**项目地址**

[**https://github.com/haoyang1166/poem2poetry**](https://github.com/haoyang1166/poem2poetry)

**（没有模型，没有中间生成文件，没有cbdb数据库（上传大小限制））**

**内容的直观理解如下所示，在做ppt和论文过程中，如果想放个图（把图片的位置空出来，写上描述（需要什么图这种），等说明文档出来之后或者之后再找我要图）**

唐诗诗人社交网络分析

基本思路就是首先把唐诗爬出来，之后用到一个CBDB数据库去查找诗人的别名，之后看其他诗人的诗中或者文章中有无相应的单词，之后根据这些权重构建社交关系网络。最后的结果具有正确性

唐诗篇章语义相似度计算

基本思路就是用gensim的包里面的词向量模型，把分词后的东西放入word2vec中训练，得到每个词的词向量，一种方法是词向量加和作为这首诗的向量，另一种是先把向量合并成矩阵，之后取每一维度最大的，最后当成这首诗的向量，暂且认为具有一定的语义信息，之后可以计算两个诗人的写作风格语义的相似度。发现社交网络分析中相近的诗人，相似度也比较高。

唐诗文本挖掘

就是先分分词thulac包，之后一些地点可以根据分词的词性选择出来，查看诗人愿意写哪些词性的词等等，统计统计词频，看出诗人的写作愿意用啥词，比如诗人喜欢用春夏秋冬哪个字更多一些就，更喜欢用哪种动物写诗等等，

诗人主题分析的一些东西

首先就是利用authortopic主题模型，输入一个（诗人，唐诗）预料，输出诗人的主题向量，同时还能生成20个主题大概，之后根据每个诗人的主题，再用t-sne降到2维算法可视化出来.

诗人地理位置挖掘

首先通过百度百科爬取诗人的生平信息，之后通过哈工大pyltp包进行分词+命名实体识别（自动识别地名），识别出去过古代的地名，之后通过一张古今地名映射表映射到一张地图上，经纬度也是有一张映射表，之后展现出每个诗人的地图（去过哪些位置的）