

手机短信网络的生长过程研究*

吴 晔 肖井华 吴智远 杨俊忠†

(北京邮电大学理学院, 北京 100876)

马宝军

(中国联合通信有限公司北京分公司, 北京 100038)

(2006 年 6 月 29 日收到, 2006 年 11 月 15 日收到修改稿)

模拟了短信网络的生长过程并研究其拓扑结构. 发现短信网络在生长过程中, 度分布、节点的度与其加入网络时间的关系、平均度随时间的变化等方面与先前的模型有所不符. 根据短信网络的数据分析, 提出了短信网络的生长机制——局部优先连接机制. 结果表明, 模拟以短信网络为代表的实际网络时, 局部优先连接模型优于其他网络模型.

关键词: 复杂网络, 短信网络, 局部优先连接

PACC: 0590, 0175

1. 引 言

1998 年 Watts 和 Strogatz^[1]指出, 许多现实世界中的复杂网络不是规则网络或者随机网络, 而是小世界网络, 即具有小的最短路径和高的聚类系数的一种网络结构. 由此揭开了对复杂网络研究的序幕. 1999 年, Barabási 和 Albert 指出^[2], 许多现实世界的复杂网络还具有无标度的特点, 进一步引起了人们对复杂网络研究的兴趣. 他们同时指出, 决定互联网、万维网和科学家合作研究网等具有无标度特性网络的两个基本性质是节点增长与优先连接. 依照这两个性质所生成的网络(BA 模型)可以用来描述许多真实的网络, 但仍然有一定的局限性. 例如, BA 模型生成的网络的度分布指数是 3, 而现实中的各种复杂网络中的标度指数则是在 1—3 范围内变化^[3—10], 有些还具有一定的非幂律特性, 如指数截止、小 k 饱和等情形^[5, 11, 12]. 所以研究人员提出了许多不同于 BA 模型的网络生长机制, 试图使其与现实情况更为相符. 已经提出各种类型的机制包括优先连接、内部演化、局部事件、竞争等机制^[13—19]. 这些机制在某些方面能与实际情况很好地相符, 但存在各自的局限性. 在本文中, 首先研究一个实际网

络——短信通信网络的生长过程. 在模拟其生长过程中, 发现这一类型的网络具有自己的特点, 即度分布并非为严格的幂律分布, 节点的度与节点加入网络时间的关系涨落大, 平均度是增长的. 过去提出的这些机制不能很好地描述这类网络的特点. 因此, 我们提出了一个新的生长机制——局部优先连接机制, 并构造了一个生长模型. 数值模拟结果表明, 这个模型在度分布、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, 比其他模型更能真实地描述以人际关系为基础的短信通讯网络.

2. BA 模型

BA 模型的生长分为生长和优先连接两个过程.

(1) 生长过程. 初始有一个节点数比较少的随机网络, 节点数为 m_0 , 边数为 n_0 , 在每个时间步增加一个新的节点, 同时以这个节点为一个端点连出 m ($m < m_0$) 条边, 边的另一端为网络中原来存在的节点.

(2) 优先连接过程. 选择另一端的规则是一个点被选择到的概率与这个端点的度成正比. 这样, 在 t 时刻网络的节点数 $N = m_0 + t$, 边数为 $mt + n_0$, 度分布 $p(k)$ 出现幂律特征

* 国家自然科学基金(批准号: 1045004, 10575016, 70431002)资助的课题.

† E-mail: jzyang@bupt.edu.cn

$$p(k) \propto k^{-\gamma},$$

其中指数 $\gamma = 3$.

图 1 为数值模拟 BA 模型得到的结果.

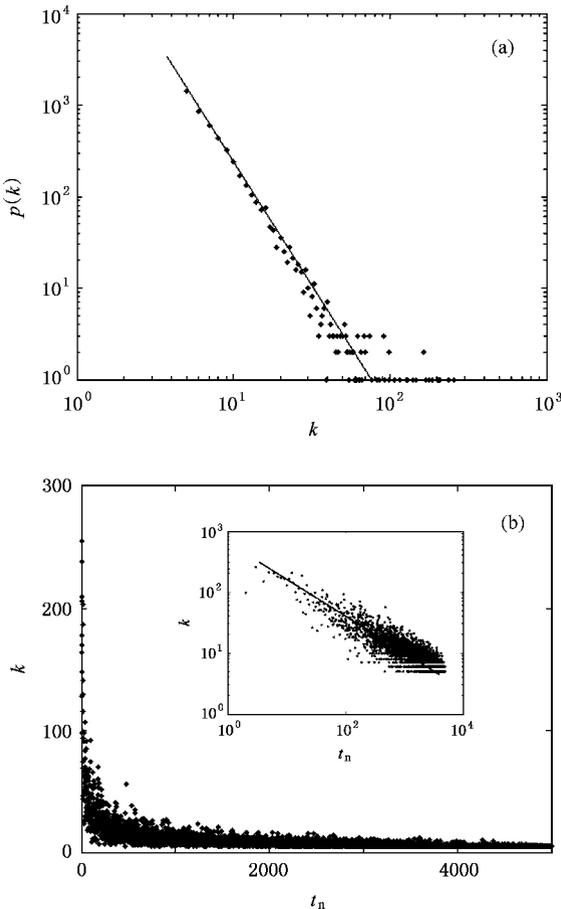


图 1 模拟 BA 模型的结果 (a) 为度分布 $p(k)$, 网络参数 $m_0 = 6, m = 5, n = 5000$ 图中直线斜率为 -3 , 表示无标度规律 (b) 在时间步为 5000 时, 节点度与该节点加入网络的时间 t_n 的关系, 内插图为双对数坐标下的情况

从图 1 可以看出 BA 模型的一些特点. 第一, 度分布 $p(k)$ 为严格幂律分布. 第二, 如果按加入网络的顺序排列, 在最后时刻网络中各个节点的度出现了很好的幂律衰减. 这意味着越早加入网络的点, 拥有越大的度. 回顾 BA 模型的生长过程可以看出 BA 模型的第三个特点, 即平均度基本不变, BA 模型每过一个时间步, 就增加了一个点, 并增加了 m ($m < m_0$) 条边. 假设初始有 n_0 条边, 那么经过 t_n 个时间步后网络的平均度可表示为

$$k = \frac{(n_0 + t_n \times m) \times 2}{m_0 + t_n}. \quad (1)$$

从 (1) 式可以看出, 当 n 增加到足够大以后, 平均度是不变的. BA 模型在许多时候能符合实际的网

络生长过程, 但 BA 模型的以上三个特点不符合一些实际网络的特征. 对比 BA 模型的生长过程和短信网络的生长过程, 我们发现 BA 模型的这三个特点明显不符合短信网络.

3. 短信通信网络

随着个人移动通信技术的发展, 中国大陆目前的手机用户数至少是 3.5 亿, 而文本短信来往已经成为用户间通信的重要方式. 据统计, 2005 年全国的短信量已经达到了 3046 亿条. 对短信网络的研究不但对经济的发展具有重要的意义, 也会使我们对移动通信网络用户间的网络结构模式和演化特点有更清晰的认识和理解. 短信网络背后的深层机制就是人际关系. 短信网络生长过程的研究对研究人际关系网的特点也很有帮助. 现在分析网络的一个有效工具就是复杂网络理论. 我们以一个集团内用户间互相发送短信的计费数据为研究对象, 用复杂网络的理论重现了短信网络的生长过程, 数据源的格式如表 1 所列.

表 1 用户话单的格式

发送号码	接收号码	发送时间
133 * * * * 0123	133 * * * * 2210	2005 * * * *

表 1 中第一列为发送用户手机号码, 第二列为接收用户手机号码, 第三列为该条短信发送的时间 (精确到 1 s). 如果把用户看作网络的点, 当用户甲和用户乙之间存在短信往来时, 甲和乙之间形成一条边, 这样就形成了一个网络^[20]. 这里, 为了简化研究同时也考虑人际关系的性质, 因而没有考虑网络的方向性以及边的权重, 故我们构造的网络为无向无权网络. 本文所使用的两组数据, 分别为 A 公司和 B 公司一个月的短信计费记录. A 公司总记录数为 398185 条, B 公司总记录数为 548182 条. 话单的最后一个是发送时间, 利用该时间点来重现网络随时间的生长过程. 按照时间的先后顺序, 将对应的点和边加入到短信网络中. 在一个月后, A 公司共有 31096 个用户, 其中有 7808 个用户构成了一个最大的连通网络, 而其他用户或为散户或构成了一个个小网络集团. B 公司共有 44430 个用户, 最大的连通网络的用户数为 20112 个. 研究上述最大连通网络集团的生长过程, 得到结果如图 2 所示.

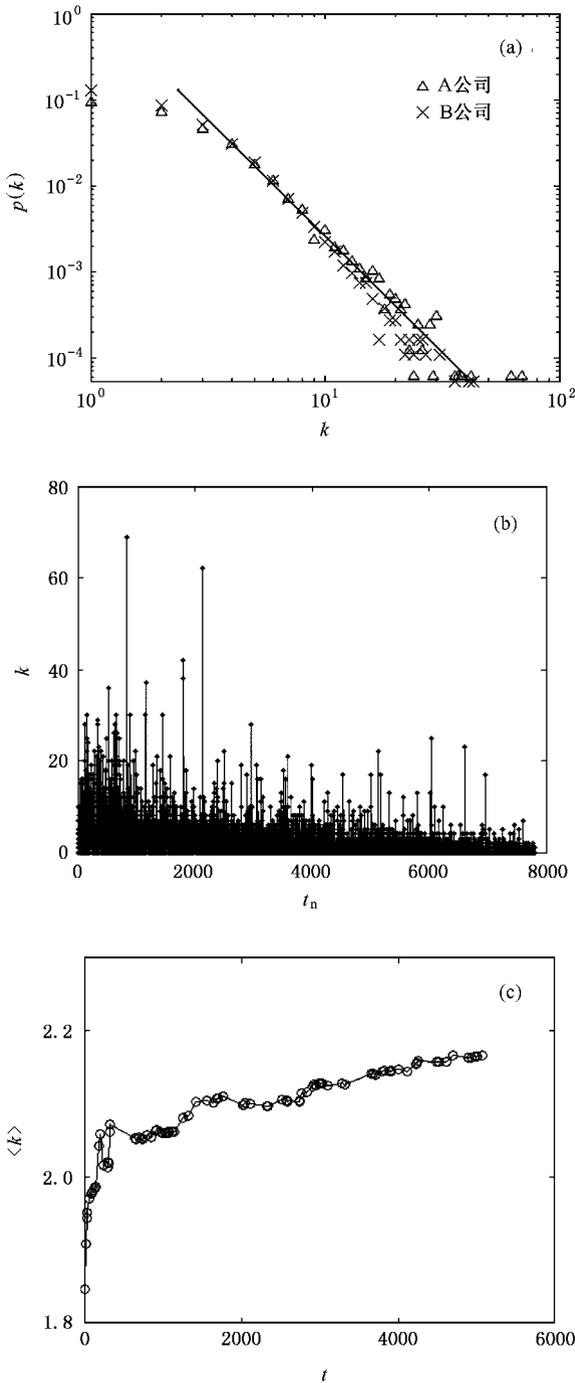


图2 所研究的短信网络的生长过程 (a)A公司和B公司的短信网络的度分布 (b)每个节点在最后时刻的度与该节点加入网络时间的关系 (c)平均度 k 增长图

对比短信网络生长过程的特点可以发现,上述BA模型生长过程的三个特点是不符合短信网络的生长过程的.首先,短信网络在度较大之处是幂律分布的,图2(a)中拟合曲线的斜率为 -2.78 ± 0.07 ,但总体度分布不是严格的幂律分布.现在已发现很多

网络的度分布不是严格的幂律分布^[5,11-19].其次,节点的度与节点加入网络时间的关系的涨落比较大.大量的事实表明,在实际网络中,节点的度不仅仅只是依赖于节点加入网络的时间.例如,在万维网中一些网页可能会在短时间内就形成大量的边^[21].第三,平均度是增长的.现在发现许多实际网络的平均度是增长的^[7,8,22].研究人员提出了很多改进的网络生长机制^[11-19]来弥补这些不足,但都不够完备,没有一种机制能完全弥补这三个不足.这里,基于人际交往关系的特点,我们提出了一个网络生长机制——局部优先连接机制,并用该机制构造了一个网络生长模型,很好地解决了BA模型的三个不足之处.

4. 局部优先连接模型

1)增长.初始有一个概率为 P_0 的随机网络,网络的节点数比较少.在每个时间步,往网络里加入一个新节点.

2)局部优先连接.新加入的点与已存在点的连接规则如下:首先随机连上一个点 A ,然后以一定的概率 P_1 与点 A 的邻居相连,接着以更小的概率 P_2 与刚连上的 A 的邻居的邻居相连.

这两个机制在很多现实网络的生长过程中也是经常看到的.比如在人际关系网中,一个人因为认识这个网络中的某个人而加入到这个网络中.同时,随着时间的演化,他以一定的概率认识该人的朋友,然后以较小的概率认识该人的朋友的朋友.

我们对局部优先连接模型做了数值模拟,得到的结果如图3所示.

从图3(a)可以看出,度分布不是严格的幂律分布,而是在度较小之处有偏离现象.对比图2可以看出,模型的度分布和短信网络的度分布基本上是一致的.从图3(b)可以看出,节点的度与节点加入网络时间的关系不像BA模型那样衰减很快,而是涨落很大,很好地符合了短信网络的生长过程.图3(c)中模型的平均度是缓慢增加的,也很好地符合了短信网络的生长过程.

图4为度分布的斜率与模型主要参数 P_0 , P_1 , P_2 的关系.

从图4可以看出,模型的度分布基本上不随着 P_0 变化,但会随着 P_1 的增大而增大,也随着 P_2 的

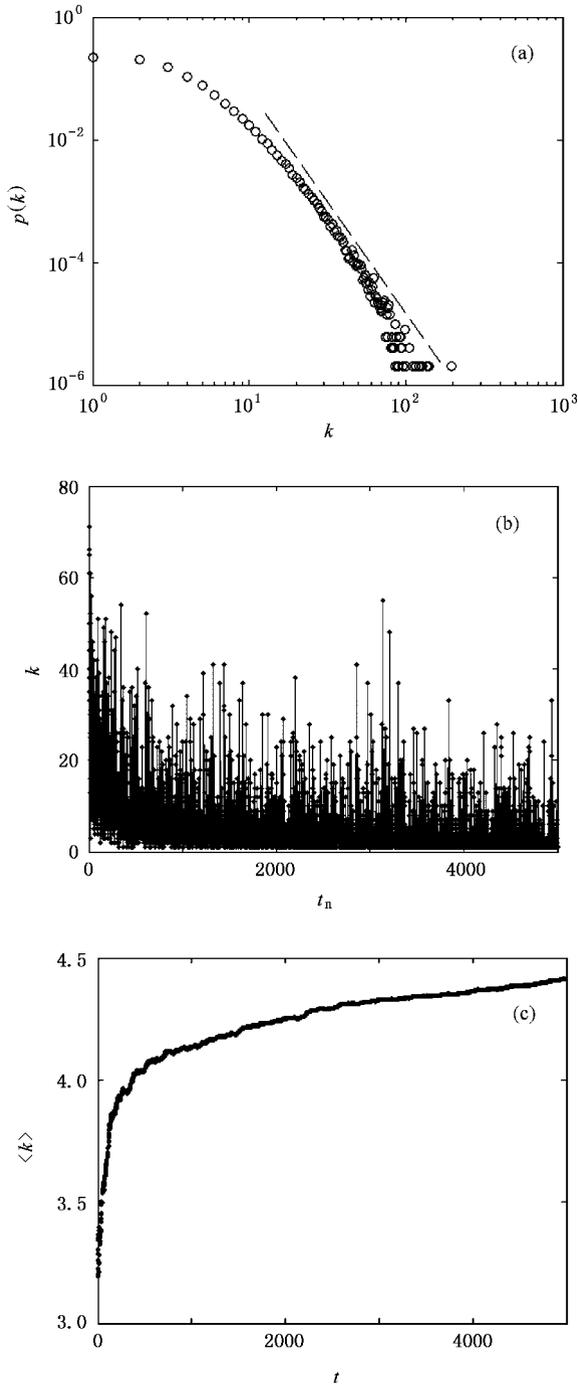


图3 局部优先连接模型模拟结果 初始条件和参数:初始网络 $m=10$,节点间以概率 $P_0=0.2$ 随机连接, $P_1=0.2$, $P_2=0.06$. (a)度分布,图中虚线斜率为-2.76 (b)每个节点在最后时刻的度(按加入网络的顺序排序)(c)平均度 k 的增长

增大而增大.这说明当一个人认识他的朋友的朋友的概率比较大的时候,整个网络里朋友多的人也就多了,亦即网络中节点的度普遍较大.在现实中也是如此的结果.

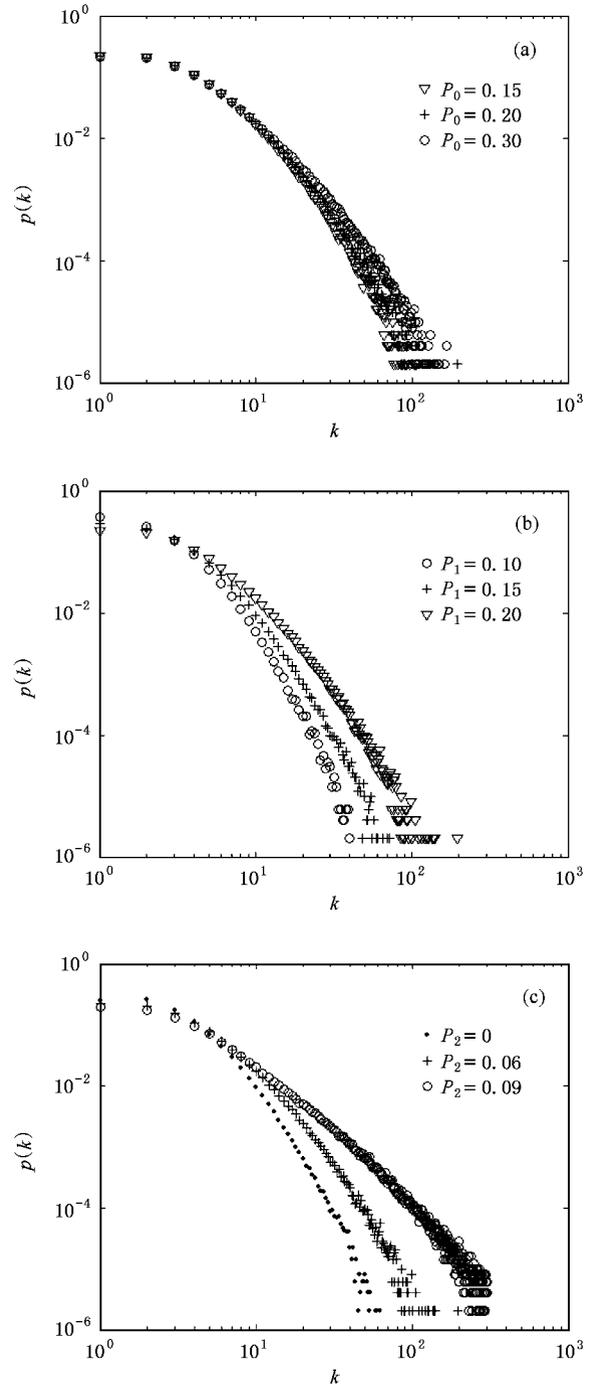


图4 局部优先连接模型的度分布与参数 P_0, P_1, P_2 的关系

(a)不同 P_0 时的度分布,初始网络 $m=10$, $P_1=0.2$, $P_2=0.06$; (b)不同 P_1 时的度分布,初始网络 $m=10$, $P_0=0.2$, $P_2=0.06$; (c)不同 P_2 时的度分布,初始网络 $m=10$, $P_0=0.2$, $P_1=0.2$

5. 结 论

本文比较了不同模型在度分布、节点的度与其

加入网络时间的关系、平均度增长等方面的不同情况. 为了弥补现有网络模型在这些方面的不足,提出

了基于局部优先连接机制的一种网络生长模型. 这种模型更好地符合了手机短信网络等许多实际网络.

- [1] Watts D J , Strogatz S H 1998 *Nature* **393** 440
- [2] Barabási A L , Albert R 1999 *Science* **286** 509
- [3] Kleinberg J , Lawrence S 2001 *Science* **294** 1849
- [4] Wang X F , Chen G R 2003 *IEEE Circ. Syst. Mag.* **3** 6
- [5] Masuda N , Konno N 2004 *Phys. Rev. E* **69** 0319173
- [6] Zhang P P , He Y , Zhou T *et al* 2006 *Acta Phys. Sin.* **55** 60 (in Chinese) [张倍倍、何 阅、周 涛等 2006 物理学报 **55** 60]
- [7] Broder A , Kumar R , Maghoul F *et al* 2000 *Comp. Netw.* **33** 309
- [8] Faloutsos M , Faloutsos P , Faloutsos C 1999 *Comp. Commun. Rev.* **29** 251
- [9] Newman M E J 2001 *Phys. Rev. E* **64** 016131
- [10] Jeong H , Tombor B , Albert R *et al* 2000 *Nature* **407** 651
- [11] Amaral L A N , Scala A , Barthélemy M *et al* 2000 *Proc. Natl. Acad. Sci. USA* **97** 11149
- [12] Newman M E J 2001 *Phys. Rev. E* **64** 016132
- [13] Dorogovtsev S N , Mendes J F F 2000 *Phys. Rev. E* **63** 025101
- [14] Bianconi G , Barabási A L 2000 *Phys. Rev. Lett.* **85** 5234
- [15] Krapivsky P L , Rendner S , Leyvraz F 2000 *Phys. Rev. Lett.* **85** 4629
- [16] Huang Z X , Wang X R , Zhu H 2004 *Chin. Phys.* **13** 273
- [17] Wang W N , Lee J Y , Mi Z K 2004 *Chin. Phys. Lett.* **21** 243
- [18] Bianconi G , Barabási A L 2001 *Europhys. Lett.* **54** 436
- [19] Jost J , Joy M P 2002 *Phys. Rev. E* **66** 036126
- [20] Ma B J , Xiao J H , Yang J Z 2006 *Dynam. Cont. Dis. B* **13** 371
- [21] Serrano M A , Boguna A , Diaz-Guilera A 2005 *Phys. Rev. Lett.* **94** 038701
- [22] He Y , Zhang P P , Xu T *et al* 2004 *Acta Phys. Sin.* **53** 1710 (in Chinese) [何 阅、张倍倍、许 田等 2004 物理学报 **53** 1710]

Research on the growing process of short message networks^{*}

Wu Ye Xiao Jing-Hua Wu Zhi-Yuan Yang Jun-Zhong[†]

(School of Science , Beijing University of Posts and Telecommunications , Beijing 100876 , China)

Ma Bao-Jun

(Beijing Division , China United Telecommunications Corporation , Beijing 100038 , China)

(Received 29 June 2006 ; revised manuscript received 15 November 2006)

Abstract

We investigate the growing process and topological features of the short message networks (SMN). Simulations show that the networks' degree distribution, average degree and the correlation between the node's degree and its time in the network are different from those of the former networks models. With due consideration of some of the characteristics of the SMN, we suggest a new type of network growing protocol, namely the local-preference connecting protocol. The numerical results show that the networks generated from our model are more representative in simulating real networks represented by SMN when compared with other network models.

Keywords : complex networks , short message networks , local-preference connecting

PACC : 0590 , 0175

^{*} Project supported by the National Natural Science Foundation of China (Grant Nos. 10405004 , 10575016 , 70431002).

[†] E-mail : jzyang@bupt.edu.cn