

人类行为的动力学与统计力学研究^{*}

汪秉宏^{1,2,†} 韩筱璞¹

(1 中国科学技术大学近代物理系 理论物理研究所 非线性科学中心 合肥 230026)

(2 上海理工大学复杂系统科学研究中心 上海系统科学研究院 上海 200093)

摘要 对人类行为的统计特性的研究,涉及研究经济、心理、交通、社会学等学科的根本问题。在最近一个时期以来,针对人类行为的统计揭示出了大量的出乎人们意料的规律,吸引了众多研究者的注意。这些规律不仅对认识人类自身和对大量的社会学科的基础研究有着重要价值,而且在实际应用中有着广阔的前景。文章作者回顾了这一新近发展起来的领域的主要研究成果,并重点对中国科学技术大学的相关研究进行了介绍。

关键词 人类行为,人类动力学,时间间隔分布,空间移动斑图,非泊松分布

The dynamics and statistical mechanics of human behaviors

WANG Bing-Hong^{1,2,†} HAN Xiao-Pu¹

(1 Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China)

(2 Research Center for Complex System Science, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract An understanding of the statistical characteristics of human behavior involves the study of many fundamental problems in economics, psychology, transportation, and sociology. In recent times, statistical studies have revealed many new unexpected patterns of human behavior which have attracted much attention. These discoveries are important for our understanding of mankind itself, and for basic studies of a large number of problems in the social sciences. There are also broad prospects for their practical applications. This article will review the main results in this newly developing research area, with emphasis on the progress obtained by the Complex System Research Group at the University of Science and Technology of China.

Keywords human behaviors, human dynamics, inter-event time distribution, mobility pattern, non-Poisson distribution characteristics

1 引言

人类的行为具有高度的复杂性。探索人类行为的统计特性,对于研究经济、心理、交通、社会学等学科以及认识人类自身有着重要的意义。在以往的一些对社会、经济系统的研究中,常常把单个人的行为简化为可以使用泊松过程描述的稳态随机过程。这种假设必然导致的推论是人的行为的时间统计特征应该是较为均匀的,两个相继行为之间存在极大的时间间隔的概率很小。但是,自 2005 年 Barabási 通过对电子邮件发送与回复以及邮件通信等人类行为的时间间隔的实际统计,人们发现这些行为存在与上述假设极为不同的特性^[1,2]:长时间的静默与短期内的高频率的爆发,同时呈现在这些人类行为中,其时间间隔分布存在满足反比幂函数的胖尾,也就是说,这些行为的

发生过程是不能用泊松过程描述的。这一出人意料的研究结论提示人们,人类的个体行为可能存在复杂的动力学机制,而随之而来的一个重要的问题是,这种非泊松特性在人类行为中是不是普遍存在的?人们对这一问题进行了极为广泛的研究。

通过各种不同的数据收集方法,人们的研究涉及市场交易^[3-6]、网站浏览^[7,8]、电影点播^[9]、欣赏网络音乐^[10]、手机通讯^[11]、在游戏及虚拟社区中的行为^[12,13]、计算机指令的使用行为^[14]等,包含了商业行为、娱乐行为、日常使用习惯等众多的人类行为,在这些行为中,普遍发现有类似的偏离泊松过程的特性。这些现象显示出,除了受到生理周期强烈影响的部分

* 国家重点基础研究发展计划(批准号:2006CB705500)、国家自然科学基金(批准号:10975126, 10635040)、高等学校博士点基金(批准号:20093402110032)资助项目

2009-05-19 收到

† 通讯联系人. Email: bhwang@ustc.edu.cn

行为外,时间间隔统计所显示的非泊松特性可能是在人类行为中普遍存在的。Vázquez 等将人类行为分为幂指数为 -1 和 -1.5 的两大普适类^[15],而近期的更为深入的实证研究发现了更为广泛的行为特性,已经突破了这两大普适类的范围^[9]。除了时间间隔分布,部分人类行为事件前后时间间隔的相关性也得到了研究者的注意。研究发现,这些人类行为相邻时间间隔的相关性并不明显,而其他同样存在爆发性和长期静默性的自然现象(如地震等)常常存在正的相关性^[16]。这一项研究初步地把人类行为和其他复杂系统中的行为特性进行了比较,暗示可能存在统一的深层机制。

上述统计特性说明,人类的众多行为是不能使用泊松过程来描述的,那么一个重要的问题是:这种胖尾分布的行为特征的来源是什么?目前的一种重要的解释是基于任务队列理论的^[1,15,17,18],该理论把人的各种日常行为视作处理一系列的任务,并根据日常生活经验假设对这些待处理任务进行优先级划分,首先处理高优先级者,指出这种具有优先权的行为模式是造成胖尾分布的重要原因。这种基于任务队列的理论模型可以合理地解释很多人类行为中的非泊松特性,例如电子邮件和水陆邮件的发送等,而且可以相当容易地推广到存在多个个体之间的交互的情况^[19],在解释人类行为时间统计胖尾分布特征方面取得了很大的成功。

此外,由于影响人类行为的因素是多种多样的,所以,有部分研究工作从不同于任务队列的方面出发,提出了多种非排队论模型。例如,有的工作考虑了人类行为中的记忆效应^[20],有的研究了行为的周期性和季节性对非泊松机制的影响^[21],近期的一种理论则从多重泊松分布的角度解释了人类的行为特性^[22]。

最后,国际上有少数工作研究了人类行为的非泊松特性对网络传播、通讯等动力学过程的影响。例如,发现与一般的泊松特性相比,这种非泊松特性可以给系统带来一些特殊性质,比如更快的传播速度等^[23]。

除了发现人类行为的时间间隔分布中广泛存在着非泊松特性,最近也发现在人类行为的空间分布中也存在非泊松特性等复杂现象。2006 年,通过统计帐单传递^[24],人们间接地发现了人类的旅行行程分布存在接近于幂律的胖尾;2008 年,Gonzalez 等通过统计移动电话用户在不同基站区域的漫游过程^[25],更进一步研究了人的旅行行程分布,同样发现该分布具有无标度特性,与早期的结果基本一致。更为直接的是,基于 GPS 数据的统计结论^[26]也支

持人类行程分布中存在无标度特性。此外,在生物学观测中也发现大量的动物物种的运动具有类似的幂律形式的行程分布^[27—29]。由于这种幂律形式的行程分布存在较高频率的远程运动,它无法通过经典的随机行走进行描述。这种行程分布的广泛性,使得人们需要去思考它背后的动力学机制是什么。虽然对于动物行为中的幂律行程分布已经提出了觅食效率优化^[30,31]、嗅觉梯度机制^[32]、确定性行走^[33]等,但目前对人类的这种行程分布模式产生机制的解释方面的研究仍然非常缺乏。

这一领域已经成为国际复杂性研究的一个新兴热点,近年来仅在 Nature, Science, Physics Review Letters 这些国际顶级学术期刊上就出现 10 余篇相关的研究论文。该领域的兴起也引起了国内研究者的注意。目前,中国科学技术大学、上海理工大学、上海交通大学等都有相关工作在国内外学术期刊发表。这些工作可以简述如下:

在实证方面,中国科学技术大学复杂系统课题组的周涛等人与韩国成均馆大学及瑞典皇家学院合作,研究了电影点播中的人类行为模式及其与个体活动性之间的关系^[9,34],洪伟等人研究了人类短消息通讯中的时间间隔分布^[35],发现了多种无标度特性;上海理工大学课题组的张宁、李楠楠和周涛合作,分析了鲁迅、钱学森等名人的邮件通讯数据^[36,37];上海交通大学的胡海波等人研究了网络在线音乐的收听行为^[10]。在理论模型方面,中国科学技术大学的韩筱璞等提出用可自适应调节的兴趣机制来解释人类行为的非泊松特性^[38,39]。此外,上海理工大学方面还发表了针对人类动力学的综述文章^[40],上海理工大学的郭进利等和中国科学技术大学的周涛等人合作编写出版了专著《人类行为动力学模型》^[41],中国科学技术大学的周涛、韩筱璞、汪秉宏也在世界科学出版社出版的专著《Science Matters: Humanities as Complex Systems》中撰写了关于人类动力学研究的一个专门章节^[42]。

在本文中,我们将回顾这一领域的主要研究成果,并重点对中国科学技术大学的相关研究进行介绍。

2 人类行为的时间统计特性研究

2.1 人类行为的时间统计特性的实证统计

目前人们已经对大量的人类行为的时间分布特性进行了统计,内容涉及工作、日常生活和娱乐等,除了部分强烈受到生理节律或工作时间安排限制的行为外,在各种具有较高自主选择性的行为中,普遍

地发现了非泊松特性。在本文中,将对 Email 的发送和回复行为、普通邮件的发送和回复行为、电影点播行为这三种典型的日常行为的统计进行介绍。

2.1.1 Email 发送时间间隔分布与回复等待时间分布

2005 年,Barabási 最早研究了 Email 用户的发送时间间隔分布与回复等待时间分布。如图 1 所示,该图显示了某个 Email 用户在 3 个月内使用电子邮件的行为特性,其中不论是相邻两次电子邮件的发送时间间隔分布,还是回复一个电子邮件所需要的等待时间分布,都满足于幂指数为 -1 的幂律关系。这种幂律分布是极不均匀的,显示出电子邮件用户的发送与回复行为同时具有强烈的爆发性和长期的静默共存的特性。这种特性显然是不能通过泊松过程有效地加以描述。

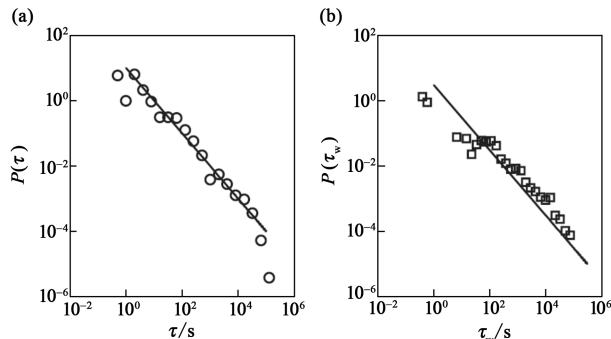


图 1 一个电子邮件用户在 3 个月内的(a)邮件发送时间间隔分布;(b)电子邮件回复时间分布((a)和(b)均满足指数为 -1 的幂律(图引自文献[1],其中横坐标 τ 是时间间隔长度, τ_w 为邮件回复时间,单位是 s;纵坐标 P 是该时间间隔出现的相对概率,注意:原作者对该概率没有进行归一化。)

2.1.2 水陆邮件回复等待时间分布

人们同样也对人发送普通水陆邮件的行为特性进行了研究。一些名人有着详细的邮件通讯记录,这为研究这一问题提供了很大的便利性。最为典型的统计结果如图 2 所示,该研究统计了爱因斯坦、达尔文和弗洛伊德的邮件回复时间分布,发现他们三人

的邮件回复时间分布都呈现出高度一致的指数为 -1.5 的幂律形式。而且,这些名人的邮件活动模式在长达数十年的观察期内是不断变化的,从开始时每年仅有不多的邮件,到后期平均每年多达数百封。因此,这一跨越整体统计时间的统计虽然显示出共同的幂律特性,但它并非是稳定的。Vázquez 等曾经基于以上这些统计结果提出了人类非泊松性行为模式的两大普适类:一个是分布指数为 -1 的,如 Email 使用等,另一个分布指数为 -1.5,如发送普通水陆邮件的行为等。

上海理工大学的李楠楠等也对钱学森的邮件通讯记录进行了统计(见图 3),统计结果表明,钱学森的邮件发送时间间隔分布满足指数为 -2.1 的幂律,同样具有非泊松特性,但已经超出了两个普适类的范围。

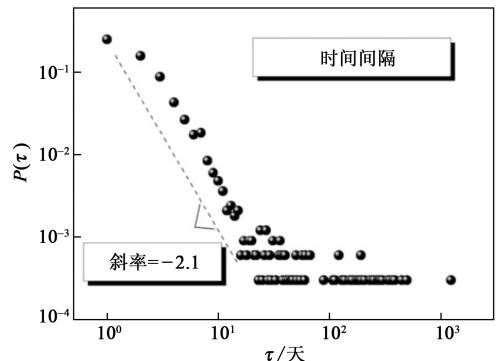


图 3 钱学森的邮件通讯时间间隔分布(图引自文献[37])

2.1.3 电影点播时间间隔分布

中国科学技术大学的周涛等统计分析了在线电影网站 Netflix 的公开数据。这一数据库包含了 17770 部电影和来自 447139 个用户的近 1 亿条点播记录。在不对用户进行区分的情况下,统计所有用户的电影点播时间间隔分布,如图 4 所示,这一分布具有指数为 -2.08 的幂律尾部,也超出了两个普适类的范围。

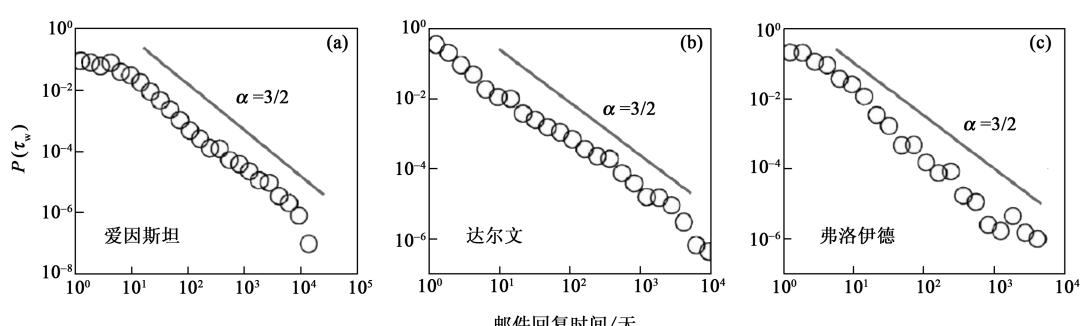


图 2 爱因斯坦(a)、达尔文(b)和弗洛伊德(c)等的邮件回复时间分布,它们均满足指数为 -1.5 的幂律(图引自文献[17],其中 α 是曲线使用幂函数拟合所得的幂指数的绝对值)

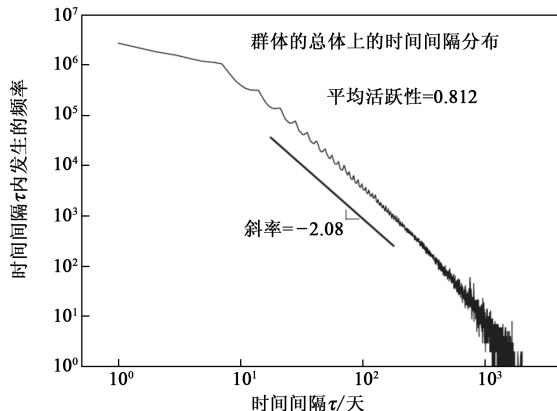


图 4 所有用户的时间间隔分布. 该分布接近幂指数为 -2.08 的幂律(分布曲线中存在周期为一星期的波动, 这一波动在电子邮件通讯行为中也曾被观察到. 图引自文献[9])

考虑到在现实中用户点播电影的行为具有活跃程度的区别, 部分用户具有更高的平均点播频率, 为了考察用户的活跃程度对点播行为的影响, 定义用户的活跃程度为在统计时间范围内该用户的单位时间内的平均点播次数, 把所有用户按照活跃程度按递减顺序排序, 然后把这些排序后的用户依次分为 20 个所包含的用户数目基本相同的群体. 统计发现, 这 20 个群体的时间间隔分布都满足幂律, 其幂指数变化范围从 -1.5 到 -2.7 , 指数绝对值与群体的平均活跃程度的关系如图 5 所示, 两者呈现单调的非线性正相关关系.

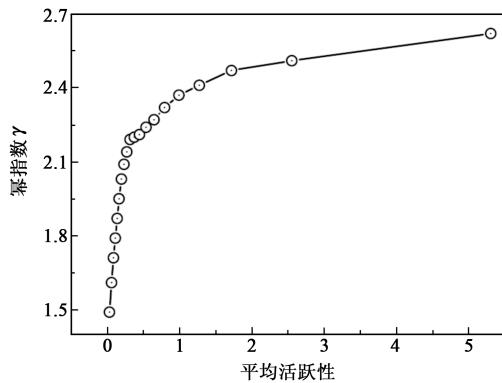


图 5 幂指数和平均活跃程度之间的关系(图引自文献[9])

2.2 人类行为的非泊松特性的任务队列理论

任务队列理论是目前研究人类行为统计特性的主流理论. 这一理论将人的日常行为视作处理一系列的任务, 而人面临的主要问题是如何从这一系列的任务中选择首先处理的对象. Barabási 排队论模型中考虑了三种情况: 第一种是按照先进先出的原则, 优先处理首先到来的任务, 这种处理方式可以导

致指数型的等待时间分布尾部; 第二种是从任务队列中任意选取一个任务进行处理; 第三种方式是优先处理任务队列中的高优先级任务. 基于这样的前提, 模型可以构造如下:

假设个体的任务列表长度为 L , 每个任务都被赋予一个优先权 x_i , 其中 $i = 1, 2, \dots, L$; 该优先权由一个均匀分布生成. 在每一个时步, 该个体执行一个任务, 并把该任务从任务队列中移除, 同时向队列中添加一个新的任务, 新任务具有一个新的优先权. 在个体选择要处理的任务时, 个体以概率 p 执行优先级最高的任务, 以概率 $1-p$ 任意挑选一个任务处理. 这样, 当 p 趋向于 1 时, 该模型对应于优先处理任务队列中的高优先级任务的情况; 当 p 趋向于 0 时, 该模型对应于从任务队列中任意选取一个任务进行处理的情况.

该模型所考察的是当一个任务进入队列到被处理所需要的等待时间分布. 模型的计算结果如图 6 所示, 当 p 趋向于 1 时, 该模型可以获得指数为 -1 的幂律尾部形式的任务等待时间分布, 这个指数同人类电子邮件回复等待时间分布的实证统计是一致的; 而当 p 趋向于 0 时, 这一分布趋向于指数形式. 这一结果说明, 人类行为中所存在的胖尾特性来自于人在日常处理事务时常常优先处理某些较重要的任务, 有效地解释了一批人类行为中的非泊松特性的机制.

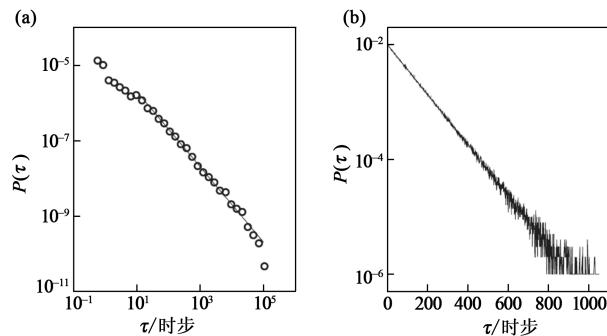


图 6 Barabási 排队论模型生成的时间间隔分布. 优先权从一个 $[0, 1]$ 之间的平均分布中抽取, 队列长度为 $L = 100$, 共模拟统计 10^6 时步((a) 双对数坐标下的时间间隔分布的尾部, 取 $P = 0.9999$, 为模型的确定性选择高优先级任务的极限情况, 其中直线对应于指数为 -1 的幂律; (b) 单对数坐标下的时间间隔分布, 取 $P = 0.00001$, 为随机选取任务的极限情况, 满足指数分布. 图引自文献[1])

2.3 人类行为的非泊松特性的自适应兴趣驱动模型

不同于排队理论, 中国科学技术大学的研究者提出了自适应兴趣模型, 来研究人类行为中的非泊

松特性^[38,39]. 在现实中, 如果观察某个人所做的某种具体的事情, 如玩游戏、看电影等, 常常可以看到一种调节机制: 即做完该事后, 人会改变此后做该事的概率或频率. 比如, 当一个人在很久没有玩游戏后, 偶尔玩一次游戏常常会激起他的兴趣, 使得他玩游戏的频率增加, 又如吃到一种很久没吃的食品, 会使人在短时间内对该食品产生很大兴趣, 而如果该频率过高(如同吃该种食品太多倒了胃口), 他常常又会出现对这种行为的节制而使得频率下降. 但是, 这种调节的范围常常是非常宽的, 最低频率和最高频率之间可能存在几个数量级的差别. 这种调节机制可能会使人的兴趣或活性出现准周期的变化, 可能会造成人类行为中时间间隔分布的标度性. 该模型假设: 对某种行为, 每做一次都会改变下一次做该事件的概率. 考虑最简单的情况, 即每次概率改变的比例是相同的. 事件的时间间隔存在两个阈值, 即间隔过小或过大都会改变做该事的概率变化的趋势: 即如果两次事件的间隔过小, 会使做该事的概率以固定的比例减小, 反之使概率以相反的比例增大.

模型规则为: 时间为离散的, 在时步 t 时某事件发生的概率为 $r(t)$; 每发生一次该事件, $r(t)$ 要进行一次更新, 更新规则为 $r(t+1) = a(t)r(t)$, 其中 $a(t)$ 只能取两个值(如果本次事件与上次的时间间隔小于或等于 T_1 , 则 $a(t) = a_0$; 如果该时间间隔大于 T_2 , 则 $a(t) = a_0 - 1$; 对于其他情况, $a(t) = a(t-1)$; 其中 T_1, T_2 为固定的正整数, a_0 为固定的参数, 分别满足 $T_1 \ll T_2, 0 < a_0 < 1$.

根据以上规则, 一般本次事件与上次的时间间隔小于或等于 T_1 的情况大多出现在当 $r(t)$ 等于或者接近于 T_1^{-1} 时. 因此, T_1 的取值决定了 $r(t)$ 所能够达到的最高值; 同样 T_2 的取值决定了 $r(t)$ 所能够达到的最小值在 T_2^{-1} 附近. 如果同时将 T_1 和模型的最长时间尺度放大同样的倍数, 并保持 T_1 和 T_2 的比值不变, 那么在该时间尺度下, $r(t)$ 的最大最小值之间的平均比率与原来相比是不变的, 因此所得到的新的时间间隔分布与原来的也是相同的. 所以, T_1 实际也表示了该模型最小有效的时间尺度, 因此, 在以下的讨论中, 为了不失一般性, 我们固定 $T_1 = 1$.

在数值模拟中, 取 $r(t)$ 的初始值固定为 1.0. 通过数值模拟, 当 T_1 和 T_2 存在有 3 个数量级以上的差别时, 可以生成接近幂律的时间间隔分布, 幂指数为 -1; 随着 T_2 的缩小, 该分布逐渐偏离幂律趋于指数, 如图 7(a) 所示; 对于相同的 T_1 和 T_2 , 较大的 a_0 也会使得分布曲线偏离幂律, 如图 7(b) 所示.

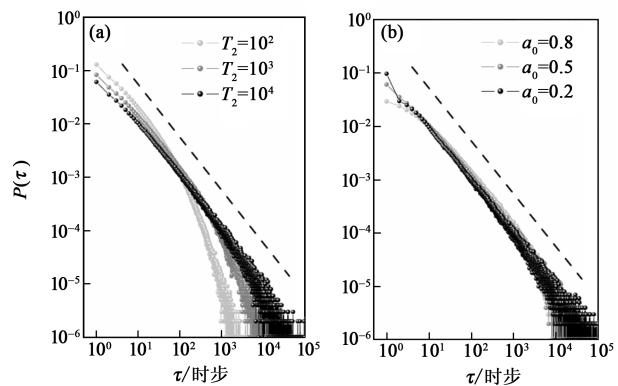


图 7 数值模拟时的时间间隔分布 (a) 固定 $T_1 = 1, a_0 = 0.5$, T_2 与时间间隔分布的关系; (b) 固定 $T_2 = 10000$, $T_1 = 1, a_0$ 与时间间隔分布的关系(数据由 100 次独立模拟, 每次模拟 10000 个事件, 取平均获得, 图中虚线表示指数为 -1 的严格幂律, 图引自文献[39])

3 人类和其他生物行为的空间统计特性

3.1 生物行为的空间统计特性的实证统计

长期以来, 在人们的常规认识中, 生物或者人类的运动行为常常被简单地视作简单的经典的随机行走. 在这种经典随机行走中, 发生远大于平均行程的长距离迁移的概率是极低的. 然而, 实证研究发现, 很多生物的运动行为的统计特性并非如此简单. 1996 年, Viswanathan 等针对信天翁的数据进行的研究^[30], 显示出信天翁的飞行时间分布接近于幂律, 具有无标度特性, 但由于该数据较为粗糙, 其幂律分布特性尚存在疑义^[43]. 但是, 对更广泛的物种的运动行程分布所作的更精确的统计, 包括对浮游微生物^[27]、蜘蛛猴^[28]等的行为的统计研究, 显示出这种标度性的行程分布行为特性是现实存在的. 特别是 Sims 等对多种海生动物(涉及鱼类、爬行类、鸟类等)的运动行为的统计^[29], 显示出清晰的幂律特性的行程分布 $p(l) \sim l^{-\alpha}$, 一般幂指数绝对值 α 介于 1.7 到 2.4 之间, 如图 8 所示. 在这些统计中, 一次运动的行程被定义为运动中相邻两次停留之间所行走的距离. 这种幂律行程分布行为特性可以用 Levy 飞行来进行描述, 说明生物在运动过程中, 存在大量的短途运动的同时, 也常常进行远距离的移动, 不能使用经典的随机运动进行描述. 与经典的随机行走相比, 这种运动模式的行程分布极不均匀, 具有较高的长程运动概率.

3.2 人类行为的空间统计特性的实证统计

最近, 人类出行行为中也存在着标度性的行程

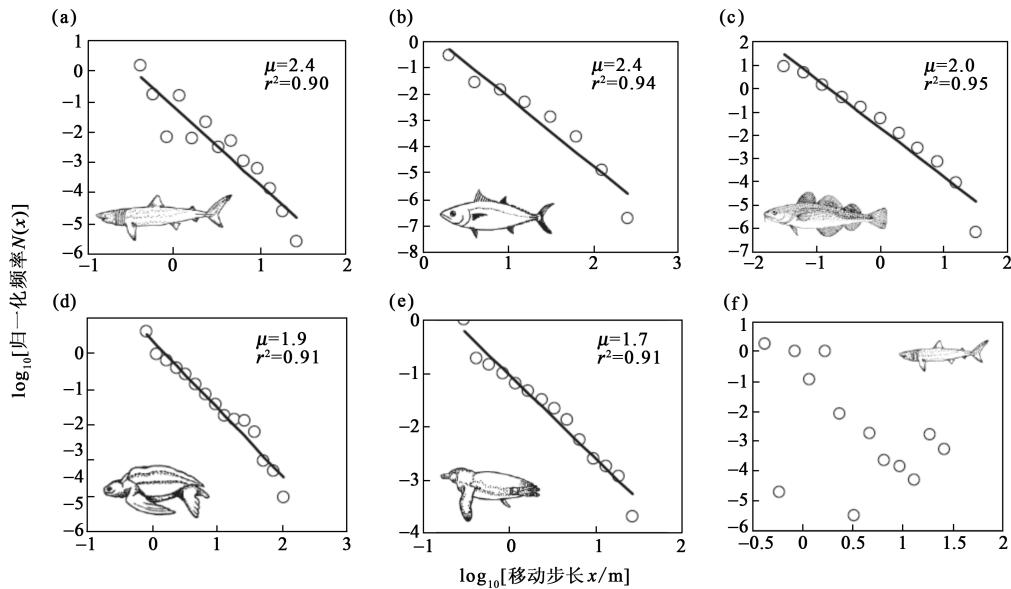


图 8 多种海生动物运动行程分布 (a)未成年和成年的姥鲛;(b)大眼鲔;(c)大西洋鳕鱼;(d)棱龟;(e)麦哲伦企鹅;(f)一岁的姥鲛. 图引自文献[29]

分布.近年来,Brockmann 等^[24]通过对帐单的统计,研究了现实中人类的空间运动行为,随后,Gonzalez 等^[25]通过统计移动电话用户在不同基站之间的漫游信息,获得了更为直接的人类空间运动数据,对人类空间运动行为进行了更为深入的探讨.这些研究发现,人类的运动行为具有迥异于经典随机行走的特性:

首先,与很多动物的运动行为类似,人类的空间行程分布是很不均匀的,其行程分布满足指数约为 -1.5 到 -1.7 的具有明显尾部截断的幂律分布,说明其远大于平均行程的长程运动的概率远高于传统的随机运动描述,例如图 9 所示.

但是,统计结果还揭示出人类空间运动具有较明显的局域性,远离某个小区域的概率随时间的衰减较慢.这一特性也可以从个体的回旋半径的增加趋势上来观察.个体 a 在 t 时间内的回旋半径 $r_g^a(t)$ 定义为

$$r_g^a(t) = \sqrt{\frac{1}{n_c^a(t)} \sum_{i=1}^{n_c^a(t)} (\mathbf{r}_i^a - \mathbf{r}_{cm}^a)^2},$$

其中 \mathbf{r}_{cm} 表示该个体在统计时间段内的平均位置, \mathbf{r}_i^a 为各个曾到达的位置, $n_c^a(t)$ 为在 t 时间段内到达的位置数,上标 a 表示对个体 a 进行的统计.如图 10 所示,对于经典的随机行走和 Levy 飞行,回旋半径随时间为幂函数增加,但是实证发现,实际人的回旋半径增加速度非常缓慢,接近于对数增长,显示出强烈的局域性.

这一特性显示,人类空间运动行为不但不能使用传统的随机行走进行描述,而且也不能使用 Levy 飞行进行描述,因为尽管 Levy 飞行同样具有较高的长程运

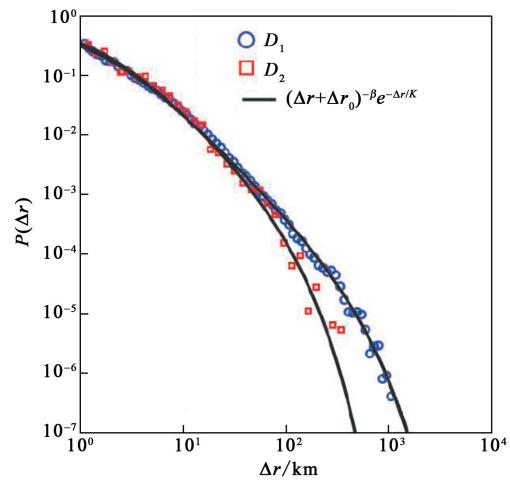


图 9 Gonzalez 等基于移动电话数据统计得到的两组数据的人类行程分布,其中指数 β 值在 1.65 附近.图引自文献[25] (图中对于分布曲线采取了函数为 $p(\Delta r) = (\Delta r + \Delta r_0)^{-\beta} e^{-\Delta r/K}$ 的拟合.其中 K 是一个拟合参数,表征了该分布函数的尾部相对于严格幂律的偏离大小.之所以采用该函数进行拟合,是因为该函数在幂律形式的基础上,可以清晰地表达出实证中常常观察到的指数尾现象. K 实际上表示了截断发生的位置:当 $\Delta r < K$ 的时候,指数项影响不明显,而当 $\Delta r > K$ 的时候,指数衰减项起主要作用. K 从数量级上指出了截断发生的位置)

动概率,但它不具有局域性,而且对于个体而言,常常存在少数几个经常前往的地点,到不同地点的平均频率也满足幂律分布,而且具有明显的各向异性.

3.3 生物运动的驱利性模型

根据生物行为所具有的两种普遍的行为准则——“趋利性”和“最小努力性”,中国科学技术大学

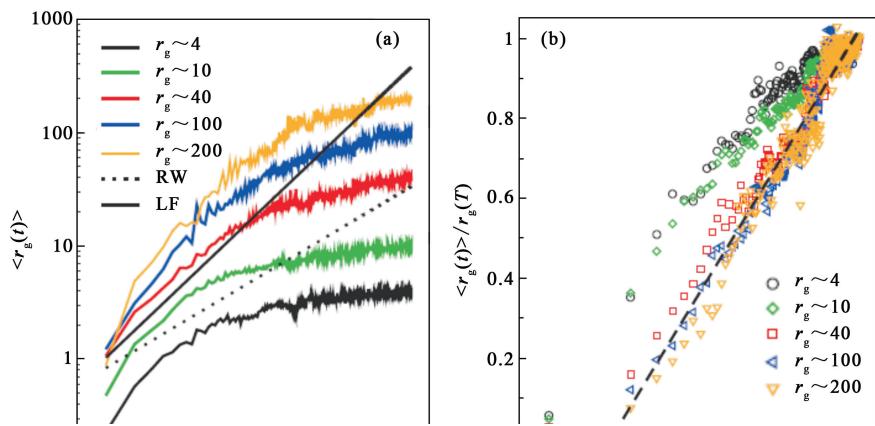


图 10 基于移动电话数据统计得到人类运动的局域性特征 (a) 双对数坐标系下, 具有不同运动活跃性的个体的回旋半径随时间的变化, 显示出对随机行走和 Levy 飞行所对应的幂函数增加律的偏离; (b) 个体的回旋半径增加趋势满足对数律(图引自文献[25]的支持文档, 图中 RW(random walk) 表示随机行走, LF(Levy flight) 表示 Levy 飞行, 即行程分布满足幂律的随机运动方式)

的研究者提出了趋利性生物运动模型^[44]. 趋利性常常表现为生物体一般倾向于向着能够提供更为丰富的生活资源、有着更为适宜的环境的区域生长或移动. 而最小努力性即生物一般在保证能够获得足够的资源的同时, 会尽可能地减小相应的代价(比如时间、营养、能量、机会等)的付出. 对于众多生物物种而言, 这两者是生物行为适应性的保证, 也是决定其各种宏观行为的基础. 在该模型中, 由于实际的生物个体一般具有较稳定的有限的生活区域, 因此设定个体在一个 $N \times N$ 大小的离散格子空间中运动. 每个格子具有一定的资源, 对于坐标为 (i, j) 的格子, 使用 $V(i, j) (0 \leq V(i, j) \leq V_m)$ 表示, 其中 V_m 表示资源的上限. 一般生物体所需的资源, 特别是食物资源, 是可有限地再生的, 因此在该模型中, 资源可以有限度地缓慢恢复. 在该空间上, M 个个体同时进行运动. 如图 11 所示, 基于趋利性原则和最小努力原则, 在每一时步个体的运动规则和格点状态更新规则如下: 各个个体寻找距离各自的当前位置最近的具有资源最大值 V_m 的格点作为其目标位置. 如果满足这样条件的格点有多个, 则随机选择其中一个作为目标位置; 个体移动到目标位置上, 并把该位置的资源 V 置为 0, 表示资源已经被消耗; 所有个体的位置更新完毕后, 对于所有资源小于 V_m 的格子, 其资源 V 增长 1, 表示资源的有限度的再生.

定义比率 $r = MV_m/N^2$, 该值表示在各个体运动时, 空间中资源小于 V_m 的格点所占的稳定的比例, 也表示资源相对于消耗的剩余程度. 当 $r = 1$

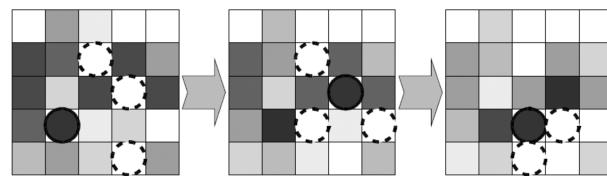


图 11 模型中一个个体在 3 个连续时步的运动示意图(图中实球表示个体当前所占据的格点, 虚球表示在下一时刻个体所可能占据的位置, 格子的颜色深浅表示资源 V 的大小, 白色表示 $V = V_m$, 颜色越深表示 V 值越小. 图引自文献[44])

时, 表示资源的再生恰好满足各个个体的消耗, 无任何多余的资源; r 越小, 资源剩余量越大, 也就是说资源越充足. 该值是这个模型的一个主要参数, 在数值模拟中, V_m 的取值由 r 来确定.

行程分布 $p(d)$ 是我们主要关心的统计学量. 模拟所得行程分布如图 12 所示, 幂指数 α 的绝对值随 r 的减小而增大. 当 $r < 0.8$ 时, $p(d)$ 的分布曲线已经偏离了常规的幂函数, 而更为接近指数分布, 此时的运动方式已经不是 Levy 飞行, 说明如果资源(比如食物等)的丰富程度增加, 生物的运动中的长程运动的概率会趋向减小, 甚至可能导致幂律运动特性的丧失. 这种的生物运动方式随食物丰富程度而改变的现象在实际观测中也有发现, 比如蜜蜂类生物在食物匮乏时行程分布的幂指数绝对值小于食物丰富时^[9], 该模型的结果与此基本一致. 同时, 这一结果也暗示, 在部分观测中所观察到的偏离 Levy 飞行模式的生物运动可能来自于资源较为丰富的区域.

个体数目 M 的大小对 $p(d)$ 也有一定影响。在该模型中, M 的大小并不表示生物种群密度, 而是表示生物活动区域的重叠程度, 或者说对资源的共享程度。如图 12 (b) 所示, 当 r 值接近 1 时, M 越大, 所得到的分布函数 $p(d)$ 的幂指数 α 就越小, 同时较大的 M 也会使分布函数出现明显的尾部截断。目前尚无实际的观测数据来对生物活动区域的重叠程度和其运动行为的关系进行定量的考察, 但是该模型的结果暗示, 当食物资源趋向临界时, 相比于一些有着严格领地的物种, 那些无明显领地区分的物种的行程分布可能存在更大的长程运动几率, 同时其运动行为也会更多地受到食物资源状况的影响。在这一方面, 目前尚无可对应的实证结果, 这一结论需要更深入的实测数据的支持。

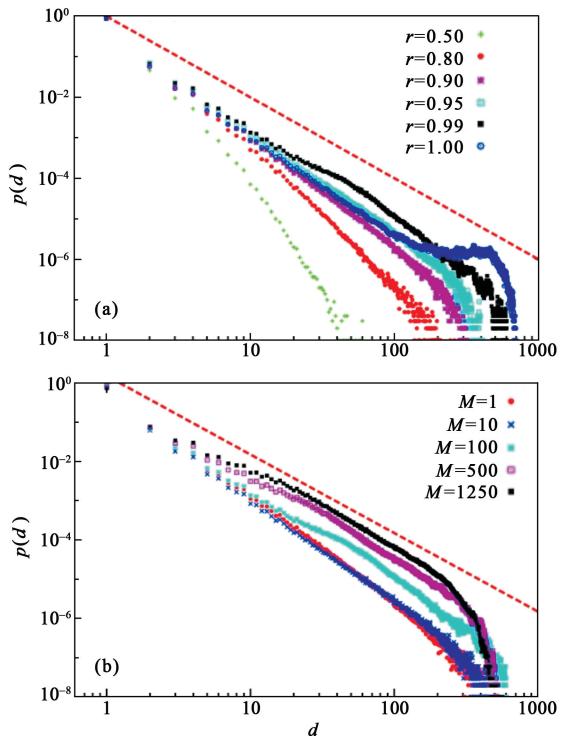


图 12 (a) 当 $M = 100, r = 0.50, 0.80, 0.90, 0.95, 0.99$ 和 1.00 时所得行程分布 $p(d)$; (b) 当 $r = 0.99, M = 1, 10, 100, 500, 1250$ 时所得行程分布 $p(d)$ (以上各组数据均是在 $N = 500$ 时, 统计了初始化后的 10^6 个移动事件, 并进行 100 次独立模拟取平均所得。虚线表示指数为 -2 的幂律分布。图引自文献[44])

3.4 人类空间运动的层次性交通模型

根据实际的行政区划以及行政区内的交通常常具有便利性等情况, 中国科学技术大学的研究者提出了基于层次性交通网络的人类运动模型^[45]。现实中, 因为行政区内的交通通常具有便利性, 人的出行常常是这样的: 如果一个人试图从属于某个中心城

市的乡村(属于 A 城市辖区的 a1 小镇)到属于另一个城市的乡村(属于 B 城市辖区的 b2 小镇)去, 那么他常常选择首先前往所属中心城市(A 城市), 然后由 A 城市到目标所在中心城市(B 城市), 再从 B 城市到达目标 b2 小镇。在大多数情况下, 这样的旅行方式常常最为便利。这样的旅行方式, 使得从个体层面来看, 交通系统具有明显的层次性。

该模型即基于这种层次性的地理交通网络结构。在现实中, 这种层次一般有 3 到 5 级, 比如在中国, 省会—市—县—镇—村可视为是 5 级结构。因此, 在模型中, 我们将所有的城市分为 3 到 5 个级别。以一个 3 级结构为例, 中心位置的城市为最高级城市(1 级城市), 整个区域被划分成 3×3 共 9 个子区域, 除了中间 1 级城市所在子区域, 每个子区域的中心位置的城市定义为各个辖区的中心城市(2 级城市)(中间区域的中心城市仍是该 1 级城市), 同时在每个 2 级城市周围分布着 8 个 3 级城市, 同样 1 级城市周围也分布有 8 个 3 级城市(相当于每个子区域又被分为 9 个次级区域)。在实际数值模拟中, 我们常常定义 1 级城市并不止 1 个。属于同一个高级城市的低级城市(在相应的同一子区域内)之间全连通, 同时全都和该高级城市直接连接。但属于不同高级城市的低级城市之间没有连边。每个小区域内的 3 级城市之间是全连通的, 同时也都和所属 2 级城市相连, 而每个 2 级城市之间(同属该 1 级城市)也是全连通的, 并也都和 1 级城市直接相连。

在这样的网络中, 一个个体沿着该层次性网络的连边(而不是地理连边)进行随机行走。显然, 该模型不允许直接进行跨子区域的运动, 除非是处在该子区域的中心城市。因此, 在该层次性网络上的随机行走具有这种特性: 人要到达不从属于同一个中心城市的目的地时需要经过两者所从属的中心城市。一般情况下, 这种特性是符合现实中人们的旅行方式的。该个体每一时步所移动的空间距离 d , 被定义为该时步移动的初始城市 A(地理坐标 (x_a, y_a)) 到终点城市 B(坐标 (x_b, y_b)) 之间的几何距离:

$$d = \sqrt{(x_b - x_a)^2 + (y_b - y_a)^2}.$$

注意: d 完全是基于地理位置计算, 与该层次网络完全无关。 d 的分布函数为 $p(L)$, 就是该模型所主要关注的。

该模型的主要参数包括: 结构层次数目 N ($3 \leq N \leq 5$), 第一级城市数 k , 每个上级区域被分为次级区域的个数 M 。在这种结构中, 第 2 级城市总数目为 $(M - 1)k$ 个, 第 3 级城市总数目为

($M-1$) Mk 个, ……, 第 n 级城市总数目为 k ($M-1$) $M^{(n-2)}$ 个 ($2 < n \leq N$). 在数值模拟中, k 固定为 9 个. 图 13 所示为取 5 级结构和 $M=9$ 时所得结果. 其中图 13(a) 表示, 尽管大多数的行走是在同一区域低级的城市之间完成的, 但仍会出现爆发性的长距离行走 (图 13). 行程分布呈幂律关系, 其指数接近于 -2. 这一结果显示, 沿这种层次性地理网络的运动可以导致标度性的行程分布.

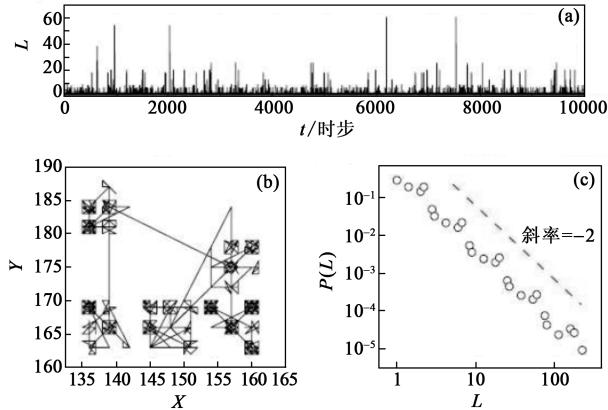


图 13 层次性交通模型所得到的空间运动斑图和行程分布 (图引自文献[45])

在现实中, 中心城市往往有着较大的城市规模和较多的人口, 一般对于居住下属城市的居民而言, 总体上前往中心城市的概率要大于到其他同级城市的概率. 因此, 在随机性城市分布模型的基础上, 引入了各个城市的权重 w , 来表示各城市对人流的影响力. 对第 n 级城市, 其权重 $w_n = w_0 r^{(N-n)}$. 其中 w_0 为常数, 固定为 1; $r \geq 1$, 是一个主要参数. 个体沿连边运动时, 到达某个相邻城市的概率 $p = w_j / (\sum w_i)$. 在这种情况下, 高级城市的权重随着 r 的增大而增大, 个体具有较高的概率向较高级城市行走, 因此也具有较高的长程运动概率. 模拟所得行程分布仍然具有肥尾特性, 当 r 在 1 到 2 之间时, 所得行程分布指数可在 -1.5 到 -2.7 之间连续变化. 该结果可以覆盖实证统计所得的指数 -1.5 的结果. 这一结果说明, 城市的差异性以及所导致的人流方向的不均等性, 是行程分布指数绝对值低于该层次系统所决定的分布指数的主要原因.

4 结束语

由于该领域的发展时间短暂, 目前在存在有大量问题有待于深入的研究: 首先, 目前已有的实证统计主要针对个体行为, 但仍然存在大量的个体行

为的特性并未被研究, 已有的研究结果尚难以根据统计特性区分个体行为的主要类别; 而针对团体行为的实证研究更几乎是空白. 事实上, 人类的行为常常受到起社会关系的影响, 在这方面定量的实证研究仍然非常欠缺. 另外, 一些最近发展的理论, 例如人类动力学的普适类假说, 受到了新的实证数据的挑战, 更清晰和令人信服的图景需要更多和更深入的实证分析. 除了人类的个体行为, 一些最新的统计也发现, 一些社会团体的宏观行为也具有类似的非泊松特性, 例如国家之间的战争的时间间隔分布等; 由于目前的实证统计有限, 对于社会团体而言, 这些特性在多大范围内存在, 是否与人类个体行为具有相似的生成机制, 都仍然是未知问题, 需要进行深入的研究. 而在研究人类行为的空间分布方面, 目前的实证数据全是根据帐单、手机漫游等数据间接获得的, 缺少对人类行为空间分布的直接观察; 而其产生机制和动力学效应方面的研究目前几乎没有. 目前的理论模型研究, 虽然已经提出了多种唯象机制来解释人类行为中的非泊松特性, 但是这些机制难以覆盖全部的人类行为中的非泊松特性现象, 需要有新的更具有普适性的模型的提出. 人类行为特性对各种社会系统的动力学效应的影响研究, 尽管已经出现了少数这方面的研究, 但这个领域的发展还处在初始阶段, 有着众多的研究空白和问题需要等待解决, 其重要理论意义和广阔的应用前景, 目前仅仅是露出了冰山一角, 有待于研究者的深入发掘.

参考文献

- [1] Barabási A L. Nature , 2005, 435: 207
- [2] Oliveira J G, Barabási A L. Nature, 2005, 437: 1251
- [3] Plerou V, Gopikrishnan P, Amaral L A N et al. Phys. Rev. E, 2000, 62: 3023(R)
- [4] Masoliver J, Montero M. Phys. Rev. E , 2003, 67: 021112
- [5] Politi M, Scalas E. Physica A , 2008, 387: 2025
- [6] Jiang Z Q et al. Physica A, 2008, 387 : 5818
- [7] Dezső Z, Almaas I, Lukács A et al. Phys. Rev. E , 2006, 73: 066132
- [8] Gonçalves B, Ramasco J J. Phys. Rev. E , 2008, 78: 026123
- [9] Zhou T, Kiet H A T, Kim B J et al. Europhys. Lett. , 2008, 82: 28002
- [10] Hu H B, Han D Y. Physica A, 2008, 387: 5916
- [11] Candia J, González M C, Wang P et al. J. Phys. A: Math. Theor. , 2008, 41: 224015
- [12] Henderson T, Nhati S. Proc. 9th ACM Int. Conf. on Multimed. ACM Press , 2001. 212
- [13] Grabowski A, Kruszewska N, Kosiński R A. Phys. Rev. E, 2008, 78: 066110

- [14] Baek S K, Kim T Y, Kim B J. *Physica A* , 2008, 387: 3660
- [15] Vázquez A, Oliveira J G, Dezsö Z *et al.* *Phys. Rev. E*, 2006, 73: 036127
- [16] Goh K I, Barabási A L. *Europhys. Lett.* , 2008, 81: 48002
- [17] Vázquez A. *Phys. Rev. Lett.* , 2005, 95:248710
- [18] Gabrielli A, Caldarelli G. *Phys. Rev. Lett.* , 2007, 98: 208701
- [19] Oliveira J G, Vázquez A. *Physica A* , 2009, 388: 187
- [20] Vazquez A. *Physica A* , 2006, 373 : 747
- [21] Ceser A, Hidalgo R. *Physica A* , 2006, 369 : 877
- [22] Dean Malmgren R, Daniel B S, Motter A E *et al.* *Proc. Natl. Acad. Sci. USA* , 2008,105: 18153
- [23] Vázquez A, Rácz B, Lukács A *et al.* *Phys. Rev. Lett.* , 2007, 98: 158702
- [24] Brockmann D, Hufnagel L, Geisel T. *Nature*, 2006, 439: 462
- [25] Gonzalez M C, Hidalgo C A, Barabasi A L. *Nature*, 2008, 453: 779
- [26] Rhee L, Shin M, Hong S *et al.* On the Levy-walk nature of human mobility. In: Proc. 27th IEEE Conf. Comput. Commun. , INFCOM, IEEE Press, 2008
- [27] Bartumeus F, Peters F, Pueyo S *et al.* *Proc. Natl. Acad. Sci. USA*, 2003,100: 12771
- [28] Ramos-Fernandez G, Mateos J L, Miramontes O *et al.* *Behav. Ecol. Sociobiol.* , 2004,55: 223
- [29] Sims D W, Southall E J, Humphries N E *et al.* *Nature*, 2008, 451:1098
- [30] Viswanathan G M, Buldyrev S V, Havlin S *et al.* *Nature* 1999, 401:911
- [31] Bartumeus F, Catalan J, Fulco U L *et al.* *Phys. Rev. Lett.* 2002,88: 097901
- [32] Reynolds A M, Phys. Rev. E, 2005,72:041928
- [33] Santos M C, Boyer D, Miramontes O *et al.* *Phys. Rev. E*, 2007, 75: 061114
- [34] 周涛. 复杂系统与复杂性科学,2008,5(1):1[Zhou T. Complex Systems and Complexity Science,2008,5(1):1 (in Chinese)]
- [35] Hong W, Han X P, Zhou T *et al.* *Chin. Phys. Lett.* , 2009,26: 028902
- [36] 李楠楠,张宁,周涛. 复杂系统与复杂性科学,2008,5(3): 43[Li N N, Zhang N, Zhou T. Complex Systems and Complexity Science,2008,5(3):43 (in Chinese)]
- [37] Li N N, Zhang N, Zhou Tao. *Physica A* ,2008,387:6391
- [38] 韩筱璞,周涛,汪秉宏. 复杂系统与复杂性科学, 2007,4(4):1 [Han X P,Zhou T,Wang B H. Complex Systems and Complexity Science,2007,4(4):1 (in Chinese)]
- [39] Han X P, Zhou T, Wang B H. *New J. Phys.* , 2008,10 :073010
- [40] 李楠楠,周涛,张宁. 复杂系统与复杂性科学,2008,5(2):15[Li N N, Zhou T, Zhang N. Complex Systems and Complexity Science,2008,5(2):15 (in Chinese)]
- [41] 郭进利,周涛,李季明等. 人类动力学模型. 香港:上海系统科学出版社,2008[Guo J L,Zhou T,Li J M *et al.* Models of Human Dynamics. Hong Kong: Shanghai System Science Publishing House,2008 (in Chinese)]
- [42] Zhou T, Han X P, Wang B H. Towards the understanding of human dynamics, In: Burguete M, Lam L eds. *Science Matters -Humanities as Complex Systems*. Singapore: World Scientific Publishing, 2008. 207—233
- [43] Edwards A M, Phillips R A, Watkins N W *et al.* *Nature*, 2007, 449:1044
- [44] Han X P, Hao Q, Zhou T *et al.* Origin of Levy Flight in Biological Mobility Pattern. unpublished
- [45] Hao X P, Hao Q, Wang B H *et al.* Origin of the Scaling Law in Human Mobility: Hierarchical Organization of Traffic System. arxiv,2009,0908:1221