20132430208 黄毅飞 以个人为起点的微博可视化分析

题目个人总感觉有点不合适。

投入时间：day day work，我应该去华东师范做毕设。

题目介绍如下

数据来源：（Python + MongoDB + Scrapy）以某一用户为起点，抓取其自身以及其粉丝的微博信息及个人信息。目标1亿条微博，同时保留用户与粉丝的关联信息 -- 时间估计10天，5天调试github开源的python weibo爬虫例子，5天抓取信息。（示例给了2000W/day条的例子，希望在抓取1亿条信息的时候不会有什么想不到的事情出现）

可视化信息思路(即问题，我想知道什么)：(D3.js + lnmp(linux,nginx,mongoDb,php)) – 30天以上

1. 某一用户按年、月的微博发表图（显示形式待定）。一维信息，变量=每天发表的微博数量，年月为辅助显示。
2. 某一用户黑夜白天微博数量对比图（类似阴阳图），直接判断是不是微博夜猫子。一维信息。
3. 某一微博转发树形图
4. 粉丝关系图
5. 整体微博发表地理位置图（国家地图上以省为基点显示数量，并辅助颜色代表数量）
6. 有时间就加入别的。。

思路来源于：

1. IEEE TVCG的一篇论文Visual Analysis of MOOC Forums with iForum（着重抄袭，他visua mooc forum，我visua 微博。只是关注点不同，不知道能不能达到答辩的要求。）
2. 北大可视化小组IEEE VAST的的D-Map: Visual Analysis of Ego-centric Information Diffusion Patterns in Social Media（find 传播模式以及核心传播者）--未见刊。

主要是想真正走遍可视化的流程，从问题，数据收集，整理，清洗，可视化

--test for youdao update