

短信网络的加权演化模型研究

刘星宏¹, 秦晓卫¹, 陈 锋¹, 骆培杰², 戴旭初¹

(1. 中国科学技术大学电子工程与信息科学系 合肥 230027; 2. 中国科学技术大学计算机科学与技术学院 合肥 230027)

【摘要】通过对实际数据的分析, 获得了短信网络的一些特点, 包括: 度分布和边权分布符合低头和重尾的幂律分布、平均点强度和点度不具有幂函数律的关联性等等。在此基础上, 将用户间短信交互次数视为短信网络的边权, 提出了一种加权短信网络演化模型, 其中新节点加入采用加权局部优先连接机制, 边权更新基于节点间亲密度及近期联系频繁度。仿真结果表明, 该生成模型较好地符合了实际短信网络的统计特性。

关键词 近期联系频繁度; 短信网络; 加权局部优先; 边权更新

中图分类号 O41; TN92

文献标识码 A

doi:10.3969/j.issn.1001-0548.2012.05.002

Research on the Weighted Evolutionary Model of Short Message Networks

LIU Xing-hong¹, QIN Xiao-wei¹, CHEN Feng¹, LUO Pei-jie², and DAI Xu-chu¹

(1. Department of Electronic Engineering and Information Science, University of Science and Technology of China Hefei 230027;

2. School of Computer Science and Technology, University of Science and Technology of China Hefei 230027)

Abstract By analyzing some factual data from short message service database, more features of short message networks (SMNs) are captured, including degree distribution as well as weight distribution demonstrating the behavior of power-law with droop-head and heavy-tail, average vertex weight with vertex degree without power-law behavior etc. And then, a weighted evolutionary model (WEM) is proposed for SMNs by taking the short message interactive times between users as weights. The proposed WEM adopts a weighted local priority mechanism for node growing and a weight updating scheme which is based on the familiarity and recent contact frequency between users. Simulation results show that WEM fits real SMNs pretty well.

Key words recent contact frequency; SMS network; weighted local priority; weight updating

小世界网络^[1]和无标度网络^[2]被视为复杂网络研究的起始标志。经过近十年来的发展, 针对各类复杂网络的研究已取得了一系列令人瞩目的成就, 如社会网络、信息网络及生物网络等等。其中, 社会网络^[3-4]作为复杂网络的一个重要代表, 反映了人的行为及其相互关系。研究社会网络的目的是挖掘隐藏在社会网络下的各种隐性关系, 将其广泛地服务于各种技术网络上的业务推广, 如电子商务^[5]、信息推荐^[6]等新型业务。

近年来, 随着移动通信技术的飞速发展及基础网络设施的逐渐完善, 移动通信用户的规模呈指数增长, 为移动增值业务的发展提供了有利条件。短信业务(short message service, SMS)因其价格低廉、准确性高、存储转发等优点, 已成为移动通信网中成熟的增值业务之一。结合复杂网络理论, 一些研

究者提出了移动通信网下短信网络的概念, 这里的短信网络(short message networks, SMNs)可理解为是一种基于移动通信系统的虚拟社会网络, 其中“节点”代表用户, “边”代表用户间有无短信往来。短信网络是一种反映人际关系的社会覆盖网络, 研究其拓扑特性及演化机制, 是实现移动通信复杂系统管理与控制的基础, 同时也对基于用户交互式的社会网络研究起着重要的推动作用。

目前, 已有文献针对短信网络展开了相关研究^[7-12], 如文献[7]提出了基于局部优先连接机制的短信网络生长模型, 并证明了人类行为特征符合双模分布, 为从个体及网络层面研究人类行为的多样性提供了一个全新的视角^[8]; 文献[9]提出了短信网络的一种动态演化模型, 在考虑节点加入和退出的情况下刻画了短信网络有增有减的动态演化特点;

收稿日期: 2012-01-28; 修回日期: 2012-09-09

基金项目: 国家自然科学基金(61101136), 中芬国际科技合作项目(2010DFB10570)

作者简介: 刘星宏(1985-), 男, 硕士, 主要从事无线业务建模、加权网络及动力学等方面的研究。

文献[10]利用联通某公司的短信数据库数据,通过仿真发现真实的短信网络是以边为生长单位,而不是以点为生长单位;文献[11]统计了千万用户数级别的大型 SMS 数据库,对 SMS 的传输可靠性及 SMNs 上的病毒传播等问题作了详尽的分析;文献[12]根据谣言短信的传播特点及方式,构建了谣言短信传播的网络模型,并通过连续域理论证明其具有无标度特性。

研究复杂网络的目标之一是理解复杂网络上的动力学行为,特别是理解网络拓扑结构对其动力学行为的影响。文献[13]讨论了加权无标度网络上的传播动力学特性,着重研究权重及其分布对传播行为的影响;文献[14]给出了一般权重分布下 SIS 模型基于边的平均场方程及其解;文献[15]研究了强弱链接对传播动力学的影响,得出强链接提高感染密度而弱链接加快传播进程的结论。因此,对实际网络进行尽可能精确且合理的建模,是复杂网络动力学研究的基础。

迄今为止,对短信网络模型的研究主要仍集中于无权网络的范畴,无权网络只能给出节点之间连接存在与否的定性描述,这种定性描述反映了网络拓扑结构最主要的信息。但在许多情况下,网络节点间相互作用强度的差异性对网络的动力学特征起着至关重要的作用,仅将实际系统抽象成无权网络往往会忽略某些重要的客观物理信息量。例如,在社会网络中,个体间接触频繁程度会对疾病传播产生重要的影响^[13]; Internet 网络上的带宽决定了数据传输速率及传输可靠性^[16];食物链网络中捕食关系及其强度的多样性是维持生态系统稳定的重要因素^[17]等等。此时,如果仅用是否连接表征节点间关系并不能准确反映实际网络的细致结构及其功能,需要引入“边权”(weight)来刻画相互作用强度的差异性。

因此,本文统计了实际 SMS 数据库记录,分析了由该用户群构成的短信网络的加权特性,将用户的短信联系人视为短信网络的节点度,将用户间短信交互的总次数视为短信网络的边权,发现该短信网络的度分布和边权分布均服从低头和重尾的幂律分布,及平均点强度和点度不具有幂函数律的关联性特征。基于此,本文提出了一种加权短信网络演化模型(weighted evolutionary model, WEM),该模型采用加权局部优先连接的增长机制,以及基于亲密度和近期联系频繁度的边权更新机制。数值模拟结果表明,WEM模型较好地符合了实际短信网

络的统计特性。

1 加权短信网络

以有权图 $G(V, E, W)$ 表示短信网络,其中 V 是节点集, E 是边集, W 是边权集, W 一般用加权邻接矩阵来表示。边权表示个体间相互作用的强度,分为相异权和相似权两种。社会网络的边权主要是相似权,即边权值越大,表明两节点的关系越亲密、交往越频繁,例如,科学家合作网的合作文章数、手机电话网的用户间通话时长等等。本文将两用户 i 和 j 之间的短信交互总次数 $w_{ij} \in W$ 视为短信网络节点 i 和 j 之间的边权值,用来表征用户间的亲密度;定义用户 i 使用短信的总次数 s_i 为该用户节点的点强度,表征用户使用短信的程度,由定义可知 $s_i = \sum_{j \in \Gamma_i} w_{ij}$, 其中 Γ_i 是节点 i 的近邻集合。

文献[18]对印度一家运营商的短信数据库进行分析,发现出、入点强度的分布及出、入度的分布均呈幂律分布,但该数据是 16 h 的短时记录,不能很好地反映出真实短信网络的生长过程。文献[19]发现在用户发送短信的行为中,用户相邻两次发送短信的时间间隔分布及相邻两次会话的时间间隔分布呈幂律分布,及一次会话中包含的短信数分布也服从幂律分布。文献[20]讨论了短信网络中的人类动力学问题,发现人类活动与时间间隔分布的幂指数存在较强的正相关,且短信系统呈现突发特性而不具有记忆效应。

本文利用韩国某城市某月 GSM 网络的 SMS 数据库中的数据^[21]和日本某城市某月 GSM 网络的 SMS 数据库中的数据^[22],分析了由各自用户群构成的短信网络的加权特性,数据库中数据的存储格式如表 1 所示。

表1 日本某SMS数据库的存储格式

发送号码	接收号码	短信交互次数
70918**3	70918**7	4
77987**8	77816**2	44

表1中第1列为发送用户的手机号码,第2列为接收用户的手机号码,第3列为两用户的短信交互次数。其中,韩国某SMS数据库格式与日本某SMS数据库格式相同,SMS数据库信息如表2所示。

表2 SMS数据库信息

数据库名称	用户数/位	记录数/条
韩国某 SMS 数据库	15 145	140 394
日本某 SMS 数据库	125 306	505 316

计算网络的主要拓扑性质,有助于直观了解该

网络的一些基本特性, 这些性质主要包括平均路径长度、平均聚类系数、模块度、匹配系数。通常, 平均路径长度 L 定义为:

$$L = \frac{1}{N(N+1)/2} \sum_{i \geq j} d_{ij} \quad (1)$$

式中, d_{ij} 是连接节点 i 和节点 j 的最短路径长度。

平均聚类系数 C 定义为:

$$C = \frac{1}{N} \sum_i c_i = \frac{1}{N} \sum_i \frac{1}{k_i(k_i-1)} \sum_{jh} a_{ij} a_{ih} a_{jh} \quad (2)$$

式中, c_i 是第 i 个节点的聚类系数; 节点 i 、 j 之间有边存在时, $a_{ij} = 1$, 否则 $a_{ij} = 0$ 。引入模块度来衡量网络划分质量的标准, 考虑某种划分形式将网络划分成 k 个社团, 定义一个 $k \times k$ 维的对称矩阵 $E = (e_{ij})$, 其中 e_{ij} 表示网络中连接社团 i 和社团 j 的节点的边在所有边中所占的比例。模块度 Q 定义为:

$$Q = \sum_i (e_{ii} - a_i^2) = \text{Tre} - \|e^2\| \quad (3)$$

式中, $a_i = \sum_j e_{ij}$ 为矩阵中各行元素之和;

$\text{Tre} = \sum_i e_{ii}$ 为矩阵中对角元素之和; $\|x\|$ 表示矩阵 x 中所有元素之和。文献[23]指出模块度大于 0.3 的网络就具有清晰的社群结构, 值越接近 1, 说明社群结构越清晰。匹配系数 r 定义为:

$$r = \frac{M^{-1} \sum_i j_i k_i - [M^{-1} \sum_i \frac{1}{2} (j_i + k_i)]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - [M^{-1} \sum_i \frac{1}{2} (j_i + k_i)]^2} \quad (4)$$

式中, j_i 和 k_i 分别为第 i 条边的两个端点的度数; M 为网络中边的条数; 匹配系数 r 反映了边连接的倾向性, 若 $r > 0$, 则网络是同配的, 表明度值相似的节点倾向于相互连接, 反之, 则是异配的。相关研究表明, 技术网络一般是异配的, 而社会网络通常是同配的。

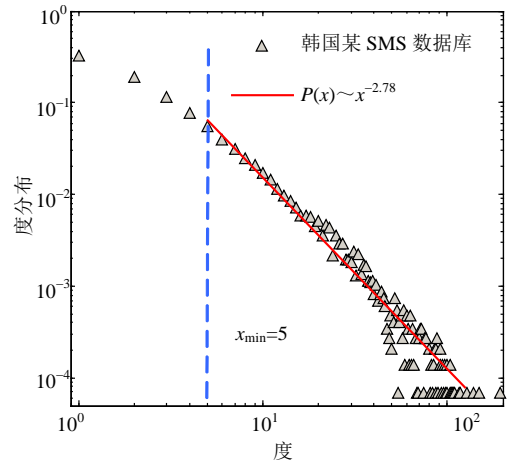
对这两个 SMS 数据库进行统计分析, 计算由相应用户群构成的短信网络的主要拓扑特性, 如表3所示。分别作出对应的度分布曲线和边权分布曲线, 如图1所示, 对于幂指数的估计, 可先假设某个分布是幂律的, 通过最大似然估计得到其幂指数, 并通过 K-S 检验^[24]判断该拟合分布是否符合幂律分布, 图1中的度分布和边权分布的幂律分布均可通过阈值为0.9的 K-S 检验。

另外, 从表3可知, 该SMS用户群构成的短信网络拓扑性质还呈现以下一些特点: 短的平均路径和大的聚类系数说明其具有小世界特性; 模块度远大于0.3说明其具有清晰的社群结构; 匹配系数为正说

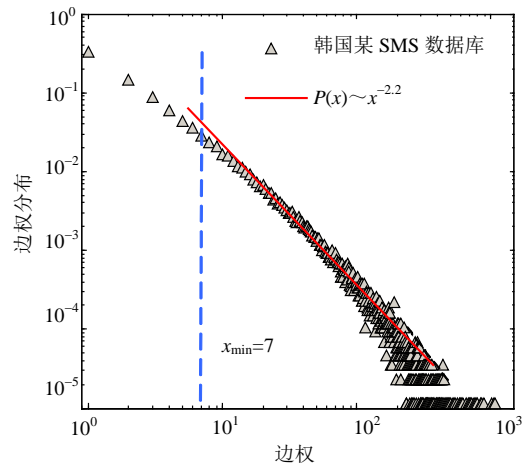
明是同配型网络, 符合社会网络的一般性特征。

表3 SMS数据库的主要拓扑特性

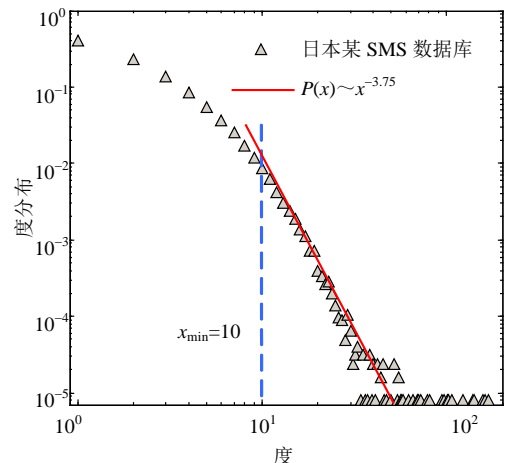
	韩国某SMS数据库	日本某SMS数据库
度分布幂指数	2.78±0.1	3.75±0.1
边权分布幂指数	2.2±0.1	2.0±0.1
平均路径长度 L	5.716 6	5.923 5
平均聚类系数 C	0.076	0.122
模块度 Q	0.757 0	0.683 4
匹配系数 r	0.465 4	0.247 5



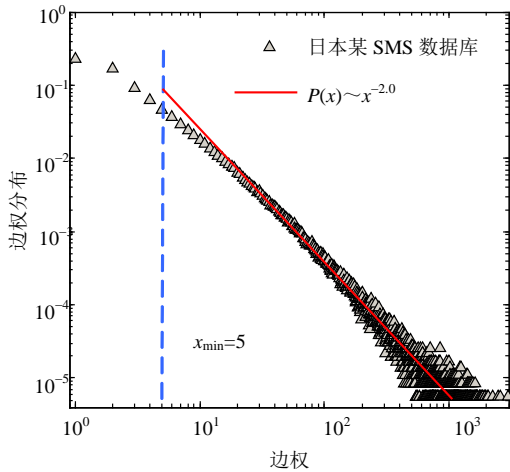
a. 韩国某 SMS 数据库的度分布曲线



b. 韩国某 SMS 数据库的边权分布曲线



c. 日本某 SMS 数据库的度分布曲线



d. 日本某 SMS 数据库的边权分布曲线

图1 韩国、日本某SMS数据库的度分布和边权分布曲线

文献[25]研究了科学家合作网(SCN)和全球航空网(WAN)的实证数据,发现平均点强度 \bar{s} 和点度 k 之间的关系具有幂函数律的关联性,即 $\bar{s} \sim k^\beta$,其中 $\beta_{\text{SCN}}=1$, $\beta_{\text{WAN}}>1$ 。然而,分析韩国某 SMS 用户群构成的短信网络的平均点强度和点度关系 $\bar{s}(k)$,如图2所示,发现该短信网络不具有形如 $\bar{s} \sim k^\beta$ 的幂函数律关联性(注:日本某 SMS 数据库也具有该特性)。

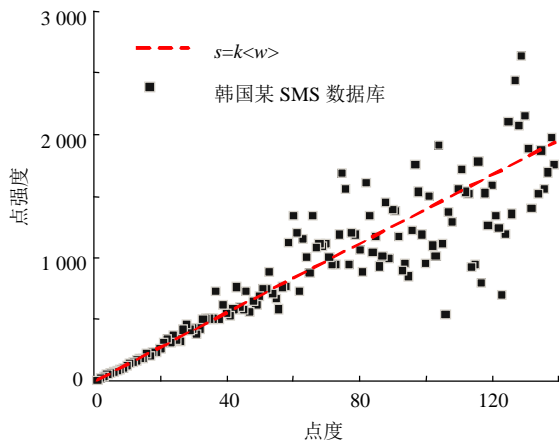


图2 韩国某 SMS 数据库的点度和平均点强度的关系

总结文献[7]的结论以及上述 SMS 数据库的分析结果,实际短信网络具有以下6个统计特性:1) 表征用户联系人数的度分布服从低头和重尾的幂律分布^[7]; 2) 节点最终度与节点加入网络时间的关系涨落比较大^[7]; 3) 网络的平均度随时间缓慢增加^[7]; 4) 表征用户间短信交互次数的边权分布具有低头和重尾的幂律特征; 5) 具有小世界特性、清晰的社群结构和同配型网络特征; 6) 平均点强度与点度不具有形如 $\bar{s} \sim k^\beta$ 的幂函数律关联性。

2 已有加权网络模型

近年来,归功于实证数据的逐渐丰富,使加权网络的研究趋于系统化。目前,加权网络的建模工作有:文献[26]提出的一种无标度网络的加权演化模型,并作了平均场分析;文献[27]提出的BBV模型,该模型是基于点强度驱动和边权逐渐加强机制建立的加权网络演化模型,可以模拟现实系统中相互作用强度的变化;文献[28]提出的流驱动TDE模型,该模型是基于点强度驱动和边权演化机制建立的加权网络演化模型,可以模拟技术网络的结构演化等等。

以BBV模型为例,其演化过程分为3个步骤:

1) 初始设定:给定一个包含 m_0 个节点的全耦合初始网络,其中每条边赋予权重 w_0 。

2) 增长:在每个时间步,带有 m 条边的新节点 n 加入到网络中,并与原网络中的 m 个节点相连,连接节点 i 的选择按权重优先选择方式进行,每条新边的权重也为 w_0 。

3) 边权值的动态演化:新边 (n,i) 的加入会导致连接节点 i 与其邻居节点 $j \in \Gamma(i)$ 之间边权值的重新调整,调整按 $w_{ij} \rightarrow w_{ij} + \Delta w_{ij}$ 的规则进行。其中 $\Delta w_{ij} = \delta w_{ij} / s_i$ 表示边权流量增量, δ 是流量调整比例系数。

BBV模型的数值模拟结果如图3所示,网络参数设置: $m_0=6$, $m=2$, $\delta=1$, $\omega_0=1$,总节点数为5000个节点。

BBV模型的度分布 $P(k)$ 和边权分布 $P(w)$ 的表达式如下:

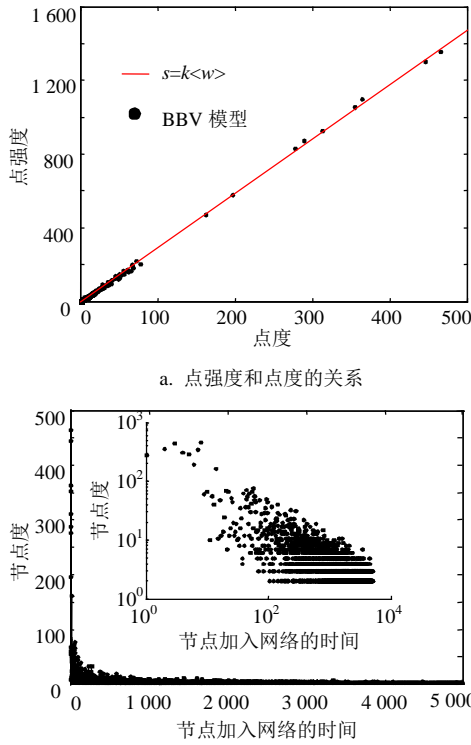
$$P(k) \sim k^{-\gamma} \quad \gamma = \frac{4\delta+3}{2\delta+1} \Big|_{\delta=1} = 2.3 \quad (5)$$

$$P(w) \sim w^{-\alpha} \quad \alpha = 2 + \frac{1}{\delta} \Big|_{\delta=1} = 3 \quad (6)$$

式中, k 和 w 分别表示度和边权; γ 和 α 分别表示度分布和边权分布的幂指数; δ 表示新边加入给节点带来的流量增量。式(5)和式(6)表明度分布和边权分布是严格的幂律分布,且幂指数只与参数 δ 有关。

图3a表明BBV模型的点强度 s 和点度 k 之间具有较好的幂函数律关联性,即 $s \sim k$;TDE模型一个典型的特点是能够给出 $s \sim k$ 的非平庸关联 $s \sim k^\beta$, $\beta > 1$,文献[29]从动态平衡流的角度也得到了 $s \sim k$ 的超线性幂律关联性。图3b表明越早加入网络的节点拥有越大的最终度 k 。通过对比发现,BBV/TDE模型的特点并不符合实际短信网络的统计特性,因

此, 针对加权短信网络的建模问题, 有必要探索驱动其生长演化的内在机制, 构建一种能较好地符合真实短信网络统计特性的加权网络模型。



b. 点度和节点加入网络时间的关系, 内插图为双对数坐标下的情况
图3 仿真模拟 BBV 模型

3 WEM模型

实际短信网络的演化过程还具有以下两个特点。

1) 用户加入网络, 其认识网络内的人具有一定的偶然性, 即新加入网络的用户无法认识整个网络的人, 同时也无法获取网络度/点强度分布的信息, 因此新加入的用户不是以点强度优先或点度优先的选择方式与网络内的用户成为“邻居”, 而是偏向于认识其社交圈中的朋友, 即选择邻居节点的方式具有随机性。这一点有别于传统的 BBV 和 TDE 加权网络模型。文献[7, 30]也指出人际关系网应具有这样的特点: 一个用户 X 因为与网络中的某人 A 相识而加入该网络, 接着会以某一概率认识 A 的朋友 B , 然后会以一个更小的概率认识 B 的朋友 C 。而现实情况与上述不同之处在于, 人际关系有亲有疏, 关系亲近的用户交往密切, 用户 X 认识了用户 A 后, 接着认识与 A 交往密切的朋友可能性大, 而认识与 A 关系一般的朋友可能性小。因此本文模型基于如下的假设前提: 一个用户 X 因为与网络中的 A 用户相识, 则该用户倾向于认识与 A 关系亲密的朋

友, 即认识 A 的朋友 B 的概率与 w_{AB} 呈正相关, w_{AB} 的值大, 相应 X 认识 B 的概率也大。

2) BBV模型认为网络的边权演化是由于新节点的加入使旧节点的邻边产生了流量增量, TDE模型认为节点间的关系是受网络交通量驱动并反过来影响网络交通流, 节点间的关系与点强度耦合, 即边权的增速与点强度乘积 $s_i s_j$ 成正比。然而在实际短信通信网络中, 用户间使用短信交流主要与用户的亲密程度及信息交流的需求有关。因此, 本文引入两个物理量来描述用户 i 和用户 j 再次使用短信交流的概率, 一个是总的短信交互次数 w_{ij} , w_{ij} 表征了亲密程度, 值越大, 说明用户间倾向于使用短信方式交流且关系亲密, 再次使用短信交流的概率也大; 另一个是最近一段时间 t 内短信交互次数 $\Delta w_{ij,t}$, $\Delta w_{ij,t}$ 表征了近期联系频繁程度, 因为 w_{ij} 的值大并不意味着用户间将会一直使用短信交流, 再次使用短信联系的概率也与他们的近期联系频繁程度有关, 即使 w_{ij} 的值比较小, 但可能由于某些特殊事件驱动造成信息交流的需求, 那么用户间再次使用短信联系的概率也大。

基于上述两个特点, 本文提出了一种新的加权短信网络演化模型, 该模型的构造分为 3 个步骤:

1) 初始化网络: 初始时刻 $t=1$, 一个连接概率为 P_0 的随机网络 $G(V, E, W)$, 初始节点数为 N_0 , 每条边赋予初值 w_0 。

2) 加权局部优先连接: 在每个时步, 向网络内加入一个新节点 n 。新节点 n 与网络内节点的连接规则: 首先随机连上一个点 A , 然后以一定的概率 $P_1 w_{AB} / \max_{X \in \Gamma(A)} \{w_{AX}\}$ 与点 A 的邻居 B 相连, 其中 $\Gamma(A)$ 表示 A 的邻居集, $\max_{X \in \Gamma(A)} \{w_{AX}\}$ 表示 A 与其邻居节点的最大边权值, 接着以更小的概率 $P_2 w_{BC} / \max_{Y \in \Gamma(B)} \{w_{BY}\}$ 与 B 的邻居 C 相连, 连边赋予边权值 w_0 。这里节点间的一次连接表征了用户间的一次 SMS 会话, 则边权值 w_0 表示一次 SMS 会话包含的短信数。

3) 边权更新: 在每一时步, 网络内已存在的连边 ($w_{ij} \neq 0$) 按如下机制更新:

$$w_{ij} \rightarrow \begin{cases} w_{ij} + w_0 & \text{以概率 } p_{ij} \\ w_{ij} & \text{以概率 } 1 - p_{ij} \end{cases} \quad (7)$$

式中,

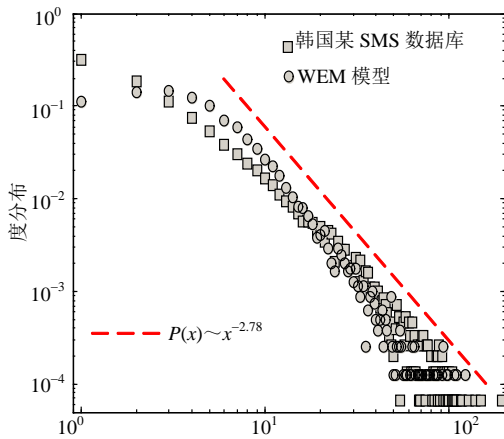
$$p_{ij} = P_3 \left\{ \alpha \frac{w_{ij}}{\max_{k,l \in V} \{w_{kl}\}} + (1 - \alpha) \frac{\Delta w_{ij,t}}{\max_{k,l \in V} \{\Delta w_{kl,t}\}} \right\} \quad (8)$$

是节点 i 和节点 j 之间的再次连边概率，与边权值 w_{ij} 及一段时间 t 内的联系频繁程度 $\Delta w_{ij,t}$ 有关； t 是近期联系频繁度的记忆长度； α 是可调参数，决定了短信交互行为的长期性和短期性的不同影响力度，取值区间为 $[0,1]$ 。

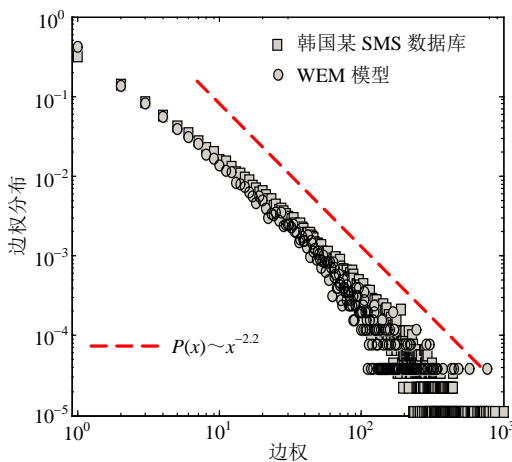
4 仿真结果及分析

4.1 WEM模型的仿真结果

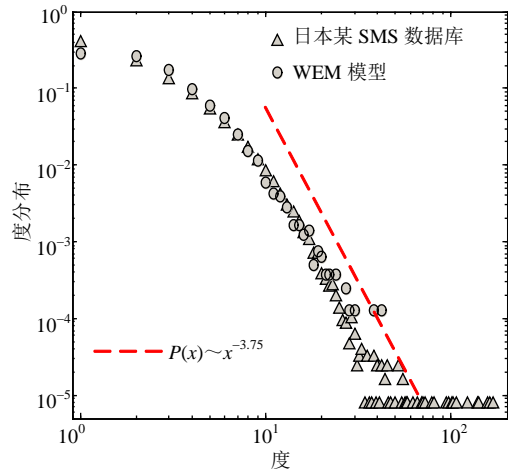
WEM模型的仿真参数设置为初始节点数 $N_0 = 10$ ，以概率 $P_0 = 0.2$ 随机连接，边权更新步骤的可调参数取值 $\alpha = 0.1$ ， $P_3 = 0.1$ ，近期联系频繁度的记忆长度为 $t = 5$ ， w_0 的分布^[19]服从幂律分布 $f(w_0) \sim w_0^{-\theta}$ ， $\theta > 2$ ，考虑到实际情况，即一次SMS会话包含的短信数至少为1，至多一般不会超过某个较大的值 $w_{0,max}$ ，因此，不妨假设 $\theta = 2.5$ ，且 w_0 的取值区间为 $[w_{0,min} = 1, w_{0,max} = 40]$ 。取两组不同的选择概率参数 (P_1, P_2) ，分别得到相应的度分布和边权分布曲线，如图4所示。



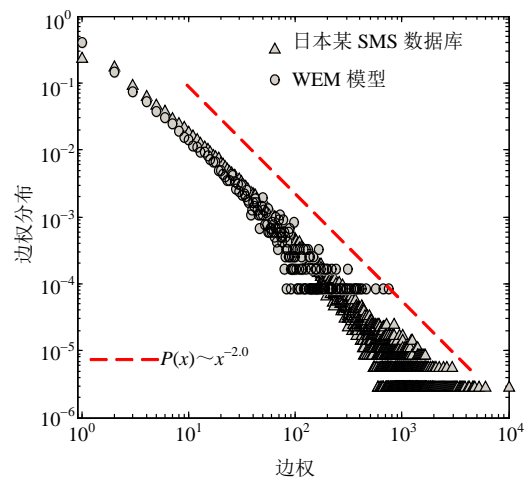
a. 第一组参数的度分布曲线



b. 第一组参数的边权分布曲线



c. 第二组参数的度分布曲线



d. 第二组参数的边权分布曲线

图4 仿真模拟 WEM 模型

第一组参数取连接概率 $P_1 = 0.6$ 、 $P_2 = 0.4$ ，度分布曲线如图4a所示，度分布的幂指数为2.78，边权分布曲线如图4b所示，边权分布的幂指数为2.2，符合韩国某SMS数据库的统计结果。

第二组参数取连接概率 $P_1 = 0.3$ 、 $P_2 = 0.15$ ，度分布曲线如图4c所示，度分布的幂指数为3.75，边权分布曲线如图4d所示，边权分布的幂指数为2.0，符合日本某SMS数据库的统计结果，以上幂指数估计均可通过阈值为0.9的K-S检验。

同时，从图4可知WEM模型的度分布和边权分布并非严格的幂律分布，而是具有低头和重尾的幂律特性，符合实际短信网络的统计特性1)和4)。

WEM模型的特性如图5所示，图5a表明WEM模型的平均点强度和点度不具有形如 $\bar{s} \sim k^\beta$ 的幂函数律关联性。图5b说明节点最终度和节点加入网络时间的关系涨落比较大，及图5c说明平均度随时间缓缓

慢增加。这3个特点分别对应实际短信网络的统计特性6)和2)、3)。

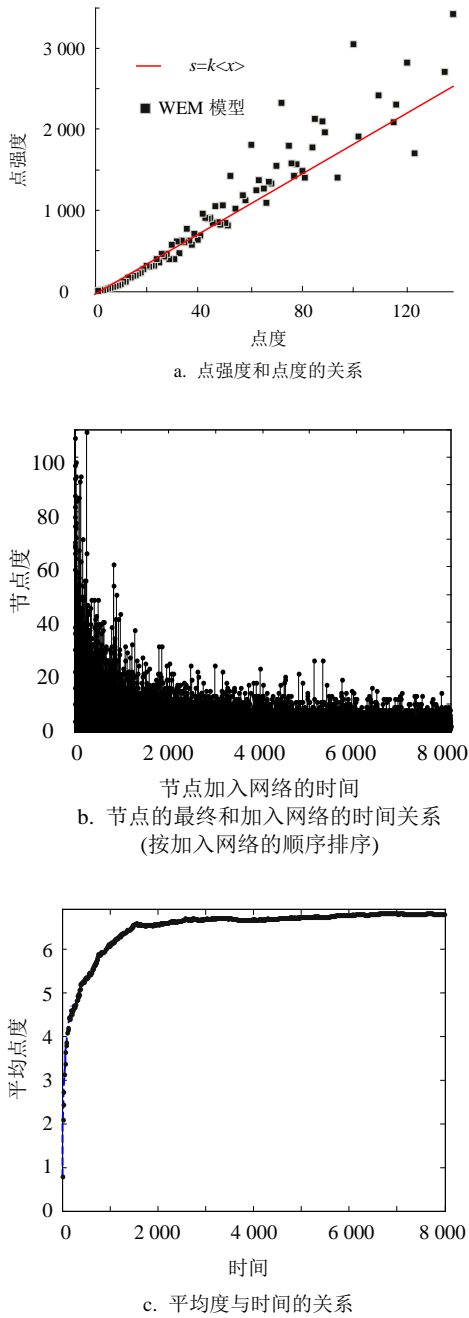


图5 WEM模型的一些特性

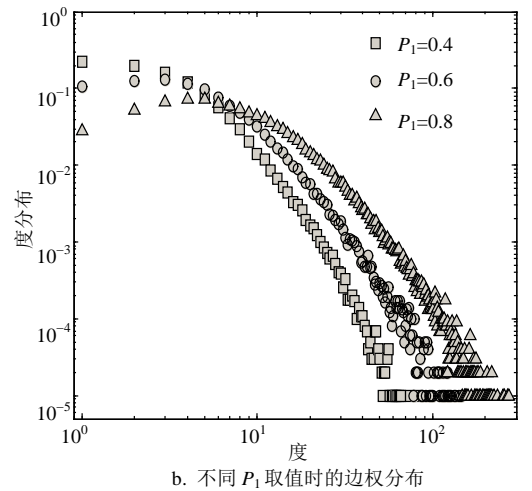
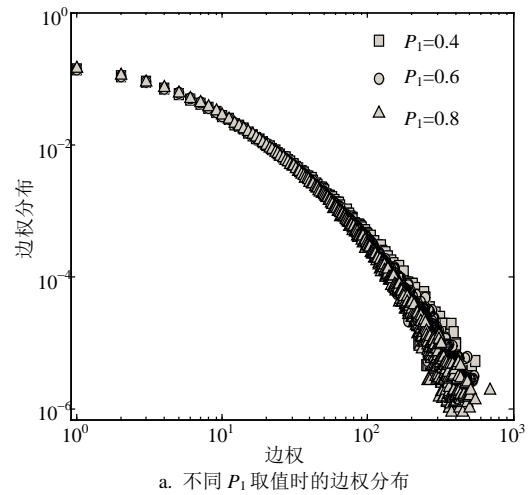
分别在不同网络规模下, 模型参数取值同第一组参数, 计算WEM模型的网络拓扑特征, 如表4所示。WEM模型构造的短信网络具有小世界特性, 即短的平均路径和大的聚类系数; 模块度接近0.7, 表明网络具有清晰的社群结构; 网络的匹配系数为正, 说明该网络是同配型的, 较好的符合了实际短信网络的统计特性5)。

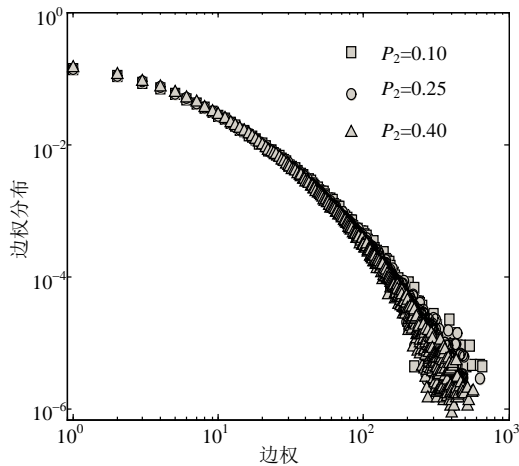
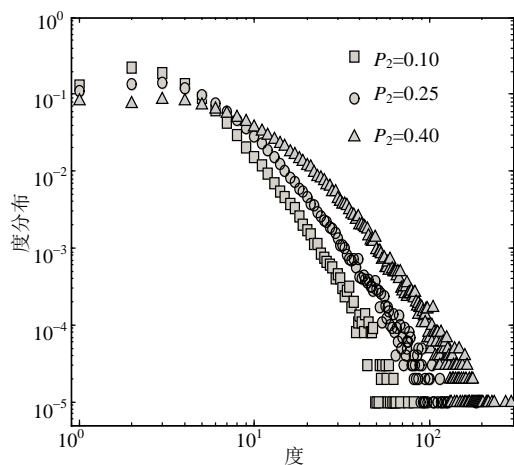
表4 不同网络规模下WEM模型的网络拓扑性质

节点数	平均路径长度 L	平均聚类系数 C	模块度 Q	匹配系数 r
1000	6.946 4	0.485 4	0.666 7	0.154 2
4000	6.662 0	0.496 1	0.670 4	0.107 0
7000	6.843 8	0.502 7	0.668 9	0.087 7
10000	7.358 3	0.511 5	0.665 6	0.077 0

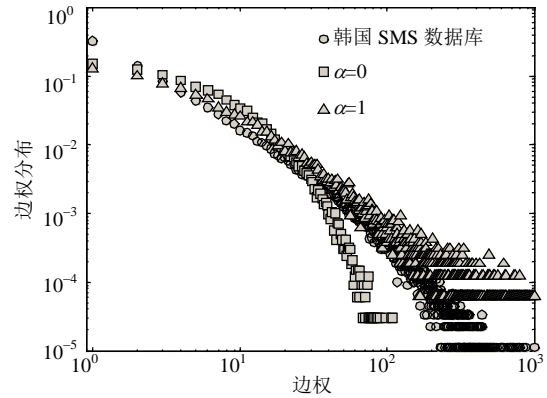
4.2 一些参数对模型的影响

首先考察网络增长步骤中加权局部优先连接参数 P_1 和 P_2 对模型的影响如图6所示。给定初始网络的参数设置: $N_0 = 10$, $P_0 = 0.2$, $P_3 = 0.1$, $\alpha = 0.1$, $t = 5$ 。当取 $P_2 = 0.4$ 时, P_1 分别取不同的值得到边权分布和度分布曲线, 如图6a和6b所示; 当取 $P_1 = 0.6$ 时, P_2 分别取不同的值得到边权分布和度分布曲线, 如图6c和6d所示。从仿真结果图可知, WEM模型的边权分布基本上不随着 P_1 和 P_2 的变化而变化, 但度分布幂指数受参数 P_1 和 P_2 的影响较大。这是由于在加权局部优先连接过程中增大连接概率 P_1 或 P_2 的值, 即提高了邻居节点被连接概率, 会增加网络中三角构成连接^[31]的数量, 导致网络中节点的度增大, 同时也提高了网络的聚类系数。



c. 不同 P_2 取值时的边权分布d. 不同 P_2 取值时的边权分布图 6 参数 P_1 和 P_2 对 WEM 模型分布的影响

接下来考察边权更新步骤中可调参数 α 对模型的影响。模型参数取值同第一组参数，如图7所示，当 $\alpha=0$ 时，边权更新概率只与边权值 w_{ij} 有关，边权分布曲线出现了严重的拖尾现象，这是由于边权值大的用户间再次联系的概率也大，导致出现“富人俱乐部”特性，没有反映出事件驱动这一短期行为因素的影响；当 $\alpha=1$ 时，边权更新概率只与近期联系频繁度 $\Delta w_{ij,t}$ 有关，忽略了人际关系的特点，没有反映出用户间亲密度的长期行为因素的影响。而实际中，人们的短信交互行为应当既包含了反映用户亲密度的长期性影响因素，也包含了反映突发事件驱动的短期性影响因素。从图7可知，实际短信网络的边权分布对应于 WEM 模型的可调参数取值为 $\alpha \in (0,1)$ ，其值反映了长期和短期共同作用的效果及各自的影响力度，模型中的近期联系频繁度的记忆长度 t 在一定程度上反映了该行为的记忆长度，间接说明了短信发送行为存在记忆性。

图7 α 取区间极值时的边权分布图

5 结束语

通过对现有 SMS 数据库的统计分析，结合已有相关文献的结论，进一步归纳总结了实际短信网络的主要特点，包括：度分布和边权分布具有低头和重尾的幂律特性、平均点强度与点度不具有幂函数律关联性、节点的最终度与节点加入网络的时间关系涨落较大、网络的平均度随时间缓慢增加、小世界特性、清晰的社群结构和同配型网络。基于上述特点，将用户间短信交互次数视为短信网络的边权，考虑到实际短信网络生长过程的一些特点，对网络增长采用加权局部优先连接的原则，着重细化边权更新过程，综合亲密度和近期联系频繁度两个因素，提出了一种短信网络的加权演化模型。仿真实验表明，该模型能够较好地与实际短信网络的统计特性相吻合。

本文提出的短信网络加权演化模型，研究了短信网络所反映出的人际关系特点及短信交互行为特征，刻画了真实短信网络的结构特征和演化规律，为分析短信网络上的谣言、舆情等信息传播行为的研究提供了有意义的参考。本文以用户间短信交互次数表征用户间的亲密程度，发现人际关系呈现幂律特性，即少数朋友关系特别亲密，而多数朋友关系较疏远。因此，研究人际关系有助于更好的了解网络动力学过程，也是比较有益的研究方向，例如加权短信网络上的谣言短信传播问题、基于用户交互式的社交网络上业务推广问题等等。

参 考 文 献

- [1] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks[J]. Nature, 1998, 393: 440-442.
- [2] BARABÁSI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286: 509-512.
- [3] HU B, JIANG X Y, DING J F, et al. A weighted network model for interpersonal relationship evolution[J]. Physica A,

- 2005, 353: 576-594.
- [4] 张培培, 何阅, 周涛, 等. 一个描述合作网络顶点度分布的模型[J]. 物理学报, 2006, 55(01): 0060-0067.
ZHANG Pei-pei, HE Yue, ZHOU Tao, et al. A model describing the degree distribution of collaboration networks[J]. Acta Physica Sinica, 2006, 55(01): 0060-0067.
- [5] 苏萌, 柏林森, 周涛. 个性化:商业的未来[M]. 北京: 机械工业出版社, 2012.
SU Meng, BO Lin-sen, ZHOU Tao. Personalization: the future of business[M]. Beijing: China Machine Press, 2012.
- [6] LU L Y, MEDO M, YEUNG C H, et al. Recommender Systems[DB/OL]. [2012-01-01]. <http://arxiv.org/abs/1202.1112>.
- [7] 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.
WU Ye, XIAO Jing-hua, WU Zhi-yuan, et al. Research on the growing process of short message networks[J]. Acta Physica Sinica, 2007, 56(4): 2037-2041.
- [8] WU Y, ZHOU C S, XIAO J H, et al. Evidence for a bimodal distribution in human communication[J]. Proc Natl Acad Sci USA, 2010, 107: 18803-8.
- [9] 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.
SUN Qin-dong, SUN Ya-hong, GUAN Xiao-hong, et al. A dynamic evolution model of short message complex network[J]. Journal of Xi'an JiaoTong University, 2009, 43(6): 5-9.
- [10] MA B J, XIAO J H, YANG J Z. Evolution of the SMS Networks[J]. Dynam Cont Dis B, 2006, 13: 371.
- [11] MENG X Q, ZERFOS P, SAMANTA V, et al. Analysis of the reliability of a nationwide short message service[C]// INFOCOM, Alaska USA: IEEE, 2007: 1811-1819.
- [12] 徐增勇, 宋运忠. 谣言短信传播网络拓扑性质研究[J]. 河南理工大学学报(自然科学版), 2008, 27(2): 217-223.
XU Zeng-yong, SONG Yun-zhong. On topological properties of propagation networks for rumor short message[J]. Journal of Henan Polytechnic University(Natural Science Edition), 2008, 27(2): 217-223.
- [13] YAN G, ZHOU T, WANG J, et al. Epidemic spread in weighted scale-free networks[J]. Chinese Physics Letters, 2005, 22(02): 501-513.
- [14] YANG Z M, ZHOU T. Epidemic spreading in weighted networks: an edge-based mean-field solution[J]. Physical Review E, 2011, 85(5): 056106.
- [15] CUI A X, YANG Z M, ZHOU T. Roles of Ties in Spreading[DB/OL]. [2012-01-01]. <http://arxiv.org/abs/1204.0100>.
- [16] PASTOR- SATORRAS R, VESPIGNANI A. Evolution and structure of the internet: a statistical physics approach[M]. Cambridge, England: Cambridge University Press, 2004.
- [17] KRAUSE A E, FRANK K A, MASON D M, et al. Compartment revealed in food-web structure[J]. Nature, 2003, 426: 282-285.
- [18] TOMAR V, ASNANI H, KARANDIKAR A, et al. Social network analysis of the short message service[C]//The 16th National Conference on Communications(NCC 2010). Chennai, India: IEEE, 2010: 1-5.
- [19] HONG W, HAN X P, ZHOU T, et al. Heavy-tailed statistics in short-messages communication[J]. Chinese Physics Letters, 2009, 26(2): 028902.
- [20] ZHAO Z D, XIA H, SHANG M S, et al. Empirical analysis on the human dynamics of a large-scale short message communication system[J]. Chinese Physics Letters, 2011, 28(06): 068901.
- [21] 中国计算机学会CCF科研数据. 韩国某城市某月SMS数据集[DB/OL]. [2012-01-01]. <http://www.ccf.org.cn/sites/ccf/ccfdata.jsp>.
China Computer Federation (CCF) Research Data Set. SMS Local call dataset of one city in Korea[DB/OL]. [2012-01-01]. <http://www.ccf.org.cn/sites/ccf/ccfdata.jsp>.
- [22] 中国计算机学会CCF科研数据. 日本某城市某月SMS数据集[DB/OL]. [2012-01-01]. <http://www.ccf.org.cn/sites/ccf/ccfdata.jsp>.
China Computer Federation (CCF) Research Data Set. SMS Local call data Set of city in Japan [DB/OL]. [2012-01-01] <http://www.ccf.org.cn/sites/ccf/ccfdata.jsp>.
- [23] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69: 026113.
- [24] CLAUSET A, SHALIZI C R, NEWMAN M E J. Power-law distribution in empirical data[J]. SIAM Review, 2009, 51(4): 661-703.
- [25] BARRAT A, BARTHELEMY M, PASTOR-SATORRAS R, et al. The architecture of complex weighted networks[J]. Proc Natl Acad Sci USA, 2004, 101(11): 3747-3752.
- [26] YOON S H, JEONG H, BARABÁSI A L. Weighted evolving networks[J]. Physical Review Letters, 2001, 86(25): 5835-5838.
- [27] BARRAT A, BARTHELEMY A, VESPIGNANI A. Weighted evolving networks: coupling topology and weight dynamics[J]. Physical Review Letters, 2004, 92: 228701.
- [28] WANG W X, WANG B H, HU B, et al. General dynamics of topology and traffic on weighted technological networks[J]. Physical Review Letters, 2005, 94: 188702.
- [29] OU Q, JIN Y D, ZHOU T, et al. Power-law strength-degree correlation from resource-allocation dynamics on weighted networks[J]. Physical Review E, 2007, 75: 021102.
- [30] 崔爱香, 傅彦, 尚明生, 等. 复杂网络局部结构的涌现: 共同邻居驱动网络演化[J]. 物理学报, 2011, 60(3): 038901.
CUI Ai-xiang, FU Yan, SHANG Ming-sheng, et al. Emergence of local structures in complex network: common neighborhood drives the network evolution[J]. Acta Physica Sinica, 2011, 60(3): 038901.
- [31] HOLME P, KIM B J. Growing scale-free networks with tunable clustering[J]. Physical Review E, 2002, 65: 026107.

编辑 蒋晓