

基于SMS/MMS的移动社会网络行为建模与分析

黄银锋,彭三城,冯梓远,郭树长,赵义,陈桂贤

(肇庆学院 计算机学院,广东 肇庆 526061)

摘要:为了有效地分析移动社会网络所具有的复杂网络特性,结合实际短信网络,研究基于短信/彩信(simple messaging service/multimedia messaging service, SMS/MMS)的移动社会网络行为建模方法及分析机制.首先,采用短信/彩信通信记录来构建社会关系图,并提出了基于社会关系图的移动社会网络行为分析机制.给出了移动社会网络行为分析的相关要素及其计算模型,采用社会网络分析的方法对出度、入度、用户朋友数、活跃度、亲密度、聚类系数等特征进行统计分析.实验结果表明:短信网络具有复杂网络的特性,活跃度能更好地刻画手机用户的移动社会网络行为.该方法能有效地用于评价移动社会网络中节点影响力的大小.

关键词:移动社会网络;SMS/MMS;社会关系图;行为分析

中图分类号:TP393.0 **文献标志码:**A **文章编号:**1672-9102(2014)01-0068-05

Modeling and analysis on behavior of mobile social network based on SMS/MMS

HUANG Yin-feng, PENG San-cheng, FENG Zi-yuan, GUO Shu-chang, ZHAO Yi, CHEN Gui-xian
(School of Computer Science, Zhaoqing University, Zhaoqing 526061, China)

Abstract: In order to evaluate the characteristics of complex network in mobile social networks, a modeling method and analytical mechanism on behavior of mobile social network were proposed based on simple messaging service / multimedia messaging service (SMS/MMS) by combining the characteristics of real short message network. At first, social relationship graph was constructed using SMS/MMS. In addition, an analysis mechanism on behavior of mobile social network was presented based on the social relationship graph. The related factors of behavior analysis and their computing model were provided, and the method of social network analysis was used to count and analyze the characteristics of mobile social network, such as out-degree, in-degree, the number of friends, activity degree, intimacy degree, and clustering coefficient. Simulation results show that the real short message network has the characteristics of complex networks and that activity degree can be used to characterize the behavior of users in mobile social networks. The method can be used to evaluate the degree of nodes' influence in mobile social network.

Key words: mobile social network; SMS/MMS; social relationship graph; behavior analysis

随着移动通信技术的迅猛发展,手机越来越得到人们的认可,成为了人们日常生活中的一个重要组成部分.据东方早报报道,截止2012年2月底,中国手机用户已达9.997亿.作为手机主要功能之

一的手机短信,以其方便、快捷、私密和价廉等优点受到了人们的青睐.短信成为了人们日常生活中一种常用的通信方式,在2012年,手机短信的发送量达到8 973.1亿条.在日常生活中,人们常常会通

过发送短信/彩信 (simple messaging service/multimedia messaging service, SMS/MMS)^[1]来进行社会交往,不仅产生了大量的短信/彩信信息,而且在一定程度上反映了人们之间的社会关系;还体现了社会网络的相关特征.因此,构成了一种特殊的社会网络,称为“移动社会网络”.它是指从移动通信数据中抽象出来的一种社会网络,每部手机表示网络的一个节点,任意2部手机之间收发短信/彩信表示网络的一条边.

移动社会网络行为是指移动用户在社会网络中通过手机进行社会交往所体现的交往模式、行为特征等.其表现形式主要有2种:1)从单个节点的角度出发所具有的表现形式有交往的范围、交互的次数、活跃度等;2)从整个网络的角度所具有的表现形式有节点在网络中的影响力等.利用社会网络理论来科学地、准确地发现和解释手机用户某些特定的社会关系,然后采用统计学方法来挖掘社会网络中各种隐含关系,为移动运营商了解客户的消费心理和消费需求提供依据,还为防治手机病毒的传播具有非常重要的价值.

本文主要对移动用户在使用 SMS/MMS 通信方式来进行社会交往的行为予以分析,主要包括交往的范围、交互的次数、活跃度 (activity degree, AD)、亲密度 (intimacy degree, ID)、聚类系数等.采用 SMS/MMS 通信记录来构建社会关系图,并对手机用户发送短信/彩信的行为进行统计和分析,来刻画移动用户之间的社交行为,从而进一步对移动用户的交往模式和行为特征进行分析.

1 相关工作

文献[2]通过对短信网络的数据进行分析,提出了一种基于局部优先连接机制的网络生长模型.该模型在度的分布、节点的度与节点加入网络时间的关系、平均度的增长等方面描述了以人际关系为基础的短信通讯网络.文献[3]提出了一种基于短信网络的动态演化模型.该模型通过对短信网络内部演化和节点退出现象进行分析以刻画短信网络的动态演化特点.文献[4]提出了一种加权短信网络演化模型.该模型在处理新节点加入时,采用加权局部优先连接机制;在处理边权更新时,采用节点间亲密度及近期联系频繁度.文献[5]采用统计分析方法并分别从交往圈和交往频度2个角度对短信社会网络进行分析,为手机短信客户细分提供了一种研究思路.还通过对出度、入度、交往圈大

小、入出边比、新联系人比、出边平均权重和出边权重方差等行为特征的统计分析,来刻画社会网络行为.文献[6]提出了一种模拟实际生活的短信息传播模型.该模型根据实际生活中手机用户的性格和转发短信息的习惯,结合复杂网络理论和传染病模型来分析短信息的传播行为.

2 社会网络图的构建

为了刻画用户间的社会交互情况和量化用户在社会网络的活跃程度,本文引入社会关系图.社会关系图可以通过有向图 $G = (V, E, W)$ 来表示, V 表示移动社会网络中的节点集合,即表示网络中的所有手机; E 表示网络中节点间的有向边集合,即表示手机用户之间发送短信的行为; W 表示有向边的权重,即表示发送短信的数量.

本文所采用的数据集是某地区电信公司在2012年9~10月中3周的 SMS/MMS 通信记录,共有约40万的手机用户、2000万条 SMS/MMS 通信记录.为了保护个人隐私,没有抽取短信内容,并对 SMS/MMS 数据中发送号码和接收号码进行伪码处理,有效地保证了电话号码不会被泄露;而且没有对个体的短信通信行为进行分析.

为了更好地阐述基于 SMS/MMS 通信的移动社会网络的构建,本文从数据集中抽取10个用户在一周内相互发送的 SMS/MMS 情况如表1所示.

表1 10个手机用户发送 SMS/MMS 情况

通信终端	周交互次数	通信终端	周交互次数
$A \rightarrow B$	2	$F \rightarrow E$	1
$A \rightarrow C$	6	$F \rightarrow G$	14
$A \rightarrow D$	4	$F \rightarrow H$	0
$B \rightarrow A$	5	$F \rightarrow J$	5
$B \rightarrow D$	15	$G \rightarrow C$	8
$B \rightarrow E$	2	$G \rightarrow D$	15
$C \rightarrow A$	4	$G \rightarrow F$	22
$C \rightarrow D$	8	$G \rightarrow J$	6
$C \rightarrow G$	12	$G \rightarrow I$	6
$C \rightarrow I$	3	$H \rightarrow E$	6
$D \rightarrow A$	9	$H \rightarrow F$	10
$D \rightarrow B$	6	$H \rightarrow J$	12
$D \rightarrow C$	3	$I \rightarrow C$	7
$D \rightarrow F$	4	$I \rightarrow G$	0
$D \rightarrow G$	13	$I \rightarrow J$	6
$E \rightarrow B$	3	$J \rightarrow F$	9
$E \rightarrow F$	8	$J \rightarrow G$	20
$E \rightarrow H$	8	$J \rightarrow H$	13
$F \rightarrow D$	8	$J \rightarrow I$	0

从表1可以得到每部手机与其它手机之间在每周的交互次数.如果把每部手机抽象成一个顶点,并根据随机图和有向图的思想,可得如图1所示的社会关系图 $G(V, E, W)$.

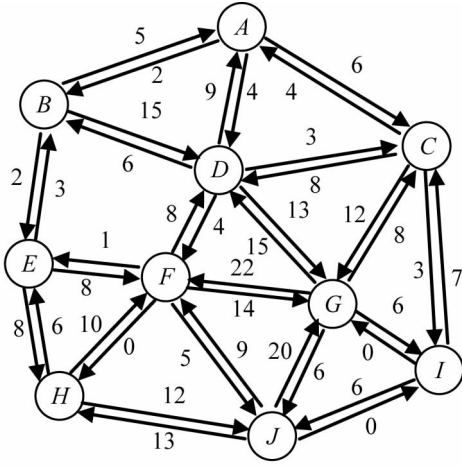


图1 带权的有向社会关系图

3 移动社会网络行为分析机制

3.1 相关要素分析

一般情况下,对移动社会网络行为的评价需要考虑每个节点的朋友个数、节点之间的交互次数、节点的活跃度等几个要素.1)朋友节点的个数.如果某节点交往的范围广,那么其朋友节点的个数就越多,也体现了与该节点对应的手机用户社会人际关系好.2)节点之间的交互次数.在单位时间内,节点之间交互次数越多,表明它们之间的关系越亲密.如果该节点与其它节点的关系都很亲密,那么该节点的社会影响力也越大.3)节点的活跃度.在单位时间内,某个节点与其交互的朋友节点个数以及该节点与其朋友节点之间相互发送短信/彩信的数量越多,说明该节点在该单位时间内越活跃,对移动社会网络的影响也越大.4)聚类系数,也称为集聚系数,是表示图的顶点聚集成团的倾向^[7-9].根据聚类系数的计算范围,可以将其细分为如下3个系数:整体聚类系数、平均聚类系数、局部聚类系数.

从以上分析可知,可通过统计每一个手机用户的朋友数、单位时间内总的有效交互数等指标来度量用户的活跃度.因此,对移动社会网络行为进行分析时,需要对朋友节点的个数、节点之间的交互次数、节点的活跃度等要素加以考虑.综合这些要素,本文提出了基于社会关系图的移动社会网络行为分析机制.

3.2 计算模型

假设 C_{ij} 表示从用户 i 发送到用户 j 的短信/彩

信数量.由于社会关系图中信息交互具有相互性,单方面的主动交互行为并不能表示该短信/彩信发送者是接收者的朋友或者他们之间关系很亲密.因此,可用 $\min\{C_{ij}, C_{ji}\}$ 表示用户 i 和用户 j 之间有效交互次数.假设 N_i 表示节点 i 的朋友数.那么节点 i 的朋友节点的个数可表示为

$$f_{ij}(t) = \begin{cases} 1, & \min\{C_{ij}, C_{ji}\} > 0; \\ 0, & \min\{C_{ij}, C_{ji}\} = 0. \end{cases} \quad (1)$$

$$N_i = \sum_{j=1}^N f_{ij}. \quad (2)$$

手机用户之间可以通过发送短信/彩信来进行留言、邀请、分享等社交活动.这些活动的频繁度可以真实地反映用户之间的亲密程度,记为亲密度 ID .在有向带权图 $G(V, E, W)$ 中,节点 i 与节点 j 的亲密度 ID_{ij} 可表示如下:

$$ID_{ij} = \frac{\min\{C_{ij}, C_{ji}\}}{\max\{\min\{C_{uv}, C_{vu}\}\}}. \quad (3)$$

其中, $i, j, u, v \in N$, ID_{ij} 表示节点 i 与节点 j 之间的亲密度.

如果某些手机用户可能拥有很多朋友,但该用户在很长时间内都没有与这些朋友们发送短信/彩信,那么该用户对移动社会网络的贡献也将较小.为了更好地刻画这种情况,引入活跃度 AD 来判断哪些节点更活跃.

假设 S_i 表示手机用户 i 与所有朋友节点的有效交互总数,节点 i 的活跃度可表示为

$$S_i = \sum_{k \in N_i} \min\{C_{ik}, C_{ki}\}. \quad (4)$$

$$AD_i = w_1 \frac{S_i}{\max\{S_u\}} + w_2 \frac{N_i}{\max\{N_u\}}. \quad (5)$$

其中, $i, u \in N$, S_i 表示节点 i 发送短信的总数; $\max\{S_u\}$ 表示在短信网络中所有节点发送短信数量的最大值; $\max\{N_u\}$ 表示所有节点中拥有朋友个数的最大值; w_1, w_2 是权重因子,且 $w_1 + w_2 = 1$,如果强调发送次数,则 w_1 取较大值;如果强调朋友的个数,则 w_2 取较大值.

由式(1)、式(2)分别可得手机用户之间的有效交互次数和每个手机用户的朋友数,具体情况如表2、表3所示.

当 $w_1 = 0.7, w_2 = 0.3$ 时,每个手机用户的活跃程度如表4所示.

由式(5)、表3及表4可知,如果手机用户的交往范围越大,则该用户的朋友数就越大,其活跃度也会相应地变大;如果2个用户之间交互的次数

越多,则他们各自总的有效交互数也越大,其活跃度也会相应地变大.

表2 手机用户之间的有效交互次数

终端	有效交互次数	终端	有效交互次数
A, B	2	E, F	1
A, C	4	E, H	6
A, D	4	F, G	14
B, D	6	F, H	0
B, E	2	F, J	5
C, D	0	G, I	0
C, G	8	G, J	6
C, I	3	H, J	12
D, F	4	I, J	0
D, G	13		

表3 每个手机用户的朋友数

终端	朋友数	终端	朋友数
A	3	F	4
B	3	G	4
C	4	H	2
D	5	I	1
E	3	J	3

表4 手机用户活跃度表

终端	活跃度	终端	活跃度
A	0.350	F	0.649
B	0.350	G	0.940
C	0.547	H	0.427
D	0.812	I	0.111
E	0.333	J	0.572

4 实验分析

首先采用 MATLAB 对手机用户短信通信行为进行了理论分析,给出了AD中2个参数的值为: $w_1 = 0.7, w_2 = 0.3$. 由于所采集的通信记录数据集比较庞大,因此,本文采用随机算法从中抽取 23 767 个用户的 SMS/MMS 通信记录来构建移动社会网络.通过采用 Visual C++ 6.0 开发了一个模拟器来进行实验验证.

首先,本文利用该数据集对移动社会的平均聚类系数进行了分析,实测值为 0.59. 表明了该数据集的平均聚类系数比较高,具有小世界网络的特性.

入度、出度的概率密度分布分别如图 2、图 3 所示.

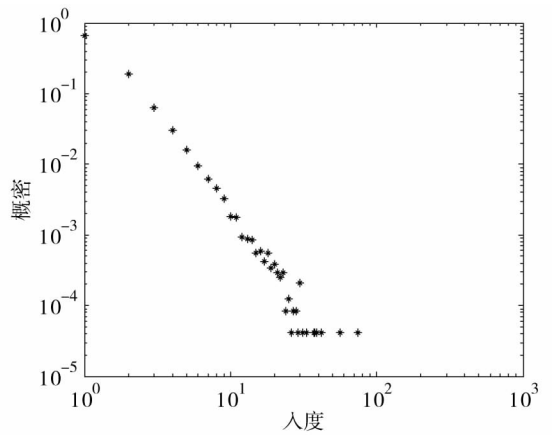


图2 入度概率密度分布

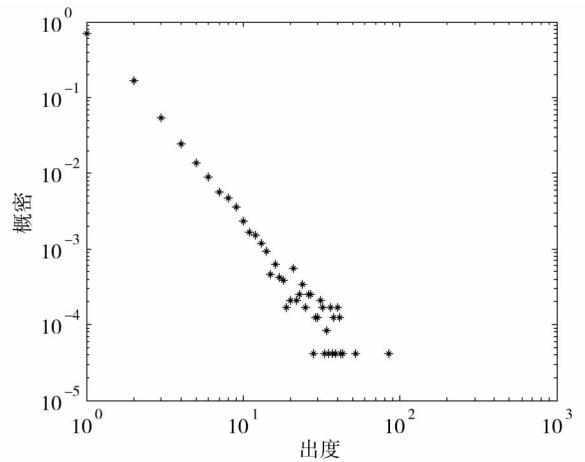


图3 出度概率密度分布

由图 2、图 3 可知,节点度数的分布情况具有幂律分布特性,表明了该移动社会网络是一个无标度网络.

通过对 3 周的 SMS/MMS 通信记录进行统计与分析,手机用户的朋友个数统计情况如图 4 所示.

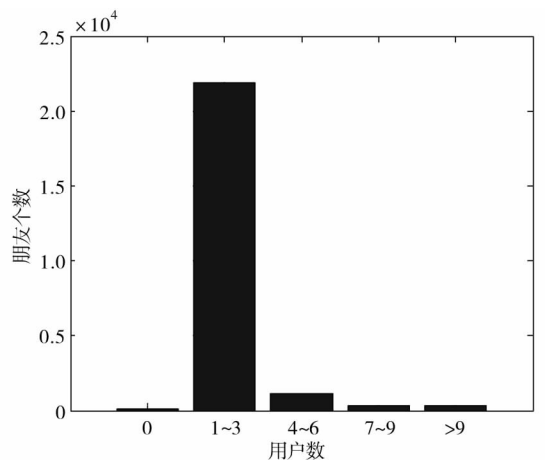


图4 手机用户朋友个数统计

由图4可知,在该短信网络中,大部分手机用户的友数为1~3个,朋友数超过9个的手机用户非常少.表明了大部分手机用户的社交范围较小,只与少数的手机用户通过短信进行社会交往.这也进一步表明了图2、图3结果是正确的.手机用户发送短信/彩信的时间分布规律如图5所示.

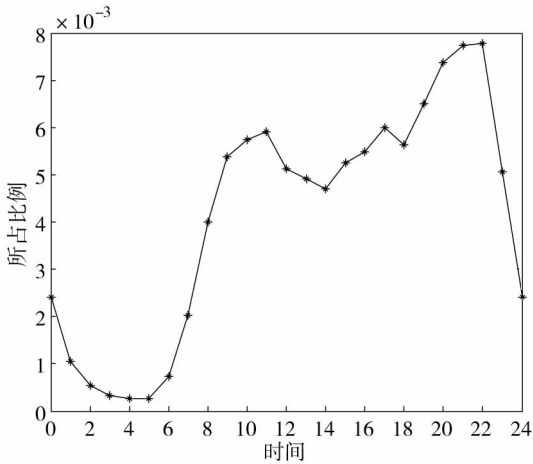


图5 短信/彩信发送时间分布

由图5可知,用户在早上0点至7点时,发送的短信数量是比较少的,因为在这个时间段内大部分手机用户还在休息;在上午9点至11点半以及在下午15点至18点时,发送的短信数量是比较多的,因为在这个时间段内大部分手机用户都已起床并开始一天的工作;在晚上19至22点时,发送的短信数量是在一天最多的,因为此时大部分手机用户都已吃完晚饭并可能在家休息,所以才有更多的时间与自己的朋友进行联系.节点活跃度的分布情况如图6所示.

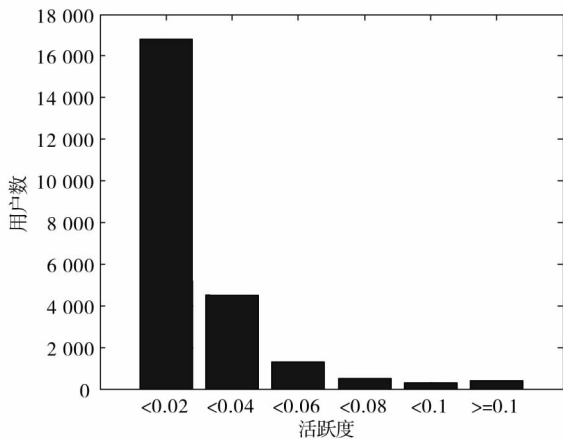


图6 节点活跃度分布

图6表示了节点活跃度分布情况.由图6可知,大部分节点的活跃度都小于0.02,活跃度大于0.06的节点是比较少的,这与节点的度具有幂律分布特性也是一致的.

5 结论

本文引入了社会网络理论对移动社会网络行为进行建模与分析,在构建社会关系图的基础上,提出了移动社会网络行为分析机制.在该机制中,探讨了移动社会网络分析的相关要素并给出了活跃度的计算方法.最后,采用了实际的短信通信记录数据集对出度、入度、用户朋友数、活跃度、亲密度、聚类系数等特征进行实验验证.本文所提出的研究方法可以进一步应用于对其它社会网络行为进行分析.在今后研究工作中,将在移动社会网络行为分析机制的基础上,对如何评价节点的信任度^[10]、如何挖掘影响力最大的节点以及如何预测短信的传播路径进行研究.

参考文献:

- [1] Peng S C, Wang G J, Yu S. Modeling malware propagation in smartphone social networks[C]// 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2013), Australia, 2013.
- [2] 吴晔,肖井华,吴智远,等.手机短信网络的生长过程研究[J].物理学报,2007,56(4):2037-2041.
- [3] 孙钦东,孙亚红,管晓宏,等.动态短信通信复杂网络演化模型研究[J].西安交通大学学报,2009,43(6):5-9.
- [4] 刘星宏,秦晓卫,陈锋.短信网络的加权演化模型研究[J].电子科技大学学报,2012,41(5):649-657.
- [5] 柳国华,谢璨,英春.基于短信的社会网络行为分析[J].计算机应用与软件,2011,9(3):117-119.
- [6] 李明杰,吴晔,刘维清,等.手机短信息传播过程和短信息寿命研究[J].物理学报,2009,58(8):5251-5258.
- [7] 杨偶,李朝奎,李拥.复杂校园网络实证研究[J].湖南科技大学学报(自然科学版),2012,27(1):79-83.
- [8] 汪小帆,李翔,陈关荣.复杂网络理论及其应用[M],北京:清华大学出版社,2006.
- [9] Watts D J, Strogatz S H. Collective dynamics of small-world networks[J]. Nature, 1998, 393: 440-442.
- [10] 彭三城,陈建平,胡忠望.移动自组网中基于多约束和协同过滤动态信任机制[J].计算机应用研究,2011,28(3):1065-1069.