原始数据处理：

1：DataAnalyze.cpp 和GenerateHash.cpp 对所有邮件进行分析，统计每个人与bush的邮件往来次数。

2：splitFile.cpp 和tfidf.cpp使用tf-idf计算每个词项的权重，形成空间向量

3：人工标注300封邮件所属分类：

生成人物关系图谱：

1：运行 pre\_process\_before\_vis.py 生成 画图需要的数据的格式

2: 将步骤1 生成的数据拷贝到1.html 中（由于画图使用的js，然而这个程序只是一段代码，不是一个完整的项目，处于安全性考虑，js机制不能读取本地文件）

3：在浏览器中打开 1.html 即可

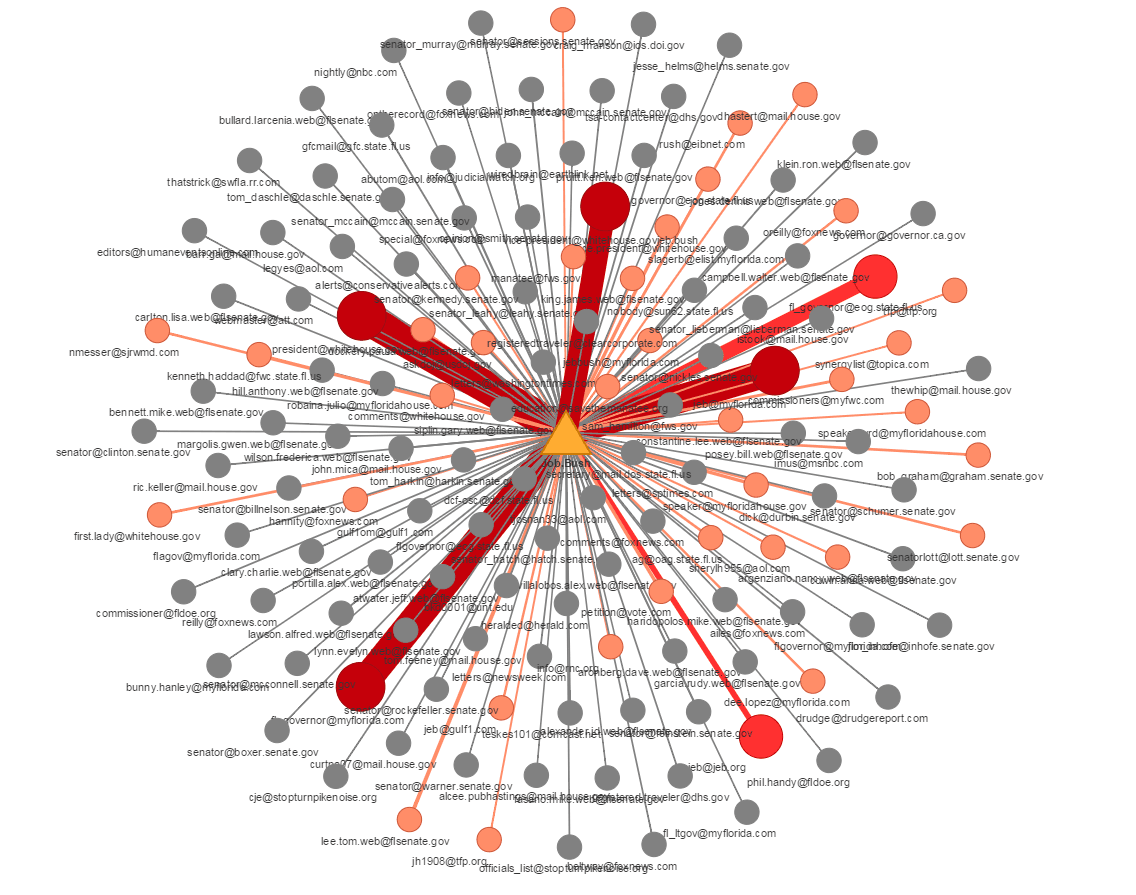
生成兴趣图谱：

1： 利用K-means算法对所有邮件进行聚类分析，初始K 点的选择可以使用人工标注的300封邮件

2：强聚类结果数据传到2.html 中，用浏览器打开即可。

结果展示：

以bush为中心的人物关系图谱，通过点和线段的颜色，大小和宽度标识此人与bush的往来密切程度。



以bush为中心的兴趣图谱如下图

