

## 手机通话网络度特征分析\*

余晓平<sup>1)</sup> 裴韬<sup>2)†</sup>

1) (新疆石河子大学信息科学与技术学院, 石河子 832003)

2) (中国科学院地理科学与资源研究所, 资源与环境信息系统国家重点实验室, 北京 100101)

(2013年5月19日收到; 2013年7月26日收到修改稿)

手机通信数据详细记录了人们的通信行为, 成为研究人们社会关系、行为模式的重要资源. 通话号码个数、通话次数和时长是手机通信网络的基本属性. 本文在复杂网络理论上, 应用统计的方法研究了中国西部某城市三百余万手机用户不同节假日和工作日的四天通话数据在不同尺度下的号码度、通话度、时长度的分布以及平均号码度、平均通话度、平均时长度的特征. 研究表明, 所有尺度下, 号码度、通话度、时长度均为幂律分布, 幂指数随尺度、日期和指标的不同而不同, 在 [1.3, 4] 范围内波动. 总体上, 号码度幂指数大于通话度和时长度幂指数, 入度幂指数大于出度幂指数; 节假日幂指数大于相应指标的工作日幂指数, 休息时段幂指数大于工作时段幂指数; 与工作日相比, 节假日的平均号码度和平均通话度较小, 平均时长度较大. 揭示了绝大多数用户每日只拨打 1 个号码的电话, 节假日期间拨打电话的用户数、次数、时长减少, 但平均通话时长增大的特征.

**关键词:** 手机通话网络, 复杂网络, 度分布, 通话模式

**PACS:** 89.75.Da, 89.75.Fb, 89.75.-k, 89.70.Hj

**DOI:** 10.7498/aps.62.208901

## 1 引言

手机通信是当前人际交流的重要方式之一. 随着手机的普遍使用, 手机通信数据已成为研究人们社会关系、行为模式的重要资源<sup>[1]</sup>. 手机通信数据不仅构成一幅社会关系网络, 而且详细记录了人们的通信行为, 引起各领域专家学者的普遍关注, 并进行了广泛研究, 但是尚缺乏对不同时间尺度下通话数据网络度特征及其行为模式的全面深入分析. 本文应用复杂网络理论, 从主被叫号码、通话次数和通话时长三个维度研究手机通话数据的网络结构及其通话行为模式. 此研究不仅有助于了解和认识人们的社交活动, 而且对异常模式识别、社区发现、人群聚集监控等应用提供研究基础.

复杂网络理论的创建<sup>[2]</sup>为人类行为研究开辟了新视角, 启发人们将各种人类行为数据(手机通话、收发短信、在线论坛、网上购物、协同写作等)的特征概括为拓扑结构, 并对复杂网络的静态拓扑结构<sup>[3-9]</sup>、结构演化<sup>[10-13]</sup>和信息传播<sup>[14-17]</sup>

等问题展开研究. 以手机通话数据为例的研究表明, 人类通话网络的节点度分布满足幂律分布<sup>[6-9]</sup>, 幂指数随统计样本和统计指标的不同而不同. Aiello 等<sup>[6]</sup>得到的一天有向长途电话数据的入度和出度幂指数均为 2.1. Onnela 等<sup>[7]</sup>得到的双向呼叫无向网络的节点度分布幂指数是 8.4, 通话时长幂指数是 1.9. 如果考虑节点的地理位置, 则两节点间连线的概率以及主被叫之间联系的强度(时长和次数)随节点间距离增大而减小, 满足重力模型<sup>[8,9]</sup>. 例如, Lambiotte 等<sup>[8]</sup>对 250 万用户 6 个月 8.1 亿次的语音和短信数据的研究发现, 顶点度分布为  $k^{-5}$  ( $k$  为节点度), 两顶点间有边连接的概率为  $d^{-2}$  ( $d$  为两节点间距离). Kang 等<sup>[9]</sup>对黑龙江省 13 个城市手机用户的通话数据研究表明, 通话次数和时长与距离关系的幂指数分别为 0.5 和 0.4. 可见, 通信网络中不考虑节点空间位置时的度分布幂指数 (2.1, 8.4, 5) 与微博<sup>[3]</sup>、网上论坛<sup>[4]</sup>、民航运输<sup>[5]</sup>等研究得到的幂指数 (多介于 [1, 3.5]) 相比变化区间较大, 说明它所代表的分布分异程度较大.

\* 国家自然科学基金(批准号: 41171345)和国家高技术研究发展计划(批准号: 2012AA12A403)资助的课题.

† 通讯作者. E-mail: peit@reis.ac.cn

自然界爆发 (burst) 模式的发现 [18] 掀起了人类行为动力学的研究热潮 [19], 并迅速在短信收发、手机通话、信件交互、网上购物、协同写作、恐怖袭击等领域展开 [20–29]. 这其中, 基于人类通信数据的研究直接全面深入地揭示了这一特征. Hong 等 [20] 对短信数据的研究表明人们相继发送短信的时间间隔和发送 - 响应的等待时间因个体不同而不同, 为多种幂律分布. Wu 等 [21] 的研究表明它们是泊松分布和幂律分布的组合. Zhao 等 [22] 的研究指出收发短信行为的爆发无记忆产生机理. Jiang 等 [23] 对通话数据的研究发现人们拨打电话的时间间隔只有 3.46% 的用户满足幂律分布, 73.34% 的用户是 Weibull 分布. 不仅人类行为的时间间隔近似服从幂律分布, González 等 [24] 的研究表明人类运动的距离也满足幂律分布, 绝大多数人在几公里的区域范围内做规则的往返运动, 且其运动模式是可预测的, 预测限度是 93% [25]. 不同于物理学家对人类行为的量化研究, 地理学家主要从人类活动的时空韵律定性研究人类行为模式. 如 Ahas 等 [30] 对不同类型单个个体时空活动韵律差异的研究和文献 [31, 32] 对运动定位的研究. 文献 [33–39] 利用基地站的 Erlang 数据, 研究城市人群时空活动韵律, 揭示城市结构和功能, 成功应用于意大利罗马 [33–35]、米兰 [36]、奥地利 Graz [37]、中国深圳等城市 [38, 39], 掀起了群体行为研究的热潮. 定性、定量研究均表明, 人类行为并非完全随机的, 而是具有一定的时

空或统计规律性.

综上所述, 尽管对手机通信数据已做了很多研究, 但是, 尚缺乏对不同时间尺度下通话数据网络度特征和人们通话行为的全面深入分析. 本文正是从这一视角出发, 以复杂网络理论为基础, 以实证统计的方法, 研究日尺度和时段尺度下手机通话网络的号码度、通话度、时长度的分布以及平均号码度、平均通话度和平均时长度的特征, 了解和认识人们的通话行为模式, 为手机通话网络的社区结构模式、突发事件下人群响应等研究奠定基础.

## 2 研究方法

通话号码数、次数、时长是手机通信网络的基本属性. 本文研究这三个属性的度分布及其平均值特征.

### 2.1 基本概念

#### 2.1.1 号码度、通话度、时长度

对顶点  $v$ , 它与多少个号码联系、联系的次数和时长分别用  $v$  的号码度  $d(v)$ , 通话度  $r(v)$  和时长度  $t(v)$  表示, 并有入度 (接听电话) 和出度 (拨打电话) 之分, 如表 1 所示.

对网络  $G$ , 其号码数、通话量和通话时长分别定义为  $G$  的号码度  $d(G)$ , 通话度  $r(G)$  和时长度  $t(G)$ , 也有入度和出度之分, 如表 1 所示.

表 1 概念的符号和含义

顶点/网络	指标名称	符号	含义
顶点 $v$	号码入度/出度	$d_{in}(v)/d_{out}(v)$	$v$ 接听/拨打的号码个数
	通话入度/出度	$r_{in}(v)/r_{out}(v)$	$v$ 接听/拨打电话的次数
	时长入度/出度	$t_{in}(v)/t_{out}(v)$	$v$ 接听/拨打电话的时长
网络 $G$	号码入度/出度	$d_{in}(G)/d_{out}(G)$	$G$ 的主叫/被叫号码个数
	通话入度/出度	$r_{in}(G)/r_{out}(G)$	$G$ 的主叫/被叫电话次数
	时长入度/出度	$t_{in}(G)/t_{out}(G)$	$G$ 的主叫/被叫电话时长

$v$  的各个入度和出度指标可由数据统计得到.  $\{\text{called}(G)\}$  和  $\{\text{calling}(G)\}$  分别表示  $G$  的主叫号码集和被叫号码集, 则  $G$  的号码入度  $d_{in}(G) = |\{\text{called}(G)\}|$ , 号码出度  $d_{out}(G) = |\{\text{calling}(G)\}|$ , 号码度  $d(G) = |\{\{\text{calling}(G)\} \cup \{\text{called}(G)\}\}|$ .  $G$  其余指标的计算如 (1)–(4) 式所示:

$$r_{in}(G) = \sum_{v \in \{\text{called}(G)\}} r_{in}(v), \quad (1)$$

$$t_{in}(G) = \sum_{v \in \{\text{called}(G)\}} t_{in}(v), \quad (2)$$

$$r_{out}(G) = \sum_{v \in \{\text{calling}(G)\}} r_{out}(v), \quad (3)$$

$$t_{out}(G) = \sum_{v \in \{\text{calling}(G)\}} t_{out}(v). \quad (4)$$

由于主被叫的每一次通话对应着一条记录, 所以有以下等式成立:  $r_{in}(G) = r_{out}(G) = r(G)$ ,

$t_{in}(G) = t_{out}(G) = t(G)$ . 但是, 因为研究对象  $G$  仅包含一家通信公司用户的通话记录, 与该公司用户联系的其他公司用户的通话记录未包含在内, 所以  $d_{in}(G) \neq d_{out}(G)$ . 又由于  $\{\text{calling}(G)\}$  中的元素与  $\{\text{called}(G)\}$  中的元素有重复, 所以  $d(G) < d_{in}(G) + d_{out}(G)$ . 但是,  $G$  中拨出的号码总数与接听的号码总数相等 (表示为  $d_{sum}(G)$ ):

$$\sum_{v \in \{\text{calling}(G)\}} d_{out}(v) = \sum_{v \in \{\text{called}(G)\}} d_{in}(v) = d_{sum}(G).$$

网络  $G$  的指标反映了  $G$  的大小 (以  $d(G)$  表征) 和活跃度 ( $r(G)$  和  $t(G)$ ).

### 2.1.2 平均号码度、平均通话度、平均时长度

网络  $G$ , 平均一个号码与其他多少个号码联系称为平均号码度  $\overline{d(G)}$ , 联系的次数称为平均通话度  $\overline{r(G)}$ . 平均时长分两种: 平均每个电话的时长  $\overline{t_r(G)}$  以及平均每个号码的通话时长  $\overline{t_d(G)}$ . 计算公式分别如 (5)—(8) 式所示:

$$\overline{d(G)} = d_{sum}(G)/d(G), \quad (5)$$

$$\overline{r(G)} = r(G)/d(G), \quad (6)$$

$$\overline{t_r(G)} = t(G)/r(G), \quad (7)$$

$$\overline{t_d(G)} = t(G)/d(G) \quad (8)$$

对网络  $G$ , 由于  $d_{sum}(G) \leq r(G)$ ,  $d(G) \leq r(G)$ , 所以  $\overline{d(G)} \leq \overline{r(G)}$ ,  $\overline{t_r(G)} \leq \overline{t_d(G)}$ .

平均度反映了网络  $G$  中通话号码数和通话强度 (次数和时长) 的总体水平.

### 2.1.3 度分布

网络  $G$  的号码度、通话度和时长度的入度和

出度分布分别用 (9)—(11) 式表示, 反映了网络中用户的行为模式. 式中  $\gamma$  表示概率分布的幂指数.

$$p_{in|out}(d_{in|out}(G)|_{d_{in|out}(v)=k}) \propto k^{-\gamma} \quad k = 1, 2, \dots, \quad (9)$$

$$p_{in|out}(r_{in|out}(G)|_{d_{in|out}(v)=k}) \propto k^{-\gamma} \quad k = 1, 2, \dots, \quad (10)$$

$$p_{in|out}(t_{in|out}(G)|_{d_{in|out}(v)=k}) \propto k^{-\gamma} \quad k = 1, 2, \dots. \quad (11)$$

## 2.2 研究对象

本文研究数据取自中国西部某城市某通信公司 3 百多万手机用户在 2012 年 1 月 1 日 (元旦)、8 日 (周日)、16 日 (周一)、23 日 (正月初一) 四天的通话记录 (每一天的通话记录构成日尺度通话网络). 通话记录包括事件类型、主叫号码、主叫归属地、被叫号码、被叫归属地、通话起始时间、通话截止时间、时长等八个属性. 本文研究结果是聚合数据特征, 不存在用户隐私泄露问题. 尽管如此, 也为每个号码重新编码以保证用户的信息安全. 数据格式如表 2 所示.

由于银行、保险、铁路、通信公司等企业的客服号码是服务型公共电话, 被叫次数非常大, 入度大于其他号码数十至上百倍, 这些号码对分析正常人际交流以及社会关系没有意义. 因此, 本文中滤去这些客服号码的相关数据. 日尺度下的四个数据集经过数据预处理 (见 2.3 节) 后的通话量是 3073255 条. 去除与公共服务号码相关的记录后是 3003829 条, 详细情况如表 3 所示, 最后一列“总和”的数字是各行数值的直接求和.

表 2 手机通话记录的数据样表

事件类型	主叫号码	主叫归属地	被叫号码	被叫归属地	起始时间	截止时间	时长
16001	189****1234	0911	151****1234	029	2012-01-01 23:46:01.000	2012-01-01 23:47:45.000	104
16002	135****1234	0912	131****1234	0910	2012-01-01 13:09:37.000	2012-01-01 13:28:53.000	1156

表 3 日尺度下四个数据集的数据详情

指标名称	2012/01/01	2012/01/08	2012/01/16	2012/01/23	总和
通话量 $r(G)$	763002	730014	989442	521371	3003829
主叫号码数 $d_{out}(G)$	283712	274431	343291	217386	1118820
被叫号码数 $d_{in}(G)$	322706	302762	368803	264459	1258730
号码总数 $d(G)$	467602	442070	538637	387829	1836138

### 2.3 研究方法

首先进行数据预处理,包括主被叫号码的一致性检查、删除重复记录和删除涉及服务号码的记录. 1) 号码一致性检查的步骤是:删除 11 位长手机号码前的 86 前缀、7 位或 8 位长固话号码前的 17911 等前缀,根据主被叫号码的归属地以及事件类型等为所有不带区号的固话号码增加区号,等等诸如此类的工作; 2) 由于带有 17911 等前缀的号码,一次通话出现两条数据,一条数据记录通话费用,一条数据不记录通话费用,删除其中的一条数据; 3) 根据通话数据和百度文库等外部资料确定了 37 个“裸”客服号码,如 10086, 118114, 12306, 95533 等,根据每个号码的长度以及连接各地长途区号后的可能长度,扫描通话数据总表中所有的主叫和被叫号码,得到 985 个带区号的客服号码,将这 1022 个号码全部存入数据表中. 选择主叫或被叫号码不在客服表中的数据记录插入到新表中,以备后续统计分析用.

在数据预处理的基础上,统计日尺度和时段尺度下每个网络三个指标的度分布及其平均度. 首先获得待研究数据集:根据通话截止时间,获得 2012 年 1 月 1 日、8 日、16 日、23 日共四个日尺度通话数据集,以及这四天每天五个时段(凌晨 0:00—9:00,上午 9:00—13:00,中午 13:00—15:00,下午 15:00—19:00,晚上 19:00—24:00)共 20 个时段尺度数据集;第二步,统计每个数据集  $G$  的主叫号码数  $d_{out}(G)$ 、被叫号码数  $d_{in}(G)$  和通话量  $r(G)$ ;第三步,统计每个数据集  $G$  中每个主叫号码的号码出度  $d_{out}(v)$ 、通话出度  $r_{out}(v)$ 、时长出度  $t_{out}(v)$ ,以及每个被叫号码的号码入度  $d_{in}(v)$ 、通话入度  $r_{in}(v)$ 、时长入度  $t_{in}(v)$ ;第四步,分组聚合不同入度、出度下的号码个数  $d_{in|out}(G)|_{d_{in|out}(v)=k}$ 、通话次数  $r_{in|out}(G)|_{d_{in|out}(v)=k}$  和时长  $t_{in|out}(G)|_{d_{in|out}(v)=k}$ ;第五步,得到不同入度和出度下的号码数、通话数和时长的概率及其度分布,并用 matlab 工具得到每个分布的幂函数拟合;第六步,对每个数据集计算其平均号码度  $\bar{d}(G)$ 、平均通话度  $\bar{r}(G)$  和平均时长度  $\bar{t}_r(G)$  和  $\bar{t}_d(G)$ . 最后,在这些统计结果的基础上,分析各个数据集的度特征和人们的通话行为模式.

### 3 日度分布

下面将系统研究 2012 年 1 月 1 日(元旦)、8 日(周日)、16 日(周一)和 23 日(正月初一)每天

的入度和出度分布以及平均号码度、平均通话度、平均时长度,分析不同节假日、工作日的度特征和通话模式差异.

### 3.1 同一日期的度分布

以 1 月 1 日为例,研究该日网络的入度 ( $d_{in}(G)$ ,  $r_{in}(G)$ ,  $t_{in}(G)$ ) 和出度 ( $d_{out}(G)$ ,  $r_{out}(G)$ ,  $t_{out}(G)$ ) 分布及其拟合幂指数. 统计结果如图 1 所示.

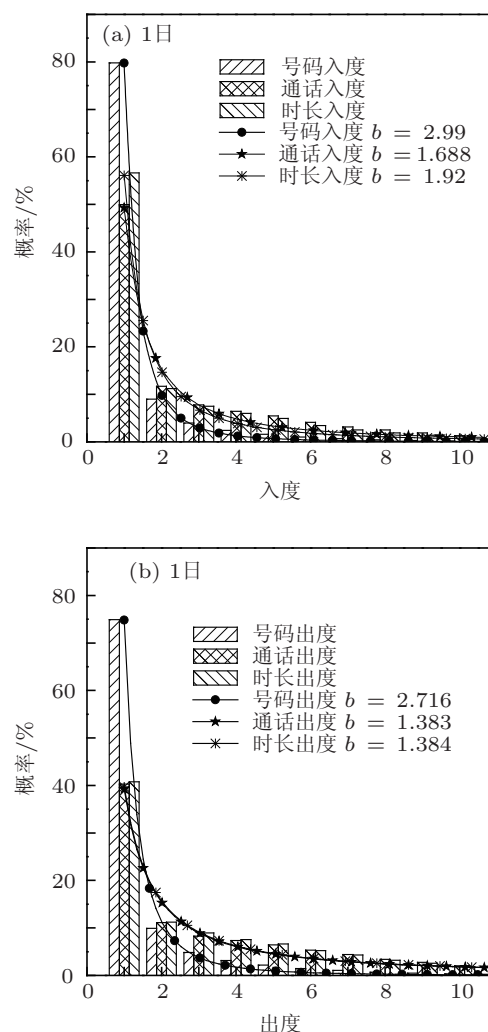


图 1 号码度、通话度、时长度的概率分布及其拟合曲线(拟合函数是  $f(x) = ax^{-b}$ ) (a) 入度分布; (b) 出度分布

图 1 说明入度、出度、号码度、通话度、时长度均为幂律分布. 度为 1 时, 号码度概率大于 75%, 通话度和时长度概率大于 40%, 远超过其他度的概率, 说明虽然 75% 以上的用户只与一个号码联系, 但是他们产生了 40% 以上的通话, 是为通信公司贡献利润不可或缺的力量. 因此, 对于通信公司而言, 客户的绝对拥有量是其赢得市场竞争的利器. 图 1(a) 和 (b) 的区别是, 度为 1 时, 入度概率(分别

为 80%, 50%, 56%) 均大于出度概率 (分别为 75%, 40%, 41%), 度为 2 时的差异较小, 当度大于 2 以后, 入度概率均略小于出度概率. 说明除度为 1 之外, 其他号码的“输出”大于“输入”.

观察入度和出度三种指标的幂指数可知, 号码入度幂指数 (2.99) 大于通话入度和时长入度幂指数 (1.699, 1.92), 号码出度幂指数 (2.716) 大于通话出度和时长出度幂指数 (1.383, 1.384). 幂指数越大, 度分布衰减越快. 上述现象说明, 相比之下, 号码度的分布衰减更快, 其分布具有“趋向小值”的特征. 另外, 入度幂指数大于相应指标的出度幂指数. 这又说明, 入度比出度更趋向于“小值”.

上述特征表明手机通话网络具有小世界和无标度性. 通过对该网络连通子图分解发现, 它由许多节点数不同的连通子图构成, 除去海量的 1 度、2 度等小度数组件之外, 尚有许多具有很大度数的 Hub 节点, 这印证了文献 6 的结论, 即当节点度 (号码度) 在 [2, 3.47878] 范围内, 网络中存在巨型组件和第二大组件. 如果删除网络中最大的连通组件, 剩余图的节点数显著减少, 幂指数增大. 因此, 通过幂指数的大小, 可估计网络中最大连通组件的大小, 并用于异常监测、社区发现算法设计等.

不仅 1 日的度分布具有该特征, 8 日、16 日、23 日的度分布均具有该特征, 仅仅是三种度的概率大小的微小差别.

### 3.2 不同日期的度分布

图 2(a) 和 (b) 展示了 1 日、8 日、16 日、23 日四天三种指标 ( $d(G)$ ,  $r(G)$ ,  $t(G)$ ) 的入度和出度幂指数, 由此分析不同日期的通话模式差异.

对度值, 表 3 展示了四天的差异: 无论是通话度  $r(G)$  ( $r_{in}(G) = r_{out}(G) = r(G)$ ) 还是号码度  $d(G)$  (包括  $d_{out}(G)$  和  $d_{in}(G)$ ), 16 日最大, 23 日最小. 其原因是, 16 日是周一, 大量的工作业务电话使当天的通话号码数  $d(G)$  和次数  $r(G)$  在四天中最多. 23 日是正月初一, 大量的流动人员离开该市回家过年, 工作电话大量减少, 虽然亲朋好友之间的问候电话有可能增多, 但仍然不抵总体人数减少带来的影响, 故总的通话号码数  $d(G)$  和次数  $r(G)$  为四天中最少. 1 日元旦和 8 日周日, 均属节假日, 情况类似, 只是不如春节显著. 可见, 随着节假日级别提高, 通话号码数  $d(G)$  和次数  $r(G)$  减少.

对度的幂指数, 图 2(a) 和 (b) 表明四天具有共同的规律: 号码度幂指数大于通话度和时长度幂指

数, 入度幂指数大于相应指标的出度幂指数, 不同日期三个指标的入度幂指数变化显著, 出度幂指数变化很小. 图 1(a) 和 (b) 中显示 1 日的情况与上述基本规律符合. 说明尽管日期不同, 但是每天的通话模式基本一致. 与度值规律相对应, 16 日和 23 日三个指标的幂指数分别是四天中的最小和最大 (时长出度例外). 说明幂指数与其相应的指标值有可能呈负相关, 如号码数  $d(G)$  或通话数  $r(G)$  越大, 号码度分布或通话度分布的幂指数越小. 同时也说明工作日通话范围广、次数多、通话总时间较长, 节假日通话范围小、次数少、通话总时间较短 (时长出度例外). 从另一个角度看, 由于 23 日号码入度和出度的幂指数最大, 所以其最大组件的度最小, 连通组件的个数也少, 进一步证明该日用户联络范围小、次数少的特点.

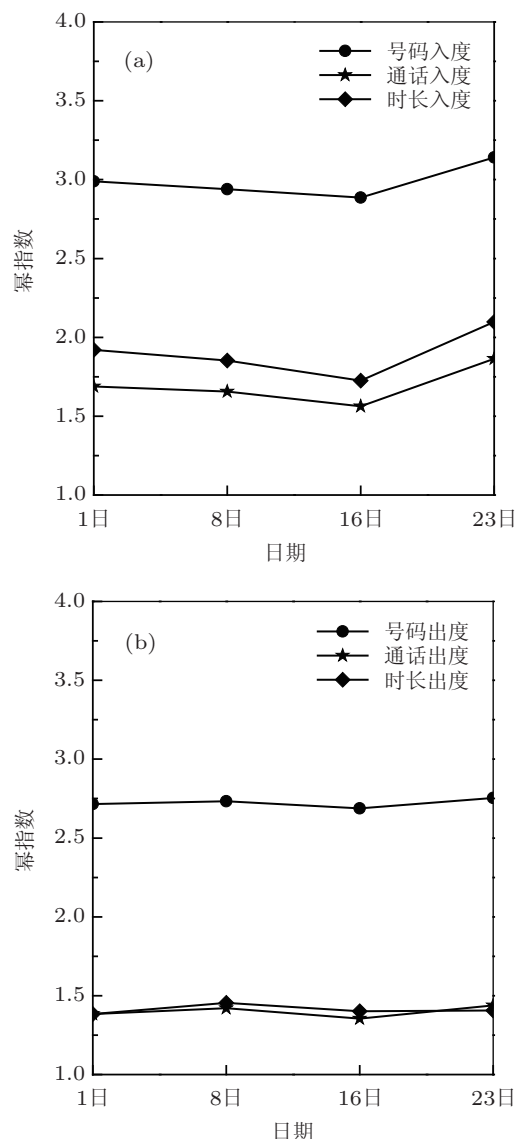


图 2 四天的号码度、通话度和时长度幂指数 (a) 入度幂指数; (b) 出度幂指数

### 3.3 日平均度

由(5)—(8)式, 四天的平均号码度 $\overline{d(G)}$ 、平均通话度 $\overline{r(G)}$ 、平均每个电话的时长 $\overline{t_r(G)}$ 和平均每个号码的通话时长 $\overline{t_d(G)}$ 分别如图3(a)和(b)所示.

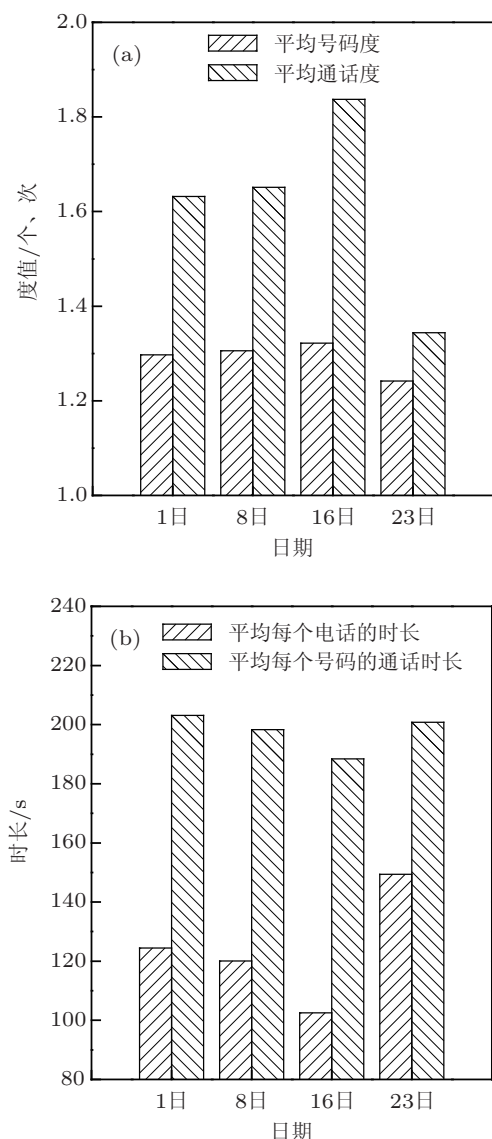


图3 平均号码度、平均通话度以及平均每个电话的时长和平均每个号码的通话时长 (a) 平均号码度和平均通话度; (b) 平均每个电话的时长和平均每个号码的通话时长

从图3(a)可知, 任何一天, 平均每个号码接听或拨打的号码数约为1.2—1.3个, 接听或拨打的电话不足两个. 其中, 16日的平均号码度和平均通话度最大, 23日最小, 8日略大于1日. 可见, 随着节日级别提高, 平均接打电话的号码范围和次数减少. 然而对于平均通话时长, 平均每电话时长和平均每号码时长则增加(图3(b), 16日最小, 23日或1日最

大). 说明节假日的平均通话模式与工作日不同, 也体现了节假日以感情交流为主的电话与工作日以业务联系为主的电话的模式差异.

## 4 时段的度分布

根据该市冬季的作息时间, 将一天24 h分为五个时段, 凌晨(0:00—9:00)、上午(9:00—13:00)、中午(13:00—15:00)、下午(15:00—19:00)和晚上(19:00—24:00). 本节从日期和时段两个角度分析2012年1月1日(元旦)、8日(周日)、16日(周一)和23日(正月初一)三个指标( $d(G), r(G), t(G)$ )的入度和出度分布以及平均度( $\overline{d(G)}, \overline{r(G)}, \overline{t_r(G)}, \overline{t_d(G)}$ )的特征.

### 4.1 不同日期、不同时段的角度分布

图4(a)—(c)展示了四天五个时段三个指标( $d(G), r(G), t(G)$ )不同视角下的幂指数. 首先, 从图4(a)可知, 同一日期、同一时段、三个指标的幂指数具有大致相同的规律: 号码度幂指数大于通话度和时长度幂指数, 入度幂指数大于相应指标的出度幂指数, 不同日期三个指标的入度幂指数变化较大, 出度幂指数变化很小. 与图2(a), (b)完全符合, 说明该特征不随统计尺度和日期的改变而改变.

其次, 从图4(b)和(c)可以看出, 除凌晨时段外, 无论入度、出度, 不同日期同一时段的号码度、通话度和时长度的幂指数是23日最大(时长出度上午时段例外), 16日最小, 1日与8日几乎重叠, 介于23日和16日的中间, 说明每天从早到晚通话模式基本稳定: 节假日的通话范围、次数和时长都比较少, 工作日比较多, 与图2(a), (b)结论基本一致. 凌晨时段的规律与此相反, 23日最小(号码入度和通话入度是1日最小, 也是个大节日), 16日最大, 反查数据记录发现节假日日凌晨时段的国内长途以及国际长途增多(可能与时差有关), 因此幂指数变小.

最后, 从图4(a)—(c)还可看到, 尽管个别时段数值差异不大, 但是, 无论入度、出度, 同一日期不同时段三个指标的幂指数从大到小排序依次为凌晨-中午-晚上-上午-下午(号码出度略有不同), 说明话务量大小与时段长度无关, 只与人们的作息规律有关, 工作时段电话使用多, 幂指数小, 休息时段电话使用少, 幂指数大.



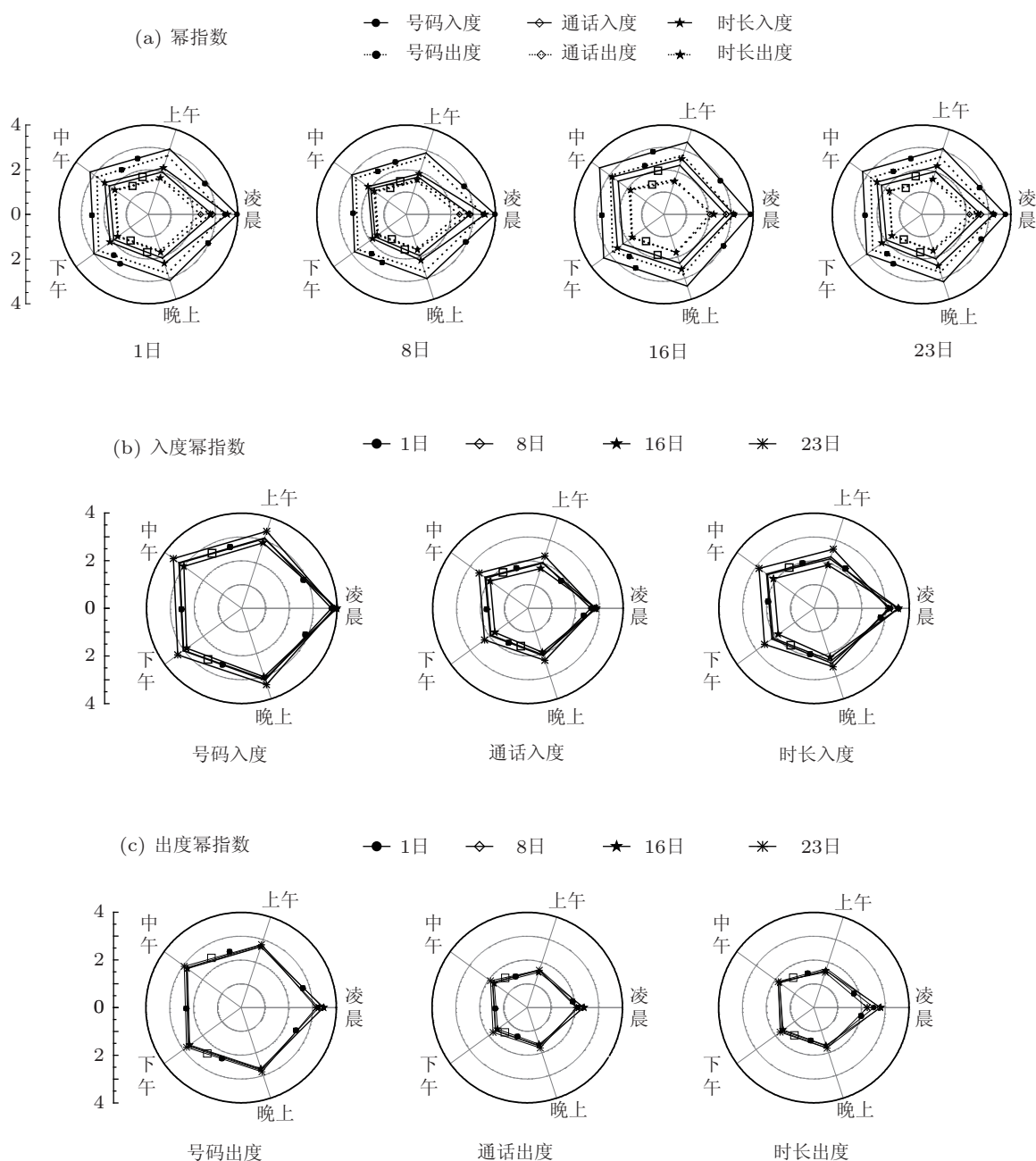


图4 不同视角下的号码度、通话度、时长度的幂指数 (a) 同一日的入度和出度幂指数; (b) 不同日的入度幂指数; (c) 不同日的出度幂指数

## 4.2 不同日期、不同时段平均度

图 5(a) 和 (b) 展示了四天五个时段平均号码度  $\overline{d(G)}$ 、平均通话度  $\overline{r(G)}$ 、平均每个电话的时长  $\overline{t_r(G)}$  和平均每个号码的通话时长  $\overline{t_d(G)}$ 。

图 5(a) 表明, 无论哪一天、哪个时段, 平均号码度  $\overline{d(G)}$  小于平均通话度  $\overline{r(G)}$ , 且二者在不同时段的变化趋势相同. 对比图 5(a) 和 (b) 可知同一日期不同时段平均度差异. 除 23 日以外, 1 日、8 日和 16 日, 平均号码度  $\overline{d(G)}$  和平均通话度  $\overline{r(G)}$  从

大到小排序依次是: 下午 - 上午 - 晚上 - 中午 - 凌晨, 刚好是 4.1 节所述的“同一日期不同时段幂指数排序”的逆序, 证明号码度和通话度幂指数与号码数  $d(G)$  和通话次数  $r(G)$  的负相关关系. 与上述规律不同, 平均每电话时长  $\overline{t_r(G)}$  和平均每号码时长  $\overline{t_d(G)}$  从大到小排序是: 凌晨 - 晚上 - 上午 - 下午 - 中午 (16 日是下午 - 上午). 说明同一日期不同时段平均度具有较大差异: 凌晨时段通话范围小、次数少, 通话时间长; 上午和下午的通话范围广、次数多, 通话时间短; 晚上时段的通话范围、次数和

时长都比较大;中午则都比较小. 23日是个特殊的日期,其通话模式除具有其他日期的共性外(如凌晨时段的平均号码度  $\overline{d(G)}$  和平均通话度  $\overline{r(G)}$  为全天最小,平均通话时长  $\overline{t_r(G)}$  和  $\overline{t_d(G)}$  却是全天最大,晚上时段拥有较大的平均号码度  $\overline{d(G)}$ 、平均通话度  $\overline{r(G)}$  和平均通话时长  $\overline{t_r(G)}$  和  $\overline{t_d(G)}$ ),还具有其特殊性:中午时段的平均号码度  $\overline{d(G)}$  和平均通话度  $\overline{r(G)}$  为全天最大,除凌晨时段外,平均通话时长  $\overline{t_r(G)}$  和  $\overline{t_d(G)}$  均大于其他三天的相应时段,而凌晨时段的平均通话时长  $\overline{t_r(G)}$  和  $\overline{t_d(G)}$  小于其他三

天的凌晨值.

图 6 显示了不同日期相同时段平均度的差异. 总体来说,对于各个时段,16日平均号码度  $\overline{d(G)}$  和平均通话度  $\overline{r(G)}$  最大,23日最小(中午和凌晨时段除外),但是不同时段平均通话时长  $\overline{t_r(G)}$  和  $\overline{t_d(G)}$  却是 23日最大,16日最小(凌晨时段除外),这一点与日总体特征(图 3(a), (b)) 基本符合. 16日凌晨时段平均通话次数  $\overline{r(G)}$  和平均通话时长  $\overline{t_r(G)}$  和  $\overline{t_d(G)}$  均大于 23日,平均通话号码个数  $\overline{d(G)}$  略小于 23日.

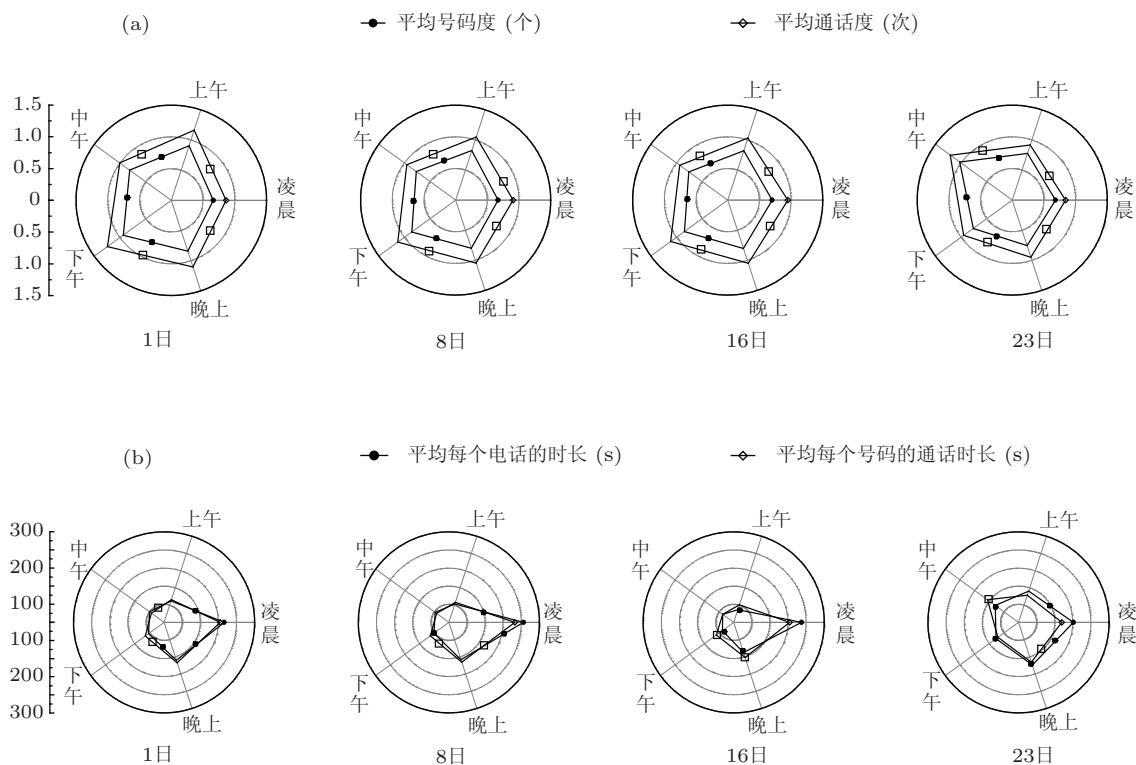


图 5 四天五个时段的平均号码度、平均通话度和平均时长度 (a) 平均号码度、平均通话度; (b) 平均每个电话的时长、平均每个号码的通话时长

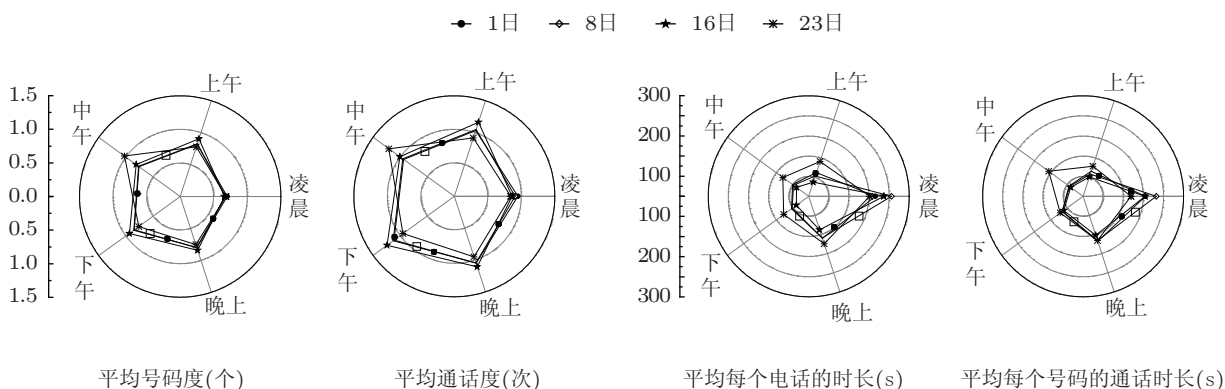


图 6 不同日期的平均号码度、平均通话度和平均时长度



综上所述,虽然不同日期、不同时段三个指标的幂指数显示,凌晨时段外工作日的通话范围、次数、时长都较大,节假日较小.然而,平均度的情况并非如此,表现在平均号码度 $\overline{d(G)}$ 、平均通话度 $\overline{r(G)}$ 与平均时长度 $\overline{t_r(G)}$ 和 $\overline{t_d(G)}$ 的变化趋势不一致.凌晨时段外,工作日的平均号码度 $\overline{d(G)}$ 和平均通话度 $\overline{r(G)}$ 较大,但平均时长度 $\overline{t_r(G)}$ 和 $\overline{t_d(G)}$ 较小,节假日的情况相反.同一日不同时段平均度特征差异较大,23日的平均度情况比较特殊.上述特征为通信公司根据不同日期和时间制定定价策略提供了科学依据.比如,针对节假日通话用户少,通话时间长特点,减少通话资费,吸引更多客户使用电话,达到公司和客户双赢.工作日工作时段相应提高资费,以减小通信公司设备负荷,降低成本,解决掉线等问题.

## 5 结论

本文使用中国西部某城市3百多万手机用户四天的通话数据研究了手机通话网络的度特征和人们的通话模式,通过对不同日期、不同尺度数据集的号码度 $d(G)$ 、通话度 $r(G)$ 、时长度 $t(G)$ 的度分布以及平均号码度 $\overline{d(G)}$ 、平均通话度 $\overline{r(G)}$ 、平均时长度 $\overline{t_r(G)}$ 和 $\overline{t_d(G)}$ 的统计,发现以下规律.

1) 总体特征是:号码度 $d(G)$ 、通话度 $r(G)$ 、时长度 $t(G)$ 均为幂律分布,号码度幂指数大于通话度和时长度幂指数,入度幂指数大于出度幂指数,不同日期三个指标的入度幂指数变化较大,出度幂指数变化较小.由于通话度和时长度依赖于号码度,所以对通信公司而言,增加客户数量才能提高企业竞争力.另一方面,如果有异常通话行为,如少量用户之间超频繁超时长(其数值接近或大于比其度小的用户通话强度之和)的通话,则使幂指数急剧下降,甚至变为负数,度分布曲线将由衰减变为上升.因此,幂指数有助于异常通话行为的监测.

2) 日尺度下,16日(周一)的号码度、通话度、时长入度(时长出度无明显规律)的幂指数最小,23日(正月初一)最大.说明工作日接打电话范围广、次数多、接听电话总时间较长,而节假日接打电话范围小、次数少,接听电话的总时间也较短.但是从日平均度视角看,16日的平均号码度 $\overline{d(G)}$ 和平均

通话度 $\overline{r(G)}$ 最大,平均每电话时长 $\overline{t_r(G)}$ 和平均每号码时长 $\overline{t_d(G)}$ 最小,23日的情况正好相反.说明节假日时虽然联系号码和次数较少,但由于主要是亲朋好友之间的感情交流电话,所以通话时间较长,而工作日时由于业务、工作需要,与较多的用户联系,导致联系的次数增多,但平均通话时间较短.体现了节假日和工作日通话模式及网络拓扑结构(表现在最大连通子图的个数和度数上)的差异,为通信公司优化网络和服务提供了依据.

3) 时段尺度下的规律如下:凌晨时段,16日的号码度、通话度、时长度的幂指数最大,23日最小(号码入度和通话入度例外).其他时段的情况正相反.说明除凌晨时段外,工作日接打电话范围广、次数多、时间长,而节假日接打电话范围小、次数少,时间也较短,与日尺度的通话模式一致.从平均度的视角看,不同日期同一时段除凌晨和中午以外,16日的平均号码度 $\overline{d(G)}$ 、平均通话度 $\overline{r(G)}$ 最大,平均时长度 $\overline{t_r(G)}$ 和 $\overline{t_d(G)}$ 最小,23日的平均号码度 $\overline{d(G)}$ 、平均通话度 $\overline{r(G)}$ 最小,平均时长度 $\overline{t_r(G)}$ 和 $\overline{t_d(G)}$ 最大.与日尺度的平均度特征一致.同一日不同时段平均度特征有较大差异,上午和下午有较大的平均号码度 $\overline{d(G)}$ 、平均通话度 $\overline{r(G)}$ 以及较小的平均时长度 $\overline{t_r(G)}$ 和 $\overline{t_d(G)}$ ,凌晨有最小的平均号码度 $\overline{d(G)}$ 、平均通话度 $\overline{r(G)}$ 和最大的平均时长度 $\overline{t_r(G)}$ 和 $\overline{t_d(G)}$ ,晚上的平均度均较大,中午均较小.23日平均度比较特殊,中午的平均号码度 $\overline{d(G)}$ 和平均通话度 $\overline{r(G)}$ 是四天中最大,凌晨的平均时长度 $\overline{t_r(G)}$ 和 $\overline{t_d(G)}$ 是四天中最小.不同时段平均度的差异体现了人们的每日生活韵律:紧张忙碌的工作(平均通话时间也短)之后随着晚上娱乐活动的增多,人们的联络电话也增多,而凌晨时段主要是亲属之间的联络,人次少,时间长.该发现将有助于通信公司制定分段计价的策略.

本文用实证统计的方法研究了4个日尺度和20个时段尺度的通话网络,探究了不同日期和时段下人们的通话行为模式.该研究不仅对通信公司开发产品,吸引和保持客户具有一定的指导意义,而且对异常通话模式监测、社区发现算法设计和人群聚集研究具有重要价值.下一步,将在此基础上,研究手机通话网络幂指数的演变过程以及突发事件下人们的响应行为.

- [1] Kwok R 2009 *Nature* **458** 959
- [2] Barabási A L, Albert R 1999 *Science* **286** 509
- [3] Yuan W G, Liu Y, Cheng J J, Xiong F 2013 *Acta Phys. Sin.* **62** 038901 (in Chinese) [苑卫国, 刘云, 程军军, 熊菲 2013 物理学报 **62** 038901]
- [4] Si X M, Liu Y 2011 *Acta Phys. Sin.* **60** 078903 (in Chinese) [司夏萌, 刘云 2011 物理学报 **60** 078903]
- [5] Chi L P, Wang R, Su H, Xu X P, Zhao J S, Li W, Cai X 2003 *Chin. Phys. Lett.* **20** 1393
- [6] Aiello W, Chung F, Lu L Y 2000 *Proceedings of 32nd Annual ACM Symposium on Theory of Computing* Portland, OR, USA, May 21–23, 2000 p171
- [7] Onnela J P, Saramaki J, Hyvonen J, Szabo G, Lazer D, Kaski K, Kertesz J, Barabasi A L 2007 *Proc. Nat. Acad. Sci. USA* **104** 7332
- [8] Lambiotte R, Blondel V D, de Kerchove C, Huens E, Prieur C, Smoreda Z, van Dooren P 2008 *Physica A: Statist. Mech. Appl.* **387** 5317
- [9] Kang C G, Zhang Y, Ma X J, Liu Y 2013 *Int. J. Geograph. Informa. Sci.* **27** 431
- [10] Zhang L, Liu Y 2008 *Acta Phys. Sin.* **57** 5419 (in Chinese) [张立, 刘云 2008 物理学报 **57** 5419]
- [11] Wu Y, Xiao J H, Wu Z Y, Yang J Z 2007 *Acta Phys. Sin.* **56** 2037 (in Chinese) [吴晔, 肖井华, 吴智远, 杨俊忠 2007 物理学报 **56** 2037]
- [12] Gong K, Tang M, Shang M S, Zhou T 2012 *Acta Phys. Sin.* **61** 098901 (in Chinese) [龚凯, 唐明, 尚明生, 周涛 2012 物理学报 **61** 098901]
- [13] Ai J, Zhao H, Carleyb K M, Su Z, Li H 2013 *Chin. Phys. B* **22** 78902
- [14] Li M J, Wu Y, Liu W Q, Xiao J H 2009 *Acta Phys. Sin.* **58** 5251 (in Chinese) [李明杰, 吴晔, 刘维清, 肖井华 2009 物理学报 **58** 5251]
- [15] Wang H, Han J H, Deng L, Cheng K Q 2013 *Acta Phys. Sin.* **62** 110505 (in Chinese) [王辉, 韩江洪, 邓林, 程克勤 2013 物理学报 **62** 110505]
- [16] Xiong X, Hu Y 2012 *Acta Phys. Sin.* **61** 150509 (in Chinese) [熊熙, 胡勇 2012 物理学报 **61** 150509]
- [17] Zhang Y C, Liu Y, Zhang H F, Cheng H, Xiong F 2011 *Acta Phys. Sin.* **60** 050501 (in Chinese) [张彦超, 刘云, 张海峰, 程辉, 熊菲 2011 物理学报 **60** 050501]
- [18] Barabási A L 2005 *Nature* **435** 207
- [19] Wang B H, Zhou T, Zhou C S 2012 *J. Univ. Shanghai Sci. Technol.* **34** 103 (in Chinese) [汪秉宏, 周涛, 周昌松 2012 上海理工大学学报 **34** 103]
- [20] Hong W, Han X P, Zhou T, Wang B H 2009 *Chin. Phys. Lett.* **26** 028902
- [21] Wu Y, Zhou C S, Xiao J H, Kurths J, Schellnhuber H J 2010 *Proc. Nat. Acad. Sci. USA* **107** 18803
- [22] Zhao Z D, Xia H, Shang M S, Zhou T 2011 *Chin. Phys. Lett.* **28** 068901
- [23] Jiang Z Q, Xie W J, Li M X, Podobnik B, Zhou W X, Stanley H E 2013 *Proc. Nat. Acad. Sci. USA* **110** 1600
- [24] González M C, Hidalgo C A, Barabási A L 2008 *Nature* **453** 779
- [25] Song C, Qu Z, Blumm N, Barabási A L 2010 *Science* **327** 1018
- [26] Li N N, Zhang N, Zhou T 2008 *Comp. Syst. Comp. Sci.* **5** 43 (in Chinese) [李楠楠, 张宁, 周涛 2008 复杂系统与复杂性科学 **5** 43]
- [27] Dong Y W, Cai S M, Shang M S 2013 *Acta Phys. Sin.* **62** 028901 (in Chinese) [董宇蔚, 蔡世民, 尚明生 2013 物理学报 **62** 028901]
- [28] Zhao F, Liu J H, Zha Y L, Zhou T 2011 *Acta Phys. Sin.* **60** 118902 (in Chinese) [赵飞, 刘金虎, 查一龙, 周涛 2011 物理学报 **60** 118902]
- [29] Zhu J F, Han X P, Wang B H 2010 *Chin. Phys. Lett.* **27** 68902
- [30] Ahas R, Aasa A, Silm S, Tiru M 2010 *Transport. Res. C: Emerg. Technol.* **18** 45
- [31] Ahas R, Silm S, Järv O, Saluveer E, Tiru M 2010 *J. Urban Technol.* **17** 3
- [32] Isaacman S, Becker R, Cáceres R, Kobourov S, Martonosi M, Rowland J, Varshavsky A 2011 *Pervasive Computing. 9th International Conference*, Pervasive San Francisco, CA, USA, June 2011 p133
- [33] Reades J, Calabrese F, Sevtsuk A, Ratti C 2007 *IEEE Pervasive Computing* **6** 30
- [34] Calabrese F, Ratti C 2006 *Networks and Communication Studies* **20** 247
- [35] Reades J, Calabrese F, Ratti C 2009 *Environment and Planning B: Planning & Design* **36** 824
- [36] Pulselli R M, Romano P, Ratti C, Tiezzi E 2008 *Int. J. Design Nature and Ecodynamics* **3** 121
- [37] Ratti C, Sevtsuk A, Huang S, Pailer R 2005 *Proceedings of the 3rd Symposium on LBS & TeleCartography* Vienna, Austria, November 28–30, 2005
- [38] Sun J B, Yuan J, Wang Y, Si H B, Shan X M 2011 *Physica A: Statist. Mech. Appl.* **390** 929
- [39] Xu Z X, Wang Y, Si H B, Feng Z M 2011 *Acta Phys. Sin.* **60** 040501 (in Chinese) [徐赞新, 王钺, 司洪波, 冯振明 2011 物理学报 **60** 040501]

# Analysis on degree characteristics of mobile call network\*

Yu Xiao-Ping<sup>1)</sup> Pei Tao<sup>2)†</sup>

1) ( College of Information Science and Technology, Shihezi University, Shihezi 832003, China )

2) ( State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China )

( Received 19 May 2013; revised manuscript received 26 July 2013 )

## Abstract

Mobile phone data record the people's communication behaviors in detail, and become an important resource for studying people's social relationships and behavior patterns. The call numbers, call volumes and call durations are the basic properties of the mobile phone data network. Based on the theory of complex networks and the statistic method, studied in this paper are the degree distribution and average values of the call numbers, call volumes, call durations, by using the mobile phone call data of different holidays and workdays and different scales of day and period, which are produced by the 3.3 million subscribers registered in a western city in China. Research shows that at all scales the number degrees, volume degrees and duration degrees are all of power-law distribution, and the power exponents are different depending on the scale, date and index, and fluctuate from 1.3 to 4. In general, the number degree exponent is greater than that of the volume degree and duration degree, and the exponents of three indexes of in-degree are greater than those exponents of out-degree; the exponents in holidays are greater than those of the workdays, and the exponents in working periods are greater than those of non-working periods. Compared with in workdays, in holidays the average number degree and volume degree are small, and the average duration degree is large. It reveals that most of the subscribers call only one number in a day or a period, and the call numbers, call volumes and call durations all decrease on holidays or working periods, however the average call duration increases meanwhile.

**Keywords:** mobile call data, complex networks, degree spatial-temporal distribution, behavior patterns

**PACS:** 89.75.Da, 89.75.Fb, 89.75.-k, 89.70.Hj

**DOI:** 10.7498/aps.62.208901

---

\* Project supported by the National Natural Science Foundation of China (Grant No. 41171345) and the National High Technology Research and Development Program of China (Grant No. 2012AA12A403).

† Corresponding author. E-mail: peit@lreis.ac.cn