



中国科学技术大学
University of Science and Technology of China

多媒体通信文献阅读

移动社交大数据——微信数据集、网络应用和机遇

孙圣印

SA19006050



01

背景介绍

02

研究内容

03

未来展望



01

背景介绍

02

研究内容

03

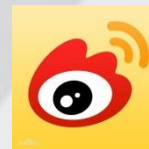
未来展望

○ 社交应用分类

☆ 单向跟帖/单向关注

① 对一个用户已发布的内容，**任何人**都可发表评论

② 发布内容的用户**不必回复**每一个跟帖



☆ 互相关注

① 对一个用户已发布的内容，**好友范围**内可见

② 支持**即时**消息传递



○ 微信特性

☆ 陌生人朋友圈不可访问

- ① 朋友圈转发路径不可见
- ② 陌生人相互浏览朋友圈需要以共同的好友为中介

☆ 朋友圈指定人可见

- ① 朋友圈只显示给指定好友
- ② 私密内容指定好友分享

更紧密的
朋友关系

☆ 群聊

- ① 陌生人间的即时通信方式
- ② 群聊中多媒体信息（文本、视频）链接共享

○研究问题

☆移动蜂窝网络的信息传播

①微信通信 \implies 社会关系 \implies 蜂窝网络中的可靠通信点

②寻找通信的**高质量连接**用户 vs 寻找**有影响力**的朋友圈传播用户

☆骨干网络流量预测

通过数据挖掘，预测用户的离线位置，根据离线人口密度分配流量资源

☆移动人口分布预测

- 分析数据集中的转发行为
- 监控社交网络结构变化
- 预测移动人口分布



01

背景介绍

02

研究内容

03

未来展望



01

背景介绍

02

研究内容

03

未来展望

○ 移动蜂窝网络中的信息传播

① 获取信息扩散图（FIBODATA平台数据）

$\langle U_1, U_2, PID, IP, t \rangle$



信息扩散记录五元组

事件发生时间

用户 U_2 的 IP

朋友圈文章 ID

朋友圈文章浏览者 ID

朋友圈文章发布者 ID

☆ 用户 U_2 以地址 IP 在时刻 t 阅读了用户 U_1 发布的朋友圈文章 PID

○移动蜂窝网络中的信息传播

②确定最有影响力的用户

问题



☆数据集庞大、扩散树巨大

☆微信好友数受限，无法通过好友数确定影响力

策略



☆吉布森抽样技术，抽取部分扩散树

☆投票策略确定影响力

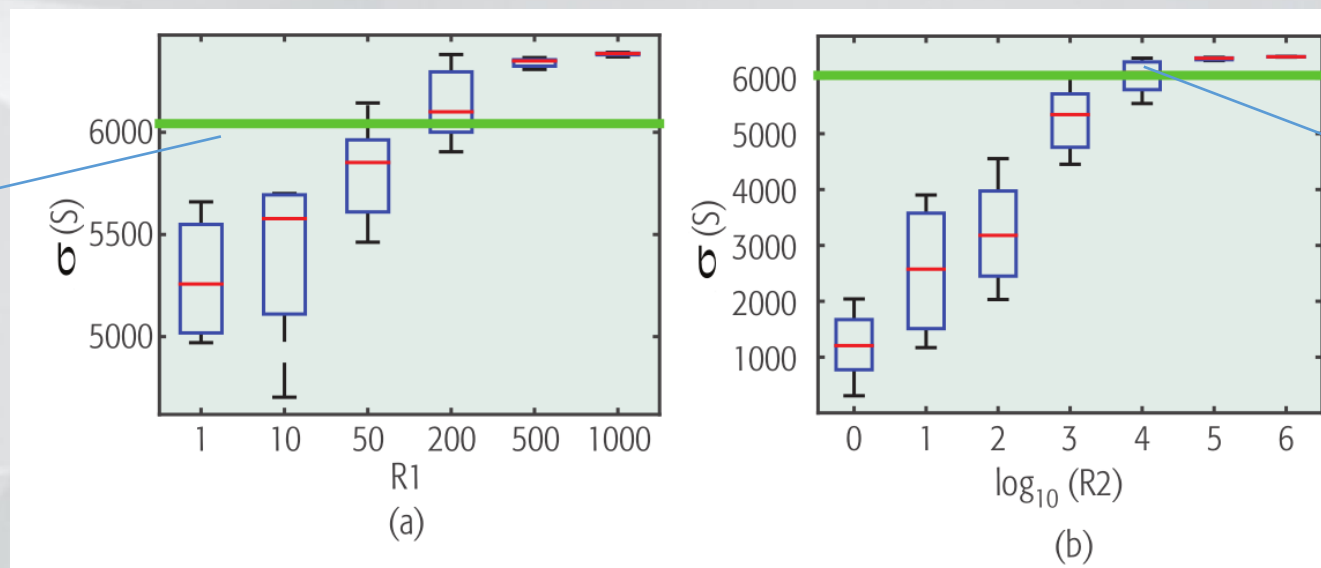
微信文章从父节点传到子节点，子节点投票给父节点



○移动蜂窝网络中的信息传播

③实验结果

贪心算法



投票策略稳定

图1 基于投票策略寻找最有影响力用户

R_1 : 抽取的决策树数目 R_2 :

执行投票次数 K :

需要选定的有影响力用户数目

② 研究内容



○ 骨干网络流量预测

① 从通信模式到流量预测

☆ 通信模式 建模为马尔科夫随机场

☆ 变量：通信距离、通信频率

☆ 预测目标：底层网络流量

② 最优服务器放置

☆ 最优服务器位置和数量，最小化流量负载

☆ 方法：反向贪婪策略

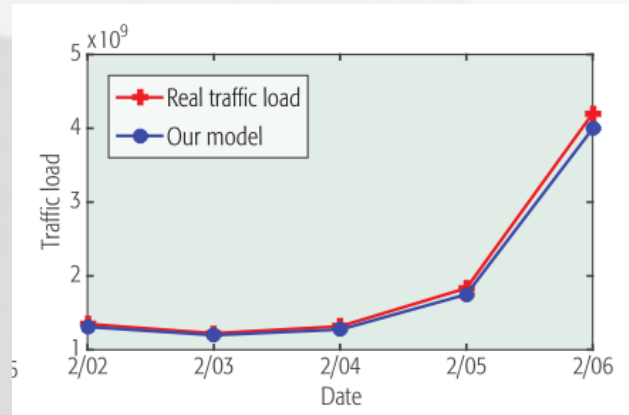


图2 底层网络流量预测

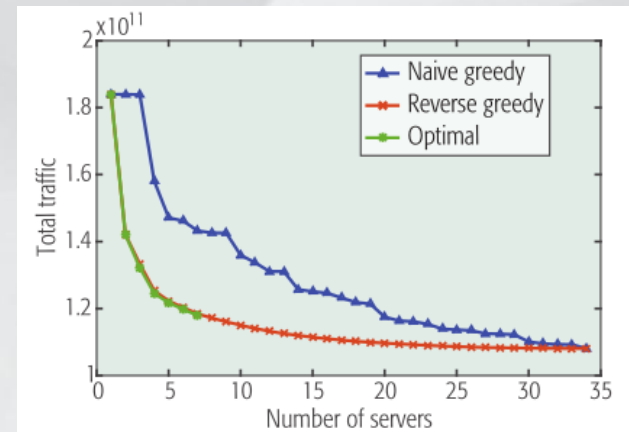


图3 最优服务器放置

○流动人口分布预测

①地理同质性模型

☆同一区域的人更容易交流

(a) 春节前

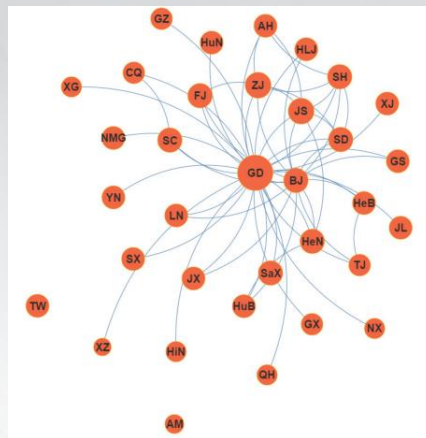
各省之间交流密切
圆圈间连线复杂

(b) 春节时

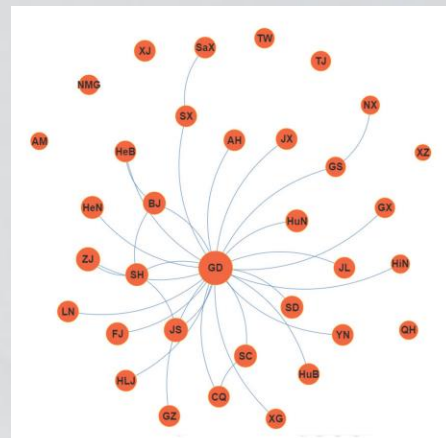
各省之间交流减少
圆圈间连线稀疏

(a) 对照

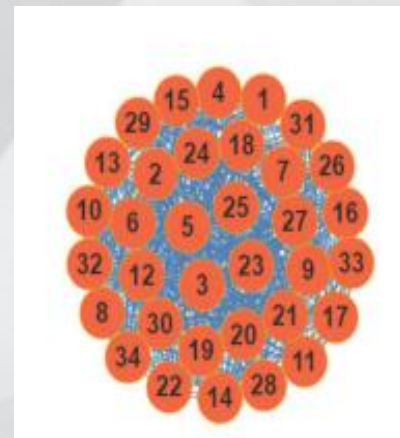
无地理同质性
混乱的结构



(a)



(b)



(c)

图4 地理同质性可视化

②狄利克雷过程推导流动人口分布

② 研究内容



中国科学技术大学
University of Science and Technology of China

○ 流动人口分布预测

③ 实验结果

转发源头区——北
京浏览转发次数最
多

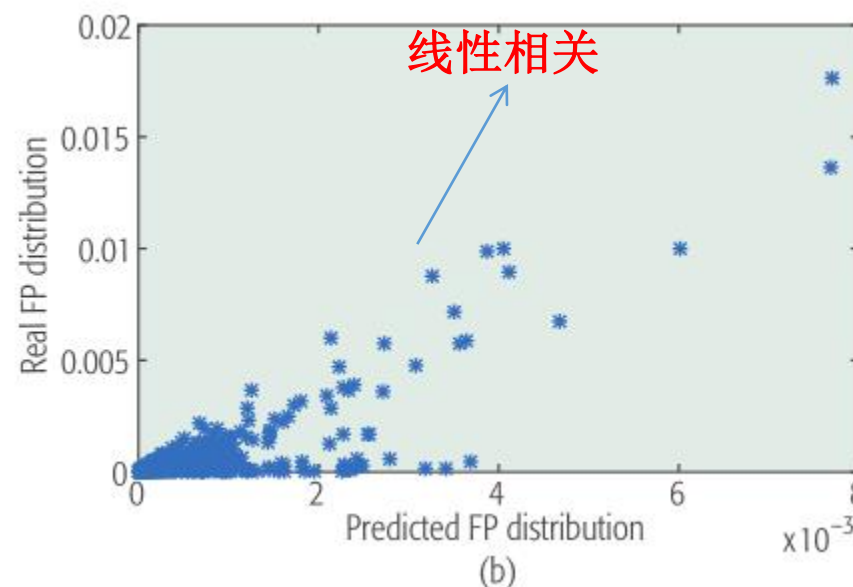
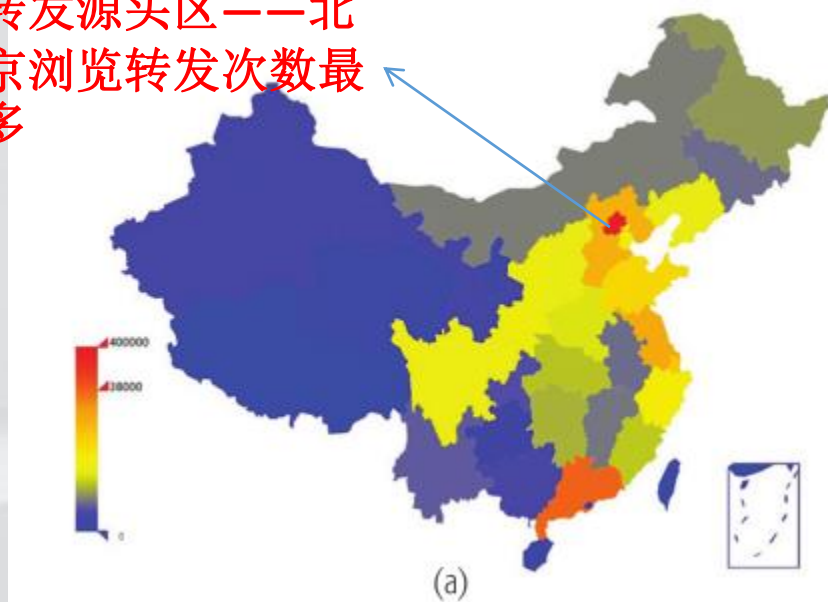


图5 朋友圈文章扩散过程、真实流动人口分布与预测流动人口分布相关性



01

背景介绍

02

研究内容

03

未来展望



01

背景介绍

02

研究内容

03

未来展望

○本文工作

- [1] 移动蜂窝网络信息传播分析
- [2] 流量预测及最佳服务器位置寻找
- [3] 流动人口预测

○潜在研究机会

- | | |
|-------------------|---------------|
| [1] 新的社交网络结构 | [2] 动态扩散图中的营销 |
| [3] 如何更好的理解人口统计结果 | [4] 垃圾邮件检测 |
| [5] 降低隐私泄露的方法 | [6] 线下促销活动策划 |

感谢聆听！
