时间：20160422

地点：北航新主楼G318

与会人：孟翰、苏若、李璇、孙敏芳

第八周

1. 为啥计划安排中只介绍了第七点：

【】因为PPT前面已经介绍全几个配置环境的工作

2. 已将改进性能改变为处理数据？

是否可以加入内核

3. 讲的太粗，是否仅仅依靠距离相近来判断

如何定位地点，是否仅用ID确定具体的精确地点

不同的ID是否都有不同的经纬度？一个ID是一个地点的ID

如果两个人经常使用相近的ID来登录/签到，那就判断他们是朋友？

【】两人经常在不同的地点重合

是否有相应的论文验证这种合理性？可能会有很多是干扰信息，恰巧也有这种特征。仅凭现有数据无法验证命题的正确性

搜索数据时没有设置检验，所以不能确定数据的可靠性

换一种说法，降低检验的难度：某些地点，某些人群可能会趋向于干某一件事？但是这样就要找到这个题点的意义。群体社会行为

原有命题的困难之处：无法检验

【---】学生提问开始【---】

4. spark计算的时候，是使用什么模块/命令计算的

textfile、

5. 搜索的数据大概有多少，有无考虑到真正放到内存中需要多大，集群规模需要多大。一定要想好，否则会出现问题

占用内存一般是数据大小的4-5杯

txt文本格式，到时候实际占用会减小【文本有一定的冗余（比如：一个树在txt中占取多个字节，但是读取进来只占四个字节），实在不行加内存】

针对以上问题，讨论了修改方向，并安排下周的分工