



校园无线网用户群体 移动行为聚集分析

周昌令 钱群 赵伊秋 尚群
北京大学 zclfly@pku

背景及相关研究

关联关系及稀疏链接区间表示

聚集关系度量(社交网络分析)

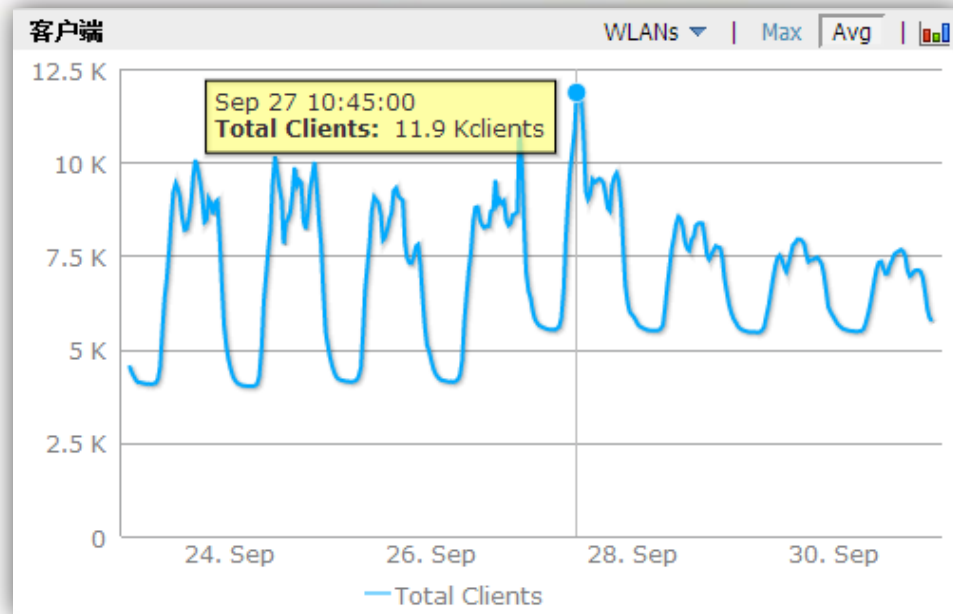
数据集及数据整理

结论与展望

- 无线网络管理的挑战
- 移动行为的相关研究
 - 比较：关联矩阵

无线网络管理的挑战

- 终端数量
 - 例：上个月
(2013/9/1 – 2013/9/30)
总MAC数：**11.6万**
- 移动特性
 - 区别于有线网络
 - 关注终端节点
 - vs 网络设备
 - 变化大，规律复杂



移动行为的相关研究

- 用户行为以及网络性能指标 (Balachandran等[2])
 - 移动性与使用习惯
- 驻留位置 (home location) (KotzD等[3,4])
- 移动模型(郑宇、谢幸等, Ghosh等, Lin, M等 [5-7])
 - 轨迹变化, 位置估计和移动轨迹预测
 - 个体或统计平均的用户移动性
- TRACE框架(Hsu W等[8-10])
 - 用户移动性带来的信息扩散能力
 - 使用关联矩阵 (association matrix)

相关研究 - 关联矩阵

- 关联矩阵

- 每天每用户

- 列：节点位置

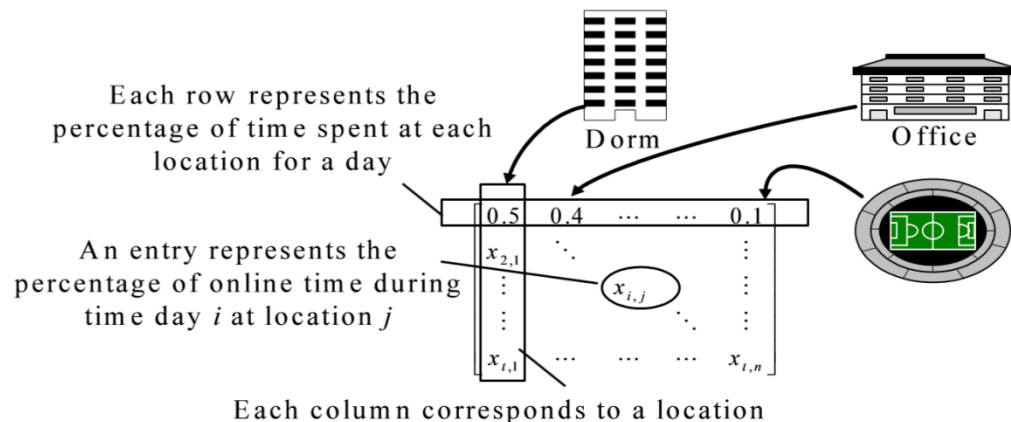
- 行：时间比例

- 不关心时长和顺序

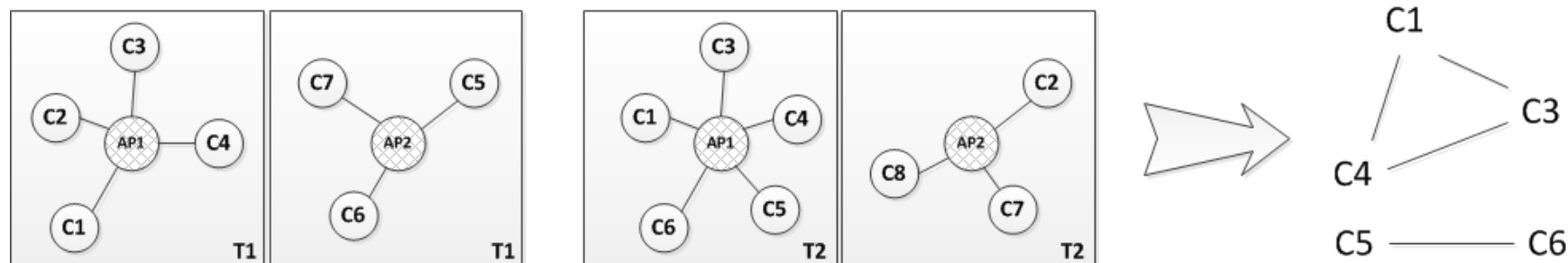
- 比例相同 (10分钟:20分钟 vs 2小时:4小时)

- 先后顺序不同(A->B vs B->A)

- 矩阵规模增长迅速，计算复杂

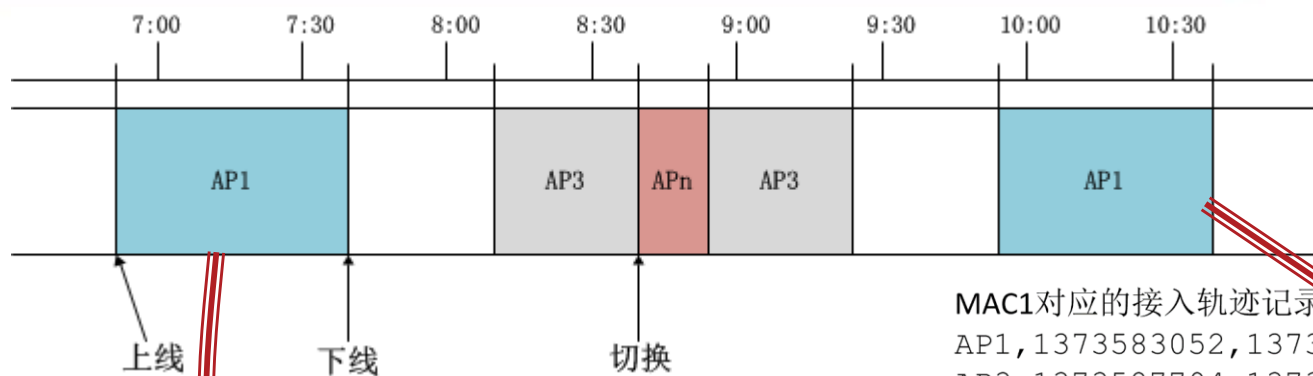


基于位置的关联关系



- 二部图投射 (Bipartite network projection)
- 用户节点与用户节点之间的关联
 - 重合：同一时间段关联到相同位置
 - 重合的时长越多，关联度越大
 - 不同用户之间重合时长分别计算

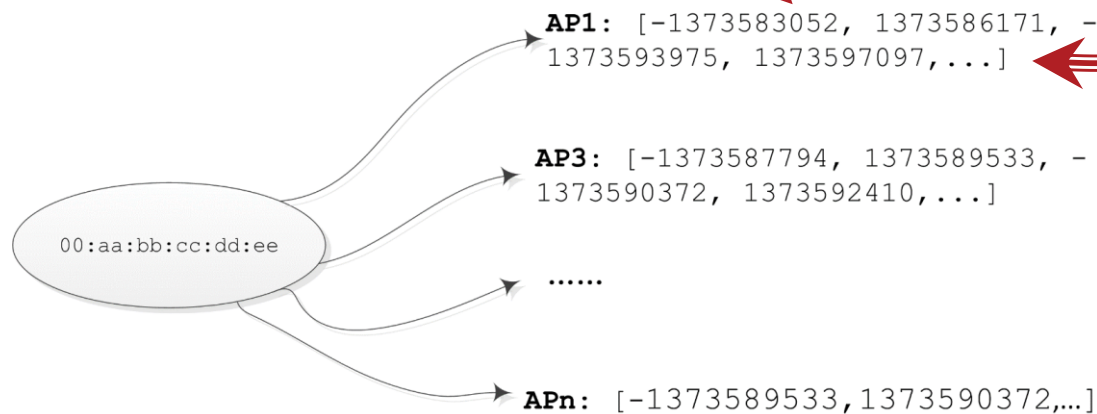
移动轨迹记录



MAC1对应的接入轨迹记录:

AP1, 1373583052, 1373586171
AP3, 1373587794, 1373589533
APn, 1373589533, 1373590372
AP3, 1373590372, 1373592410
AP1, 1373593975, 1373597097

稀疏链接区间—SLI



- SLI(Sparse Linked Intervals)的特点
 - 链接是稀疏的
 - 因为几乎所有的用户都只会接入很小比例的部分AP
 - 每用户节点(MAC)对应一系列记录
 - 按接入AP进行分组，每组的时间按绝对值顺序排列
 - 区间中的起始时间用**负数**表示
 - 计算区间的重合长度
 - 数值求和为区间长度
 - 只需计算用户关联的A P 相关的其它MAC
 - 区间合并，按绝对值排序，连续的负数后面的区间为重合的区间

稀疏链接区间-举例

- 例：
 - 区间 A[-1, 3, -5, 8, -9, 10] 区间 B[-2, 4, -5, 7]
 - 合并后排序: [-1, -2, 3, 4, -5, -5, 7, 8, -9, 10]
 - 相似度计算
 - 对重合的元素求和，此例的长度为3
- 计算复杂度为 $O(n \cdot k \cdot m \cdot \log m)$
 - n 为节点总数， k 为节点平均接入的AP数量， m 为节点的平均记录数量
 - 直接计算 $O(n^2 m^2)$

- 2个MAC相似度定义为

$$sim_{i,j} = \frac{\sum_{ap} len(overlap(I(i), I(j)))}{\sum_{ap} len(I(i))}$$

- $I(i), I(j)$ 分别为2个MAC在不同AP上的接入时间区间。 $overlap()$ 为计算重合区间， $len()$ 为计算区间的长度
- 超过预设值的才认为存在强关联关系
- 此度量值是非对称的

聚集关系度量(社交网络分析)

- 采用社交网络分析 (SNA) 的方法来理解无线终端间关联的关系
 - 用户节点地址对应社交网络图中的节点，存在关联关系的节点间具有边相连
 - 聚集系数、频度中心、间接中心、亲近中心
 - 社交网络图分析软件Gephi(<https://gephi.org/>)
 - 本文取阈值为：在线总时长24 h，关联度不小于50%

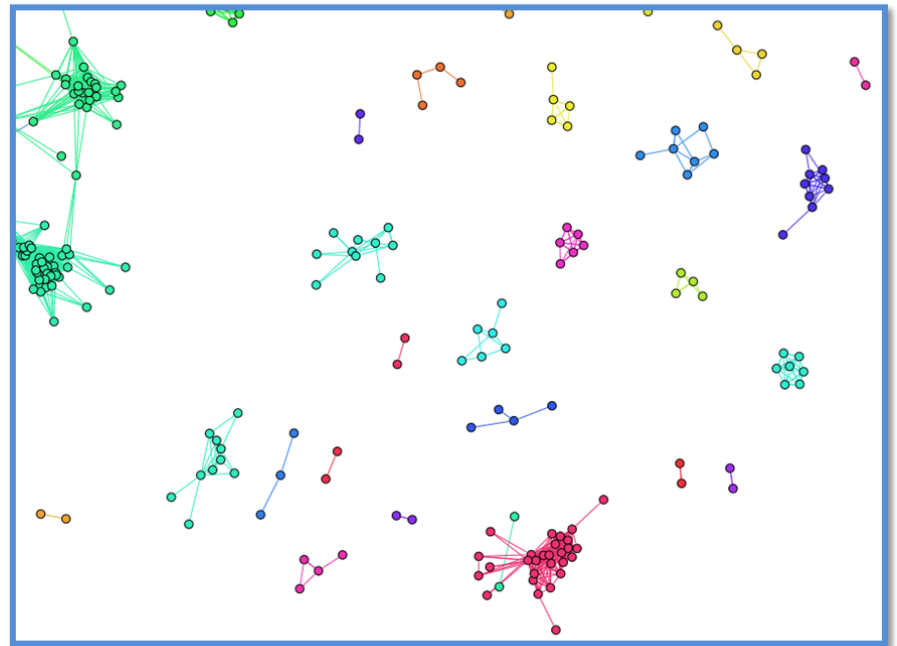
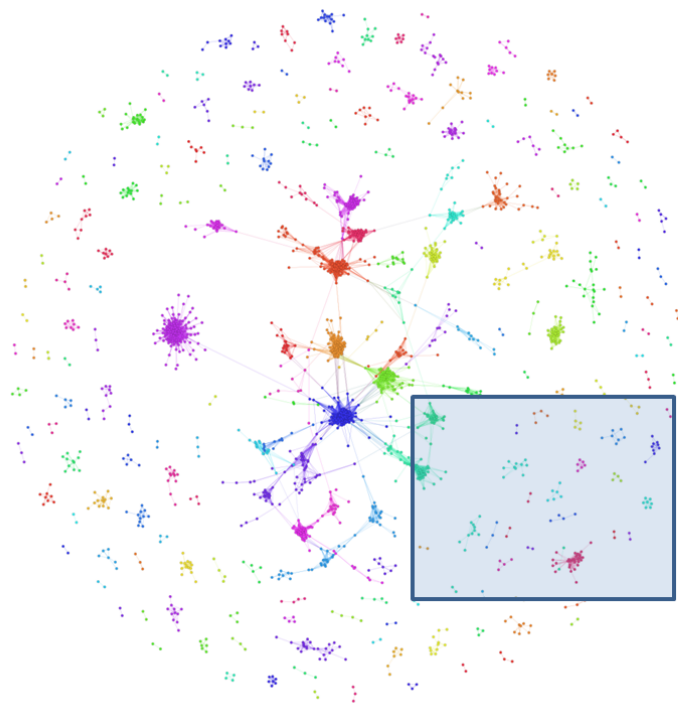
各数据集的统计信息

数据集	AP数量	区域数	MAC总数	强关联节点数	强关联边数	聚集系数	采集时间
北京大学PKU1	1 388	-	52 297	4 060	83 350	0.225	2012.6.1~2012.6.30
北京大学PKU2	1 134	-	37 776	2 072	29 601	0.218	2013.7.15~2013.7.31
南加州大学USC ^[19]	-	137	25 381	20 484	362 368	0.260	2006.1.25~2006.4.28

- PKU采用定时轮询方法，使用SNMP协议从无线控制器获取信息，每10 min取一次
- USC使用syslog日志

- 在线总时长少于1天或出现天数少于2天的节点
 - 不是典型的校园网用户
- 持续在线
 - 终端是一个常驻节点，没有移动行为。另一个原因是AP本身的缓存故障
- 只关心有移动特征的节点
 - 关联超过 2 个 A P

数据集

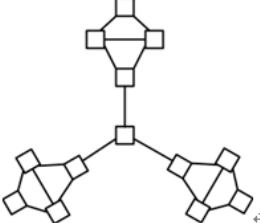
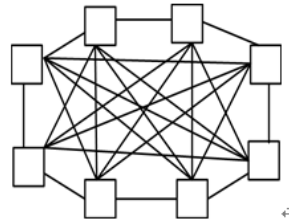
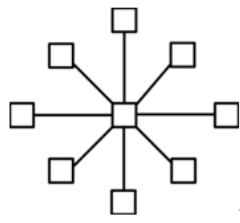
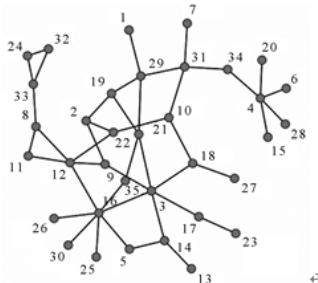
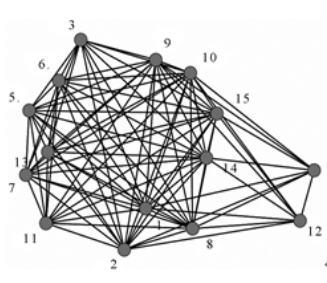
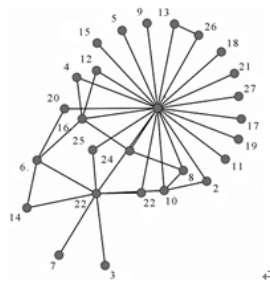


代表子图



表 3

多节点的聚集模式

类别	均衡分散型	关系密集型	中心散射型
子图形状			
代表性的子图			
特征	节点数很多，节点的平均度数不大，3种不同的中心度量指标差异较大。节点的聚集指数较小。	节点数比较多，节点平均度数大，两两节点间几乎都互连，3种中心度量指标值几乎相同。节点的聚集指数较大。	节点数比较多，节点的度数不均衡，有明显的频度中心节点。中心节点度数最大，平均聚集指数较小。
聚集产生的原因	用户节点有明显的移动行为，节点通常出现在多个位置，没有驻留位置或各个节点驻留位置不同。没有固定的出现区域，通常出现在教学区。	用户节点倾向于较少的移动，选择了共同的驻留位置。终端用户通常是同一个实验室的或在学生宿舍区邻近的位置。	中心节点与其他节点行为模式有差别，中心节点在线时间长，并且有驻留位置。而其他节点移动性较大。常出现在图书馆和大的自习室。
对无线网络管理和规划的启示	用户之间关联度相对较弱，位于间接中心的节点往往对多个区域的网络都有体验，无线规划时可以考虑参考它的意见。	用户关联度高，建议网管系统做节点故障诊断时把相关节点也列出。规划时应有针对性地对此类用户的需求进行优化。	频度中心用户是代表性用户，排查故障和做规划应了解此节点的信息和此终端用户的需求。

- 移动终端轨迹的稀疏链接区间 (SLI) 方法能有效地分析用户节点的关联
- 用社交网络分析的方法来帮助分析校园无线网用户的移动行为
- 结合真实运行数据，理解常见的聚集子图模式的特征和产生原因
- 可以应用到无线网络优化、信息推送、协议设计等领域

• 谢谢！

