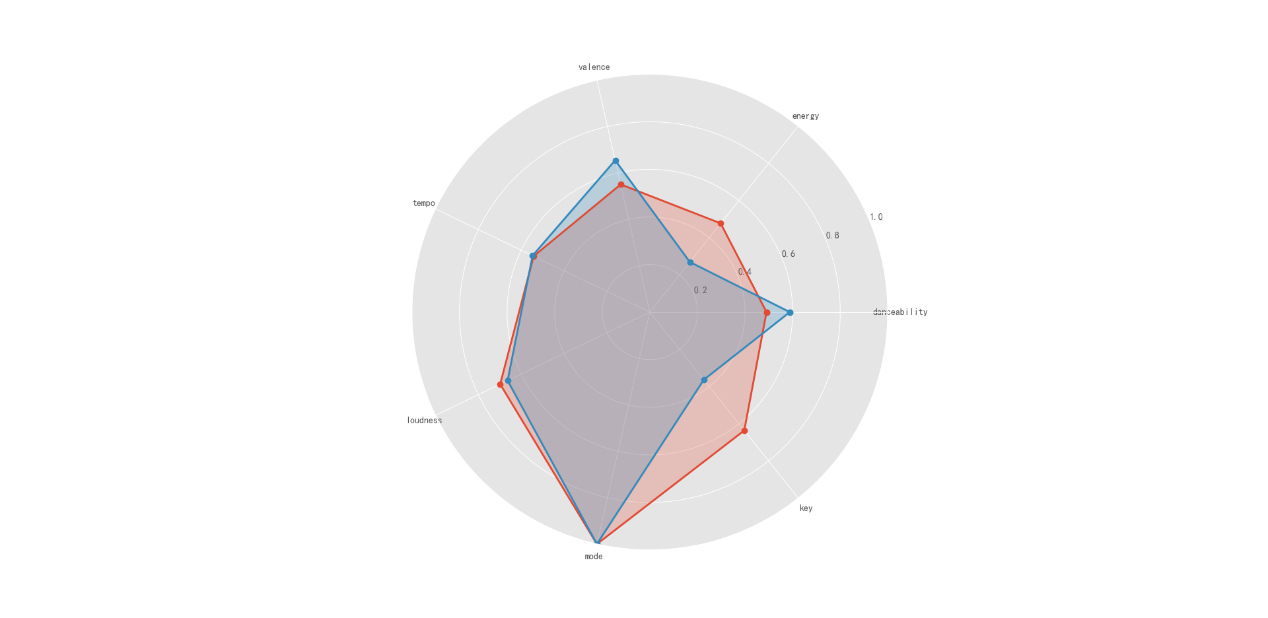
对66915 id的artist（”Bob Dylan”）

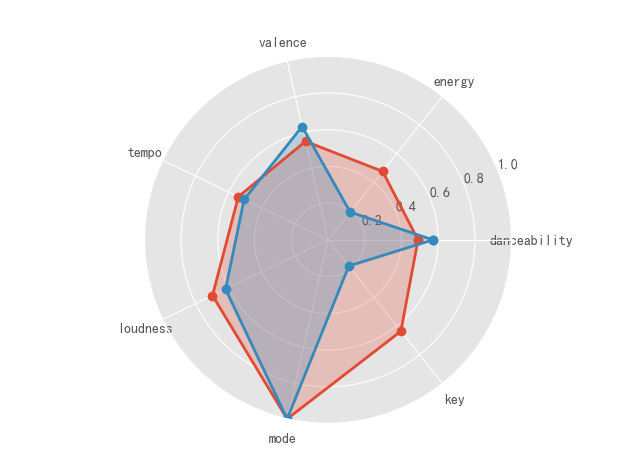
id=81083 ,genre=Country

id=549797,genre=Country

id=577531,genre=Folk

(从左到右, 从上到下的id号及genre)





可以发现，若取0.1为差距的阈值，那么loudness, tempo,danceability这几个指标more ‘contagious’ 。而相对而言，energy，valence，key等差异较大，说明这几个特征的影响较低。