

# 阵发性影响因素的研究

顾亦然, 李金发, 谢鸿飞

(南京邮电大学 控制与智能技术研究中心, 江苏 南京 210046)

**摘要:**人类自身的复杂性和多样性, 导致人类行为具有高度的复杂性。经典的人类动力学分析认为人类行为时间性可以用泊松过程来描述, 而 Barabási 发现人们从事特定活动的行为具有阵发和胖尾的特性。为验证人类行为时间间隔分布具有阵发性及探寻阵发特性的影响因素, 根据移动通信网络中人与人之间发短信的时间间隔分布, 运用统计分析方法, 进一步证实了人类行为的时间间隔分布具有阵发性, 发现该特性受到作息时间、节假日、地理位置、风俗习惯以及偶发性事件等因素的影响。

**关键词:**人类动力学; 复杂网络; 阵发; 时间间隔分布

**中图分类号:** TN929.5

**文献标识码:** A

**文章编号:** 1673-629X(2010)09-0168-04

## Study on Influence Factors of Characteristic of Burst

GU Yi-ran, LI Jin-fa, XIE Hong-fei

(Center for Control & Intelligence Technology, Nanjing University of Posts and Telecommunications, Nanjing 210046, China)

**Abstract:** The complexity and diversity of human beings result in a high degree of complexity of human behavior. Classic analysis of human dynamics think that the timing of human behavior can be used to describe by the Poisson process, however, Barabási finds that the behavior of the specific activities has the characteristics of bursts and heavy tails. In order to verify the time interval distribution of human behavior has the characteristic of burst and explore its influence factor, this paper which based on the short message of the mobile communication network among people tells us the distribution of interval time has the property of bursts with the method of statistical analysis. Further analysis, this distribution affected by the holidays, location, customs, schedule and so on.

**Key words:** human dynamics; complex network; burst; time interval distribution

## 0 引言

人类作为一种高等生物, 由于本身的复杂性和多样性, 从而导致其行为高度复杂。早期的人类行为研究者认为, 人类行为在时间上是随机分布的, 可用泊松过程来描述。这种假设必然导致的结论是<sup>[1]</sup>: 人的行为的时间统计特征应该是较为均匀的, 两次相继事件的间隔时间分布服从指数分布, 事件之间存在极大时间间隔的概率很小。Barabási 等人的实证研究和理论分析表明<sup>[2]</sup>: 人类行为所出现的规律并不与泊松过程一致, 而是其间隔时间遵循非泊松过程, 时间间隔分布具有胖尾且接近于幂律, 胖尾过程允许长时间的静默和短时间内集中式的爆发, 表现出一定的阵发性。

为了验证这一结论, 国内外很多学者都进行了实证研究, 从市场交易<sup>[3]</sup>、网页浏览<sup>[4]</sup>到电影点播<sup>[5]</sup>、在线游戏<sup>[6]</sup>等。研究表明: 它们的间隔时间近似符合幂律分布, 而不能用传统的泊松过程来描述。

人类活动遵循非泊松过程的事实表明观察到的阵发性反映了人类动力学一些基本的和潜在的一般特征。为了解释人类动力学中的这一现象, Barabási 提出了一种基于排队论的模型<sup>[2,7]</sup>: 人类在处理各类事物时, 常将待处理任务进行优先级划分, 首先处理最高优先级, 然后处理较高优先级, 依此类推, 这种行为模式是造成胖尾分布的重要原因。而韩筱璞等人提出了一种基于自适应调节的人类动力学模型<sup>[1]</sup>: 人类行为由自身的兴趣或欲望(即进行该行为的“活性”)所驱动, 这是造成胖尾分布的因素之一, 并进行了数值模拟和解析研究, 该模型可以获得指数为 -1 的幂律形式的时间间隔分布, 能与一些实证结果相吻合。

目前, 手机短信已成为继 Internet 之后的“第五媒体”, 成为人们日常交流的主要方式之一。最近的一项

收稿日期: 2009-12-26; 修回日期: 2010-03-05

基金项目: 国家自然科学基金项目(60874091)

作者简介: 顾亦然(1972-), 女, 江苏金坛人, 副教授, CCF 高级会员, 研究方向为通信网络的性能分析与控制, 复杂网络理论在通信中的应用, 嵌入式系统开发等。

调查表明,我国手机用户已突破 7 亿户,手机普及率越来越高,使越来越多的人习惯于使用短信这一便利的交流方式。无论在节假日还是平时,短信已经成为手机用户沟通信息、传递情感最快捷的途径之一。短信网络的研究不但对经济的发展具有重要意义,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。

文中根据移动通信网络中人与人之间发短信的时间间隔分布,运用统计分析的方法,进一步证实了人类行为的时间间隔分布具有阵发性,发现该特性受到作息时间、节假日、地理位置、风俗习惯以及偶发性事件等因素的影响。

1 数据来源

文中数据来源于朋友提供的短信清单,记录了 12 个人(编号为 A 到 L)最近几个月的短信通信结果。这些人包括 4 个(C、D、E、I)工作的,其他几个为在读研究生;他们中有的在北京、广州,还有的在浙江、山东和江苏。每条短信都记录了发送人发送短信的日期和时间,详细记录如表 1 所示,由于个人隐私,文中没有列出详细的短信清单。

表 1 数据记录						
人员	A	B	C	D	E	F
时间跨度(月)	6	6	6	6	6	6
短信总量(条)	822	694	1907	551	1149	1336
人员	G	H	I	J	K	L
时间跨度(月)	6	6	5	6	5	4
短信总量(条)	432	1116	642	2089	2135	984

2 统计分析

文中定义的时间间隔为两次连续发短信的时间差额,统计的最小时间单位为 1 小时。图 1 给出了时间间隔分布(这里的数据是 12 个人的统计平均),星号为

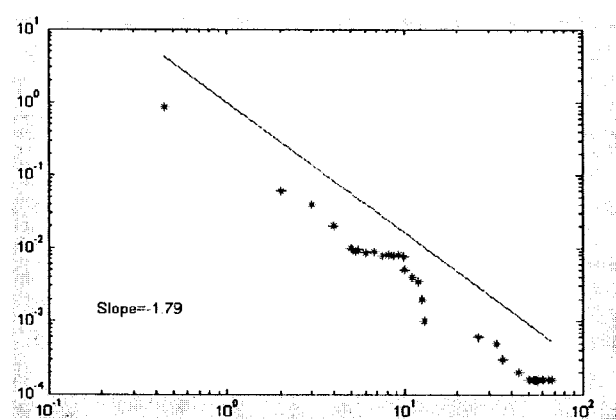


图 1 时间间隔分布

实证数据,直线为拟合后的结果,横坐标为间隔时间(小时),纵坐标表示相应的概率,即单位时间内短信占总短信的比值。该分布可用  $p(\tau) \propto \tau^{-\tau}$  (幂律分布)很好地拟合,幂指数为 -1.79,与洪伟<sup>[8]</sup>等人的研究结果相符,说明该时间间隔分布具有阵发和胖尾特性。

2.1 阵发性受作息时间的影响

人类发展至今,很多事情都变得有规律、有组织,只有在有限的时间里做自己的本职工作,才能使效率提高,而闲暇时间的娱乐则是为了更好的工作。短信作为人类传播信息的媒体,在生活中起着越来越重要的作用。

逐渐增加的数据表明,手机短信已成为继报纸、广播、电视、互联网之后名副其实的“第五媒体”。文中根据表 1 的数据将其划分为 5 个时间段,即 6:00—10:00、10:00—14:00、14:00—18:00、18:00—24:00 和 0:00—6:00,统计每个时间段内 12 人所发短信的总量,如图 2 所示。

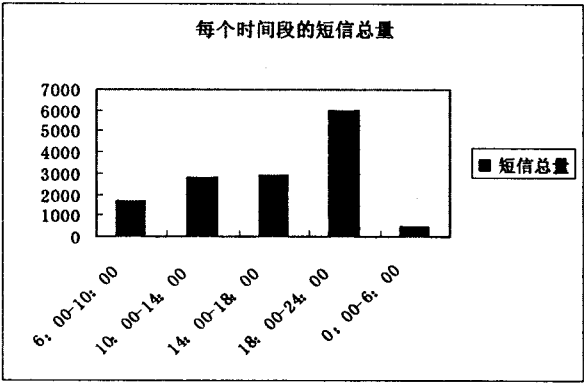


图 2 每个时间段的短信总量(条数)

由图 2 可以看出,人们发短信的时间多集中在晚上 6 点以后,即下班后和晚上无课时间,而这个时间段内 10 点左右会出现一个高峰,达到一天中的最大值。零点以后短信发送量锐减,这与人们的生活规律有关,但由于现在的工作制度,工作人员有可能晚上工作,因此也会有一定的发送量。10:00—14:00 时间内 12 点左右会出现一个高峰,此时正是下班时间和午休时间。此外,周末的短信发送量明显高于其它时间。综上所述,阵发性受到人们作息时间、工作规律的影响,人们在休息时间比较喜欢发短信,以此来放松心情、交流情感、进行事物联系等。

为分析作息对阵发性特性的影响,文中定义以下两个影响度函数:

个体影响度函数:

$$\alpha_i(x) = \frac{\alpha_i(x)}{\alpha_i} (1 \leq x \leq 5, x, i \in N+)$$
 (1)

平均影响度函数:

$$\langle a(x) \rangle = \frac{\alpha_1(x) + \alpha_2(x) + \cdots + \alpha_i(x)}{\alpha_1 + \alpha_2 + \cdots + \alpha_i} \quad (1 \leq x \leq 5, x, i \in N+)$$

(1) 式中  $x$  表示时间段,  $\alpha_i(x)$  表示第  $i$  天  $x$  时间段内的短信总量,  $\alpha_i$  表示第  $i$  天内的短信总量。 $\alpha(x)$  越大, 说明作息时问对阵发特性的影响越大, 反之亦然。

(2) 式为(1)式的统计平均, 宏观上更好地反映作息时问对阵发特性的影响。

## 2.2 阵发特性受节假日、风俗习惯和地理位置的影响

随着移动通讯技术的飞速发展和手机的迅速普及, 手机短信越来越受到大家的青睐。信息产业部的统计数据显示, 2007 年, 全国手机用户数达到 5.47 亿户, 全国手机短信发送量达到 5921 亿条, 日均发送短信超过 16 亿条, 而今, 手机用户已突破 7 亿大关, 短信的发送量也越来越多。为了探索节假日对阵发特性的影响, 文中收集了 12 人在五一和端午期间的短信总量, 如图 3 和图 4 所示。

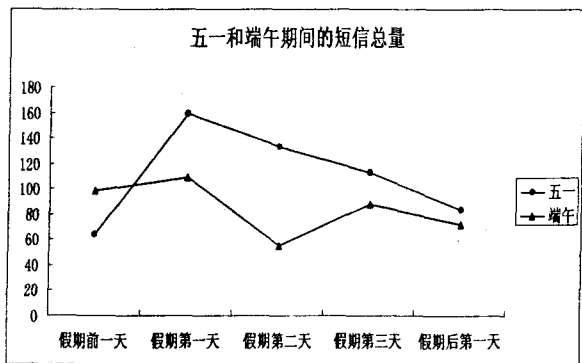


图 3 五一和端午期间的短信总量(条数)

由图 3 可知, 节假日期间, 用户的短信发送量明显高于其它时间, 尤其是在五一和端午节当天, 短信发送量剧增; 而假期第二天则有所下降, 相对五一而言, 端午节第二天下降的幅度要大于五一; 另外, 周六的短信发送量明显高于周日(五一假期第二天和端午假期第三天为星期六)。数据说明, 人们在节假日期间喜欢发短信, 以此来传递祝福、交流消息等, 从而验证了阵发特性受到节假日的影响。

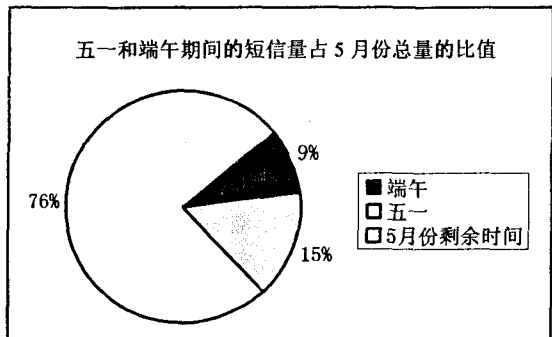


图 4 五一和端午期间的短信量占 5 月份的比值

科学技术的发展导致人类可以方便地交流, 而其中最具有代表性的当属校园里的大学生。节日问候祝福短信随着大学校园节日的增多, 也变得多姿多彩, 他们过的节日不仅包括传统节日如春节、端午节、中秋节等, 还包括情人节、愚人节、圣诞节等西方节日。在节日的问候祝福中, 手机短信成为主要传播的媒体。图 4 数据表明, 五一期间的短信发送量明显高于端午期间, 说明五一这个西方节日受到人们很高的重视。五一和端午期间短信总量占 5 月份的比例说明节假日的短信发送量明显高于其它时间, 进一步说明阵发特性受到节假日的影响。定义节假日对阵发特性的影响度函数如下:

个体影响度函数:

$$\beta(y) = \frac{\beta_i(y)}{\beta_i} (1 \leq i \leq 12) \quad (3)$$

平均影响度函数:

$$\langle \beta(y) \rangle = \frac{\beta_1(y) + \beta_2(y) + \cdots + \beta_i(y)}{\beta_1 + \beta_2 + \cdots + \beta_i} (1 \leq i \leq 12) \quad (4)$$

(3) 式中  $y$  表示节假日,  $\beta_i(y)$  表示  $i$  月内某一节假日的短信总量,  $\beta_i$  表示  $i$  月内的短息总量。 $\beta(y)$  越大, 则节假日对阵发特性的影响越大, 反之亦然。

(4) 式是(3)式的统计平均, 同样是从宏观上反映节假日对阵发特性的影响。

为理解作息时间和节假日同时对阵发特性的影响, 定义关联性函数:

$$\rho(z) = \frac{z_i(x \cap y)}{z_i} (1 \leq i \leq 12) \quad (5)$$

上式中  $z$  表示某一节假日内的时间段,  $\rho(z)$  越大, 即作息时间和节假日的关联性越大, 表明其对阵发特性的影响越大。

尽管一年中只有一个端午节和一个五一节, 但短信拥堵却不是这一特定日子的偶然现象。调查显示, 春节、中秋节、情人节、“十一”等节日, 短信发送量也会在短时间内剧增, 表现出一定的阵发特性, 进一步说明阵发特性受到节假日的影响。与中国人相比, 美国人发短信的数量很少, 调查显示, 美国人一年的发送量还不如中国的一周。美国人认为, 用短信的方式向人们表示祝贺是不礼貌的, 他们更倾向于直接通电话或电话留言; 而中国人表述感情比较含蓄, 短信的优势恰好能发挥此功能, 这与国家的风俗习惯和地理位置有关。从而说明阵发特性不仅与人们的风俗习惯有关, 而且还和地理位置有关。

人们都有这样的经验, 每逢节假日, 无线网络、互联网、通信网络都经受着巨大流量的冲击, 往往平常很容易做的事, 而此时需要较长时间才能做完, 甚至不能

完成。短信拥堵与流量有直接关系,这一方面说明短信业务量的可观,另一方面也告诫和提醒运营商,系统容量和管理方面要更上一层楼,从而保护手机用户的消费热情,提高服务质量。另外,进行必要的分析流量预测,也是移动通信部门需要考虑的课题。例如,节假日期间增加临时短信平台,加强维护或采用高性能、高处理能力的短信平台等。

### 2.3 阵发性受到偶发性因素的影响

2006年一号台风“珍珠”和四号强热带风暴“碧利斯”登录时,福建省电信、移动、联通共发送公益短信1800多万条,提高了广大民众的知晓率和防台意识。五号台风“格美”登陆时,三大运营商两天内向民众发送了848万条防台风公益短息。在自然灾害到来的时候,给民众发送一条公益短信,不仅让民众倍感温馨,而且提高了民众对灾害的防备意识,从而让民众在自然灾害面前做到不慌不乱,减少自然灾害带来的损失。据悉,目前,我国每天接受气象服务的公众超过10亿人次,手机气象短信用户已达3300多万,1.5万多个乡村建立了信息服务站。

2008年5月12日,四川汶川发生大地震,地震等级为里氏8.0级,受灾地区达10万平方公里,直接经济损失8451亿元人民币。汶川地震发生后,为了帮助人们获得灾情信息、亲人消息以及自救常识,当地及附近省份的运营商迅速制定措施,组织人员修复基站、通信线路,调拨通信应急车,保证灾区人民能够与外界联系。根据工业和信息化部统计,截止6月1日,全国已累计发送应急提示和抢险指导等方面的短信约7.4亿条,这一方面显示了我国人民的团结,另一方面也让通信企业经受了一次考验。

综上所述,短时间内短信量的激增说明阵发性受到偶发性因素的影响。

## 3 短信网络的研究意义

短信网络背后的深层机制就是人际关系,短信网络的研究对研究人际关系网的特点很有帮助。现在分析网络的一个有效工具就是复杂网络理论,现今,复杂网络已经被运用到很多领域,如社会领域中的电影演员合作网络、信息领域的引用网络、技术领域的电力网络以及生物领域的蛋白质网路等。国内很多学者<sup>[9,10]</sup>运用复杂网络理论对短信网络进行了分析,发现该网络具有小世界和无标度性质。尽管这些研究大多还停留在理论阶段,但是它带来的潜在价值却是不可估量的。

文中针对于人类行为时间间隔分布的研究不仅具有理论价值,而且具有应用价值,使得短信服务商能够

根据人们发短信的规律而调节短信平台的使用情况或扩容,从而满足人们的心理需求,提高服务质量。对人类通讯时间特性的统计分析是理解电话、手机短信和网络即时通讯中突发性拥塞形成机制并设计相应对策的基础。因此,对于人类行为时间和空间特性的定量理解,有着重大的科学价值和应用前景。

## 4 结束语

文中统计了12个人的短信通信结果,发现其时间间隔分布服从幂律分布,幂指数为-1.79,表现出一定的阵发性,进一步分析可知,该特性受作息时间、节假日、风俗习惯、偶发性事件等因素的影响。由于文中收集到的数据有限,因此幂指数上存在一定的偏差,从而使更加细致的分析受到限制。另外,人类行为动力学研究是建立在大量实证数据基础之上的,文中的分析可能存在不足,更加详尽的分析还有待数据的支持。

人类行为动力学的研究具有重大的科学和商业价值,一方面模拟人类行为对于建立诸如虚拟社区<sup>[11]</sup>、谣言传播和市场行为<sup>[12]</sup>等大规模社会组织模型非常重要;另一方面信息服务提供商、运营商可以根据人类行为的统计特性进行资源分配和业务定价,从而做到因时因地的改变。

### 参考文献:

- [1] 韩筱璞,周涛,汪秉宏.基于自适应调节的人类动力学模型[J].复杂系统与复杂性科学,2007,4(4):1-6.
- [2] Barabási A L. The origin of bursts and heavy tails in human dynamics[J]. Nature,2005,435:207-211.
- [3] Plerou V, Gopikrishnan P, Amaral L A N, et al. Economic fluctuations and anomalous diffusion[J]. Phys Rev E,2000,62(3):3023-3026.
- [4] Dezs Ö Z, Almaas E, Lukács A, et al. Dynamics of information access on the web[J]. Physical Review E,2006,73(6):1-6.
- [5] Zhou T, Kiet H A T, Kim B J, et al. Role of activity in human dynamics[DB/OL]. 2007-11-27[2007-11-30]. <http://www.iop.org/EJ/abstract/0295-5075/82/2/28002>.
- [6] Henderson T, Nhatti S. Modelling user behavior in networked games[C]//Proc 9th ACM Int Conf on Multimed. New York: ACM Press,2001:212-220.
- [7] Vázquez A. Exact results for the Barabási model of human dynamics[J]. Phys Rev Lett,2005,95(24):1-4.
- [8] Hong W, Han X P, Zhou T, et al. Scaling behaviors in short-message communication[DB/OL]. 2008-02-03. <http://arxiv.org/abs/0802>.
- [9] 王静,孔令江,刘慕仁.小世界网络上的手机短信息传播模型[J].广西师范大学学报:自然科学版,2006,24(3):2-4.

值的数量级均是  $10^4 \mu\text{g}/\text{m}^3$ , 均值分别为  $191.97 \mu\text{g}/\text{m}^3$ 、 $472.03 \mu\text{g}/\text{m}^3$ 、 $583.68 \mu\text{g}/\text{m}^3$ 、 $348.97 \mu\text{g}/\text{m}^3$ , 而多环芳烃在大气中的含量最大值为  $103 \mu\text{g}/\text{m}^3$ , 均值为  $30.13 \mu\text{g}/\text{m}^3$ , 亦比二恶英的含量均值少一个单位的数量级。

在水体中, 有机氯农药与多溴联苯醚、多氯联苯, 含量基本都在一个数量级, 而多环芳烃的含量最大值竟然达到了  $16489 \text{ ng}/\text{g}$ , 其均值也在  $103 \text{ ng}/\text{g}$  这个数量级上。

(3) 污染物主要集中在珠江流域的广州和深圳段, 其中广州的多环芳烃和二恶英含量较高, 含量的均值分别在  $10^3 \mu\text{g}/\text{m}^3$  和  $10^4 \mu\text{g}/\text{m}^3$ 。珠江深圳段多氯联苯含量较其它地方偏高。

(4) 多氯联苯随着时间的推移, 从 96 年开始含量逐步减少, 而二恶英、多溴联苯醚呈上升趋势, 多环芳烃和有机氯农药的含量变化不显著, 但有机氯农药污染一直比较严重。

## 4 结束语

(1) 基于数据库、可视化和 Internet 技术, 利用 Microsoft 公司的 SQL Server 建立了珠江三角洲 POPs 数据库系统, 并利用中科院计算机网络信息中心科学数据中心开发的 VisualDB 成功实现了珠江三角洲 POPs 数据库属性数据的发布, 无需大量编程即可完成系统的建设。基于本系统, 做了珠江三角洲 POPs 污染物的时空分布分析, 实现了系统的初步应用。

(2) 本系统的初步结果表明: 珠江三角洲地区环境中的土壤、大气、水体和生物体等各种介质中均不同程度存在着 POPs 污染, 特别是农药的污染, 部分地区更是存在潜在的生态风险。区域内有机氯农药的生产使用是环境中污染物的主要来源, 需要开展大量的相关研究调查。

(3) 通过数据库的分析与可视化功能, 发现了一些 POPs 污染物的分布与演化规律, 说明数据库技术、可视化技术以及 Internet 技术在 POPs 研究方面已展现出很好的应用前景。

(4) 本系统需改进之处:

(a) 由于客观原因的限制, 现有的珠江三角洲的

POPs 数据还不是很全面, 而且数据不够标准化, 影响本系统的完整性, 有待不断完善;

(b) 进一步研究与建立珠江三角洲的 POPs 空间数据库, 并不断完善系统的各种功能, 使用户能更方便更智能化地利用该系统。

## 参考文献:

- [1] Wania F, Mackay D. Tracking the distribution of persistent organic pollutants [J]. *Environment Science & Technology*, 1996, 30(9): 390-396.
- [2] 余刚, 黄俊, 张俊义. 持久性有机物污染物: 倍受关注的全球性污染问题 [J]. *环境保护*, 2001, 11(4): 37-39.
- [3] Simonetta C, Romeo T, Nicoletta A, et al. POPs in key species of marine antarctic ecosystem [J]. *Microchemical Journal*, 2002, 73(1-2): 187-193.
- [4] 刘建国, 唐孝炎, 胡建信. 持久性生物累积性有毒污染物与国际相关控制策略和行动 [J]. *环境保护*, 2003(4): 52-56.
- [5] 安太成, 陈嘉鑫, 傅家谟, 等. 珠三角地区 POPs 农药的污染现状及控制对策 [J]. *生态环境*, 2005, 14(6): 981-986.
- [6] 杨燕红, 傅家谟, 盛国英, 等. 珠江三角洲一些城市水体中微量有机污染物的初步研究 [J]. *环境科学学报*, 1998, 18(3): 271-277.
- [7] 杨清书, 麦碧娴, 罗孝俊, 等. 澳门水域水体有机氯农药的垂线分布特征 [J]. *环境科学学报*, 2004, 24(3): 428-434.
- [8] 康跃惠, 麦碧娴, 黄秀娥, 等. 珠江三角洲地区水体表层沉积物中有机污染现状初步研究 [J]. *环境科学学报*, 2000, 20(增刊): 164-170.
- [9] 骆世昌, 余汉生. 珠江口及附近海域生物体中 BHC 和 DDT 的含量研究 [J]. *海洋通报*, 2001, 20(2): 44-50.
- [10] 方展强, 张润兴, 黄铭洪. 珠江河口区翡翠贻贝中有机氯农药和多氯联苯含量及分布 [J]. *环境科学学报*, 2001, 21(1): 113-116.
- [11] MAI B X, FU J M, SHENG G Y, et al. Chlorinated and PAHs in riverine and estuarine sediments from pearl river delta, China [J]. *Environmental Pollution*, 2002, 117(3): 457-474.
- [12] 中国科学院计算机网络信息中心. 可视化关系数据库管理与发布工具 VisualDB1.0 [EB/OL]. 2008-05. <http://www1.csdb.cn/prohtml/0.projects.download/pages/0313.html>.

(上接第 171 页)

- [10] 吴晔, 肖井华, 马宝军, 等. 手机短信网络的生长过程研究 [J]. *物理学报*, 2007, 56(4): 2037-2041.
- [11] Eubank S, Guclu H, AnilKumar V S. Modelling disease outbreaks in realistic urban social networks [J]. *Nature*, 2004,

429: 180-184.

- [12] Caldarelli G, Marsili M, Zhang Y C. A prototype model of stock exchange [J]. *Europhys Lett*, 1997, 40: 479-484.