

新的任务

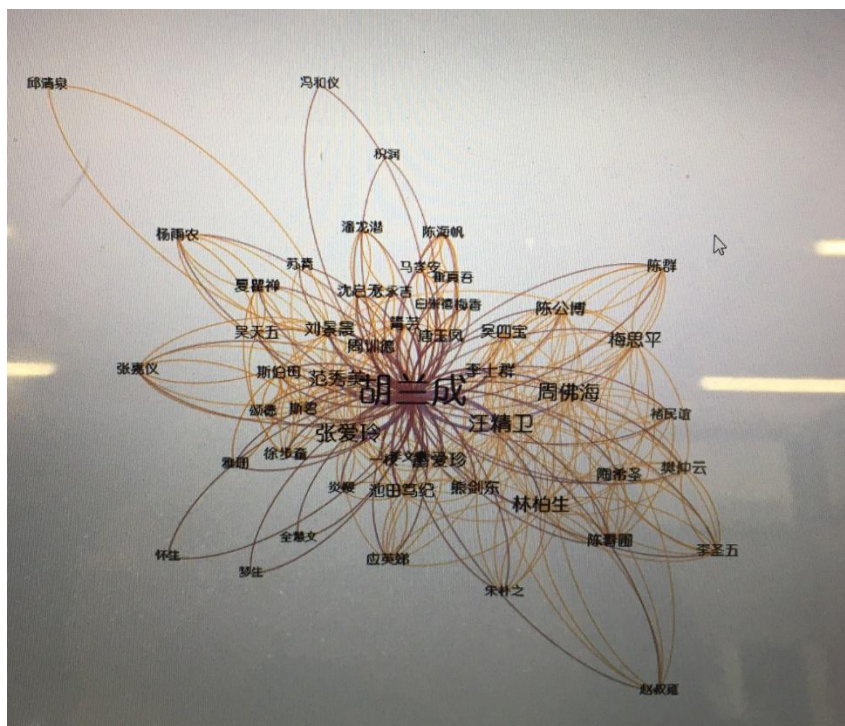
运行环境 Python2.7 这三个图麻烦输出在一个文件夹中。

1. 这回的所有数据是在一个 CSV 文件中。这个文件较大，大概 100mb. 需要预处理 comments 那一列（能否在使用 nltk 包的 stopwords, 之外留出可以自己编辑的 stopwords），然后：

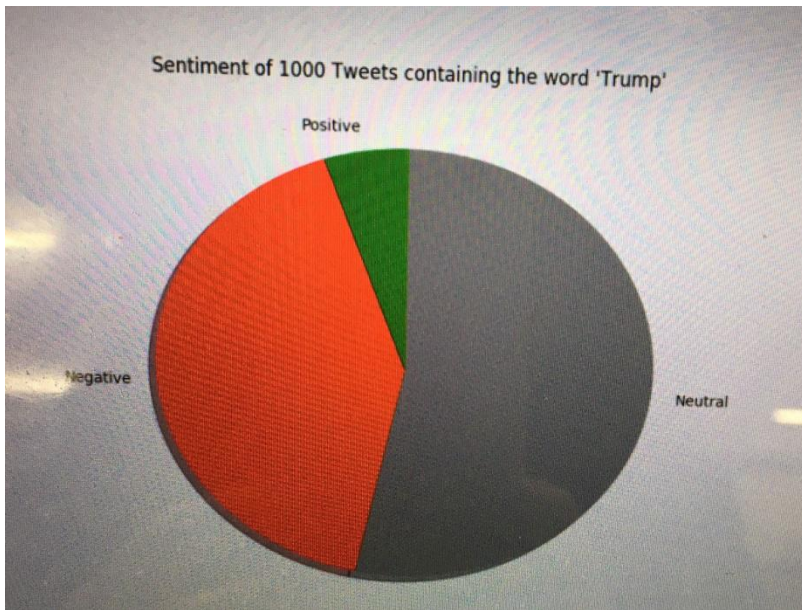
1. 我需要词云图（如图例所示）颜色最好以深色为主，由于数据较多，云图上显示大概 50-70 个词就行。同时输出前 50 的高频词的频数（表格形式）。



2. 词关系图：词关系图的词目前选择出现频率最高的 10-20 个词（由于我们最终出现在这个图中的词还没有确定，能否在代码中显示出哪里可以做这些词改动的地方）。



3.还有情感饼状图”（也是需要正负中情感的分类）。不需要建立分类器，大概用 nltk 中带的词典，或者其他包的词典可以，bag of words 也可以。



4. 麻烦写一下 topic modeling 的代码，还是根据统一数据。Topic 暂定 5 个

这三个图的代码在不在一个代码档中也可以。

如果分开的话，麻烦每一个代码都要提前预处理一下文字部分。