

人类行为时空特性的统计力学

周涛^{1,2}, 韩筱璞^{2,3}, 闫小勇^{4,5}, 杨紫陌¹, 赵志丹¹, 汪秉宏^{2,6}

- (1. 电子科技大学互联网科学中心 成都 611731; 2. 中国科学技术大学近代物理系 合肥 230026;
3. 杭州师范大学阿里巴巴商学院 杭州 310036; 4. 石家庄铁道大学交通运输学院 石家庄 050043;
5. 北京师范大学系统科学学院 北京 海淀区 100875; 6. 上海理工大学复杂系统科学研究中心 上海 杨浦区 200093)

【摘要】人类行为的定量化分析,特别是时空统计规律的挖掘和建模,是当前统计物理与复杂性科学研究的热点。对人类行为的深入理解,有助于解释若干复杂的社会经济现象,并在舆情监控、疾病防治、交通规划、呼叫服务、信息推荐等方面产生应用价值。该文综述人类行为时间和空间特性方面的研究进展,内容包括人类行为时间特性的实证分析和建模,人类行为空间特性的实证分析和建模,以及人类行为统计分析的应用研究。该文还将评述当前研究存在的亮点和不足,指出若干亟待解决的重大理论和实际问题。

关键词 人类动力学; 人类行为; 非泊松统计特性; 标度律

中图分类号 N94

文献标志码 A

doi:10.3969/j.issn.1001-0548.2013.04.001

Statistical Mechanics on Temporal and Spatial Activities of Human

ZHOU Tao^{1,2}, HAN Xiao-pu^{2,3}, YAN Xiao-yong^{4,5}, YANG Zi-mo¹, ZHAO Zhi-dan¹, and WANG Bing-hong^{2,6}

- (1. Web Science Center, University of Electronic Science and Technology of China Chengdu 611731;
2. Department of Modern Physics, University of Science and Technology of China Hefei 230026;
3. Alibaba Business College, Hangzhou Normal University Hangzhou 310036;
4. Department of Transportation Engineering, Shijiazhuang Tiedao University Shijiazhuang 050043;
5. School of Systems Science, Beijing Normal University Haidian Beijing 100875;
6. Research Centre of Complex Systems Science, University of Shanghai for Science and Technology Yangpu Shanghai 200093)

Abstract Quantitative analysis on human behavior, especially mining and modeling temporal and spatial regularities, is a common focus of statistical physics and complexity sciences. The in-depth understanding of human behavior helps in explaining many complex socio-economic phenomena, and in finding applications in public opinion monitoring, disease control, transportation system design, calling center services, information recommendation, and so on. This review article summarizes the state-of-the-art progresses of human dynamics, including the empirical analysis and modeling on human temporal activities, the empirical analysis and modeling on human mobility patterns, and the applications of human dynamics. Furthermore, this article reviews the advanced and insufficient points of related works, and points out some unsolved challenges in both theoretical and practical aspects.

Key words human dynamics; human behavior; non-Poissonian statistics; scaling laws

霍金认为,复杂性科学是二十一世纪的科学。对复杂系统的深入研究,不仅产生了新的理论体系,如耗散结构理论、突变论、微循环论、协同学等,而且带来了从还原论到系统论这一研究理念和方法论上的革新。复杂系统的研究成果,极大地拓展了人们理解世界的广度和深度,并在若干问题中发挥了巨大的实用价值。复杂性科学也是最广泛最深入的一门交叉科学,其中物理学、数学、信息科学、

系统科学、管理学、生物学、化学、地球科学、经济学、金融学等等我们能够想象到的学科,几乎都涉足其中。

大量典型的复杂系统,都直接或间接和人发生关系。经济金融系统和社会系统是由人的意志和行驱动为驱动的;如计算机互联网这样的大规模技术网络是由人建造的;被很多学者认为最复杂的系统之一的神经系统,是人身体的一部分。从人的社会属性

收稿日期: 2013-05-15

基金项目: 国家自然科学基金(11222543, 11275186, 10975126, 70871082, 70971089, 90924011, 91024026); 安徽省高校自然科学研究重大项目基金(KJ2011ZD07); 中组部拔尖人才计划; 教育部新世纪优秀人才计划; 四川省百人计划; 四川省创新平台资助项目

作者简介: 周涛(1983-), 男, 教授, 主要从事网络科学与工程、信息挖掘、人类动力学和群集动力学等方向的研究。

出发,研究社会经济系统中通过人的相互作用和策略博弈涌现出来的复杂性;从人的生物属性出发,研究人体内的神经系统、代谢系统、基因调控系统等等。尽管这方面的研究已经是硕果累累,还是注意到前者把人看得太小,后者把人看得太大,对于与人相关的复杂系统的整体认识,尚缺失一个环节,就是人自身行为,特别是个体行为,在时间和空间上表现出来的复杂性。

对人类行为时空复杂性的分析,一直不乏零星报道,但未能引起学术界集中关注。转折点来自于2005年在《自然》上刊登的一篇开创性的文章^[1],清晰地揭示了人类行为在时间上对泊松过程的偏离,并提出了一个简单的基于任务优先级的排队论模型,从而引起了广泛的相关研究。一年以后,文献[2]发表了关于人类行为空间标度律的工作,揭示了人类长程旅行中表现出来的既不同于随机游走,又不同于Lévy飞行的独特性质。受这两篇开创性文章的影响,大量文章相继在《自然》、《科学》、《美国科学院院刊》等期刊发表,掀起了有关人类行为动力学研究的热潮^[3-9]。

追根溯源,对人类行为的思索古已有之。举例来说,现代心理学和经济学热衷的关于群体压力和从众行为的研究,早在蔡松年诗赋“槽床过竹春泉句,他日人云吾亦云”中便可觅得踪影,而近年传播学关于面对谣言和恐慌情绪时人类非理性行为和判断的报道研究,亦可视作三人成虎、曾子杀人这些典故的科学注脚。对人类行为进行科学而系统的研究,或始于华生的行为主义流派^[10]——如此算来,也不过百年历史。尽管这一百年里,对人类行为的理解一直是社会学、心理学和经济学共同关注的焦点,但人类自身的复杂性和多样性,对一切科学的尝试来说都是巨大的挑战。本文所介绍的“人类动力学”的交叉学科,主要是由统计物理学家发起并推动的,与社会心理学所关注的人的行为有显著区别。从研究对象上讲,社会心理学对人的行为的研究主要集中于某些非常态的特异行为,例如犯罪行为、成瘾行为、变态性行为等等,而人类动力学关注的是人日常的行为模式。从研究目标上讲,社会心理学希望挖掘特异行为背后具体而细微的心理学和社会学机制,而人类动力学则力图挖掘新的统计规律并建立相应的动力学模型。以个体间通信行为为例,社会心理学的研究希望回答的问题是人为何要和其他人通信,这中间既有心理学的原因——渴望交流,又有社会学的原因——信息的共享与资

源的整合;而人类动力学则是希望揭示个体通信行为中展现出来的特定的统计规律——如通信时间间隔和回复时间的幂律分布,并挖掘相应的动力学机制,如重要信件优先处理,容易回复的信件优先处理,有截止日期的信件优先处理,不一而足。从研究方法上讲,社会心理学的结论主要来自于有主观意图参与的实验和临床数据,通过分析得到定性或者简单统计形成的半定量结果,而人类动力学则是通过无干预的客观数据,从一个外在观察者的角度给出量化的分析结果。从技术路线上讲,社会心理学主要是通过“观察-推断-实验设计-实验验证-修正推断”循环开展,其中实验设计中往往不可避免地掺杂很多研究人员的主观判断,而人类动力学则是遵从“观察-数据获取与分析-统计规律挖掘-建模再现数据规律”的循环。后者几乎不对数据的产生过程、获取过程和分析结果进行人工的干预。

作为一门新兴的交叉学科,人类动力学的提出不过寥寥数年,但其重大的理论和应用意义已初露峥嵘。在理论层面,人们已经观察到很多由人的行为驱动的复杂系统的各类现象,其中人类动力学的研究提供了新视角和新解释。举例而言,对于部分计算机邮件病毒以极低感染比例长期存在这个问题,文献[11]提出了基于人处理电子邮件时间上的非泊松特性的解释,该解释完全独立于Pastor-Satorras和文献[12]基于接触网络度分布异质性的解释。与此同时,人类动力学的研究还提出了很多新的理论问题,如对于到达时间间隔分布二阶矩发散的服务请求,以及有优先级限制的服务请求,如何用排队论的方法进行处理?又如具有向心性和局域化倾向的出行行为能否纳入某种扩散动力学框架内加以解释?人们有理由相信,人类动力学的研究会推动随机过程和排队论经典理论的丰富和发展^[13]。人类行为时空统计特性的挖掘,还具有相当的应用价值。对人出行,特别是长途旅行规律的理解,有助于理解、预测和控制全球性传染病的流行^[14-15];对人出行和网络活动中表现出来的阵发性和记忆性的分析,可以应用于异常行为检测,如疑似恐怖分子非正常的频繁跨越国境的旅行^[9]和处于非正当目的的网络舆情操控行为的检测^[16]等;对于人类群体出行行为时空图的分析可以为交通路网设计和道路流量规划,包括公共服务设施和商业设施的选点提供借鉴^[17];对人网上行为的分析可以为更好地开展互联网信息服务提供借鉴^[18]等。

为了帮助我国学者尽快了解这个新兴方向的前

沿进展, 并作出具有原创意义的贡献, 本文总结了近五年来人类动力学的相关研究进展, 内容包括人类行为时间特性的实证分析和建模, 人类行为空间特性的实证分析和建模以及人类行为统计分析的应用研究。本文的特色在于对很多有争议的问题, 例如人类行为时间特性排队论模型中的普适类问题, 时间间隔胖尾特性的泊松解释、空间出行模式中个体标度律的存在性问题等, 给出了深入的讨论。并且, 对可能成为未来研究焦点和突破口的挑战性课题, 进行了展望和初步分析。希望本文能对我国学者在相关方向开展前沿研究提供助益。

1 人类行为时间特性的实证分析

人类行为时间特性所分析的是人多次从事某特定事件表现出来的时间上的统计规律。如一个用户经常上网, 那么可能会出现几个问题: 这个用户两次上网的时间间隔的分布有没有什么特征? 这个用户访问他最喜欢网站的时间间隔的分布是不是也有类似的特征? 这个用户平均要隔多久才会访问一个新网站, 这个间隔时间是越来越长还是越来越短? 这样的问题还可以问很多。对这些问题的回答, 就组成了本节的主要内容。

本文为大家呈现近几年来在理解人类行为时间特性方面快速而激动人心的进展。其间, 我们不仅可以亲经典泊松范式的塌陷, 见证幂律分布如何崛起并迅速成为主流, 还可以听到夹杂其间对数正态分布和双峰分布非主流的呐喊。正餐之外, 诸君还可以欣赏到人类动力学普适类假说的惊艳与脆弱, 阵发和记忆分析的威风与罩门, 周期和波动效应一经提出就不绝于耳的争议……

1.1 从泊松到幂律

泊松分布最初是由泊松在他审判管理案件的工作中引入概率概念时提出的^[19], 现在被广泛应用于人类活动的量化模型中, 用于描述单位时间内或指定范围内特定事件出现次数的统计规律性, 如模拟交通流量的模式或事故发生频率^[20]、商业话务中心的配置^[21]、库存管理^[22], 以及估计移动通信中占线的电话数量^[23]。特别当对人类特定活动的收集能力有限时, 泊松过程作为经典的刻画人类活动模式的手段, 经常被不加仔细评估地应用于实际问题中。

最初被普遍使用的人类活动的随机模型假设:

1) 在不相重叠的时间区间内行为发生的次数是相互独立的, 即无记忆性和无后效性。2) 对充分小的 Δt , 在时间区间 $[t, t + \Delta t)$ 内有1个行为发生的概率

与 t 无关, 而与区间长度 Δt 成正比, 即:

$$P_1(t, t + \Delta t) = \lambda \Delta t + o(\Delta t) \quad (1)$$

式中, 当 $\Delta t \rightarrow 0$ 时, $o(\Delta t)$ 是关于 Δt 的高阶无穷小, 可忽略, $\lambda > 0$ 是常数, 它表示单位时间有一个行为发生的概率 λ 为概率强度。3) 对充分小的 Δt , 在时间区间 $[t, t + \Delta t)$ 内多于一个事件发生的概率极小, 以至于可以忽略, 即:

$$\sum_{n=2}^{\infty} P_n(t, t + \Delta t) = o(\Delta t) \quad (2)$$

式中, $P_k(t, t + \Delta t)$ 表示在时间区间 $(t, t + \Delta t)$ 内发生 k 个事件的概率。

在上述条件下, 可知事件到达的随机动力学过程可以用泊松过程很好地描述^[24]。这种情况下两个随机事件到达的间隔时间 τ 的概率密度函数具有指数形式

$$p(\tau) = \lambda e^{-\lambda \tau} \quad (3)$$

实际上, 泊松过程可以看作具有负指数间隔的计数过程^[25], 在实践中意味着人类活动模式是随机和平稳的, 在时间上显示了一些规律, 如相邻事件平均的时间间隔大体相差不大, 很长的时间间隔指数般地罕见。由泊松分布中标准差 $\sigma = \sqrt{\langle \tau^2 \rangle - \langle \tau \rangle^2} = \langle \tau \rangle$ 可知。图1a展示了由泊松过程产生的事件发生序列, 可以看出, 该序列中相继事件发生的时间间隔大体上是均匀的。

最近几年, 通过对记录了人们活动历史的海量数据库的分析, 从通讯、工作到娱乐, 越来越多的证据显示人们的很多行为的时间统计特性无法用泊松过程刻画。特别地, 这些行为所对应的间隔时间分布具有明显偏离指数分布的肥胖的尾部, 可以用幂函数更好拟合, 即:

$$p(\tau) \propto \tau^{-\alpha} \quad (4)$$

在下一节介绍具体的实证系统之前, 首先直观认识一下泊松分布和具有典型胖尾特征的幂律分布的区别, 如图1所示^[1]。从形态上讲, 泊松过程所对应的分布尾部按指数下降, 致使每两个行为之间的时间间隔长度几乎是规则的, 长时间的等待是不存在的。而幂律分布的尾部落降缓慢, 允许在非常长的时间内无事发生, 在这些长长的空白与空白之间是犹如爆发一般密集的活动。通过对真实系统的统计分析, 发现很多人类行为模式都具有胖尾的特性。

图1a中, 泊松过程给出的事件发生序列, 其中任意时刻有一个事件发生的概率是 q 。横坐标表示时间, 每一条竖线代表在那个时刻有一个事件发生。注意到任意两个连续事件之间的间隔时间长度都是

差不多的, 遥遥无期的等待是罕见的。图1b中, 1 000个连续事件的间隔时间 τ , 其中线段在垂直方向的长度就对应于图1a中的两个连续竖线的空白长度。同样, 观察不到很长的间隔时间。图1c中, 在给定的时间间隔里面正好发生 n 个事件的概率是 $P(n; q) = e^{-qt} (qt)^n / n!$, 由此两个连续事件的时间间隔分布是 $p(\tau) = qe^{-q\tau}$ 。本图给出了图1a、图1b两图中数据点在半对数坐标下的分布, 其曲线可以用一条负斜率的直线较好地拟合。图1d表现了具有胖尾分布的连

续事件。图1e中, 1 000个连续事件的间隔时间 τ , 其中间隔时间的平均值和图1a~1c中泊松过程的平均值相同。注意那些像长钉一样扎在水平轴上的竖线, 它们对应于那些很长的间隔时间。由于图1b和1e的纵坐标是一致的, 所以可以通过比较看出泊松分布的规则性和胖尾分布的阵发性直观上的不同。图1f中, 具有胖尾的时间间隔分布 $p(\tau) \approx \tau^{-2}$, 该分布对应于图1d和图1e中的数据点。分布曲线在双对数坐标下显示为斜率为-2的一条直线。

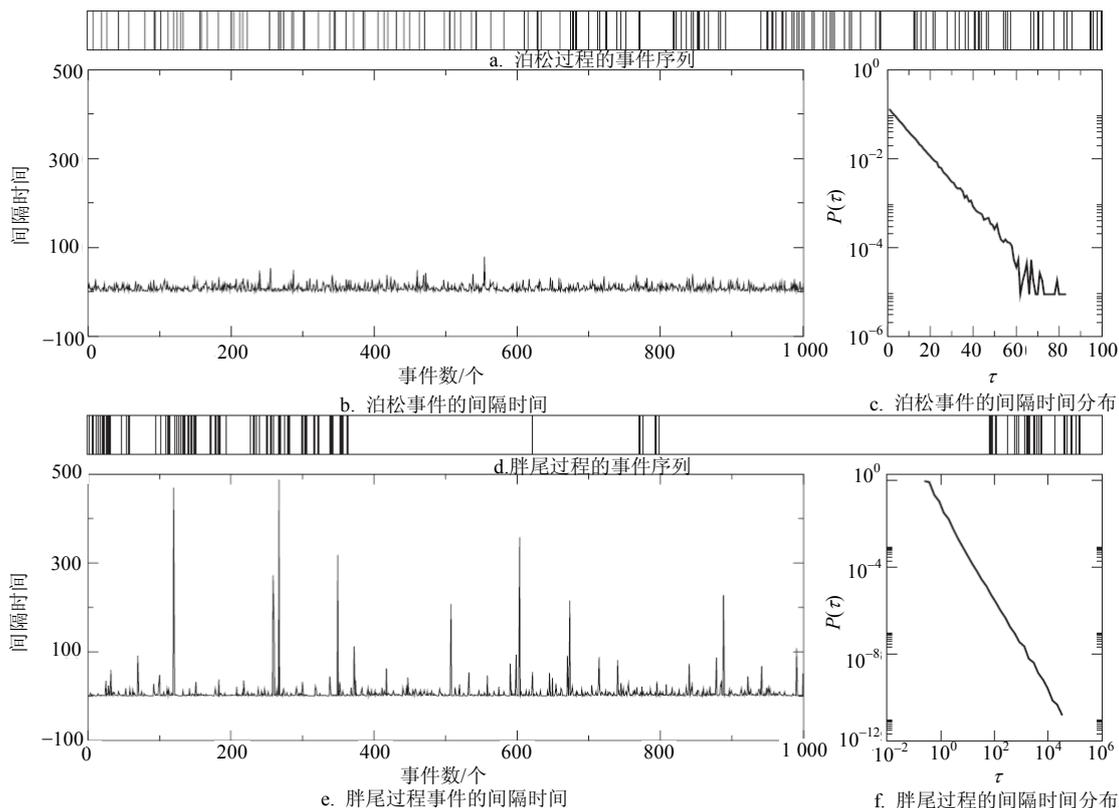


图1 泊松过程所预测的人类活动模式与真实的人类动力学中观察到的胖尾分布之间的区别^[1]

1.2 实证结果

表1总结了最近几年研究人员通过分析人类行为时间规律的真实数据所得到的关于行为时间间隔和任务等待时间的分布规律, 其中后者是指一项任务从收到到执行完毕所需要花费的时间, 如电子邮件或水陆信件从收到到发出回信的时间。表中“个体层面”是指对单个个体的行为进行统计分析的结果, “群体层面”或“全体”是指把每个个体相应的间隔时间数据放到一起进行统计。从表1可以看出, 林林总总的人类行为在时间上具有惊人相似的统计规律, 即间隔时间和等待时间分布在绝大多数情况下具有胖尾的特性, 很多可以用幂函数较好地刻画。除了针对间隔时间和等待时间分布规律的分析、经济学中的Gibrat定律^[81]、波动分析中的Taylor定律^[82]

和时间序列分析中的分形特征提取方法^[83]都可以用来刻画人类行为的时间特性, 限于篇幅, 本文不做详述。

尽管这些实证研究所采用的分析手段非常简单, 并且时常针对相近甚至相同的数据集, 在发现很多共同的特征之外, 也包含了大量的争论——这些争论不仅仅在于大家可以从表1看到的幂指数的不同, 还包括对于人类行为时间特性是否能够用幂律分布刻画这一问题从根本上的质疑, 以及与此相关的对这些统计特性内在驱动因素截然不同的解释。在本节的后面3个小节中, 本文会针对一些特别重要的专题——如阵发与记忆, 周期与波动——展开详细的讨论, 并尽可能全面地向读者介绍目前存在的未解决的问题和争端。另有一部分有争议的内

容, 将在第3节讲述。

表1 人类行为时间统计特性实证研究结果总结。幂指数若无特殊说明指间隔时间分布

分析对象	数据来源	幂指数
电子邮件	大学里以3个月为周期的3 188位用户收发共129 135封信件 ^[1,26-27] ;	个体回复时间和间隔时间分布幂指数均为1; 不符合幂律, 建议用级联非齐次泊松过程刻画;
	数据同上 ^[28-29] ;	不符合幂律, 建议用对数正态分布刻画;
电子邮件	数据同上 ^[30] ;	邮件从收到到回复的时间服从一个很广的分布——从立刻回复到超
	一个大型咨询公司在2006年7月~2007年1月间, 其1 052个管理人员之间发送的1 493 441封电子邮件 ^[31] 。	过1 000 h不等, 且接受者倾向于在较短时间内回复和自己存在社交关系的友人。
水陆信件	爱因斯坦、达尔文和弗洛伊德一生收发的30 801封、14 121封和5 858封信 ^[32-33] ;	回复时间分布幂指数为3/2;
	钱学森通信记录, 来自涂元季主编的《钱学森书信》, 跨越时间为1955年~2000年。包括了记录有写作时间的回复时间和间隔时间分	3 335封信件, 其中1 819封信注明了此布幂指数均为2.1; 信所回复的信件收到时间, 单位是天 ^[34-35] ;
手机短信通讯	从十六世纪中叶至二十世纪中叶, 16位著名作家、演员、政治家和科学家联非齐次泊松过程的通信记录 ^[36] 。	不符合幂律, 建议用级联非齐次泊松过程刻画;
	志愿者的短消息通信记录, 统计时间长度从3~6个月不等 ^[37] ;	幂指数在1.2~1.7之间, 针对个体; 发送时间间隔分布在30~20 000 s之间符合幂律, 指数为1.188; 回复时间间隔分布在60~20 000 s之间符合幂律, 指数为1.148; 不符合幂律或截尾幂律, 建议用双模分布刻画, 前段为幂律, 后段为指数。
手机短信通讯	2006年新年期间6 326 713位用户37 577 781条短信记录 ^[38] ;	时间间隔分布服从带指数截断的幂律, 幂指数为0.9, 截断值为48天。
	来自3家公司147 672位用户1 589 869条短信记录 ^[39] 。	间隔时间分布, 幂指数在1附近分布, 针对个体; 个体借阅时间分布近似为幂律, 指数集中在1.5左右; 群体借阅时间分布形式复杂, 有明显的截止时间效应; 个体借还书间隔时间分布指数在0.6~1.6之间, 与借阅量正相关。
图书馆借阅	圣母大学三年间2 247名教工从图书馆借阅的48 409条记录 ^[33] ;	
图书馆借阅	中国两所不同大学图书馆数据库, 包括13 866名学生和教师共计772 504条借书记录和647 048条还书记录 ^[41] 。	
	中欧银行在1999年6月~2003年5月间由一个股票经纪人发起的所有54 374次交易 ^[33] ;	幂指数为1.3, 但是有一个指数的截断;
金融活动	在2002年的3月、6月和10月GSK和VOD证券交易中将近800 000次订购和540 000次交易 ^[42] ;	胖尾, 但不符合幂律分布;
	2006年12月9日台北市长选举前以5位候选人为对象设计的在线期货交易实验, 实验共持续了30天, 有400多名志愿者参加 ^[43] 。	全体层面间隔时间分布服从指数截断的幂函数律, 幂指数为1.3。

(续表)

分析对象	数据来源	幂指数
网页访问	匈牙利新闻娱乐网02年11月8号~02年12月8号期间每个访问者每次下载文件的时间纪录 ^[33] ;	不同个体略有不同, 幂指数以1.1为中心;
	数据同上 ^[44] ;	全体用户群集数据对应的分布为1.2; 个体均值为1.14;
网页访问	美国乔治亚州亚特兰大市埃默里大学从2005年4月1日~2006年1月17日其门户(www.emory.edu)及所属网页访问记录 ^[45] ;	同一用户访问同一网页间隔时间分布指数为1.0, 统一用户访问不同网页间隔时间分布为1.25;
	上海理工大学15天内访问Internet的记录 ^[46] ;	个体分布在2.1~3之间, 全体用户群集数据对应2.82;
网络通信	2004年12月23日~2008年10月8日, 7 565 401位用户登录维基百科的17 531 208次记录 ^[47] 。	群体层面间隔时间分布指数约为1.2。
	1998年~2008年, Ebay上748 282位用户所发出的149 087 003条反馈信息 ^[47] ;	群体层面间隔时间分布指数约为1.9;
网络搜索	MSN即时通信2006年6月, 2.4亿用户间发生的约300亿次交谈 ^[48] ;	群体层面交谈时间间隔分布指数为1.53;
	五位志愿者在QQ上4个半月至1年不等时间内发信息的记录 ^[49] 。	个体层面幂指数在2.0~2.5之间。
网络搜索	2006年3月1日~2006年5月31日, 美国在线657 426位用户36 389 566次搜索记录 ^[47] ;	群体层面间隔时间分布指数约为1.9;
	2007年4月1日~2007年4月8日手机终端上498 872个搜索需求 ^[50] 。	全体层面上24 h内的时间间隔分布接近幂律, 24 h以外按指数衰减。
论坛博客活动	2005年8月~2005年9月间, 45 000位博主发表的220万篇博文 ^[51] ;	群体层面上博文发表的时间间隔分布指数为2.7;
	南京大学小百合BBS截止至2009年9月1日20 379名用户发布的1 627 697个帖子 ^[52] ;	群体数据幂指数为1.98, 作者实际上采用的Mandelbrot分布 ^[53] 进行拟合;
论坛博客活动	数据同上 ^[54] ;	同一用户连续两次发表帖子的间隔时间分布都近似服从幂函数律, 但在分钟、小时和天的不同尺度上指数不同;
	科学网博客4名博主发布的博文, 数量分别为588、191、536和772篇 ^[55] ;	时间间隔服从伽玛分布 ^[56] , 不严格地讲, 该分布可以看作带指数截断的幂函数和Mandelbrot分布 ^[53] 的结合;
社交网络活动	用户在博客和微博上的行为 ^[57] ;	都服从幂律, 指数分别是1.3和2.0;
	随机抽取的2006年注册用户的30%在Boards.ie下面的9个论坛上的行为 ^[58] ;	群体层面幂指数约为1.7;
社交网络活动	用户在新浪微博上发微博行为 ^[59] 。	幂指数为1.4。
	2003年6月~2005年10月, 在韩国最大的在线社交网络Cyworld上17 788 870位用户在留言本上留言的记录 ^[60] 。	群体层面间隔时间分布出现三段幂律: 在间隔小于36 min时, 指数为1.696; 间隔在36 min到1天, 指数为0.910; 在间隔大于1天时, 指数为2.276。
任务执行	2009年1月~2010年4月, 亚马逊推出的名为土耳其机器人的再现任务解决系统, 共包含9 436个任务(总奖金的等待时间分布符合幂律, 幂指数为1.48; 任务 ^[61] ;	
任务执行	超过10 000个文件中自动记录的用户两次中断/暂停之间的在使用计算机软件时出现的中断/暂停时间间隔分布符合幂律, 指数为1;	

(续表)

分析对象	数据来源	幂指数
任务执行	6位用户在服务器上产生的7个历史文件所记录的Linux命令日志文件, 每个文件包含命令数在12 000~93 000之间 ^[63] ; 一个大规模协作实验, 要求参赛队在尽可能少的时间里在美洲大陆十个指定地点放上红色气象球, 最终麻省理工代表队借助社会网络的力量在8小时52分完成该任务, 获得冠军。文献[64]分析了该队数千名参加者接到任务到回复的时间间隔(这个间隔不能直接得到, 因此用一个参与者和他所发展的下线注册时间差代替);	两次提交命令的间隔时间分布符合幂律, 指数在1.47~1.74之间;
在线服务	在线共享电影网www.netflix.com电影观看数据, 包含了17 770部电影, 139位用户和近亿条记录 ^[65] ; 上海交通大学在线音乐共享网站上超过4万师生在105天内的登陆记录 ^[66] ; Delicious网站上从2004年4月1日~2007年11月1日220 867个用户共计54 204 641个书签收藏记录 ^[67] ; 2008年12月2日~2010年1月12日274 148位用户针对京东商城上28 620件商品的购买数据 ^[68] ;	群体数据指数为2.08, 个体分布形式不定, 但有胖尾; 群体数据显示用户两次听音乐的时间间隔分布具有胖尾特征, 但分布后部分波动剧烈, 难以用幂函数拟合; 群体层面在一天以内和一天以外, 间隔时间服从不同的分布, 幂指数分别为1.07和2.41;
战争叛乱暴力	54 679个有报导的暴力事件 ^[69] , 有趣中国历史上有记载的战争 ^[70] ; 从MIPT网站上获得的在伊朗发生的8 627次恐怖袭击数据和在阿富汗发生的772次恐怖袭击 ^[71] ;	每天发生的暴力事件数, 本文在《自然》期刊第911页发表, 应该是来自编辑的冷幽默; 性; 间隔时间服从广延指数分布; 伊朗和阿富汗幂指数分别为2.61和2.41。
写作	2005年2月11日~2005年10月2日利用谷歌RSS阅读器得到的68 022篇博文的内容 ^[72] ; 中文维基百科在2002年10月26日~2009年6月7日81 823名用户9 641 842次编辑记录 ^[54] ;	总出现次数小于214次的单词相续出现的间隔时间(单位: 天, 注意此处间隔不是按照词语数衡量, 而是真实时间)分布指数约为1.5; 同一用户连续两次编辑的间隔时间分布都近似服从幂函数律, 但在分钟、小时和天的不同尺度上指数不同;
物理接触	大文学家苏轼和活跃科学家Stanley Newman, Barabasi发表文学作品和学术论文的时间 ^[73] ; 2005年5月在迈阿密召开的IEEE两位志愿者之间产生接触 ^[74] ; 数据同上 ^[75] ; 100位麻省理工学生通过内置蓝牙设备的手机记录的长达9个月的空间位置信息 ^[75-76] ; 2008年10月13~17日, 50余名与会人员在明为“面对疾病传播的挑战”的研讨班上, 佩戴RFID胸牌, 可以记录近距离两两交互 ^[77-79] ;	发表时间间隔服从幂律分布, 幂指数相差很大。 具有很强截断的幂函数, 幂指数约为1.4; 幂指数约为1.6; 两位志愿者之间产生接触的间隔时间分布满足具有很强截断的幂函数, 幂指数约为1.49; 一个志愿者与其他志愿者两次物理接触之间的间隔时间和一对有多次接触的志愿者连续两次接触之间的时间间隔都符合相近的幂律分布, 幂指数未给出;
	日本一家公司163位志愿者73天中通过佩戴传感器记录的51 879次面对面的交谈行为 ^[80] ;	幂指数为2.52。

除了本文讲述的对于人类行为时间统计特性的直接分析以外, 还有一些工作针对的是由人群作用所驱动的系统的时间统计特性, 如一条微博被评论或转发的时间间隔, 一段视频被点播的时间间隔等。这些分析可以提供人类行为特征的间接证据, 作为人类行为时间统计分析的有益补充, 以及相关建模研究的评判标准。文献[84]分析了500万YouTube视频点播的记录, 发现这些视频上线伊始, 会有一段很短的活跃期, 之后吸引力呈幂律衰减。文献[85]发现维基百科上一篇文章被多个用户编辑的时间间隔服从多标度分布, 其中前两段可以用幂函数刻画, 并且, 编辑之间会出现相互删除的行为——这种典型的“编辑战争”行为也被文献[86]观察到。文献[87]发现一个新闻从上线到收到第一个回复的间隔时间分布服从幂函数律。文献[88]的分析显示, 不管是Wikipedia还是各类网页, 其被访问的时间间隔分布都呈现明显的胖尾特征。文献[89]发现在一些群集计算服务器上任务到达的时间间隔分布非常广。文献[90-91]分析了Condensed Matter Physics和Physica A两个期刊的投稿和发表记录, 发现(接收发表的)稿件的等待时间分布是一个单峰但尾部较泊松分布宽广的形式。文献[92]对天涯社区的研究发现, 在同一个主题下两个相继评论出现的时间间隔分布可以用幂函数刻画。

1.3 阵发与记忆

用幂指数刻画时间间隔分布, 固然是一种常见的方法, 但是这里面隐含了一个假设, 就是人类行为的时间间隔分布符合幂函数律, 而这个假设并没有得到学术界广泛的接受^[28-30]。本节介绍两种定量刻画时间间隔分布的方法, 这些方法并不需要先假设该分布满足某种特殊的函数形式。

从图1中可以直观看出, 当人类行为的时间特性可以用泊松过程描述时, 事件发生频率比较均匀; 反过来, 当时间间隔分布服从幂函数律的时候, 有时候很多事件会在较短时间内密集发生, 然后又会出现一个很长的空档期, 把后面这种情况叫做阵发性。从时间间隔分布的角度来看, 阵发性很强的系统大部分时间间隔都小于平均时间间隔, 但是时不时会出现非常大的时间间隔——这种情况往往意味着该分布的标准差很大。文献[93]提出了如下的指标来刻画阵发性:

$$B = \frac{\sigma_{\tau} - m_{\tau}}{\sigma_{\tau} + m_{\tau}} \quad (5)$$

式中, σ_{τ} 和 m_{τ} 分别表示分布 $p(\tau)$ 的标准差和平均

值, B 的取值显然在 $-1 \sim 1$ 之间。对于泊松分布, 均值和标准差相等, 因此阵发性恰为0, 可以看作Goh和Barabási设定的平衡点。对于严重的胖尾分布, 标准差远大于均值, B 靠近1; 对于周期性规则信号, 时间间隔分布实际上是一个 δ 函数, 标准差为0, B 值为 -1 。

除了阵发性以外, 事件发生的时间序列还可以用记忆性刻画。其实, 自然界中极端气候时间和地震等灾害发生的时间序列就具有相当强的记忆性, 并且这种记忆性被认为对观察到的很多标度律有解释作用^[94-95], 在人造的大技术系统, 如互联网流量时间序列中, 也同时观察到了记忆性和标度性^[96]。类似的方法也可以被用来刻画人类行为发生的时间序列。简单地说, 一个行为发生的时间序列被认为具有记忆性, 如果长的时间间隔后容易跟着一个也较长的时间间隔, 而短的时间间隔后面容易跟着一个也较短的时间间隔。如果把所有的时间间隔按发生先后排成一个序列(注意, 不是行为发生时间的序列, 而是两次相继行为时间间隔的序列), 这个序列一共有 n_t 个元素(意味着一共有 $n_t + 1$ 个行为发生), 指定前 $n_t - 1$ 个元素构成序列1, 后 $n_t - 1$ 个元素构成序列2, 如式(6)所示, 这两个序列的Pearson关联就可以用来衡量该序列的记忆性。

$$M = \frac{1}{n_t - 1} \sum_{i=1}^{n_t-1} \frac{(\tau_i - m_1)(\tau_{i+1} - m_2)}{\sigma_1 \sigma_2} \quad (6)$$

式中, m_1 和 m_2 分别是序列1和序列2的均值; σ_1 和 σ_2 分别是序列1和序列2的标准差。显然, M 的取值范围在 $-1 \sim 1$ 之间: $M > 0$ 意味着记忆效应, $M < 0$ 意味着反记忆效应。

图2给出了对于序列阵发性和记忆性直观的示意。从中可以看到, 尽管阵发性和记忆性之间存在关联(譬如阵发性 B 值为 -1 的序列记忆性 M 值肯定为0), 但它们的确从两个不同维度刻画了事件发生时间序列的特征——尽管这两个维度之间存在非平凡的关联。图2a是通过一个泊松过程生成的时间序列, 阵发性和记忆性都接近于0, 分别为 $B = -0.05$ 和 $M = 0.02$ 。图2b和图2c是典型的具有强阵发性和强规则性的序列, 分别由一个指数为 -1 的幂函数律和一个高斯间隔时间分布产生, 其记忆性均接近0, 阵发性 B 值分别为0.44和 -0.81 。保持时间间隔取值不变, 通过交换顺序, 可以增强或降低记忆性。图2d和图2e分别对应于取值和图2a相同, 但是记忆性很强和反记忆性很强的两个序列, 对应的 M 值分别为0.90和 -0.74 。注意, 图2d看起来似乎具有很强的阵发性,

但实际上没有。

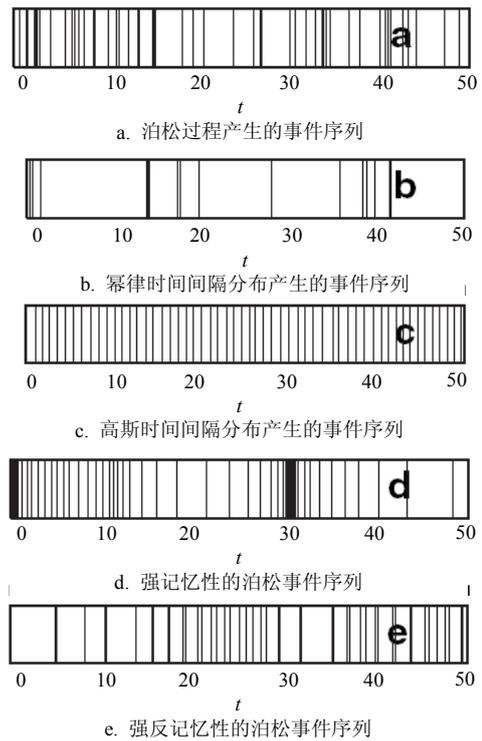


图2 事件发生的时间序列示意图^[93]

文献[93]分析了电子邮件发送、图书馆借阅、大学里打印任务提交记录、银行里的电话记录和手机通讯记录5组数据, 认为人类行为在发生时间上具有“强阵发弱记忆”的特性。图3的粉色区域给出了这5组数据在 B - M 图上的位置。类似的现象在论坛帖子的张贴和回复行为中也有报道^[97]。

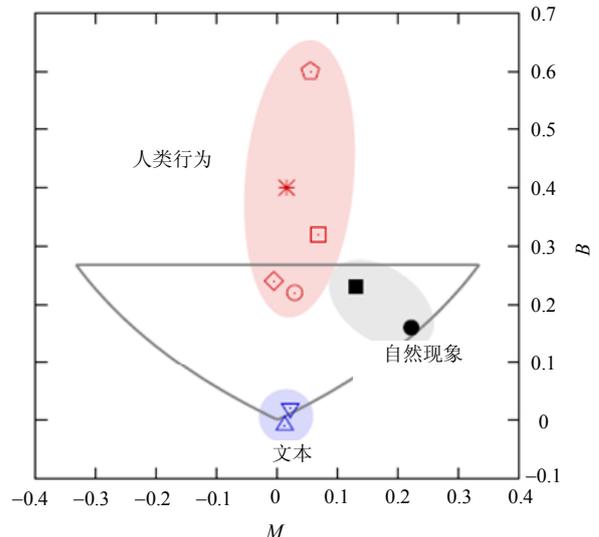


图3 若干真实系统中事件发生时间序列在 B - M 图上的位置^[93]

注意到式(5)和式(6)的定义中直接使用了标准差, 而真实人类行为时间间隔分布的标准差是发散

的。尽管在有限系统中永远不会出现真正意义上的发散,但是这种数学定义上的缺陷让人担忧。如对于具有很大的 B 值的系统,对应分布的二阶矩很可能是发散的,所以式(6)的分母非常大,这可能 B 值很大而且数据量非常大的系统(发散情况严重) M 值一般都较小的原因。文献[98]所提出并被广为采用的网络同配性指标也具有类似的问题^[99]。又如如图4所示,可见其中每一个蓝色圆点代表一个用户,红色点是所有用户的平均值: $B=0.263$, $M=-0.065$ 。得到系统平均的 B 值和 M 值不一定有价值,因为个体在 B - M 图上可能分布非常广阔^[38]。

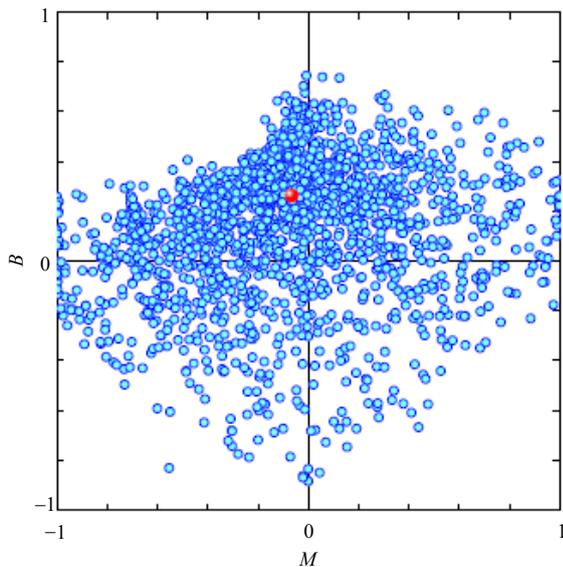


图4 6326713位手机用户发送短信的时间间隔的阵发性和记忆性图^[38]

1.4 周期与波动

如果定义个体的活跃性为单位时间内一个个体发出特定行为的频数,直观地讲,人类的日常行为应该有明显的波动性和周期性。举个例子,作为一个市场销售主管,上班时间你可能每小时平均接到2个电话,打出去11个电话,可以认为你总的活跃性是13次通话每小时。然而,你希望这样的事情发生在凌晨3点吗?如果把电话通讯的活跃性画在时间轴上,会发现平均而言,早上10点的活跃性可能是凌晨3点活跃性的数百倍。这就是活跃性随时间的波动性。这种波动性还具有明显的周期性——日复一日,白天永远不懂夜的黑。图5给出了同一个时区“美国在线”356610名用户活跃次数,可以明显看到巨大的波动和以天为单位的周期^[100]。本图给出了10天的数据,分辨率为min,其中虚线划分的每个区域是一天的数据(1440 min)。

活跃性的周期和波动在很多人类行为中都可以

观察到,包括用户访问互联网网页^[45]、用户编辑维基百科^[101]、玩家登陆在线游戏^[102]、研究人员向Linux服务器提交计算任务^[63]、用户通过手机进行搜索^[50]……以上所举只是部分典型的例子,类似的现象数不胜数。这种周期性有时候还直接表现在时间间隔的分布上。如有的家庭总是安排在周日礼拜或者带孩子出去野炊,那么这个家庭成员礼拜或者野炊的时间间隔主要会是7天、14天、21天……

图6给出了在线电影共享网站Netflix上面447139位用户在线观看电影的时间间隔分布,其中,时间分辨率为天,和刚才的讨论一致,在整数周的位置,分布密度会明显高于邻近点^[65]。

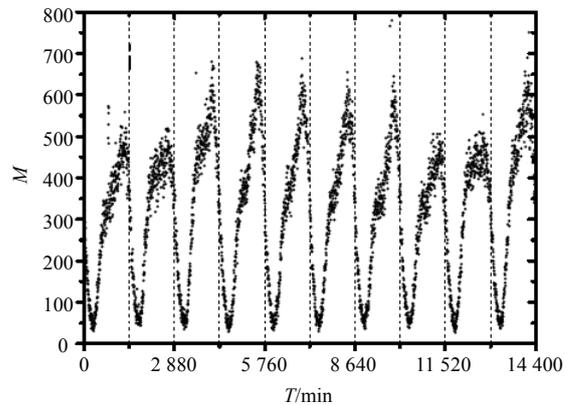


图5 “美国在线”(AOL)用户活跃性随时间的变化规律^[100]

周期和波动分析中最重要的一个问题就是:波动和周期是否是幂律时间间隔分布产生的根源?这个问题第3节会展开更详细的讨论,所以在这里只是简单提到。2006年,文献[103]从数学上证明了把一群具有不同一阶矩(服从一个均匀分布)的泊松个体聚集在一起,可以在群体层面上产生幂律分布(具有指数截断)。文献[28]认为如果把一个人一天的活动按照活跃性相近的办法拆分成若干段,则每段都可以用一个泊松过程拟合,因此与Hidalgo类似的方法可以用来解释幂律的涌现——这个时候个体就是一小段时间,全体就是整个时间轴,每个小段都可以看成是一个泊松分布。这个方法多多少少有一些无赖手段,所给出的解释也是平凡而从感情上难以被接受的。

最近文献[100]提出了一种方法,用个体相继两个行为之间,同一时区其他个体发出的行为总数来度量时间间隔长度。如果某用户在其天早上11点20发了一条短信,下午2点51分又发了一条短信,以前的时间间隔是绝对时间,也就是211 min;该方法是考虑除了这个用户之外的其他所有同一时区用户在11点20分~14点51分之间发出的所有短信数目,把

这个数目作为新的时间间隔的定义。在这个定义下, 群体活跃性很低的时段中的一个小时长度要远远短于群体活跃性很高的时段中的一个小时。如果幂律仅仅源于周期和波动, 可以证明在这种相对时间钟表系统中, 幂函数律会消失, 取而代之的是指数尾部^[100]。事实上, 真实系统中, 看起来像幂函数形式的时间间隔分布依然存在, 因此幂函数律的涌现不能完全用周期和波动来解释。

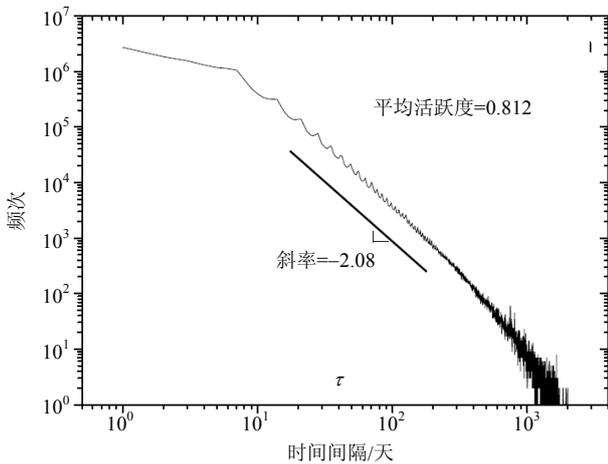


图6 Netflix用户在线观看电影的时间间隔分布^[65]

1.5 小结与讨论

本节简单概括了最近几年人类动力学研究在行为时间特性分析方面的进展。实证分析显示, 尽管不同系统的人类行为时间间隔分布形式有所不同, 但是都有较泊松过程所预测的分布更胖的尾部。人类的行为还具有“强阵发弱记忆”的特性, 并且表现出非常明显的波动性和周期性。尽管本文报道的人类动力学都着眼于对间隔时间和活跃性的简单统计分析, 但是, 其方法论——从人类活动的历史记录中挖掘统计规律, 并不仅仅局限于此。在后面的章节中, 读者可以领略到其间的绚丽风光! 下面, 简单介绍若干尚未解决或存在争议的研究问题, 其中有些争议, 如“阵发-记忆”图的适用性、周期和波动是否是幂律间隔时间分布的起源等, 前文已经有了介绍, 此处不再赘述。

首当其冲的问题是应当用什么样的函数形式来刻画间隔时间或者等待时间分布。尽管绝大多数文献中对人类行为的时间间隔都采用幂函数进行拟合, 但是这方面的争议一直不断。文献[30]认为对数正态分布能够更好地刻画时间间隔; 文献[39]提出短消息通讯中的时间间隔分布是双段的, 前段是幂律, 后段是指数; 文献[54]也发现了双段的时间间隔分布, 不过两段都是幂律, 交叉点正好在一天; 文献

[85]还报道了更复杂的三段分布。幂函数到底是不是最佳的选择, 是否还有更好的候选者, 分段分布中蕴涵了什么机制……尽管这些都是非常基本的问题, 但是到现在也没有完全让学界信服的答案。

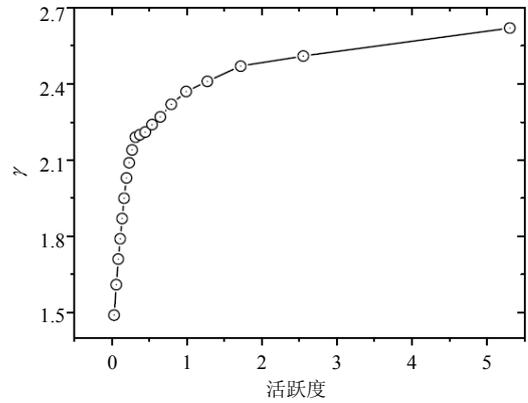


图7 将Netflix上447 139位用户按照活跃度大小分为20个组分别进行时间间隔分布统计分析^[65]

通过对Barabási模型的解析研究, 文献[33]发现该模型只能产生或为1或为1.5的两个离散的幂指数(详情请见下一节)。基于此, 文献[33]猜测人类动力学行为属于一些具有离散指数的普适类。但是, 目前越来越多的实证研究表明, 这种猜测可能是不正确的。如将Netflix用户按照其每天观看电影平均数排序, 并将排序后的用户均分为20个组, 每个组的平均活跃程度各不相同, 不同的活跃程度的组对应于不同的幂指数, 且指数随活跃程度单调连续变化^[65]。这一结果说明, 至少在群体水平上, 人类动力学系统不属于离散的普适类。文献[37]关于短消息通讯的研究, 在个体水平上给出了类似的结论。文献[47]对不同活跃性对应的时间间隔分布进行重定标, 发现重定标后曲线塌缩到一条主曲线, 从而认为不同活跃性的个体活动模式还是相同的。最近, 针对大量在线人类行为研究显示仅有少量数据可以通过文献[104]的重定标方法塌缩。更有意思的是, 文献[36]认为电子邮件通讯和信件通讯其实可以统一到同一个普适类, 而在文献[33]的观点中这正好是两个不同普适类的代表! 这方面的争论与猜想仍在继续: 尽管统计物理学家关于普适类的猜想异乎寻常的优美, 却可能终归只是镜花水月的幻觉。

最后讨论活跃性的问题。对于同一类型的行为, 活跃性差别很大的两个人, 其行为的动机和相应可观察的统计特征都可能存在本质的不同。举个例子, 对于一个很活跃的吸烟者(譬如每天吸烟两包以上), 正常情况下很难想象他能够连续一天不吸烟; 反过来, 一个一天平均吸三四支烟的人, 可能很长一段

时间不会吸烟。类似地,网络游戏瘾君子日夜操劳,勤耕不辍,一般爱好者却可以张弛有度并懂得避让更重要的任务。如果一般而言,针对同一类型的行为,个体之间活跃性差别不大,这方面的讨论价值就不大,但是越来越多的证据显示个体之间的活跃性相差非常大^[66,104-107],所以本文认为,活跃性在个体间分布的异质性对于群体层面涌现出来的标度律的影响,应该仔细慎重地加以考虑。如果按照活跃性的大小把群体分成若干组,那么不同组所得到的时间间隔分布的幂指数不同,而且随着活跃性单调变化——图7给出了针对在线电影共享网站Netflix上面447 139位用户的实验结果^[65],每个小组时间间隔都符合幂函数律,其幂指数随该小组平均活跃度单调变化。这种单调关系在其他系统中也得到验证^[37,104]。有趣的是,最近文献^[67]报导了一个非单调的例子。Taylor标度律^[82]可以用来分析活跃性和阵发性之间的关系,这方面的结果将来也可以成为人类行为时间特性模型研究的标尺。

2 人类行为时间特性建模

通过大量的实证研究,人们发现,对于各种有着较强的支配自主性的人类行为,非泊松性的时间统计特性可能是它们的一个普遍性质。那么是什么机制导致了这种非泊松特性的产生,就成为更为深入理解人类行为特性的一个重要问题。在最近的几年里,人们对这个问题进行了全方位的探索,先后提出了数十种动力学或非动力学的模型。这些模型大多数有着深刻的现实背景,其基本假设往往植根于日常生活中影响人们行为的实际因素,从而有望真正揭示人类行为中非泊松特性的产生根源。从研究思路划分,这些典型模型大致可以分成3类:

1) 基于任务队列理论的模型, 2) 基于记忆、兴趣、节律或其他因素的模型, 3) 基于社会交互影响的模型。本节将简要回顾这段研究历程,对每种类别中的几种代表性模型的规则和主要结论进行介绍,并对该领域的未来研究方向进行讨论和展望。

2.1 经典任务队列模型

首先介绍的是当前人类动力学建模的一个主流理论:任务队列理论。这种理论建立在这样一种生活经验上:一个人在日常生活中往往需要面对很多任务,如发送电子邮件或者水陆信件、打电话、阅读书报、写文章等。一般来说在生活中,人们是按照一定的顺序一件一件来做这些工作的。在针对人类行为的建模中,可以把这样一些日常活动概括为

需要处理的任务。基于这种想法,文献^[1]提出了一种基于任务队列理论的模型。在这个模型中,每人被分配了一个可容纳 L 个任务的列表。任务列表的长度是模仿人类等待执行任务的记忆而设定的。在每一个时步,个体选择执行其中的一个任务。该任务完成后,将其从任务列表中去除,并加入一个新的任务。每个任务标记一个优先级参量 $x_i(i=1, 2, \dots, L)$,该参数由给定的分布函数 $\eta(x)$ 随机生成。这时个体对于这些任务有3种可能的处理方案:

1) 先进先出方案,是个体按照其获得任务的顺序执行任务。这种方案在很多服务过程中普遍存在^[21]。在这种情况下,一个任务的等待时间为排在它前面的所有任务的执行时间之和。如果执行每个任务的时间满足一个有界分布,则任务从接受到执行完成之间的等待时间是均匀分布的。

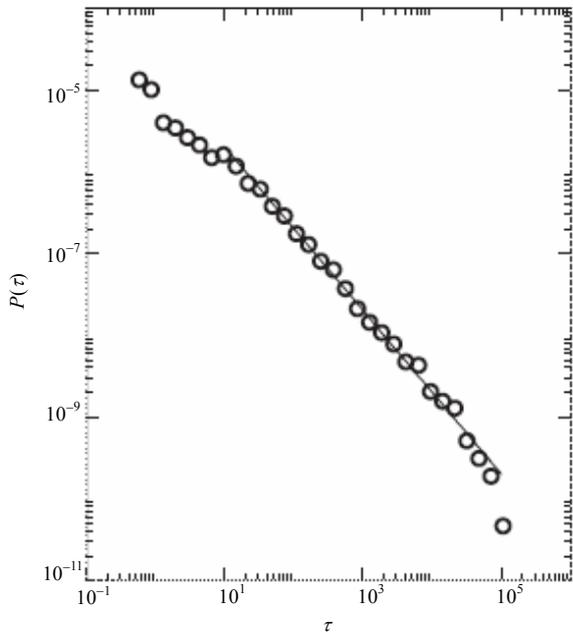
2) 不受任务优先级和接受时间约束的随机执行方案,每次从列表中随机抽取一个任务执行。任务的等待时间分布呈现出指数特性^[108]。

3) 按照任务优先级执行是最重要的一种方案。在这种方案中,优先级最高的任务最先执行,即使它是后加入任务列表的。所以优先级低的任务可能会等待很长时间才能被执行。这种处理方案在人类日常行为中很常见,如通常会最优先做那些重要的或者急切需要完成的工作,然后再做其他的。

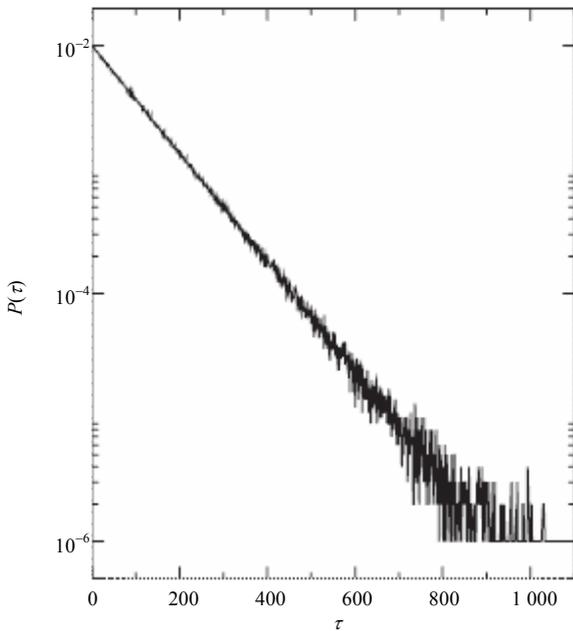
在Barabási模型中,对于每个时间步,个体执行最高优先级任务的概率是 p ,随机选取一个任务执行的概率是 $1-p$ 。容易得出,如果 $p \rightarrow 0$,则模型退化为随机选择方案;如果 $p \rightarrow 1$,则是高优先级优先执行的方案。

当 $\eta(x)$ 符合 $[0,1]$ 之间的均匀分布时,模拟结果如图8所示。任务优先级均匀分布在区间 $x_i \in [0,1]$,任务列表长度 $L=100$ 且模拟超过 $T=10^6$ 个时间步长。在图8a中,双对数坐标中概率 $P(\tau)$ 与任务等待时间 τ 的在 $p=0.999\ 999$ (p 为首先执行最高优先级任务的概率)时的图像。实线斜率为 -1 ,符合数值结果和理论预言。图8b中, $p=0.000\ 01$,近似随机选取方案的 $P(\tau)$ 分布。数据显示 $P(\tau)$ 满足指数分布。 $p \rightarrow 0$ (随机选择方案)时,等待时间分布 $P(\tau)$ 为指数分布; $p \rightarrow 1$ (按优先级执行)时, $P(\tau)$ 符合指数为 -1 的幂律分布,这与电子邮件通信的实证研究结果相一致。图8a为任务列表长度 $L=100$ 的数值模拟结果,等待时间分布 $P(\tau)$ 的尾部与 L 无关,即使在 $L=2$ 的情况下,都很好满足胖尾特性,接近于幂律。在文献^[33,109-110]中,有对 $L=2$ 的不同 p 的分析。如图9所示,个体保持

最少有两个任务, 则等待时间的胖尾现象就会被充分地显露出来。这个结果表明按优先级执行方案可能是导致人类行为为非泊松过程的重要原因。其中, 新任务优先级函数 $\eta(x)=1, 0 \leq x \leq 1$, 实线为精确解, 符号为对应的模拟数据: $p=0.9$ (方块), $p=0.99$ (菱形)和 $p=0.999$ (三角)^[33]。中间插图 $\tau=1$ 所对应 $P(\tau=1)$ 的值随 p 的变化, 其中实线是精确解。



a. $p = 0.999999$ 时双对数坐标中概率 $P(\tau)$ 与任务等待时间 τ



b. $p = 0.00001$, 近似随机选取方案的 $P(\tau)$ 分布

图8 排队模型的等待时间分布^[1]

值得注意的是, 该模型在 $p=1$ 的情况下(绝对执行最高优先级任务)是没有稳定状态的。此时, 停留在任务列表里面的任务优先级越来越低, 无限趋近

于0, 其等待时间也会趋向无限大。此时, 对系统等待时间分布的讨论就失去了意义。其详细的相关讨论, 可见文献[33,109,111]。

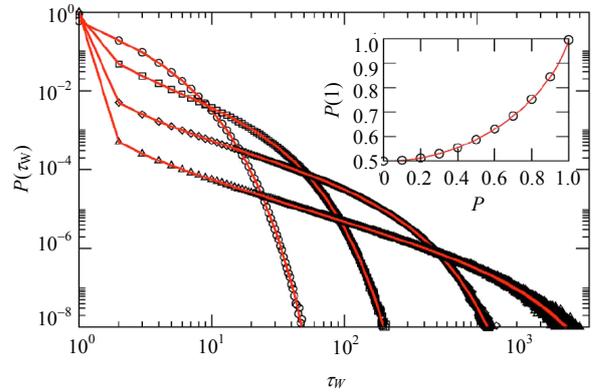


图9 $L=2$ 时Barabási模型等待时间分布图^[33]

Barabási模型^[1]代表了任务队列长度固定时的情况。此处, 可以引入一个自然的扩展: 假设接受任务的速率为 λ , 完成任务的速率为 μ , 允许任务列表长度随时间变化。令 $\rho=\lambda/\mu$, 以下有3种不同的情况需要讨论^[33]:

1) 亚临界状态。此时 $\rho < 1$, 即任务的接受速率小于完成速率。在这种情况下, 任务列表经常是空的, 大多数任务在它们接受到后不久就被执行了, 也就是说长时间的等待是很有限的。模拟得 $\rho \rightarrow 0$ 时等待时间分布为指数分布, $\rho \rightarrow 1$ 时近似的指数为 $-3/2$ 的幂律分布。

2) 临界状态。此时 $\rho = 1$, 任务列表长度随时间随意变化。不同于固定 L 的情况, 任务列表的长度的波动会影响任务等待时间分布。模拟结果显示等待时间分布满足指数为 $-3/2$ 的幂律分布。

3) 超临界状态。此时 $\rho > 1$, 意味着任务的接受速率大于任务的执行速率。即这任务列表长度不断增加, 其中 $1-1/\rho$ 的任务将永远不能完成。模拟结果是等待时间分布也满足指数为 $-3/2$ 的幂律分布。应该注意的是, 在这种情况下任务列表会一直持续增长下去, 该模型实际上没有一个最终的稳态。如何理解这越来越大的任务列表也是一个问题。以回复信件为实例, 人们将收到的信件放在办公桌上, 堆积在早期信件之上, 通常选择从中选择一封信件来回复。如果收到的信件太多以至于没有足够的时间回复所有的信件, 那么这些信件将会堆积得越来越多。人们不需要记住这些信件, 因为它们本身就在那里。模拟结果与达尔文、爱因斯坦和弗洛伊德的实际信件通讯情形相一致。这3位科学家都有许多信件从未被回复。数据结果如图10所示^[33], 该结果很

好地满足上述分析。

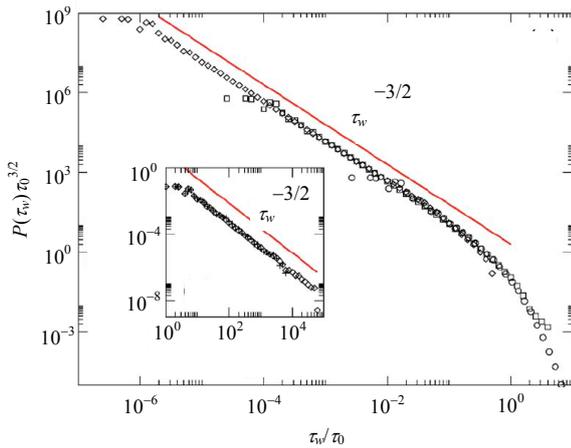


图10 连续优先级排序模型的任务等待时间分布图^[33]

数值模拟过程如下。每一个时间步长，根据速率 λ 和 μ 的指数分布分别产生一个任务抵达时间 τ_a 和服务时间 τ_s 。如果 $\tau_a < \tau_s$ 或任务列表中无任务时，新的任务被添加进列表，具有优先级 $x \in [0, 1]$ ，时间修正为 $t \rightarrow t + \tau_a$ 。此后去除列表中优先级最高的任务，并将时间修正为 $t \rightarrow t + \tau_s$ 。图中3种符号对应不同的 $\rho (= \lambda/\mu)$ 分布为：0.9(圆形)、0.99(方块)、0.999(菱形)。重定标的横纵坐标分别为 τ_w/τ_0 和 $P(\tau_w)\tau_0^{-3/2}$ ，其中 $\tau_0 \sim (1-\rho_{1/2})^{-2}$ 。插入图部分为 $\rho=1.1$ 的等待时间分布，又为执行104个任务得到的模拟结果，菱形为执行105个任务的数据。数据表明即使在 $\rho > 1$ 时(超临界状态)也等待时间也满足幂律分布。应该注意到在这种状态下，大量的任务是永远不会被执行，任务优先级列表随时间的推移直线增长，实际数据模拟有一个适合的等待时间截断。

根据以上所述，在任务队列长度动态变化的情况下，任务等待时间分布满足指数为 $-3/2$ 的幂律。结合Barabási提出的任务队列固定情况下的原始模型，可以将人类行为根据其幂指数区分为 -1 和 $-3/2$ 两大普适类^[33]：任务队列长度固定情况下为 -1 的普适类，任务队列动态可变的情况下为 $-3/2$ 的普适类。在该理论提出的当时，人们所研究的实证结果绝大部分是可以粗略地归入这两大普适类的，如电子邮件通信可以被归入幂指数为 -1 的普适类，普通邮件通信可以归入幂指数为 $-3/2$ 的普适类。但实证研究发现，大量人类行为的统计特性，对不同个体而言，往往存在一个连续变化的谱；同时，随着模型研究的发展，也使得人们认识到影响人类行为的因素的多样性和复杂性，任务队列理论也只能适用于其中一部分情况。在这种情况下，简单把人类行为归于两大普适类的观念，已经显现出其局限性。

任务队列理论作为人类动力学研究领域的第一种重要模型，到目前为止，仍然是人该领域研究的一种重要思想。针对早期模型所存在的各种问题，近年来研究者也从各种现实情况出发，更深入地挖掘了它的内在机制，提出了多种改进型模型。如研究者深入讨论了它的严格数学解析^[109,111]，提出了其背后的等效渗流理论^[110]，引入了任务本身的时间特性^[112-113]、任务处理优化^[114]等各种现实因素，并且更进一步将其拓展到社会网络之上^[39,115-116]，使之成为了当前人类动力学研究思想中的主流。

2.2 基于人类活动的记忆、兴趣和节律等解释

任务队列理论成功解释了相当部分人类活动统计特性的产生机制。但是，这一理论基于“任务队列与优选”的基本假设，导致其自身存在多种缺陷，使之并不能成为解释人类行为的普适性模型。如相当一部分人类行为，特别是娱乐等有着较强自主性的行为^[44-65]，并不能简单被视为一种任务；再如在这个理论框架下，深刻影响人类行为的生理周期因素和工作时间因素并未被考虑，而前后所做事情的相关性因素的影响也难以引入。要深入理解人类行为中的非泊松统计特性，还需要有任务队列理论以外的新方法和模拟方案，因此研究者也从排队论以外的多个角度对人类行为的统计特性进行了模型研究。这些研究涉及到历史记忆特性^[52,117]、兴趣变化^[118-119]、生理周期和工作周期的影响^[28,103]、所做事情之间的关联^[120]等。本文将对其中几个典型模型进行介绍。

首先来看基于人类活动中的记忆特性的模型。记忆是人类的基本神经活动之一，对历史事件的记忆常常可以深刻影响一个人后面的行为。文献^[117]提出了一个简单而重要的基于记忆性的模型，在这个高度简化的模型中，记忆被最简练地表达为某件事情前后发生频率和次数之间的相关性。

在该模型中，基本的假设是：人们根据自己过去进行某个活动的频率的高低，来增大和减小当前的活动频率。记 $\lambda(t)dt$ 为在 $t \sim t + dt$ 时间内某个人做某件事的概率。基于这个假设，关于 $\lambda(t)dt$ 的方程可写为：

$$\lambda(t) = a \int_t^1 \lambda(t') dt' \quad (7)$$

式中， $a > 0$ 是这个模型中的唯一参数。当 $a = 1$ 时， $\lambda(t) = \lambda(0)$ ，这时整个过程是处于稳态的。另外，当 $a \neq 1$ 时，该过程不处于稳态，做该事情的概率将要增大($a > 1$)或减小($a < 1$)。

应该注意式(7)对初始时刻有个隐含的假设: $t = 0$ 。正如文献[117]所指出的, 这是对人们记忆底线的反映, 意思是说人们不会记得或者思考在那之前发生过什么。举个例子, 无论前一天做什么, 我们每天工作时都会打开邮箱收发新邮件。

从式(7)出发, 可以通过数学推导得到时间间隔的分布形式的严格解; 具体的推导过程可以参考文献[117], 这里只给出其基本的结果: 当 $a=1$, 时间间隔的分布满足幂律形式。设 $\tau_0=(a\lambda_0)^{-1}$, 其中 λ_0 是考虑的时间段 T 内的做相应事件概率的平均值。当 $a > 1$ (概率增大), $\tau_0 \ll \tau < T$ 时, 这个模型生成的时间间隔分布接近于指数为 $-2-(a-1)^{-1}$ 的幂律分布。另外, 当 $0 < a < 0.5$ (概率减小)时, 时间间隔分布接近于指数为 $-1+a/(1-a)$ 的幂律分布。在 $0.5 < a < 1$ 时, 系统则不能显现幂律特性。研究者采用达尔文和爱因斯坦的普通邮件通讯的累积数量增长过程的实证统计结果来确定模型参数 a , 如图11所示, 其中, 图11a为达尔文和爱因斯坦的累积通讯次数, 直线显示指数为2.7的幂律增长过程; 图11b为相应通讯时间间隔分布, 其中圈点为达尔文的, 方点为爱因斯坦的。其中直线表示指数为-2.37的幂律, 该指数可由模型的解析分析得出, 其中所采用的 a 值由图11a确定。将这样一个简单的模型得到的结果与实证统计结果进行对比, 发现两者可以基本吻合。其详细的讨论见文献[117]。

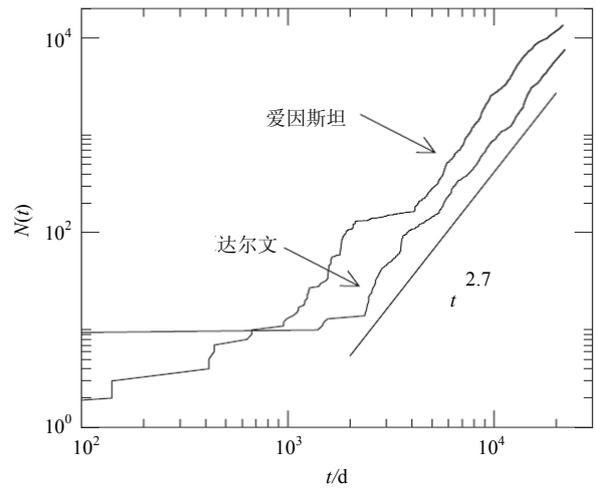
另一个涉及历史行为影响的动力学模型是从人进行某项事情的兴趣或者欲望的自适应变化角度出发的^[118]。这种自适应的兴趣变化在日常生活中很常见, 如: 如果一个人很久没有浏览网页了, 一个偶然的上网会唤起他(她)对上网的兴趣; 此后, 由于兴趣的持续, 上网的频率也不断增加, 但一段时间后就会出现兴趣的衰退, 从而减少上网的频率。此外, 还可以在日常生活中找到其他具有类似过程的例子, 如玩游戏、看电影等。简单的说就是人们对某件事情的兴趣是不能长期保持不变的, 但能够进行自适应调节。基于此, 该模型提出以下假设: 1) 对于一个给定的兴趣驱动行为, 即当行为的频率依赖于兴趣时, 每做一次该行为都会改变对该行为的兴趣; 2) 假设时间间隔 τ 有两个阈值, 当 τ 很小时(即事件的发生频率很高时), 兴趣将减弱, τ 增加; 时间间隔太大时, 模型保证事件将以很小的概率随机发生。因此, 该模型的规则如下。

1) 时间离散, 即 $t = 0, 1, 2, \dots$; 在 t 时刻, 某事发生的概率记为 $r(t)$; 两个相邻事件 i 和 $i+1$ 发生的时

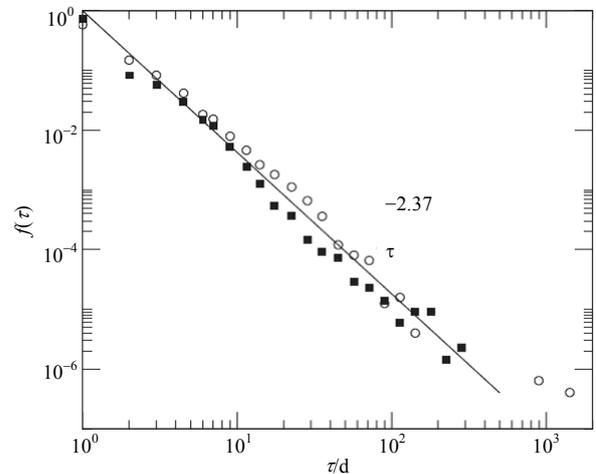
间间隔记为 τ_i 。

2) 第 $(i+1)$ 件事发生在 t 时刻, 此时 r 值表示为 $r(t+1) = a(t)r(t)$ 。其中, 当 $\tau_i \leq T_1$ 时, $a(t) = a_0$; 当 $\tau_i \geq T_2$ 时, $a(t) = a_0^{-1}$; 当 $T_1 < \tau_i < T_2$ 时, $a(t) = a(t-1)$ 。

如果在 t 时刻没有事件发生, 就规定 $a(t) = a(t-1)$ 。这里 T_1 和 T_2 是两个阈值, 满足 $T_1 \ll T_2$ 。 a_0 是表示事件发生概率变化率的一个参数($0 < a_0 < 1$)。如果没有事件发生, r 的值不发生变化。可以看出同时扩大(相同倍数) T_1 、 T_2 和最小感知时间, 将不会改变系统的统计特性。因此, 不失一般性, T_1 取值为1。



a. 达尔文和爱因斯坦的累积通讯次数



b. 通讯时间间隔分布

图11 达尔文和爱因斯坦邮件通讯统计特性^[117]

在模拟中, 文献[118]设初始值 $r_0 = r(0) = 1$, 这也是 $r(t)$ 在整个过程中的最大值。如图12所示, 可以看出在相间的爆发中会有一些较长的静默期, 相应的 $r(t)$ 也显示出周期性变化的特性。图13表示出在不同的 T_2 和 a_0 值下模拟的结果, 两图中的黑色虚线的斜率都是-1。图中所有数据由超过100次的独立模拟, 每次试验 10^4 个事件, 最后取平均得到的。当 $T_2 \gg T_1$

时,取 $a_0=0.5$,此时在这个模型下时间间隔分布呈指数为-1的幂律分布形式。当 T_2 不足够大时, $P(\tau)$ 的分布不再符合幂律形式,将在尾部出现一个截断。相应地,当 T_2 足够大时, a_0 的影响是非常小的,可以忽略。在 $T_2 \gg T_1$ 时,该模型可以得到解析解,其形式接近于指数为-1的幂律。详细的数学推导可以参阅文献[118]。

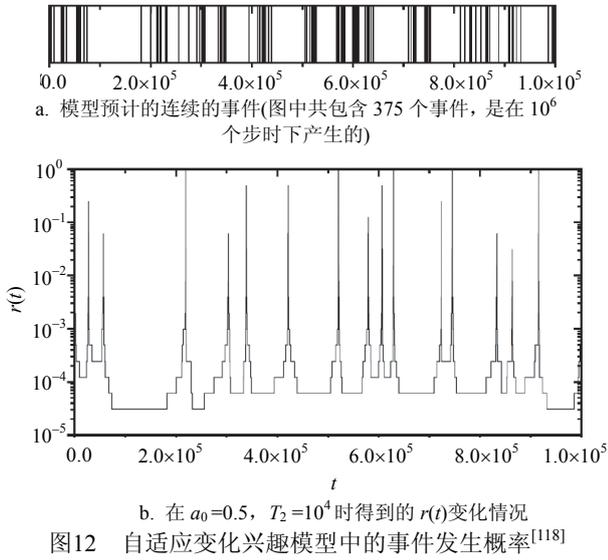


图12 自适应变化兴趣模型中的事件发生概率^[118]

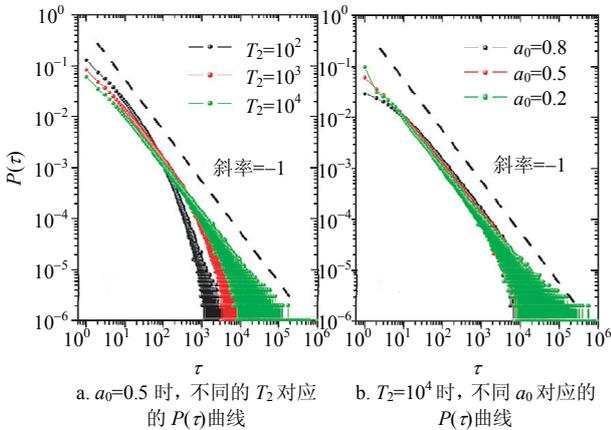
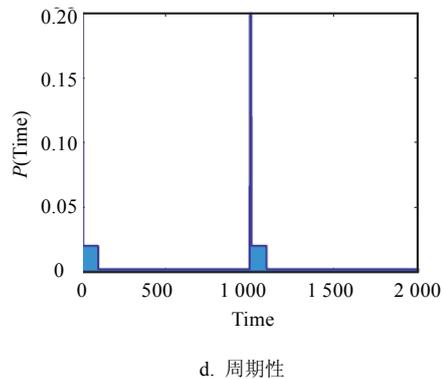
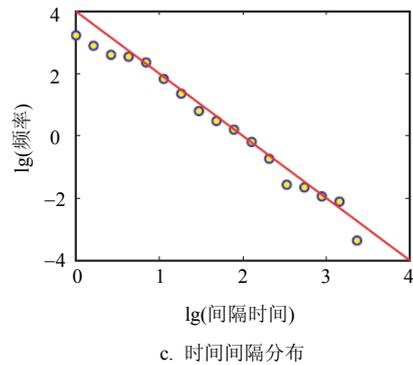
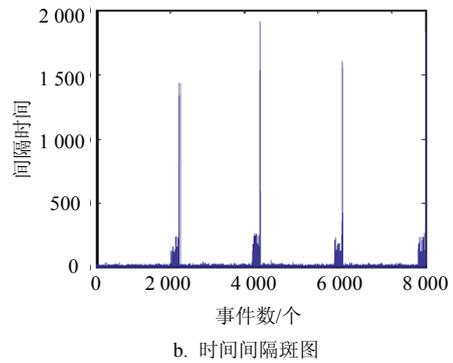
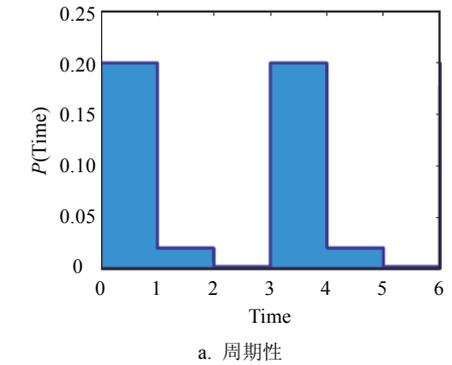


图13 间隔时间分布的双对数图^[118]

在该模型中行为是由人们的自适应变化的兴趣驱动的,这不同于前面讨论的排队论模型。模型中事件发生的频率是由兴趣决定的,同时事件的发生对兴趣也有反作用。这个简单的行为模式可以应用在其他很多人类行为的模拟上。

各种生理周期和工作节律也对人类活动有着重大影响,如休息会把白天的工作时间分割为若干段,而夜晚和周末也会迫使人们把工作性的活动停下来。因此,在现实生活中,人们并不是在任何时候都按照优先级或兴趣选择事情(熬夜打牌毕竟是少数),这些来自外部和内部的节律性制约着人的实际

行为。近年来,学者们提出了多种模型来研究这种节律性对人的活动特性的影响。首先对这个方面进行的讨论是在文献[103]中,该文构建了一个发生概率可以随时间进行周期性变化的随机过程,这种周期性变化可以相应于现实中的生理周期或者工作周期,并发现这种过程可以产生出具有标度性特征的时间间隔分布,如图14所示。



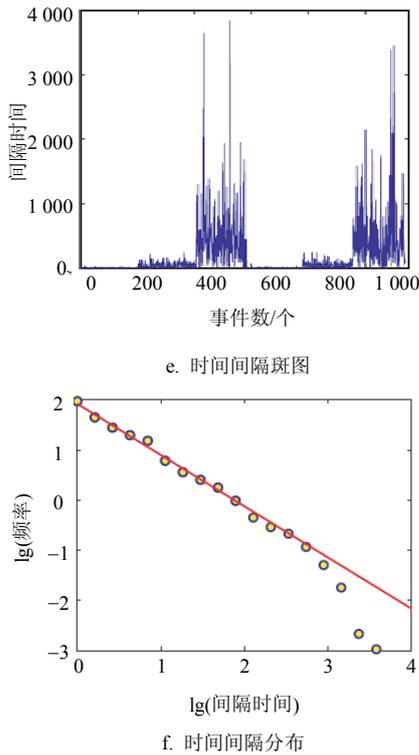


图14 基于节律的人类动力学模型示意图^[103]

图14a~图14c分布表示在该模型中发生概率的周期性变化、产生的时间间隔斑图和相应的时间间隔分布, 红线表示指数为-2的严格幂律。图14d~图14f分布表示另一种发生概率较低而持续时间较长的情况下的模型结果, 红色直线表示指数为-1的幂律。

文献[28]基于对电子邮件使用行为的实证统计提出了一种周期性级联泊松过程, 来研究这类节律性的影响。在电子邮件用户的行为中, 研究者观察到这样的一类特征: 一个人使用电子邮件的活动常常具有依赖于睡觉和工作时间的周期性变化, 并且具有明显的级联式阵发特点。相比于传统的发生概

率 ρ 恒定不变的均质泊松过程, 该模型定义了一个可随时间变化的发生概率 $\rho(t)$, 其变化满足周期 W , 即 $\rho(t) = \rho(t + W)$ 。研究者根据实证观察, 定义这样非齐次的泊松过程的发生率 $\rho(t)$ 的形式为:

$$\rho(t) = N_w p_d(t) p_w(t) \tag{8}$$

式中, 周期 W 定义为1星期, 比例常数 N_w 表示用户在每周的平均活跃次数; 分布 $p_d(t)$ 和 $p_w(t)$ 分别是基于实证数据统计获得的用户在一天和一星期的时间尺度内的活动间隔分布。

该模型进一步假设每一个由初级过程发生的事件(活动), 均会启动一个事件发生概率为 ρ_a 的次级过程。在此次级过程中, 平均事件发生次数 N_a 随机取自于某种分布 $p(N_a)$ 。当此次级过程的 N_a 个事件发生后, 系统的活动性将返回到由式(8)所定义的初级过程。图15展示了这样一个过程的基本斑图。这一个双重的随机过程还反映了实际生活中人们电子邮件使用时的基本行为: 一个人常常习惯零星的发送一些邮件, 但是当检查收到的邮件时, 他常常会连续地发送很多邮件。这样的过程被称作“级联非齐次泊松过程”。

为了将该模型与实证进行对比, 研究者基于实证数据, 对各个电子邮件用户分别估计了该模型的3个重要参数函数: $p_d(t)$ 、 $p_w(t)$ 和 $p(N_a)$, 并采用这些估计参数计算得到了模型结果, 如图16所示, 两者几乎是高度吻合。这样一个机制非常简单的模型说明, 在生理周期和工作时间周期影响的情况下, 即使人的基本行为按照泊松过程的模式发生, 也会导致整体的统计特性出现对于泊松过程的偏离。这个工作大大丰富了对人类行为的非泊松特性来源的认识, 也反映出人类行为本身所具有的高度复杂的特点。

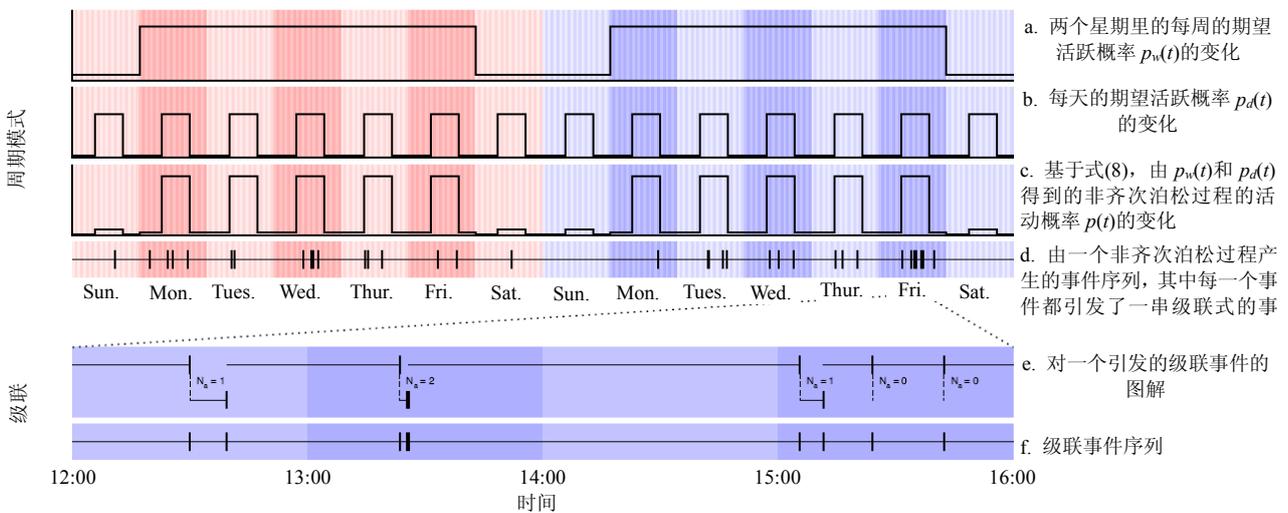


图15 周期性级联随机过程的例子^[28]

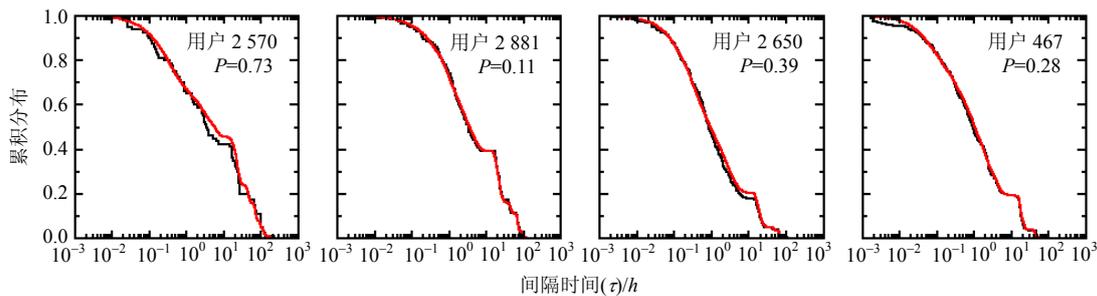


图16 对几个用户的行为,用级联非齐次泊松过程模型预测的时间间隔的累积分布 $P(\tau)$ (红色线)与实证数据(黑色线)进行的比较,两者达到了高度的一致^[28]

需要注意的是,这种基于周期性或季节性影响的思路,实际上是把人类行为中的阵发性归结于外部环境的影响。要识别实际人类行为是否真正存在内禀的阵发性,就要剔除这种来自周期性或季节性的影响。最近文献[100]提出了相对时钟的方法,有效地解决了这一问题。相对时钟,是指在计算某个体的两次事件之间的时间间隔分布的时候,并不直接采用两次事件之间的时刻差,而是采用在这两次事件之间,同一时区其他个体所发生的事件的总计数。也就是说,“相对时钟”是用事件计数间隔取代了绝对时间间隔。这样一些可作用在全体个体上的外部因素,例如昼夜影响、周末效应等,就会被有效地剔除。

文献[100]提出了一个模型来展示相对时钟对显个体内禀阵发性的效果。这个模型共包含 N 个个体,模拟了10天的时间尺度,每一时步表示一秒的时间单位,共864 000时步。对第 i 小时(包含3 600时步),有其活动性值 λ_i 。对第一天内的24小时, λ_i 的取值满足一个确定的分布 $\psi(\lambda)$,后面各天重复第一天 λ 的取值,即 $\lambda_{i+24}=\lambda_i$ 。模型考虑了4种典型情况: 1) 全部个体都是发生率为 r 的泊松个体,即对每个个体,在每秒发生事件的概率是 $r\lambda_i$,其中 i 标记当前的小时区间,并且在任何时候 λ_i 取值保持不变,此时系统中事件的发生满足纯粹的泊松过程; 2) 与情况1)唯一不同的是, $\lambda_i (i = 1, 2, \dots, 24)$ 是取自于 $0 \sim 1$ 范围内的均布随机数,即 $\psi(\lambda) = U(0, 1)$; 3) 所有个体具有内禀的时间间隔分布满足 $\Phi(t) \sim t^{-\beta}$ 的阵发性,在模型中,初始时对每个个体首先从分布 $\Phi(t)$ 中抽取一个随机的 t 值,然后令 t/λ_1 为第一个事件发生的时刻,此后每次独立抽取一个 t 值, t/λ_i 作为该事件与上一个事件之间的时间间隔,此处固定 $\lambda_i (i = 1, 2, \dots, 24) = \lambda$; 4) 与情况3)不同之处在于 λ_i 取自于均匀分布 $\psi(\lambda) = U(0, 1)$ 。总之,1)和2)的情况中个体不存在内禀阵发性,情况3)和情况4)存在内禀阵发性;

而且情况2)和情况4)还受到外界周期性因素的影响。

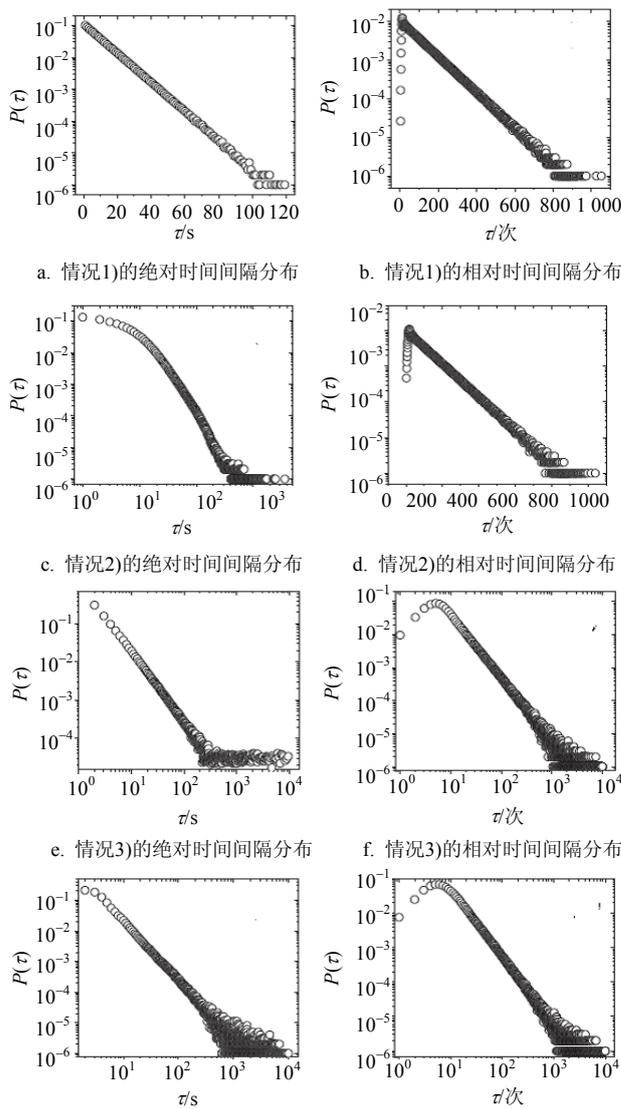


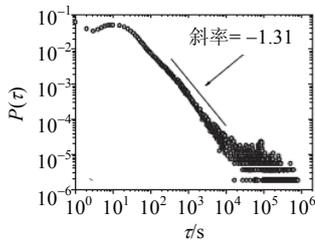
图17 若干人工生成的序列的绝对时钟时间间隔分布与相对时钟时间间隔分布的对比^[100]

这4种情况下的绝对时钟时间间隔分布和相对时钟时间间隔分布如图17所示,模型参数为: $N=100, r=0.2, \lambda=0.5, \beta=2$ 。显然,存在内禀阵发

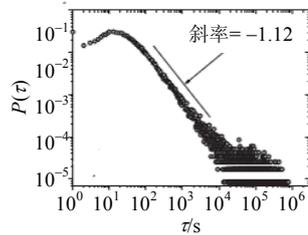
性的3)和4)的相对时钟间隔分布仍然存在标度性;而对于不存在内禀阵发性的2), 尽管因为周期性因素的影响导致其绝对时间分布显现胖尾的特征, 但是其相对时钟时间间隔分布则显露出了其泊松性的实质。

更进一步, 通过对美国在线、美味书签、短消息通讯和推特网的用户行为的绝对时间间隔分布和相对时钟时间间隔分布的比较, 如图18所示, 发现实际的用户行为在剥离外部因素影响后, 均呈现出内禀的阵发性特征, 说明试图把人类行为中的阵发性归结于周期性或者季节性等外部因素的思路是远不足以说明这些行为特性的根本来源的。

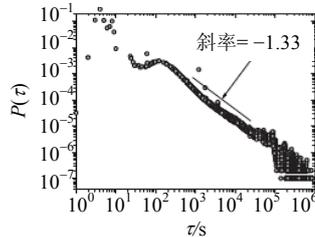
此外, 基于空间结构上的随机运动也可以用来描述人类行为时间上的统计特性。在这一方面, 文献[120]基于对用户浏览网页行为的实证分析, 提出了超级链接上的自吸引行走模型, 对此进行了初步的讨论。



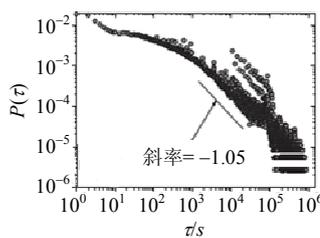
a. 美国在线的用户行为的绝对时间间隔分布



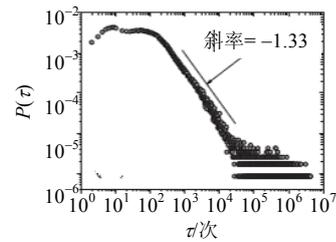
b. 美味书签的用户行为的绝对时间间隔分布



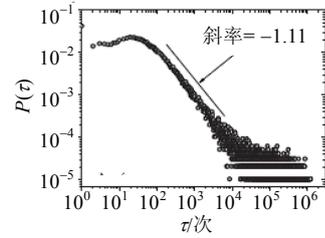
c. 短消息的用户行为的绝对时间间隔分布



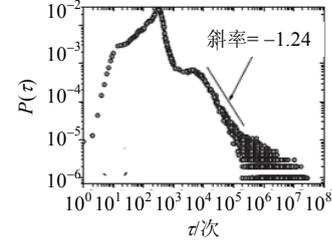
d. 推特网的用户行为的绝对时间间隔分布



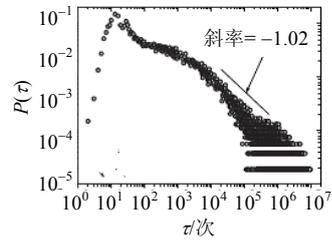
e. 美国在线的相对时钟间隔分布



f. 美味书签的相对时钟间隔分布



g. 短消息的相对时钟间隔分布



h. 推特网的相对时钟间隔分布

图18 若干实际用户行为序列的绝对时钟时间间隔分布和相对时钟时间间隔分布对比^[100]

门户网站的页面超链接网络如图19所示, 其主页链接了大量的子页面, 而这些子页面也各自链接了一批更下级的子页面; 这些子页面以及次级子页面又都有链接可以直接返达主页和其上层页面, 从而构成了一个层次性的有向网络结构。该模型基于实际门户网站的页面超链接网络, 通过一个有偏好性的随机行走过程, 来模拟人们浏览网页的行为特性。该偏好性的随机行走过程的基本规则是: 个体从主页节点出发, 按照一个概率 p_{ij} 随机移动到当前位置 i 节点的一个邻居 j 节点上; 其中 p_{ij} 的取值正比于从 i 到 j 的连边的权重 w_{ij} , 而权重 w_{ij} 正比于实证统计获取的通过该边的双向点击量。当模型运行初始两个时步之后, 个体将以 p^* 的概率返回两个时步前的节

点,以 $1-p^*$ 的概率继续以上述规则随机选择邻居继续运动。数值模拟结果与实证结果的对比如图20所示。其中,蓝色数据点表示实证结果,其意义为在用户连续 n 次的网页浏览中,曾经被用户所浏览过的可识别的子页面数 Z 的最大值;黑色点为在 n 次浏览中,用户浏览过的子页面数 Z 的平均值,而灰色区域表示了用户在 n 次浏览中所浏览过子页面数 Z 的误差区间;红色数据点由上述模型进行300时步的数值模拟获得。其中图20a中的实证数据为基于一天的数据统计获得,图20b为基于一周的数据。可以看出,这样一个简单的模型可以大致再现用户的网页浏览行为。

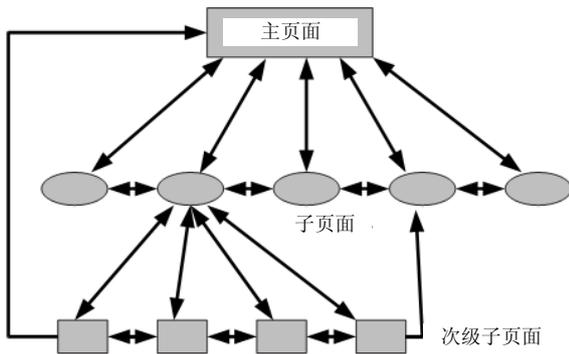
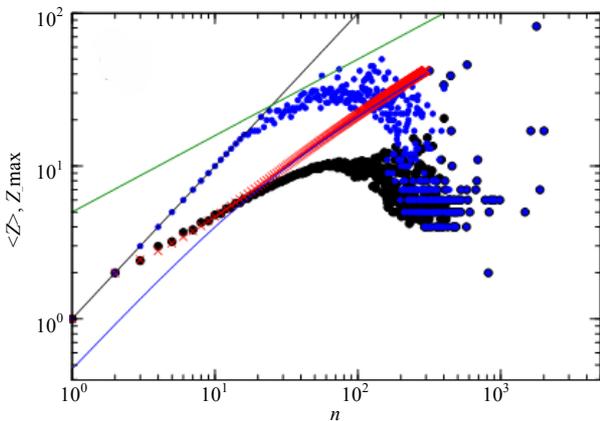
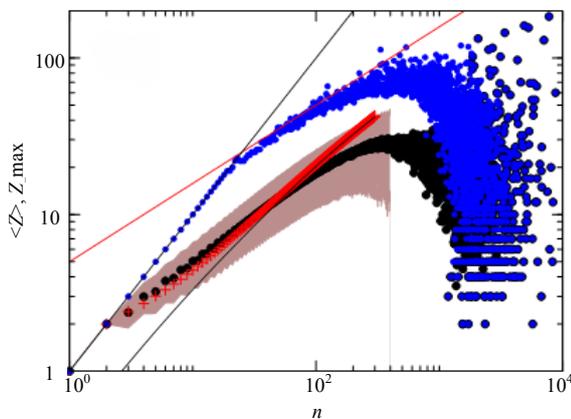


图19 门户网站的超链接结构示意图^[120]



a. 一天数据



b. 一周数据

图20 实证与模型数值模拟获得的浏览网站时网页的子页面数量与跳转次数的关系^[120]

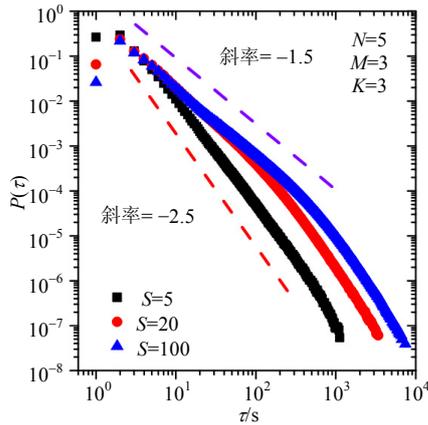
文献[120]的模型实际上是这类建模思路针对网页浏览行为的一个特例。该类建模思想的核心是,考虑现实中人类各种行为之间所具有的关联或条件关系所构成的网络,把人类的行为变化视作在这样的关联网络上按照某种规则进行的随机运动。如当一个人使用电脑时,他会有更大的可能点播网络音乐、进行网络通信等,而进行线下购物、参加聚会等行为的可能会非常小。也就是说,使用电脑、点播网络音乐和网络通信等行为直接构成了高度相关的一个集团,而这个集团的条件就是使用电脑;其他类别的行为,诸如旅行、购物等行为类别中也能观察到明显的关联和条件关系。同时也有部分行为几乎可以随时发生,如休息、收发短信等,游离于各种行为集团之外。该类思想的一个更一般化的模型是模拟人类空间运动行为的层次性交通模型^[121],这个模型在第5节会详细介绍,此处只抽象出与时间特性有关的规则:在此基础之上,通过考虑任务关联中明显具有的集团化特性,提出了如下规则:

1) 网络层次数目为 N ,第一层个体数目为 K ;第 n 层个体数目为 KM^{n-1} ;第一层个体之间全连通;在生成网络时,第 n 层个体随机选择一个第 $n-1$ 层个体进行连接,并与所有的连接到同一个 $n-1$ 层节点的第 n 层节点进行连接。

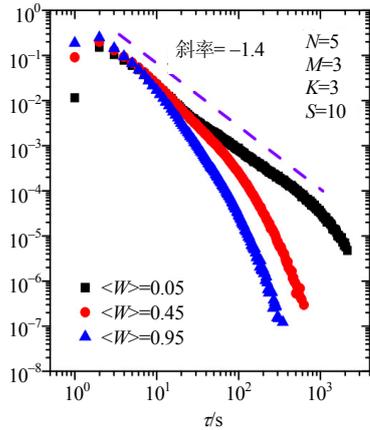
2) 所有这些节点共包含 S 种类别。每种类别具有权重 w_i (w_i 在 $0\sim 1$ 之间等概率随机取值, $i=1,2,\dots,S$),节点为第 i 种类别的概率正比于 w_i 。

3) 一个个体以随机的初始位置在该网络中沿着连边进行随机运动,统计相邻两次经过第 i 类别的时间间隔 τ_i 的概率密度 $P(\tau_i)$ (注意,不是相继两次通过同一节点,而是通过同一类别的节点的时间间隔)。

通过数值模拟,可以获得该模型的如下基本结论:首先,节点的类别数目 S 对平均时间间隔分布 $P(\tau)$ 有着明显影响: S 越小, $P(\tau)$ 的斜率越高,其幂指数介于 $-1.5\sim -2.5$ 之间,如图21a所示;在 S 较大时会出现明显的指数尾。其次,当 S 相同时,权重 w_i 越大的类别对应的 $P(\tau_i)$ 的斜率越大,并且出现明显的相对幂率的偏离,呈现出指数尾的情况,如图21b所示。这说明,该模型可以产生基本符合实证研究结论的人类行为的时间间隔分布特性。



a. 不同节点种类数S下的平均时间间隔分布 $P(\tau) = \langle P(\tau) \rangle$



b. 不同类别的权重 w_i 所对应的平均时间间隔分布 $P(\tau)$ ($\langle w \rangle = 0.05$ 表示对全部 w_i 取值在0~0.1之间的节点取平均, 而 $\langle w \rangle = 0.45$ 表示对全部 w_i 取值在0.4~0.5之间的节点取平均)

图21 层次性模型产生的访问时间间隔分布^[121]

需要注意的是,除了人类行为,在对Linux内核程序代码的统计分析中也发现,计算机程序的关键词出现的间隔行数分布也存在胖尾特性。而在计算机程序中,这种层次性集团化(模块化)的结构是其程序设计的主要原则。这一模型同样也适用于对这种计算机程序关键词间隔行数分布的解释。总而言之,结合下文中将要提到的针对人类空间运动行为中的标度性出行特性的层次性地理网络模型,该模型揭示出这种层次性结构在标度性时间空间特性的涌现中扮演了非常重要的角色,也暗示出,系统结构的层次性与体现出的标度性统计特性之间可能存在更为紧密的联系。

除了以上所介绍各种基于动力学演化的非任务队列理论模型,最优化思想也在一些模型当中得以体现,如文献[122]发现采用非泊松性的时间间隔分布策略,可以有效地回避未来事件的不确定性,削减这种不确定性带来的代价。

2.3 基于社会交互的人类动力学模型

以上所叙述的模型都是仅仅考虑了孤立的个

体。在现实生活中,每个人都是生活在一个与其他人互相交流影响的社会环境中。除了那些纯粹的私人活动,大部分的日常行为会经常受到来自他人活动的影响,特别是各类直接或间接的社交、通讯行为。如在电子邮件通讯中,我们常常是等待对方回复后再进行进一步的通信联系。这种人与人之间的交互影响,把整个社会连接成为活性的整体。如何有效地阐述社会交互对人类个体与群体行为的统计特性的影响,成为了人类动力学模型研究的一个重要挑战。

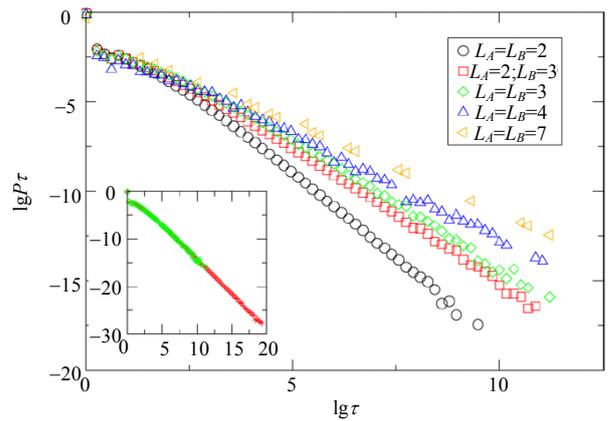


图22 OV 模型中 I 型任务的间隔时间 τ 的概率分布^[115]

在这一问题上,研究者首先在最简单的两个个体交互模型上取得了突破。文献[115]提出一种拓展的任务队列模型(OV 模型),该模型仅仅考虑两个个体: A 和 B ,每个个体都有两种类型的任务:1)交互任务(I),必须由两人共同完成;2)非交互任务(O),可由个体单独完成。每个任务根据概率密度函数 $\eta(x)$ 得到一个随机优先级 $x_{ij}(i=I, O; j=A, B)$ 。每个时间步,每个个体都选择执行在其任务列表中优先级最高的任务(A 的任务列表长度为 L_A , B 的任务列表长度为 L_B)。如果双方都选择 I 型任务,则他们可以共同完成;否则他们都只能选择 O 型任务执行。图22为该模型的数值模拟结果,可以看出 I 型任务的间隔时间分布很接近幂律分布,并且可以产生不同的幂指数。

这一模型尽管极为简单,但它扩展了Barabási模型的研究范围,提供了一个研究人类社会交互行为特性的成功范式。随着大量针对社会交互行为的实证统计研究的涌现,研究者在 OV 模型的基础上,成功地将其推广到了更为实际的条件之下,并取得了一系列的成功。其中一个典型的基于双个体通讯的模型针对的是手机用户之间的短消息通信的行为特性^[39]。

不同于前面大量讨论的幂律时间间隔,在针对

手机用户的短消息通讯行为中，研究者观察到短消息发送时间间隔分布存在明显的双模特性^[39]：其分布尾部出现了明显偏离幂律的指数性的衰减。为了阐明在这类交互行为存在的双模特性，研究者提出了一个基于交互行为影响的模型。这个模型的基本规则包括两种过程：

1) 个体对一系列任务采用基于优先级的方案执行。一系列任务一个一个地根据概率 $\Pi = x^\alpha$ 被执行；此处 x 是一个介于0~1之间的随机数，表示该任务的优先级， α 是该模型的一个可调参数。相比于标准的任务队列理论模型，这个模型又进行了几项扩展：首先，依据任务处理时间 t_p 引入了一个时间尺度，每经过 t_p 秒，就有任务从队列中被移除或者新加入该队列；其次，所有任务被区分为两类：交互类任务(I类任务)和其他任务(O类任务)；最后，在每一个处理时间步，I类任务以一个小概率 $\lambda_p = \lambda t_p$ 随机加入队列，这一点表示了实证数据中观察到的任务激发所具有的泊松特性。

2) 个体之间的相互作用过程。这种相互作用出现在当个体A或者B执行一个I类任务时。此时，一个I类任务将以 $P_B(P_A)$ 的概率被添加在B(A)的任务列表中，其中 $P_B(P_A)$ 就是个体B(A)的短信回应概率。所有的由个体自发启动的或者相应交互所产生的I类任务都将获得一个随机的优先级 x ，并加入等待队列中。对每个个体，该模型有3个重要的参数： λ_i 、 α_i 、 P_i ($i = A, B$)，分别相应于该模型中所存在的泊松机制、执行的决定过程和相互作用。

研究者首先基于实际的短消息数据来拟合该模型的参数，并将基于这些参数所获得的模型预测结果与实证结果进行比较。其中，参数 λ_A 、 λ_B 、 P_A 、 P_B 可以通过区分独立爆发的事件来获得；参数 α_i 可以通过 $\alpha_i = 1/(\gamma_{wi} - 1)$ ，其中 γ_{wi} 表示实证数据中爆发区域内的等待时间分布的幂指数，这个公式来自于任务队列理论的解析式 $\gamma_{wi} = 1 + 1/\alpha$ 。研究者采用不同的 t_p 进行了模拟，并计算了所得时间间隔累积分布 $F(\tau)$ 实证数据之间的相对差异 E 。图23展示了模型获得的时间间隔分布 $P(\tau)$ 和累积分布 $F(\tau)$ 与实证结果的比较，模型结果为红色线(点)，实证结果为蓝色线(点)。曲线 $E(t_p)$ 表示了累积分布 $F(\tau)$ 和实证数据之间的相对差异 E 随 t_p 的变化，对于两个个体，在 $t_p=10$ 附近 E 都趋于最小值。可以看出两者结果吻合很好。这里所涉及的技术细节，在文献[39]的补充信息中进行了详细的介绍。

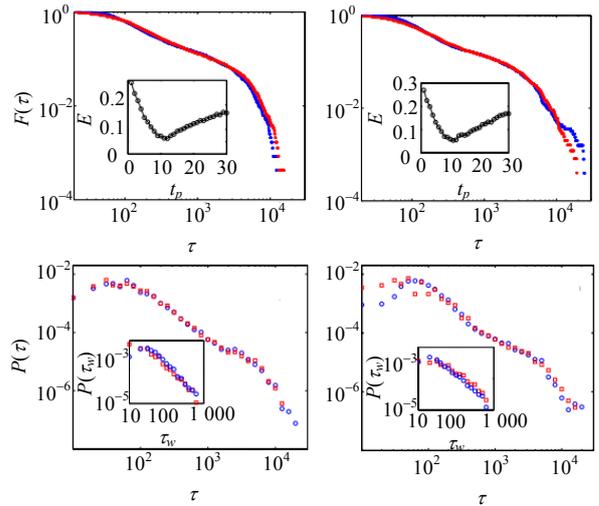


图23 模型获得的时间间隔分布 $p(\tau)$ 和累积分布 $F(\tau)$ 的实证结果比较^[39]

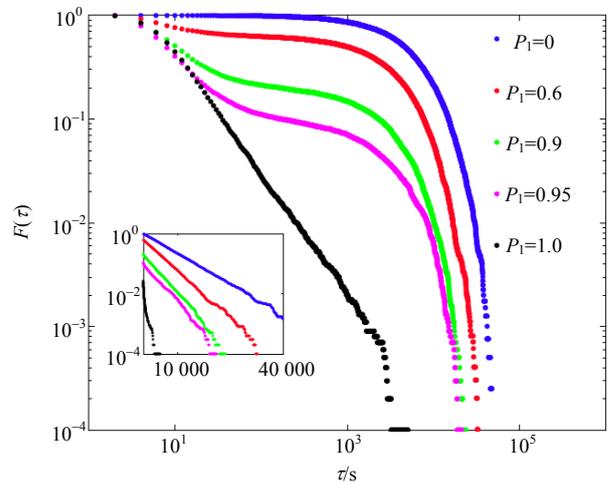
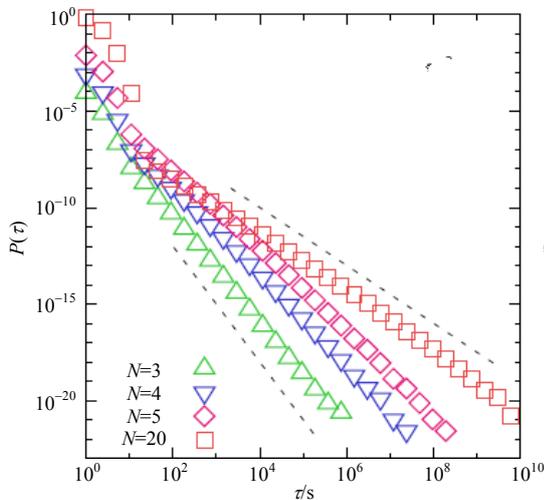


图24 相互作用对人类活动特性的影响^[39]：不同短信回应概率 P_1 所对应的累积分布 $F(\tau)$

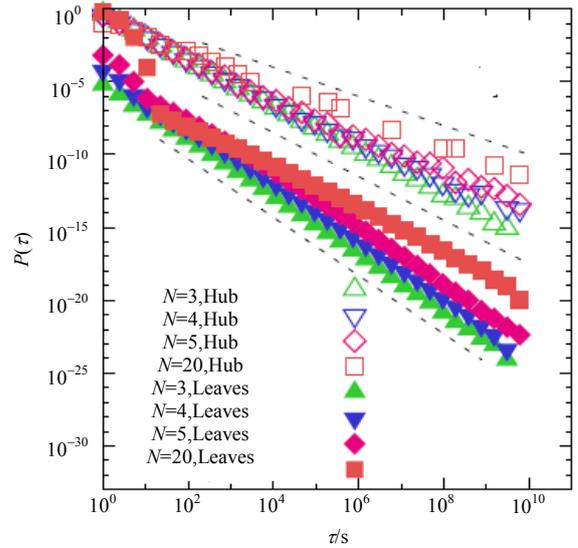
另一方面，为了考察个体之间的相互作用过程对模型结果的影响，设定两个个体相应的参数完全相同，即 $P_B = P_A = P_1$ ，并固定 $t_p = 1$ 来考察 P_1 对时间间隔分布的影响。图24展示了不同的短信回应概率 P_1 时所得到的累积分布 $F(\tau)$ ，其他参数为 $\lambda = 1.5 \times 10^{-4}$ ， $\alpha = 1.0$ ， $t_p = 1$ 。其中插图表示在单对数坐标下的指数尾。可以看出，模型结果敏感依赖于 P_1 的取值。当 P_1 略微小于1.0时，时间间隔分布曲线就会出现相对幂律的明显偏离——出现与短消息通讯中所呈现的双模特性相一致的指数尾偏离。

除此之外，近期一些工作也把OV模型范式的社会交互模型成功推广到了社会网络之上，在这种框架下讨论了社会网络结构对人类行为特性的影响。在这方面，文献[116]的模型是一个典型，相比于基

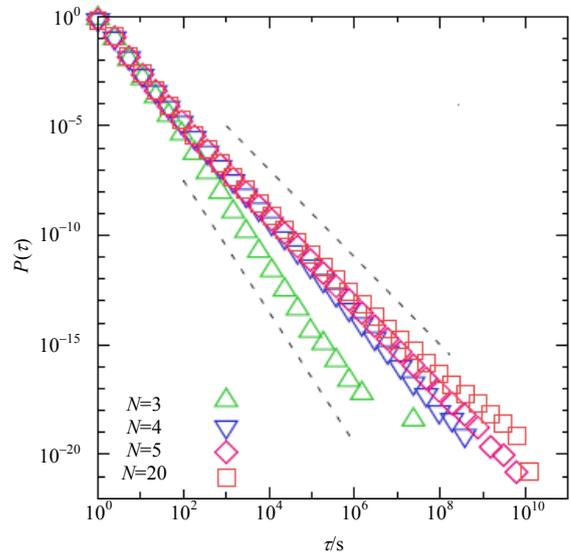
本的OV模型, 提出了任务队列网络的概念: 网络上每一个节点都是一个任务队列(其实相应于社会网络上的一个个体); 在每一个节点的任务序列中, 它对每一个邻居均有一个I类任务, 再加上本身的一个O类任务, 因此对一个度为 k 的节点 i (队列), 其任务队列长度为 $L_i = k_i + 1$ 。用 I_{ij} 表示节点 i 和 j 之间的I类任务, 用 O_i 表示节点 i 的O类任务。初始时, 对每个任务随机赋予一个在 $0 \sim 1$ 之间的值来表示其优先级。在每个时步, 每个节点选择其队列中优先级最高的任务执行。同基本的OV模型类似, 如果任务 O_i 被选择, 就直接被执行; 而任务 I_{ij} 只有在它同时是节点 i 和 j 的最高优先任务时才能被执行, 如果该条件不满足, 则选择一个最高优先级的O类任务执行(与模式)。可以看到, 这一模型在基本规则上并没有对基本的OV模型进行更多的拓展, 仅仅推广到可以描述社会接触网络中的相互作用。研究者首先在一个星型网络上考察了该模型产生的时间统计特性, 并区分考察了网络结构中枢纽节点和普通节点之间的行为特性的区别, 发现这种模式下I类任务被执行的时间间隔分布满足标度性, 但是中枢和普通节点之间行为没有明显的区别。同时, 研究者也考虑了另一种情况, 即或模式: 在每一个时步随机选择一个节点 i 进行任务执行; 如果 I_{ij} 任务有着最高优先级, 则不管相应 j 节点的任务优先级如何, 均同时执行 I_{ij} 和 I_{ji} ; O_i 任务的执行规则不变; 所有被执行过的任务都被随机更新其优先级。在这种情况下, 模型所得到的I类任务的时间间隔分布敏感依赖于网络规模, 同时中枢节点和普通节点之间的行为特性也出现了明显的差别: 对中枢而言, 其O类任务被执行的时间间隔分布具有较小的幂指数, 如图25所示。



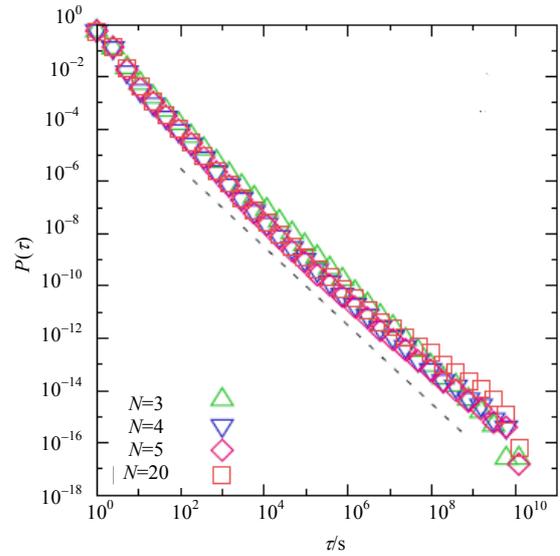
a. 不同网络规模时的I类任务的时间间隔分布(其中总节点数 $N = 3$ 时分布幂指数为 -3 , 而 $N = 20$ 时幂指数为 -1.5)



b. 不同网络规模时, 中枢节点和普通节点相比, O类任务的时间间隔分布(3条虚线分别对应斜率1、1.5和2)



c. 在不同规模的全连通网络上得到的I类任务的执行时间间隔分布



d. 在不同规模的全连通网络上得到的O类任务的执行时间间隔分布

图25 在“或模式”下基于星形网络得到的模型结果^[116]

这类基于社会交互的模型还可以进行更进一步的简化。最近文献[123]提出了一个高度简化的人类社会交互模型，并获得了丰富的非泊松的时空统计特性。该模型的基本规则如下。

1) 在一个结构固定的网络中(节点最小度为 k_{\min})，每个节点 i (个体)被随机赋予一个活跃度 A_i ， A_i 是 $0 \sim 1$ 之间的一个随机数。

2) 在每一时步，从网络中选择一个个体作为信息发送者，选择方法是以概率 p 选择当前时步网络中活跃度 A 最高的个体，或者以 $1-p$ 的概率随机任选一个个体。

3) 选出信息发送者后(假如是个体 i)，从其邻居中抽取 n 个个体作为接收者，并发送消息(如个体 j 、 k 等；此处 $n \leq k_{\min}$)。抽取规则是：假如最近一次向发送者 i 发送消息的个体为 j ，那么以概率 q 选定 j 为其中一个接收者，或者以 $1-q$ 的概率随机抽取一个接收者，再另外随机抽取 $n-1$ 个邻居作为接收者；若发送者 i 此前没有接收到过任何信息，则直接随机从其邻居中抽取 n 个个体作为接收者。这里概率 q 表示了个体对以往信息的应答回复行为的概率。

4) 发送消息者(个体 i)和这 n 个接收者的活跃度 A 分别进行随机更新——更新为 $0 \sim 1$ 之间的新随机数。随后进行下一个时间步的演化。

该模型固定每个时步发送消息的个数为1，因此该模型所统计的时间间隔分布实际上是系统事件计数间隔分布，因此，该模型所统计的时间间隔分布

本身就是处在相对时钟意义上的，可以与前文所述的相对时钟间隔统计^[100]进行直接的对比。

可以看出，该模型的规则已经高度抽象化，仅仅保留了发送或者接收消息对个体活跃性的改变这样一个关键的机制。当网络为规则网络时，并且 $n = k$ 时(k 为规则网络节点的度)，此时 q 的取值对系统没有影响，若取 $p = 1$ 该模型就转变为 Bak-Sneppen 断续平衡临界性模型^[124]。而当 $n = 0$ 时，就完全忽略了个体之间的直接相互作用，该模型可趋近于固定任务队列长度的 Barabási 任务队列模型^[1]。数值模拟显示出，经过长时间演化，个体的平均活跃度趋向于一个稳定值，并呈现出断续平衡特性。当 $n > 0$ ， $p \rightarrow 1$ ， $q \rightarrow 0$ 时，在规则网络上，该模型可以获得非泊松形式的事件计数间隔分布(即个体相邻两次发送消息之间其他个体所发生的发送消息事件次数)，如图 26a 和 26c 所示，以及具有标度性的相继两个发送者空间距离分布(即相邻两次发送事件的发送者之间的空间距离分布)，如图 26b 和 26d 所示。当 q 增大时，模型的时间间隔分布会显现出类似双峰特性的特征，暗示在短消息通信中所观察到的双峰特性可能也与通信应答机制有关系。总而言之，该模型既显示了长时间的关联性，也显示了系统的长程关联性，系统可能处在一种统计性质稳定但个体变动剧烈的临界状态上，并进而暗示这种临界性也是一种可能导致人类社会交互行为中的非泊松特性的动力学机制。

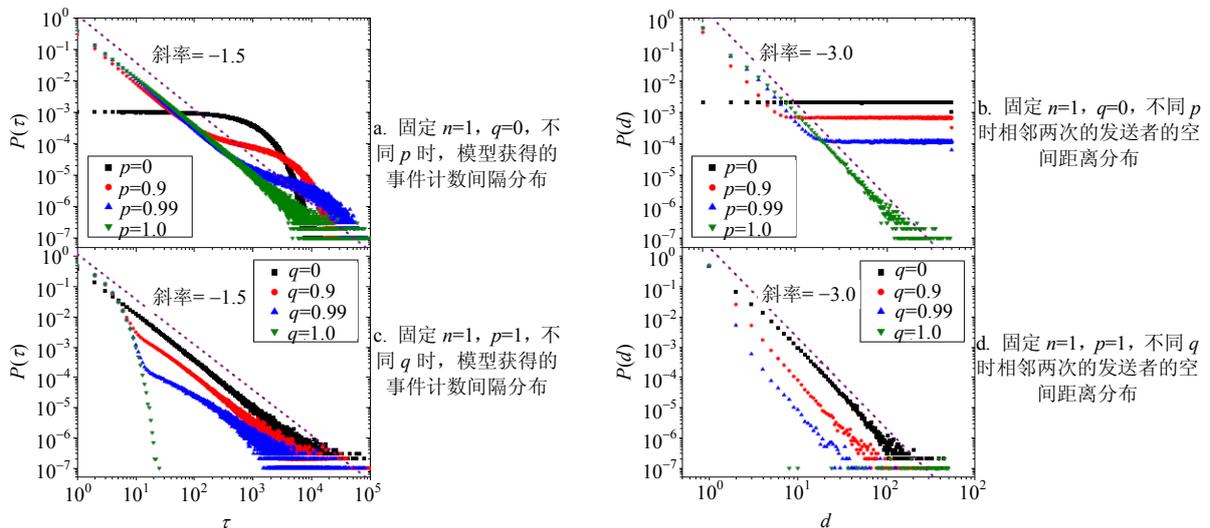
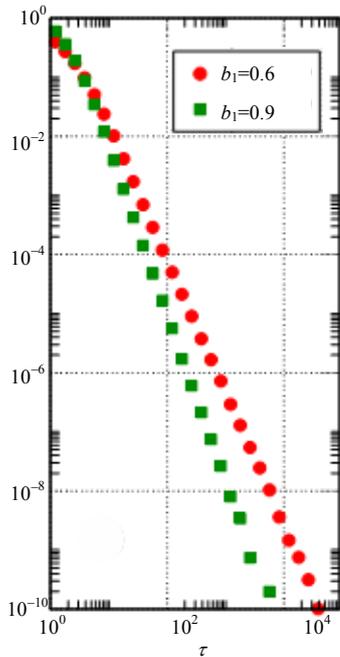


图26 人类社会交互模型的时空统计特性

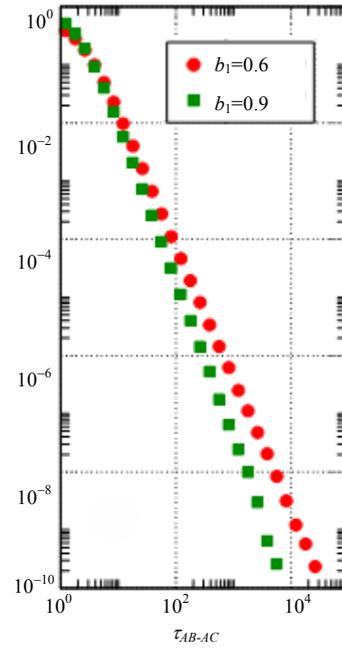
以上这些模型都是基于固定的社会网络结构。但是，现实的社会接触过程常常是暂态存在且不断演化着的，而且这种社会结构上的演化也会影响到

人类的社会接触行为特性。因此，有必要构建模型来研究个体的行为模式与社会网络结构之间的协同演化过程。文献[125]提出了一个高度简化的基于动

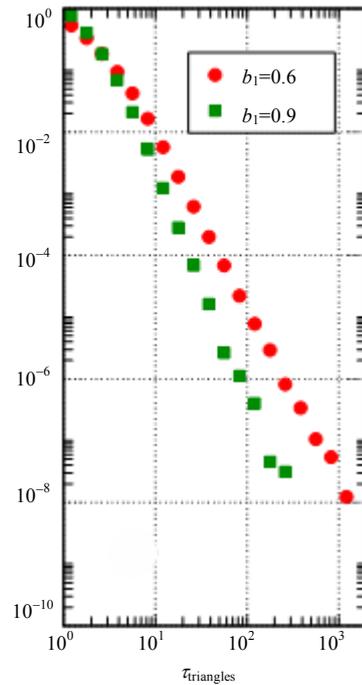
态变化的小规模社会群组的交互行为模型。在该模型中, 总个体数目 N 固定不变; 每个接触群组由一个瞬态的接触网络所构成, 而每个个体可以依照一定规则加入或者离开一个群组。具体的规则如下: 每一个个体(以个体 i 为例)由以下两个变量所标示: 正在接触的其他个体数 p_i 和 p_i 上一次改变的时间 t_i 。在每一个时步, 随机选择一个个体 i , 若该个体是孤立的($p_i = 0$), 则该个体以概率 $b_0 f(t, t_i)$ 随机改变其状态。此时, 以一个概率 $\Pi(t, t_i)$ 随机选择另一个孤立的个体 j , 并与 j 组成一个接触对, 即 $p_i \rightarrow 1, p_j \rightarrow 1, t_i \rightarrow t, t_j \rightarrow t$ 。若个体 i 不是孤立的, 而是在群组 G 中(G 的规模大于1, 即 $p_i > 0$), 则 i 的状态以概率 $b_1 f(t, t_i)$ 改变。此时, 个体 i 会以概率 λ 离开该群组成为一个孤立的个体, 或者以概率 $1-\lambda$ 把另一个孤立的个体引入该群组, 此时所涉及各个体的 p 值和 t 值相应发生变化。显然, 在该模型中, 个体行为主要由参数 b_0 和 b_1 所决定, 参数 λ 控制了个体的群组特性, 而该模型的动力学行为主要由 f 和 Π 所决定。考虑一个相当自然的假设: 随着时间的延长人们会更加倾向保持原有状态, 因此对 f 和 Π 采取了简单的单调衰减形式: $f(t-t_0) = \Pi(t-t_0) = (1 + \tau)^{-1}$, 此处 $\tau = (t-t_0)/N$, 其中 t_0 表示该个体上一次改变状态的时刻。在这种情况下, 该模型可以产生多种标度性的时间分布特性, 既包含了个体交互行为的时间分布, 也包括小型交互群组的存在时间分布, 如图27所示, 揭示了社会群体变化和个体行为之间的互动演进过程。



a. 两个个体之间连边的持续时间分布



b. 某个体A与两个个体B和C分别产生连边的起始时间的间隔分布



c. 三角形群组的持续时间分布

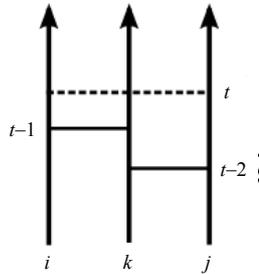
图27 Stehlé模型所得时间统计特性^[125], 模型参数为:

$b_0 = 0.7, \lambda = 0.8, b_1 = 0.6, 0.9, N = 1000, T = 105 N$

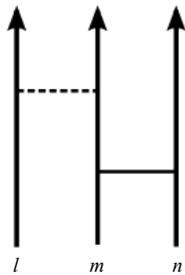
文献[126]提出了一个考虑了多种社会交互模式的模型, 来研究社会交互行为模式与社会网络结构的协同演化过程。在该模型中, 两个陌生个体之间的连接可以通过两种方式产生: 1) 随机与陌生个体接触, 称为全局性接触过程(GA), 该过程仅仅涉及到两个体的相互作用; 2) 局域接触过程(LA), 表示了个体通过已经存在的邻居与次近邻个体发生接触

的过程,其具体形式如图28a所示,其中箭头表示了时间方向,横的实线表示已经在 t 时刻前发生的接触事件,虚线表示 t 时刻将发生的接触。这一过程涉及到了3个个体的相互作用。该文讨论了两类模型:

- 1) 三方交互增强模型(TI),即优先进行三体交互行为;这一类型实际上是文献[127]提出的含权网络模型的拓展。
- 2) 过程平衡模型(PE),即同等考虑了以上各种形式的相互作用,其处理过程类似于Barabási任务队列模型的形式。



a. 局域性接触过程(LA),即个体 i 与个体 j 之间,只有当他们在时间区间 $[t-1, t-2]$ 范围内分别同其共同邻居 k 产生过接触,才能在 t 时刻进行接触



b. 全局性接触过程(GA),即个体可以随机地与任选的个体产生接触

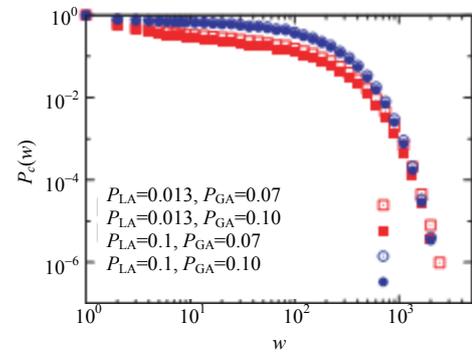
图28 模型中局域性与全局性接触过程示意图^[126]

TI模型的演化规则同时包含了3种基本过程:

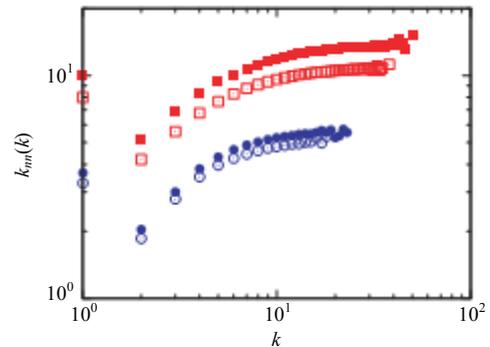
- 1) 三体交互行为:如图28a所示若两个个体在过去的两个时步中分别与同一个个体发生了接触,则看这两者间是否存在有连边,如果有连边,则发生一次接触事件;如果没有,则以概率 p_{LA} 产生一条连边;
- 2) 两体接触过程:若节点在上述三体接触行为中未被涉及到,则以概率 p_{GA} 随机选择一个陌生节点进行接触并建立连边,或者以概率 $1 - p_{GA}$ 随机选择一个邻居进行接触;
- 3) 记忆损失过程,即对每一个个体,会以概率 p_{ML} 损失掉其连边成为一个孤立节点。考虑到交互的方向性与否,该模型还包含了TI-OR和TI-AND两种类别,其中TI-OR模型不考虑交互行为的方向性,而TI-AND模型则假设两个个体之间不同方向的交互是分时交替进行的。

通过数值模拟,如图29所示,TI模型即可以获得接近实际的个体接触行为的时间统计特性,也可以获得在部分主要统计参量上与实证比较接近的社

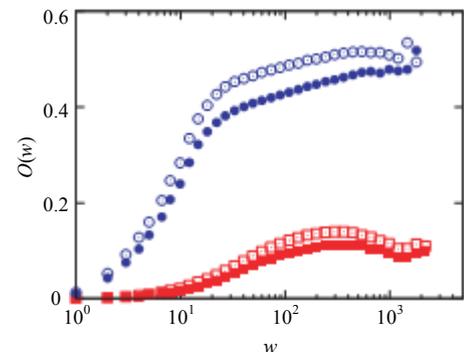
会网络结构。该模型是一个比较典型的个体行为-社交网络结构协同演化模型。因为实际社会交互中行为模式的多样性,该模型的规则相对前面所述各模型显得略为繁琐,但模型规则和模型结果显示,社会交互中的基本模式已经被该模型成功地描述了出来。



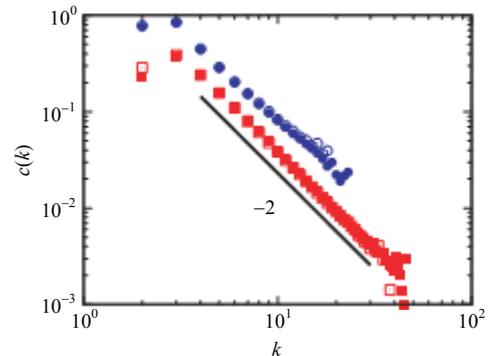
a. TI-OR模型的累积权重分布 $P_c(w)$



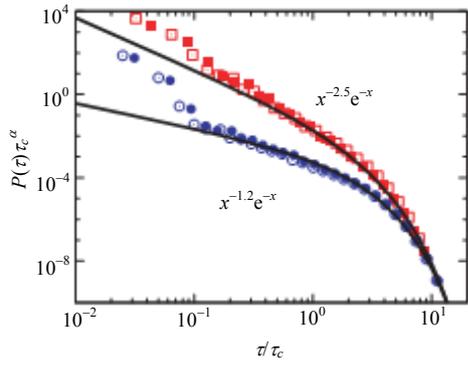
b. TI-OR模型的平均次近邻居数目 $k_{nn}(k)$



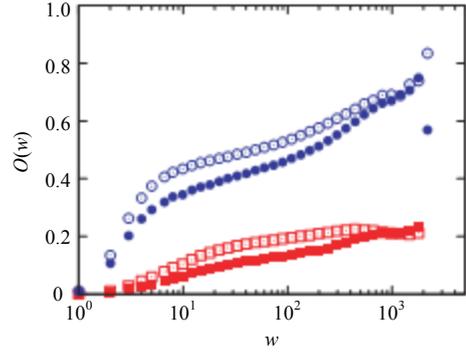
c. TI-OR模型的平均重叠数 $O(w)$



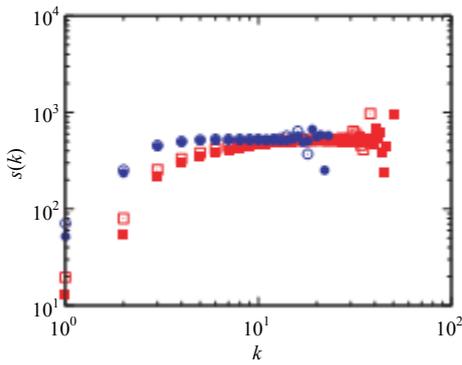
d. TI-OR模型的局域簇系数 $c(k)$



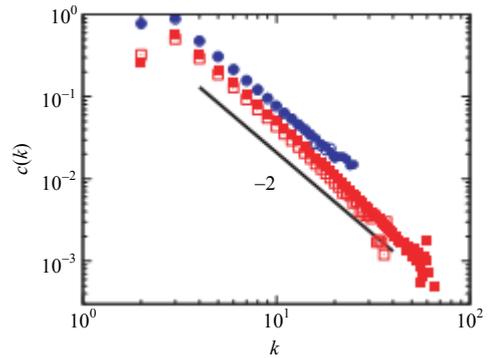
e. TI-OR模型的时间间隔分布 $P(\tau)$



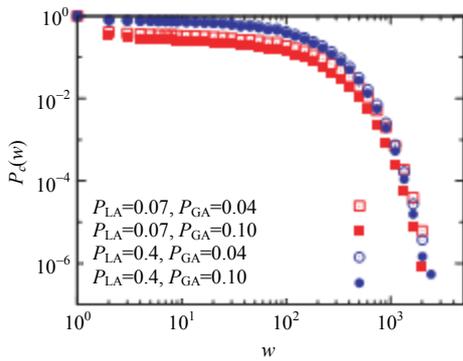
i. TI-AND模型的平均重叠数 $O(w)$



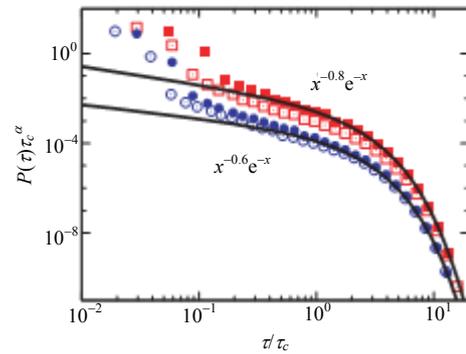
f. TI-OR模型的平均强度 $s(k)$



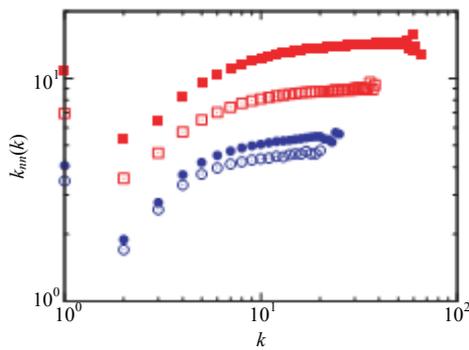
j. TI-AND模型的局域簇系数 $c(k)$



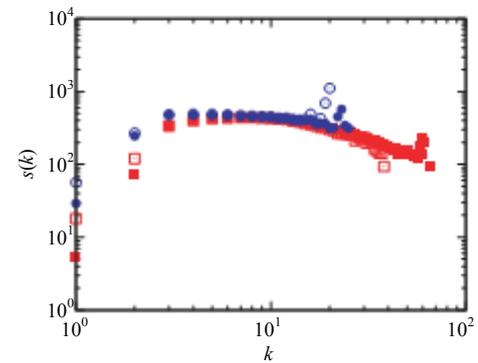
g. TI-AND模型的累积权重分布 $P_c(w)$



k. TI-AND模型的时间间隔分布 $P(\tau)$



h. TI-AND模型的平均次近邻居数目 $k_{nn}(k)$



l. TI-AND模型的平均强度 $s(k)$

图29 TI-OR和TI-AND模型^[126]

2.4 小结与讨论

综上所述,经过数年来的深入探讨,目前所提出的人类行为特性机制已经触及到了影响人类行为的各个方面,不仅涉及到人类行为决策时的内部因素,也考虑了各种来自外部环境和所处社会群体的影响。更进一步地,研究人员已经开始关注社会系统结构与人类行为模式的协同演化。由于人类行为的高度复杂性,可能影响人类行为的因素极其多样,使得针对人类行为非泊松特性产生机制的研究丰富多彩。近年来,甚至还有工作从完全不同于传统建模的角度出发,采用生命科学方法来探索人类行为特性的生理机制来源。如文献[128]发现,野生果蝇的内源性活跃开关的活动间隔满足具有胖尾特性的Weibull分布,而且该分布的特性可以随着一些生理因素(如多巴胺信号、蘑菇体、决策神经回路等)的变化而改变。这些研究从不同的方面探讨了人类时间行为中的非泊松特性,大大深化了人们对自身行为规律的认识。这些被揭示出的机制,涉及到了日常生活和社会交往的各个方面,一步步揭示出人类行为特性背后所隐藏的高度的复杂性和丰富的动力学机制,其中每一类模型都有其具体的适用范围和针对的行为类别,其所采用的模型规则和作用范式各不相同。从目前的研究来看,要给出对这些模型的理论是非常困难的——把人类行为中的各类非泊松特性归结到一个简单原理上几乎是不可能的。而对于各种模型所讨论机制的归类,以及对实证研究结果和各类模型结果比较,在对人类行为机制的进一步的深入探讨中,就显得十分必要。可以说,在研究如同人类行为这样的复杂系统的动力学机制时,相比于传统物理,模型研究和实证研究需要更为紧密的结合。

由于针对人类行为统计特性的动力学模型,也可被用作对人类行为整体特性进行预测,因此这方面的研究也可望对实际应用产生价值。随着当前社交网络、电子商务的迅速发展,个性化、精准化、社会化的信息挖掘与推荐方法的应用正在急速拓展。而对人类行为统计特性及其机制的深入理解,可以在很大程度上提高对用户行为预测的准确性,有助于实现更为精准的用户信息挖掘。这类深度结合了大规模的实际用户数据的建模和预测研究,在未来将会有长远的发展,将在第6节展开进一步的叙述。

3 人类行为空间特性的实证分析

理解人类行为的空间特性对于研究受人类空间

运动驱使的各种复杂现象具有重要意义。如某些传染病正是由于人的旅行和相互接触才发生传播,因此研究人群的空间运动规律有助于研究者们深入理解这些疾病的传播机制,进而提出有效的预防和控制措施,这部分内容在第5节还会介绍。另一个更直接的例子是交通流的研究,人在不同地点间的移动直接导致交通网络上的各种复杂流动现象,只有在了解人类空间运动规律的基础上,才能合理规划交通设施,预防和控制交通拥堵。事实上,人在城市内和城市间的移动一直是交通工程学研究的主题。交通工程师很早就开始用问卷调查等方法收集城市居民的日常出行数据,并根据这些数据建立模型来预测交通流量^[129]。但这些传统的交通调查方法往往成本较高,难以大规模、长时间地对人类空间运动行为进行观测记录。随着现代电子技术的发展,已有越来越多的手段(如手机定位系统、GPS等)可以长期记录人类个体的空间运动数据,这些数据为采用统计方法研究人的空间运动行为提供了可能。从2006年起,在Nature、Science等顶级期刊上已发表了多篇关于人类空间运动特性研究的文献,开启了人类动力学中的一个新的研究方向。本节对此方向的主要实证研究成果进行介绍。

3.1 美元流通数据

文献[2]利用美元流通记录网站www.wheresgeorge.com的数据间接地对人类空间运动行为进行了实证统计。该数据集包含美国境内100多万条美元流通数据,每条数据记录了钞票的编号、使用地点和使用日期,可以由此推算出每张钞票的空间移动轨迹。由于钞票是被人携带的,因此钞票的移动轨迹可以在一定程度上反映人的空间运动轨迹。

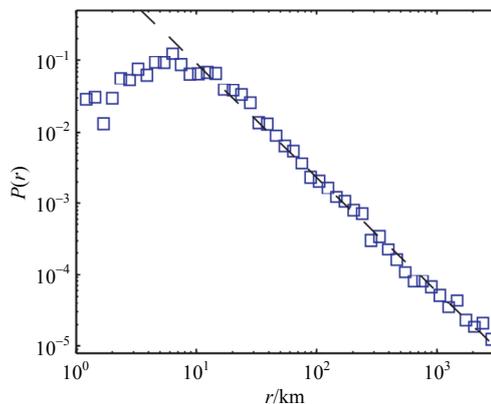


图30 钞票移动步长 r 的概率密度分布^[2]

对这些数据的统计分析结果显示,钞票的移动步长分布曲线具有明显的幂律尾部,如图30所示,图中直线是幂指数为-1.59的幂律函数。在10~3 500 km

的范围内, 钞票在距离为 r 的地点间移动的概率近似服从 $p(r) \sim r^{-(1+\beta)}$, 其中指数 $\beta \approx 0.59$ 。这说明钞票的空间移动具有类似Lévy飞行的特征: 存在大量短途移动的同时, 也能够观察到远距离的移动。

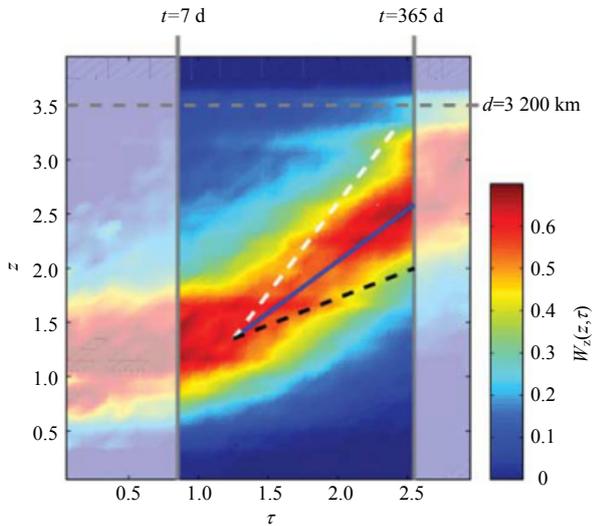


图31 钞票时空扩散概率密度函数 $W(z, \tau)$ 的图形

上述结果似乎表明钞票的空间移动可以用Lévy飞行模型描述。但本文发现, 仅用步长分布为 $p(r) \sim r^{-1.59}$ 的纯Lévy飞行模型并不能完整地再现钞票的空间移动特征。因为按照这一模型的预测结果, 钞票在68天内就会扩散到美国全境, 而实证统计却显示钞票的扩散速度要慢很多。本文用一个概率密度函数 $W(z, \tau)$ 刻画钞票的时空扩散特征, 该函数的含义是 τ 时刻在距离钞票移动起始点距离为 z 的位置再次观测到该钞票的概率。根据实证数据绘制的概率密度函数 $W(z, \tau)$ 的图形如图31所示, 图中, 横坐标 τ 是观测时间 t 的对数; 纵坐标 z 是钞票扩散距离的对数。图中实线是CTRW模型(参数根据实证数据估计)预测的典型扩散距离增长曲线, 白虚线是幂指数为 -1.59 的列维飞行模型预测的典型扩散距离增长曲线, 黑虚线是随机游走模型预测的扩散距离增长曲线。从图中可以看出钞票的扩散速度远低于具有相同幂指数的Lévy飞行模型的预测结果。为解释钞票的这种低速扩散的特性, 文献[2]统计了钞票在两次移动之间的停留时间分布, 发现它也服从幂律分布, 幂指数约为 -1.6 。为此, 该文献建议采用同时具有幂律的步长分布和幂律的停留时间分布的连续时间随机游走(CTRW)模型^[130]来模拟钞票的空间移动。该模型能够很好地再现钞票低速扩散的特征。

需要指出的是, 钞票位置的变化不能反映个体进行空间移动的统计规律, 因为一张钞票在登记过

程中会多次经手不同的人。事实上, 钞票位置的变化是否能够真正地反映人的空间移动本身也是存在疑问的。因为一张钞票两次登记中可能已经经过了很多人, 即便是同一个人, 也可能忽略了很多他自己认为不重要的旅程, 从而使得每次移动有可能是多个旅程的加和。尽管如此, 文献[2]是最早用统计物理方法对人类空间运动特性所进行的实证研究, 对后续的人类空间运动行为的统计和建模研究具有重要的启发意义。

3.2 手机数据

与在不同人之间流通的钞票不同, 手机是许多人每天都会随身携带的通信设备。当手机用户接打电话或收发短信时, 承担通信服务的基站被记录下来, 这使得研究者可以从手机通信记录中提取出用户的空间运动轨迹。因此, 相对于钞票流通记录数据, 手机数据能够更直接地反映出人的空间移动行为。文献[131-133]最早利用手机数据开展人类空间运动行为的实证统计工作。文献[131]分析了10万名匿名手机用户在6个月内的通信记录, 发现手机用户的移动步长近似服从带有指数尾的幂律分布:

$$p(r) = (r + r_0)^{-\beta} \exp(-r/\kappa) \quad (9)$$

式中, $\beta = 1.75 \pm 0.15$; $r_0 = 1.5 \text{ km}$; $\kappa = 400 \text{ km}$ 。

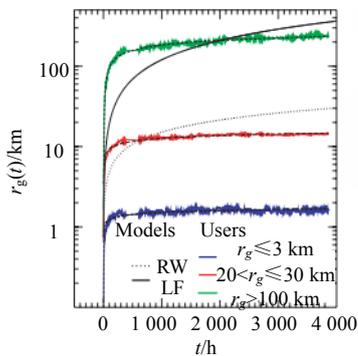
从幂指数看, 这一结果与对钞票流通数据的统计结果^[2]并无太大差异, 说明二者可能都反映了人类空间运动模式中的某种共同机制。但群体水平上人类移动步长的幂律分布可能是移动模式各不相同的若干个体混合所致的, 并不能据此推断每个个体的步长也服从式(9)的分布。为辨别这一点, 文献[131]进一步对个体移动轨迹的回转半径随时间变化的规律进行了分析。个体 a 在 t 时刻的回转半径定义为:

$$r_g^a(t) = \sqrt{\frac{1}{n_c^a(t)} \sum_{i=1}^{n_c^a(t)} (r_i^a - r_{cm}^a)^2} \quad (10)$$

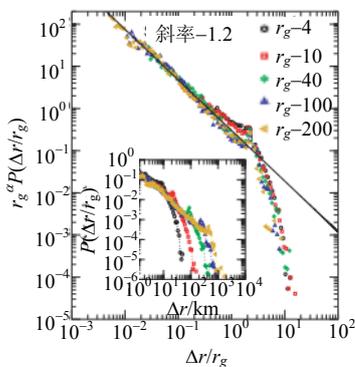
式中, r_{cm} 表示该个体在统计时段内所有轨迹点质心的所在位置; r_i 为各个轨迹点; $n_c(t)$ 为在 t 时刻的轨迹点数量, 上标 a 表示对个体 a 进行的统计。从图32a中可以看出, 个体的回转半径会很快趋于一个稳定值 r_g 。当把具有相同 r_g 值的群体归为一类, 并用 r_g 对每类群体的步长进行重新标度后, 所有类型群体的步长显示出近乎相同的分布形式: 在 r/r_g 小于1时服从幂律分布, 幂指数约为 -1.2 ; 而在 r/r_g 大于1时具有明显的截断效应, 如图32b所示。文献[131]据此认为每个个体的移动步长都服从一个与 r_g 无关、幂指数为 -1.2 的截尾幂律分布。但需要指出的是, 图32b中的步长分布依然是在群体水平上(若干具有类似 r_g 值

的个体)得出的统计结果,并不能据此推断出其中的每个个体步长都服从同样的分布。

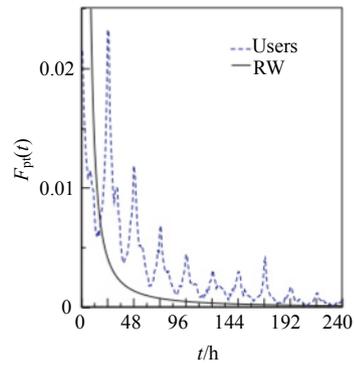
图32a对回转半径的统计结果还说明人类空间运动具有高度的有界性特征,这与经典的Lévy飞行模型具有本质区别:Lévy飞行模型的回转半径会随时间幂律增长(见图32a中的实线),而人类运动轨迹回转半径的增长异常缓慢,说明人更倾向于在一个有限的区域内活动。为揭示导致人类个体空间运动范围有界性的原因,文献[131]计算了用户在 t 时刻返回他之前被观测地点的概率,结果如图32c所示。图中的返回概率分布曲线在每24小时出现峰值,说明了人类空间运动在时间上具有强烈的周期性特征。如果以访问次数对个体访问过的地点进行排序,可以发现个体在一个排序为 L 的地点出现的频率符合指数为-1的幂律分布,如图32d所示。该分布与用户访问地点的总数无关,反映出人类具有返回之前访问过地点的强烈倾向。这些结果说明人类空间运动行为具有高度的规则性——大多数人日复一日地在少数几个地点(如家和工作地等)之间移动,偶尔会去访问其他地点。这种规则性可能是导致人类个体空间运动范围有界性的根源。



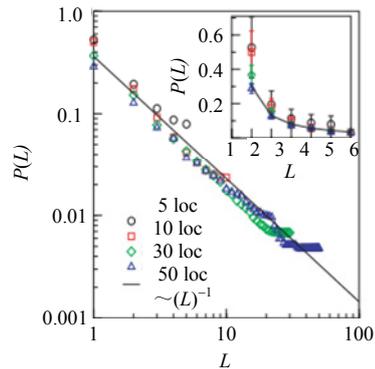
a. 回转半径随时间增长情况。三组手机用户的回转半径随时间变化的律,实线是Lévy飞行模型预测的回转半径增长曲线。虚线是随机游走模型预测的回转半径增长曲线



b. 按回转半径稳定值 r_g 对群体分类,并用 r_g 对每类群体的步长进行重新标度的移动步长分布。图中直线是幂指数为-1.2的幂律函数。插入图是重新标度前的步长分布



c. 用户在 t 时到返回之前地点的概率。从图中虚线是手机用户返回之前访问过的地点的频率分布,实线是随机游走模型预测的返回概率曲线



d. 个体访问地点的频率分布,实线是幂指数为-1的幂律函数。插入图是在线性坐标下表示的同样内容

图32 手机用户移动行为的时空统计特征^[131]

对手机数据的实证研究还发现,不同个体的空间运动模式具有高度的相似性,这可以通过统计个体访问区域的空间概率密度分布 $\Phi_a(x,y)$ 来说明—— $\Phi_a(x,y)$ 表示在位置 (x,y) 找到个体 a 的概率。图33a是3个手机用户的空间概率密度分布图形,从中可以看出,无论个体的日常活动范围是几十公里还是上千公里,他们的空间运动模式都是非常相似的——每个个体都在一个以家和工作地点为中心的有限范围内活动,距离家或工作地越近的地点被个体访问的频率就越高,且访问区域具有明显的各向异性。当去除了空间概率密度分布中的各向异性并对距离进行重新标度后,不同个体的空间概率密度分布变得难以区分,如图33b所示。这说明,尽管每个人的出行历史是千变万化的,但人类可能遵循着某种简单的、可预测的空间运动模式。

基于同样的手机通信记录数据,文献[133]进一步研究了人类空间运动模式的可预测性问题。每个个体的空间运动轨迹可以表示为一个由每小时间隔内个体所在地点所构成的序列,通过计算该序列的信息熵(计算方法详见第5节),就可以衡量个体空间运动轨迹的可预测程度。在计算5万名手机用户的访问地点序列的信息熵后,发现熵分布的峰值出现在

0.8左右。这说明在平均意义上, 预测手机用户在下一小时所在地点的不确定性只有2的0.8次方(不到2个地点)。文献[133]还用信息论中的Fano不等式计算了每个用户的访问地点序列能够被准确预测的概率上限 Π^{\max} , 发现群体的 Π^{\max} 分布在0.93处有峰值, 而且方差很小。这说明平均意义上人类空间运动模

式具有93%的可预测性, 而且不同个体的可预测性差别并不大。这种高可预测性同样根源于人类空间运动行为的高度规则性(人每天都在重复着近乎相同的出行序列), 同时说明了开发精确的人类空间运动行为预测模型是有科学依据的。

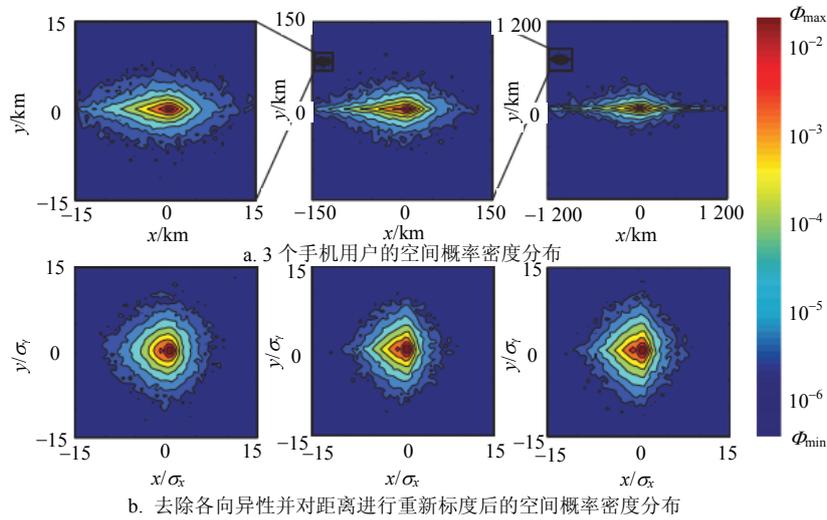


图33 个体空间运动模式的相似性^[131]

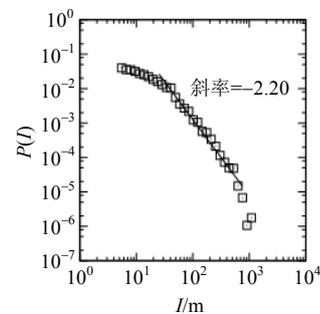
综上, 文献[131]对手机通信记录所进行的实证统计研究表明, 人类除了可能具有近似幂律的移动步长分布和停留时间分布之外, 在空间运动模式上还具有有界性、周期性、规则性和高可预测性等特征。这些结果不仅深化了人们对人类空间运动行为的理解, 也对人类空间运动行为的模型研究留下了许多开放性问题。

3.3 GPS数据

手机虽然能够长期记录个体的移动轨迹, 但其时空分辨率与GPS相比还有差距。首先, 手机定位精度取决于通信基站的覆盖半径, 而民用GPS的定位误差一般在几十米以内; 其次, 如果用户没有拨打电话或收发短信, 他们的交通活动就无法被采集到, 而GPS则可以实时地记录个体的活动轨迹。因此, GPS被普遍认为是一种能更直接地反映人类空间运动模式的数据采集手段。

文献[134]研究了校园、公园、城区和展览馆等小范围场景中携带GPS终端的44名志愿者的空间运动行为。由于GPS终端所记录的轨迹点中, 有些是个体实际停留的地点, 而另一些轨迹点则只是个体移动过程中的途径点。为了识别出GPS轨迹数据中个体的停留地点及移动步长, 文献[134]引入了矩形识别、角度识别和停留时间识别3种不同的方法。矩

形识别是从某个起始轨迹点开始, 将该点与其后续的 n 个轨迹点之间连接一条直线, 如果两者之间的 $n-1$ 个轨迹点都处于以这条直线为中线、以某个阈值为宽度的矩形内, 则认为该直线的两个端点是停留地点; 角度识别是用连续的线段串起所有轨迹点, 依次判断两相邻线段之间的夹角是否大于某个给定的阈值, 如果大于则两线段共有的轨迹点就是一个停留点, 否则认为其只是途经点; 停留时间识别是对轨迹先应用矩形识别, 去除一部分轨迹点后, 再运用角度识别得到的结果。发现在各个场景下群体的移动步长都近似服从幂律分布, 但幂指数却因轨迹识别方法的不同而有所不同, 如图34所示^[134]。



a. 矩形识别方法提取出的移动步长

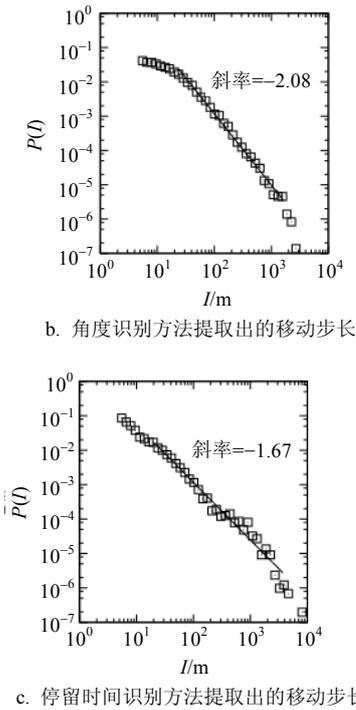


图34 某大学校园中4名学生在一天中的移动步长分布^[134]

文献[134]统计研究还发现人在小范围内运动的停留时间分布也是幂律的，与文献[131]的实证统计结果类似。通过进一步统计群体移动的均方位移(MSD)随时间变化的规律，发现人类的运动具有反常扩散现象——在约30 min以内的时间范围内是超扩散的，而超过30 min后则是亚扩散的，如图35所示。这一结果暗示了时空扩散的局域化特征不仅在人的大范围空间运动行为中存在^[2,131]，在小范围的空间运动中也同样存在。当然，文献[134]提供的数据从样本规模和采样时间都还不够充分，场景也比较局限，更坚实的结果还有待更多数据的分析。

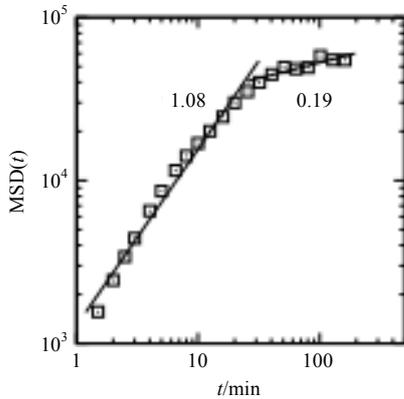


图35 某展览馆内人群的均方位移(MSD)随时间变化的规律^[134]

3.4 单一交通方式出行数据

前述实证研究所分析的数据中并没有区分出行

者所使用的交通方式，所统计的空间运动模式是混合了多种交通出行方式的结果^[135]。针对单一交通方式下出行者的空间运动模式，近年来也发表了一批实证研究成果，包括利用GPS数据研究出租车乘客、私家车驾驶者和航空乘客的出行模式^[136-142]、利用客流记录数据研究地铁和公交乘客的出行模式^[143-144]等。本小节对此方面的研究进行简要介绍。

文献[136]收集了瑞典4个城市中50辆出租车在6个月内行驶轨迹的GPS记录数据，并通过出租车计价器上所记载的乘客上下车时间，将这些轨迹分解为7万余条独立的乘客乘车记录。通过对这些数据进行统计分析，认为出租车乘客的乘车距离服从双段幂律分布，如图36所示^[136]。图中，上方插入图为3~23 km范围内的乘车距离分布，图中虚线为幂指数为-2.5的幂律函数。下方插入图为大于23 km范围内的乘车距离分布，图中虚线为幂指数为-4.5的幂律函数。在3~23 km的范围内幂指数约为-2.5，在大于23 km的范围内幂指数约为-4.6。但图36中的分布曲线在双对数坐标下整体上呈现出向下弯曲的形状，而且尾部的幂指数如此之大，整个分布曲线似乎更适合用指数函数进行拟合。

文献[137]基于更大规模的GPS数据研究了出租车乘客的乘车距离分布特征，所分析的一个数据集包含了北京市万余辆出租车在3个月间的GPS轨迹数据，其中记录的出租车乘客出行共有12 028 929次。通过严格的统计推断和检验，发现出租车乘客的乘车距离很好地服从指数分布，如图37所示^[137]，而不是之前报道的幂律分布。类似的呈指数形式的乘客出行距离分布在对上海市出租车GPS轨迹数据的研究中也被报道^[138]，说明对于出租车乘客出行距离来说，指数分布可能是比幂律分布更为普遍的一种模式。

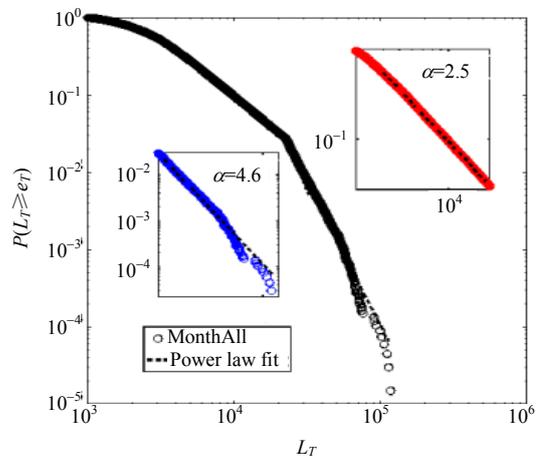


图36 瑞典若干出租车乘客的乘车距离分布^[136]

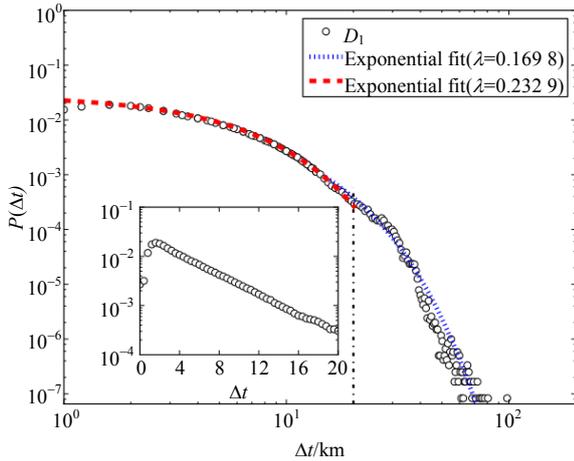
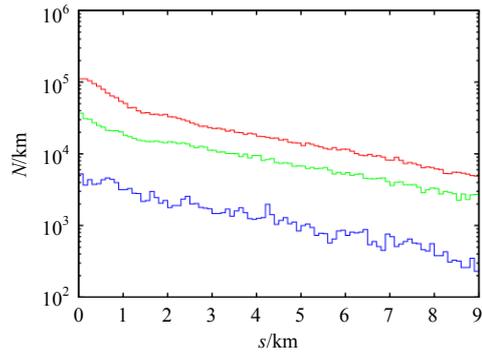


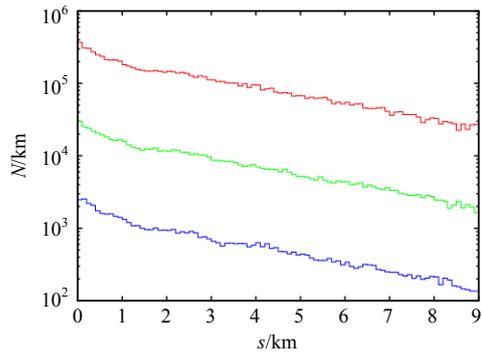
图37 北京市出租车乘客的乘车距离分布^[137]

文献[139-141]通过车载GPS数据对意大利罗马、博洛尼亚、西尼加利亚和佛罗伦萨等城市的私家车驾驶者的出行行为进行了统计研究。通过汽车发动机的运行状态识别,从汽车行驶轨迹中计算出了单次行驶的距离。统计结果显示3个城市的车辆单次行驶里程都近似服从指数分布,如图38所示^[139],而且这一分布并不随观测时间段的不同而变化,说明对于私家车驾驶者群体,行驶距离的指数分布规律是相当普遍和稳定的。此外,还对车辆全天行驶总里程分布(服从指数分布)、车辆全天出行次数分布(服从指数分布)、车辆停留时间分布(服从幂律分布)、各地点访问频率分布(服从幂律分布)等进行了实证统计,详见文献[140-141]。文献[139-141]的这一系列工作是目前为止对私家车驾驶者出行行为所进行的最全面和细致的实证分析。

文献[142]研究了美国国内航班在11天内的飞行记录数据。这一数据集是用飞机上的GPS终端记录的,一共包括205 662次飞行。统计结果显示,美国国内航班的飞行距离在9 000 km以内呈现指数分布,在超过9 000 km的部分有一个幂律尾部。但考虑到美国国内相距最远的两个州之间的距离也不过是9 000 km,超过该距离的飞行不应该被认为是有效的数据(事实上这部分数据只有128条,仅占总记录数的0.06%)。因此可以认为,航班的飞行距离分布也是指数的。需要指出的是,航班的飞行距离分布并不等同于航班乘客的旅行距离分布,因为不同航班的乘客数量并不都相同。但在缺乏详细的乘客航空旅行记录的情况下,用航班的数据研究航空乘客的旅行行为也不失为一种可行的方法。



a. 意大利3个城市的私家车单次行驶里程分布,从上到下的3条曲线依次是罗马、博洛尼亚和西尼加利亚的分布数据



b. 博洛尼亚不同月份的的私家车单次行驶里程分布,从上到下的3条曲线分别统计自2006年6月、9月和10月的数据。为便于比较,6月和10月的数据在纵轴上分别乘了系数10和0.1

图38 意大利3个城市私家车单次行驶里程分布^[139]

文献[143]收集了伦敦地铁200万名乘客在一周内的IC卡刷卡数据,并统计了地铁乘客的乘车距离分布。研究发现地铁乘客的乘车距离分布不是幂律也不是指数的,而是近似负二项分布,如图39a所示^[135],并猜测这种有峰值的分布是由于城市的多中心结构所导致的。图中曲线是负二项分布函数 $f(d) = C_{d+r-1}^{r-1} p^r (1-p)^d$, 其中 $r = 2.61$, $p = 0.0273$ 。文献[144]对石家庄市公交乘客乘车记录所进行的实证研究发现,公交乘客和地铁乘客的乘车距离分布形式非常接近,同样呈现出带有峰值和指数尾部的负二项分布,如图39b所示^[144],图中曲线是负二项分布函数 $f(d) = C_{d+r-1}^{r-1} p^r (1-p)^d$, 其中 $r = 4.96$, $p = 0.51$,插入图是在线性坐标下绘制的相同内容。但石家庄市是一个典型的单中心结构城市,文献[143]中的观点并不能完全解释公交和地铁乘客的这种独特的出行模式。事实上,观察图39中的分布曲线可以看到,两个分布的峰值都出现在5 km附近,而这一距离一般被认为是步行出行的极限距离。与使用出租车、私家车等交通方式的出行者不同,对于使用公交或地铁为主的出行者来说,当出发地和目的地距离很近时(如几百米),他们一般会选择步行而

不是乘车；在步行允许的范围内，距离越远，选择公交或地铁的可能性就越大；而在超过了步行极限距离后，公交乘客的乘车距离分布则会表现出与其他单一交通方式使用者出行距离分布非常类似的规律，即呈现指数分布。也就是说，在距离小于5 km出现的向下弯头很可能来源于公交地铁出行与步行的竞争。这是对公交和地铁乘客出行距离分布模式的一个更为合理的解释。

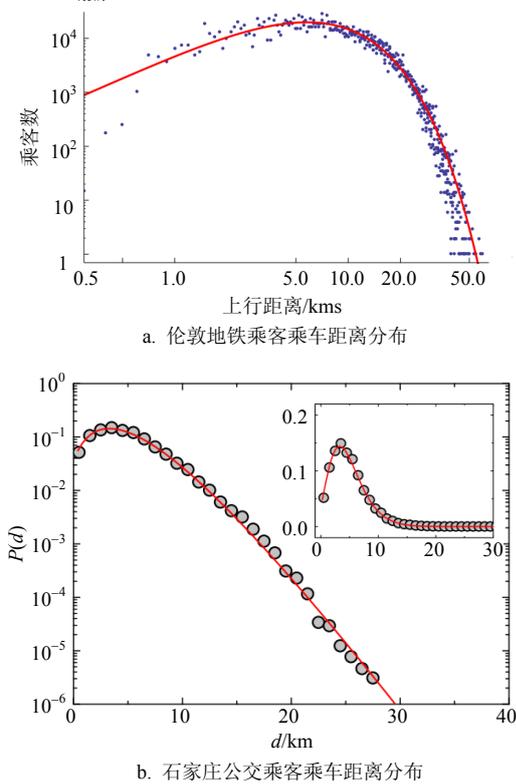


图39 地铁和公交乘客乘车距离分布^[143-144]

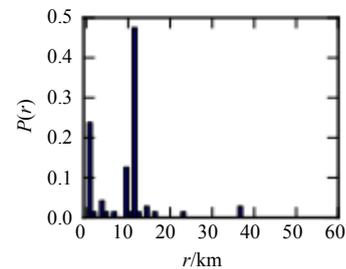
总结以上结果，针对单一交通方式的出行模式，如果不考虑其他选择项的竞争，空间位移分布一般而言服从指数形式。

3.5 个体层面的实证分析

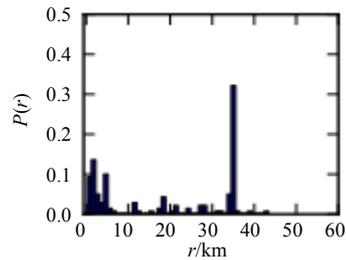
前述介绍的关于人类行为空间特性的实证研究几乎都是在群体水平上进行的，对于人类个体空间运动模式的实证研究目前仍比较缺乏。文献[132]暗示根据在群体水平上统计得到的人类出行距离服从幂律分布，可以推断出群体中的每个个体的出行距离也同样服从幂律分布。但文献[103]已从理论上证明了把一群具有不同一阶矩的泊松个体聚集在一起可以产生群体水平上的幂律分布，也就是说仅根据群体水平上的统计结果并不能推断出个体的行为，这一点在动物的空间运动实证研究中已经获得证实^[145]。为了更清晰地理解人类个体的出行模式，该文献在个体层面上对人类行为的空间特性进行了分

析^[146-147]。

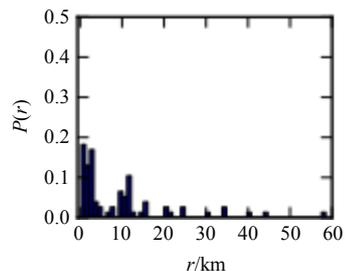
本文的数据来源于瑞士弗劳恩菲尔德市230名志愿者填写的日常出行日志^[148]，该数据集记录了6周时间内志愿者们每次出行的起点、终点、距离等信息，同时还提供了志愿者的年龄、性别、职业等个人信息。对数据集中每个个体的出行距离分布进行统计，发现绝大多数个体的出行距离并不符合幂律分布。特别是发现学生、职员、退休者等不同个体有着完全不同的日常出行模式(见图40a~图40c)：相对于退休者，学生和职员更容易在固定的地点(如家、学校或工作地)之间频繁的移动，将每个个体最频繁的出行称为“支配性出行”(对于学生而言，一般是连接家和学校的行程；对于职员而言，一般是连接家和工作地点的行程)。支配性出行的存在使得个体出行距离呈现有峰值的分布，除非个体的支配性出行距离很短，否则出行距离分布很难具有幂律特性。有趣的是，在群体水平上对数据集的出行距离分布进行统计后发现，群体的出行距离近似服从一个带有指数截断的幂律分布，如图40d~图40f所示，与手机用户所进行的实证研究结果^[131]非常接近。这一研究结果反驳了从群体行为模式中推断出个体行为模式的观点，同时深化了人们对于人类个体空间运动行为的理解。



a. 某学生的出行距离分布



b. 某职员的出行距离分布



c. 某退休者的出行距离分布

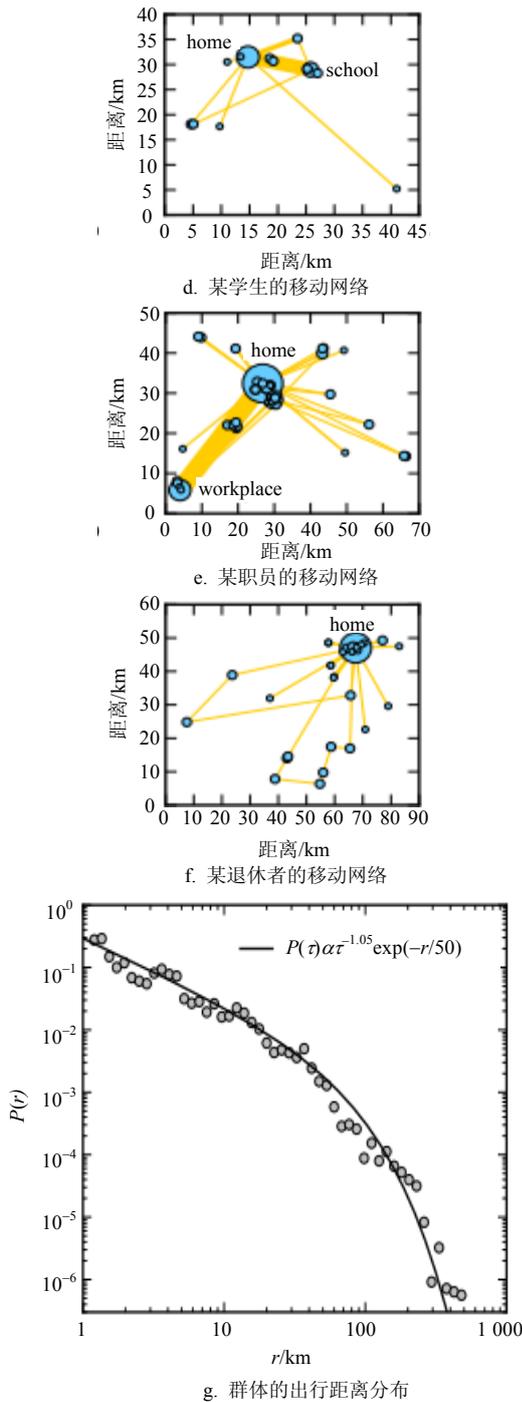


图40 个体和群体的出行模式^[147]

3.6 小结与讨论

前述实证研究揭示了人类空间运动行为中的大量统计特性, 主要包括: 1) 在群体水平上, 混合交通方式使用者的移动步长分布服从幂律或截尾幂律分布, 而单一交通方式使用者的移动步长分布则服从指数或近似指数分布; 2) 在个体水平上, 人类的移动步长分布呈现不规则的多样化特征, 并不服从某种特定的分布形式; 3) 人类的运动范围具有局域化的特点, 无论是轨迹的均方位移、回转半径

还是时空概率密度函数, 增长速度都慢于具有相同参数的Lévy飞行模型; 4) 人类的运动模式具有时空规则性, 如以天为周期频繁返回少数几个地点, 不同个体的空间概率密度分布定性上高度相似, 移动轨迹具有高可预测性。

整体看来, 目前人类空间运动行为的实证研究已经取得了长足的进展, 相关研究成果不断深化人类对自身空间运动行为的理解, 但仍有大量的研究工作有待开展。首先, 现有的研究普遍比较关注移动步长分布、扩散速度等偏物理特征的统计特征, 而对于人类空间运动轨迹的规则性、周期性、相似性等与行为预测问题密切相关的统计特征的研究还很不充分, 这方面的工作应该是下一阶段研究的重点。其次, 目前的实证研究大多是孤立地挖掘人类空间运动行为的统计特征, 而对于社会化因素对人类行为的影响考虑不足。人类独特的空间运动模式根源于人类高度的社会性, 这使得人比随机扩散的物理粒子具有更为复杂的空间运动模式^[149]。要真正理解人类出行模式的产生根源, 就不能脱离对人类社会活动背景的考虑, 但这方面定量的实证研究仍相当欠缺。最后, 在数据来源方面, 已有的实证研究或是根据钞票流通、手机漫游等数据间接获得的, 或是局限在某一种交通方式的数据上, 缺少对人类行为空间分布特性直接的、全面的观察, 而且根据不同来源和分辨率的数据所得到的统计结果之间存在着一定的争议性, 需要更进一步的实证研究来进行判定。总之, 由于人类社会活动和自身的复杂性, 人们距离真正理解并预测人类的运动行为还很远, 更清晰和令人信服的图景仍需要更多和更深入的实证分析。

4 人类行为空间特性建模

前述对不同数据所进行的统计实证发现人类空间运动模式中存在许多共同的统计特性, 暗示着可能存在支配人类空间运动行为的普适性机制。通过建立模型描述人类空间运动行为, 有助于揭示各种现象和标度律形成的底层机制, 深化人们对人类空间运动行为的理解。本节对人类空间运动模型研究方面的一些主要进展进行介绍。

4.1 偏好返回模型

文献[2]用具有幂律步长分布和幂律停留时间分布的连续时间随机游走(CTRW)再现了钞票空间运动的低速扩散特征。但是, 从对手机数据所进行的统计实证研究^[131-132]中发现人类的运动模式中存在一些标度异常现象, 具体包括: 1) 个体访问

地点的数量随时间呈幂律增长，幂指数低于具有相同步长和停留时间分布的CTRW模型所预测的结果；2) 个体访问的地点中，各点被访问的频率呈幂律分布，而在CTRW模型中各点的被访问概率是相等的；3) 个体访问点的均方位移(MSD)随时间增长的速度比对数增长还要慢。这些标度异常现象无法用CTRW模型进行解释。

文献[132]认为人类同时具有探索未知地点和返回之前熟悉地点的倾向，这两种机制的共同作用可能是导致前述几种标度异常现象的根源，为此，建立了一个新的人类空间运动模型来解释个体的空间运动行为。该模型的基本规则是：在初始时刻 $t=0$ 时，个体在事先给定的若干初始地点中的某一点上停

留。在停留 Δt 的时间后(停留时间服从分布 $P(\Delta t) = |\Delta t|^{-1-\beta}$)，个体移动到下一个地点，并不断重复这一过程。在个体每次选择下一步移动地点时，有两种可能：

1) 以 $P_{\text{new}} = \rho S^{-\gamma}$ 的概率移动到一个之前未访问过的地点，其中 S 是当前已访问过的地点总数， ρ 和 γ 是两个参数，具体取值可根据实证数据估计。移动的步长 Δr 服从分布 $P(\Delta r) = |\Delta r|^{-1-\alpha}$ ，移动方向是随机选择的。当个体到达新地点后，令当前已访问过的地点总数 $S = S + 1$ 。

2) 以 $P_{\text{ret}} = 1 - \rho S^{-\gamma}$ 的概率返回之前已访问过的地点。在之前访问过的 S 个地点中，地点 i 被选择的概率 Π_i 等于该地点已被访问过的频率 f_i 。

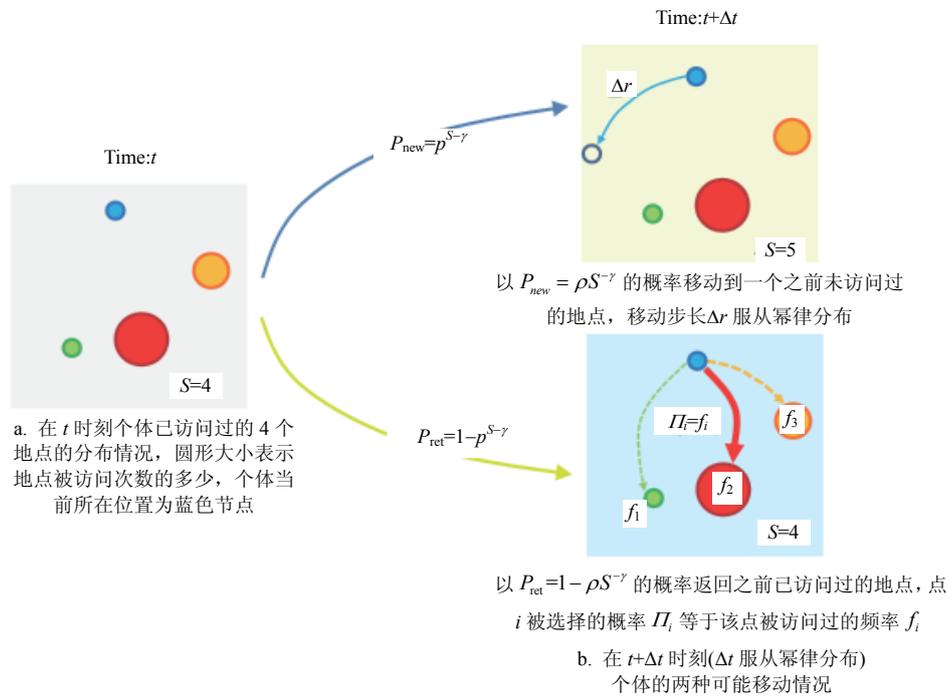


图41 偏好返回模型规则示意图^[147]

偏好返回模型规则示意图如图41所示^[132]，该模型与CTRW模型的主要区别是引入了探索新地点和偏好返回两种机制，其中 ρ 和 γ 是两个主要参数，它们决定了个体探索新地点的倾向程度。尽管 ρ 对于建立模型是重要的，但根据模型的预测结果，各种标度现象的指数却与 ρ 的具体取值无关。从探索新点的概率随访问点总数幂律下降的关系 $P_{\text{new}} \sim S^{-\gamma}$ 出发，可以解析地得到标度指数 γ 与其他3种异常标度指数之间的定量关系：首先是访问地点数量的低速增长现象，模型中访问地点随时间增长的关系为 $S(t) \sim t^{\beta/(1+\gamma)}$ ，增速低于CTRW模型预测的 $S(t) \sim t^\beta$ 。其次是地点访问次数分布的无标度性，模型中访问次数第 k 高的点被访问的频率 $f_k \sim$

$k^{-1-\gamma}$ 。而在CTRW模型中各点被访问的频率 $f_k \sim 1$ 。最后是MSD的超慢增长现象，模型中个体的MSD会先以对数速度增长，而后增速逐渐减缓，直到达到近似稳定的状态。此外，作者还定量分析了回转半径的分布与步长分布之间的关系等，这些解析结果都通过实证数据得到了验证。限于篇幅，本文不介绍模型解析细节，有兴趣的读者请参考文献[132]及其补充材料。

偏好返回模型解释了人类空间运动模式中各种标度异常现象，并且能够用指数 α 、 β 、 γ 定量预测其他标度现象中的相关指数，为人类空间运动行为建模研究提供了一个基础的、可扩展的概念框架。但该模型也有一些不足之处。首先，文献[132]从异

常标度现象(1)推出了个体探索新地点的趋势随时间幂律下降, 但反过来又用后者解析得到前者的标度指数, 实际上并未从本质上揭示这些标度现象产生的原因^[150]。其次, 模型所用的个体步长分布幂律指数 α 是根据群体的数据估计的, 相当于假定个体与群体具有相同的步长分布。但是在该小组之前的研究工作中, 基于同样的实证数据已明确得出个体与群体步长分布指数存在巨大差异的结论^[131]。尽管指数 α 只对MSD的预测结果产生影响, 但如果个体与群体步长分布指数的确存在差异, 那么模型对MSD预测结果就值得商榷。最后, 正如本文所指出的, 模型适用于再现人类运动行为中的长期的时空标度特性, 但不适用于预测个体短期内的访问地点序列。因为人类个体空间运动行为的可预测性根源于个体活动的周期性特征和访问地点的时空关联性^[133], 而这些因素并未包含在模型中。当然, 如果只关心个体空间运动模式的渐进特性, 这些细节的缺乏并不影响模型可用性。

4.2 基于层次性交通系统的人类运动模型

人类的出行行为受到所处的交通系统的影响和制约, 而人的活动同时也影响着交通系统的发展, 两者之间呈现出一种共演化的关系^[134,136]。一般来说, 实际的交通系统常常是以城市作为依托, 如在一个行政区中(如一个省), 常常有一个最大的中心城市(如省会); 这个中心城市周围一般环绕着若干个较小的城市(如县城); 而这些小城市周围也被更小的居民点(如乡镇)所包围。由于大城市之间的交通便利性, 人们进行日常的长途旅行时, 经常通过大的交通中心进行中转。在这种情况下, 实际的交通系统往往具有明显的层次性。

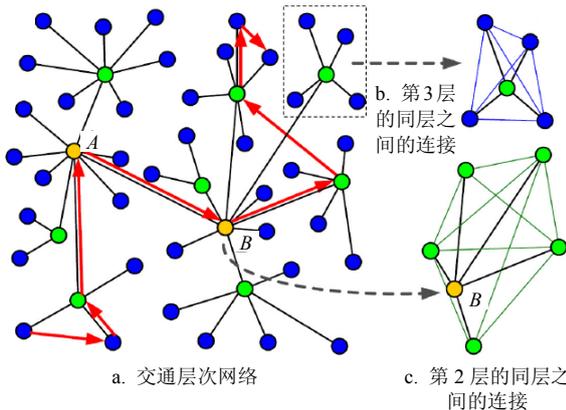
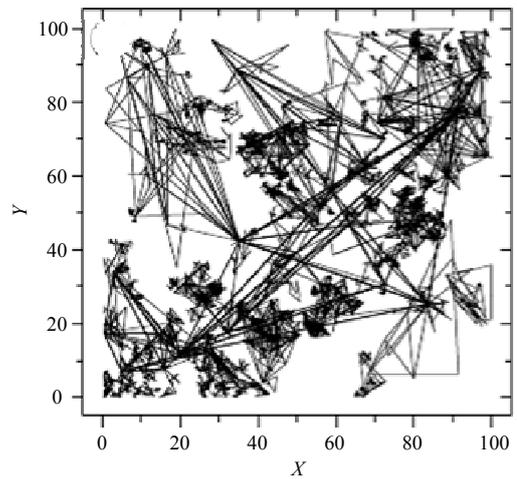


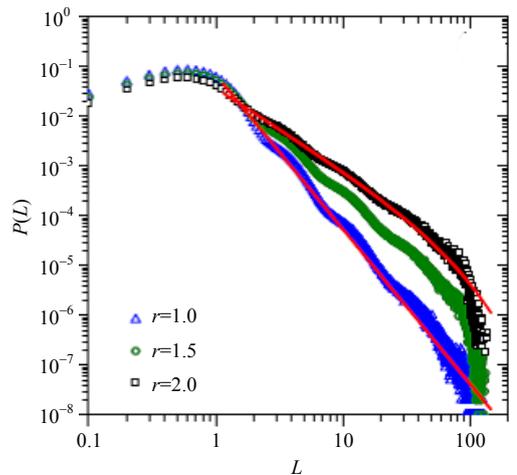
图42 一个 $K=2, N=3, M=5$ 的层次性网络的示意图^[121]

文献[121]考虑了这种层次性的交通系统对人类出行行为的影响, 并建立了理论模型[]]。在该模型中, 所有城市被区分为 N 个层次(此处把所有的节点称作“城市”)。定义 K 为第1层(最高层)的城市数目, M

为下一层与当前层次的城市数目之比, 这样, 一共有 KM^{N-1} 个城市(包含 K 个第1层的城市, $K(M-1)$ 个第2层城市, \dots , $KM^{N-2}(M-1)$ 个第 n 层($n \geq 2$)城市), 它们均匀地随机散布在一个 $S \times S$ 的连续性平面空间内。每一个第 n 层($1 < n \leq N$)城市都和离它最近的一个较高级的城市连接; 如果两个第 n 层城市都和同一个较高级的城市相连, 那么这两个城市也相连; 第1层的 K 个城市相互之间完全连接, 如图42所示^[121], 图中箭头表示个体在8个时步内的运动。假设一个个体在这样的层次网络上进行随机行走, 在每一个时步, 个体从当前所在城市处, 随机选择一个相连的邻居城市并跳跃过去作为下一时步的起始位置。由于现实交通系统中的中心城市对于人们的旅行常常有着更大的吸引力, 因此引入了城市的权重, $w_n = r^{N-n}$, 其中 n 是该城市所在层, $r \geq 1$ 是该模型的一个主要参数。当个体在相连城市中选择下一步的位置时, 城市被选择的概率正比于其权重。



a. 当 $r=2.0$ 时模拟5000时步所得到的运动轨迹



b. r 分别为1.0, 1.5和2.0时的行程分布。红色线分别为 $f(L) = 0.07L^{-3.11}e^{-0.0L}$ 对 $r=1.0$ 和 $f(L) = 0.038L^{-1.63}e^{-0.015L}$ 对 $r=2.0$ 时的拟合结果

图43 层次性交通网络模型模拟结果^[121]

该模型动力学机制的核心是在含权地理网络上的随机行走。当 $r=1$ 时,个体的行走时间趋向无穷长时,历经各条边的概率完全相同,此时个体的行程分布将等同于在该网络中各条边的长度分布。当 $r>1$ 时,访问较高层次的城市概率较高,相应长程运动的概率也较高。个体在 $r=2.0$ 时获得的5 000时步内的运动轨迹如图43a所示,存在较多的长程运动。图44b展示了不同 r 值时在5层网络上所获得的行程分布,这些分布可以很好地用一个带有指数截断的幂函数 $P(L) = cL^{-\beta}e^{-\lambda L}$ 进行拟合。当 r 从1.0变化到2.0时,通过最大似然估计方法拟合得到的幂指数 β 从3.11变化到1.63,包含了实证结果^[2]所观察到的范围。 M 值对行程分布也有影响:当 $r=1.0$ 时, M 对行程分布没有明显影响;而当 $r>1.0$ 时,分布的幂指数 β 随 M 的增大而增大,相应的长程运动的概率也随之降低。其他参数对行程分布的影响并不明显。

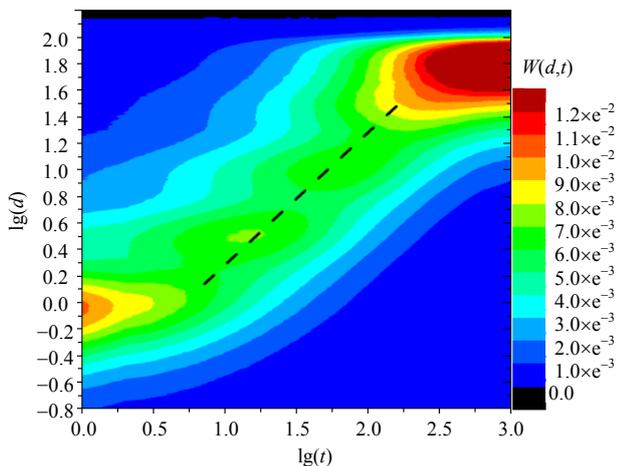


图44 当 $r=2.0$, $K=9$, $N=5$, $S=100$ 时模拟所得的概率密度 $W(d,t)$ 的变化斑图^[121]

该模型可以通过个体在相邻层之间运动的概率流密度平衡的方法进行解析,其方法见参考文献^[121]。最终的解析结果为:

$$P(L) = L^{-(3-4\log_M r)} \quad (11)$$

这一结果同数值模拟结果基本吻合。为了进一步比较该模型与实际观测结果,在该模型中计算了概率密度 $W(d,t)$ 。 $W(d,t)$ 表示经过时间 t 后,个体距离出发点的距离为 d 的概率。这一模型获得的 $W(d,t)$ 的变化斑图如图44所示,当 $r=2.0$ 时其峰值增长曲线的指数也同样在1.0附近,与实证结果^[2]一致(参考图31),说明该模型也成功地描述了实际的人类远距离出行活动中所具有的局域性特性。

综上所述,这一模型基于对实际交通系统所具有的层次性的描绘,通过最简单的随机行走假设,

再现了多种实证统计所观察到的人类空间运动模式。这一模型所主要针对的是人类的长程出行活动,揭示出交通系统自身所具有的层次性可以直接导致人类出行活动中行程分布的标度特性,有效地解释了在人类长程活动中的若干统计特性。文献^[136]中也表达了类似的观点,认为城市街道的结构是造成行程统计规律的主要原因,即便每一个乘客的出发点和目的点都是随机给定的,也会观察到类似的行程分布。

4.3 信息熵优化模型

移动步长分布是刻画人类空间运动模式的一个关键统计特征。实证研究显示,人类个体移动的步长分布近似服从指数 α 为1.2的幂律分布^[131]。而早期对几种动物空间运动行为所进行的实证统计^[151]显示出,动物的移动步长也具有幂律特性,但指数 α 介于1.7~2.4之间。为解释导致人与动物空间运动行为中步长分布标度指数差异的原因,文献^[152]构建了一个带有返家机制的Lévy飞行模型。该模型的基本规则是:假定个体在 $L \times L$ 的晶格中,从中心点(即“家”)出发,如果当前所在点是中心点,则进行一次步长服从 $P(l) = l^{-\alpha}$ 的Lévy飞行;否则,以 $1-r$ 的概率进行步长服从 $P(l) = l^{-\alpha}$ 的Lévy飞行,以 r 的概率返回中心点,直到消耗的总能量达到预先给定的总限制值 W 。

如果认为个体在游走的过程中希望访问更多不同的地点以获取尽可能多样的信息(信息的多样化程度可以用访问地点的熵值 $E = -\sum_x^{L \times L} (p_x \ln p_x)$ 度量,其中 p_x 是各点被从中心点出发的移动直接访问的频率),那么模型就可以转化为“在 W 和 r 给定的情况下, α 取何值时,个体所能获取的信息熵最大”这样一个优化问题。

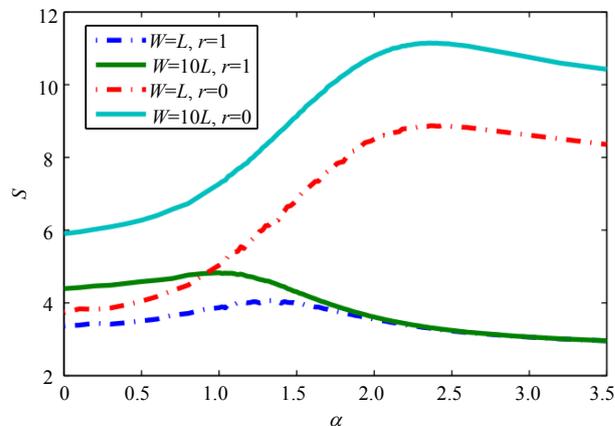


图45 熵值 S 随优化指数 α 变化关系的仿真结果^[152]。仿真所用晶格尺寸 $L=10\ 000$

如图45所示, 图中各曲线分别代表在 W 、 r 取不同组合情况下的结果。模型的仿真结果表明, 在 r 取1时(即个体更频繁地返家), α 取值在1附近时个体所能获取的信息熵最大; 而 r 取0时(即个体几乎不返家), α 取值在2~2.5之间时个体所能获取的信息熵最大。这一结果说明, 返家强度的差异可能是导致人与动物移动步长分布标度指数差异的主要原因——因为相比动物, 人更多拥有一个需要“常回家看看”的地方。该模型从优化角度探讨步长分布标度指数问题, 在人类与动物的空间运动行为研究之间架起了一座桥梁。人与动物在空间运动模式上具有许多相似之处, 借鉴动物空间运动研究领域的思路和方法有助于提高对人类空间运动行为的认识。模型存在的一个缺陷是, 人们很难理解“为什么动物希望自己访问各点的信息熵最大”, 实际上, 即便是人的活动也不是冲着优化信息熵的目的展开。

4.4 周期性随机游走模型

周期性和规则性是人类个体日常空间运动模型的关键特征^[131]。这些特征不仅是人类个体的空间运动行为具有高可预测性的根源, 也直接导致了城市交通中的“律动”现象^[153]。在多个不同城市所进行的居民出行日志调查结果^[153]显示, 城市居民的日常出行行为模型具有高度的相似性。大多数居民在工作日都遵循类似的序列: 早上从家到工作地点(对于学生则是上学地点), 下午下班后进行购物、娱乐、聚会等活动(这些活动统称为“休闲”活动), 最后再返回家中。尽管在这种“家→工作→休闲→家”的出行序列中, 个体每天进行休闲出行的次数和目的地有所差异, 但是通勤和返家两次出行在工作日中几乎是不变的。类似的规则性日常出行模式在文献^[154]对手机用户所进行的实证研究中也报道。

受上述实证研究结果的启发, 文献^[155]构建了一个模拟城市居民个体日常出行活动的周期性随机游走模型。这个模型的基本假设包括: 1) 每个个体具有固定且唯一的家和工作地点, 两地之间的距离(称为通勤距离)为 $2c$; 2) 个体每天在家和工作地停留的总时间具有固定的下限 t_m ; 3) 个体在出行中具有固定的平均速度 v ; 4) 个体每天休闲出行的次数固定为1次。

根据前述假设, 该模型将个体每天的交通活动描述为3步移动: 1) 从家移动到工作地点; 2) 从工作地点移动到一个随机选择的休闲地点, 但必须保证能够及时返家; 3) 从休闲地点移动到家。由于个体每天可用于休闲活动的时间是有限的, 且个体有

固定的出行速度, 容易计算出个体每天可用于休闲出行(包括返家)的最大距离不会超过:

$$2a = v(24 - t_m) - 2c \tag{12}$$

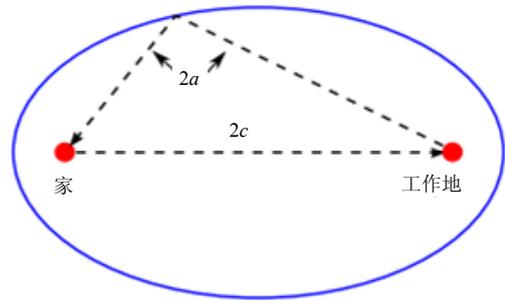


图46 个体日常出行的椭圆范围示意图^[155]

利用几何知识, 可知个体的空间运动范围处于一个椭圆形之内, 如图46所示^[155], 图中, 椭圆的两个焦点分别位于个体的家和地点所在的位置, 焦距为 $2c$ 。椭圆的长轴长为 $2a$, 它的含义是个体每天在休闲出行中可能行走的最长距离。人类出行活动中近似椭圆形的出行范围不仅在对城市居民出行日志所进行的实证研究^[153]中被观察到, 在对手机用户所进行的实证研究中也发现了类似的现象(见图33)。

回转半径是反映人类个体出行空间有界性的一个关键物理量^[131], 对模型产生轨迹的回转半径进行了解析。当模型中个体出行的次数足够多时, 可以解析地得到其轨迹的回转半径是一个常数:

$$r_g = \sqrt{\frac{7c^2 + 2a^2}{12}} \tag{13}$$

从这一结果中可知, 个体的回转半径与两个因素有关: 一是个体家到工作地点的距离; 另一个是个体每天可用于休闲出行的距离(取决于休闲时间和出行速度)。这说明在通勤距离一定的条件下, 休闲时间越多、交通能力越强的个体, 其出行轨迹的回转半径就越大。

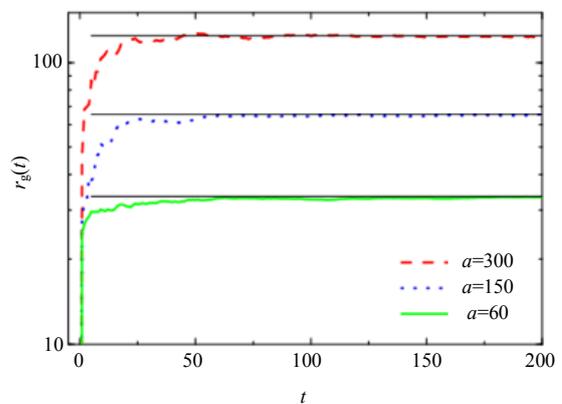


图47 模型中个体回转半径随时间变化规律的仿真结果^[155]

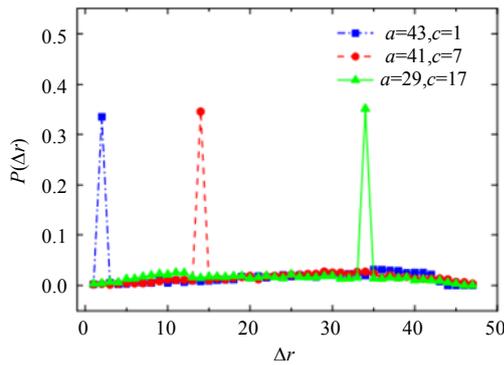


图48 具有相同 r_g 值($r_g = 15.12$)但不同 c 值的3个个体出行步长分布的仿真结果^[155]

解析结果显示个体的回转半径 r_g 是一个常数,而实证统计却发现 r_g 会随时间先快速增长,而后才趋于稳定(见图32a)。为解释二者之间的差异,本文对模型所生成轨迹的 r_g 进行仿真分析,结果如图47所示。图中,直线是相对应的解析结果。从图中可以看出, r_g 会很快增长到一个稳定值,该稳定值与得到的解析结果一致,而在 r_g 达到稳定值之前会有一个快速增长的过程,与实证统计^[131]中观察到的现象相符。

此外,从式(13)可知具有相同 r_g 值的个体的 c 值可能是不同的。由于个体每天有1/3的出行发生在家和工作地这两点之间,使得模型产生的步长分布在个体的家和工作地之间的距离上出现峰值,如图48所示^[155],3个个体的步长分布分别在各自的 $2c$ 处出现峰值。这与对个体所进行的实证研究结果^[147]是符合的。

当然,该模型是一个极端简化的人类空间运动模型,它不能被期望再现人类空间运动的全部统计特征。但模型中描述的个体以天为周期重复“家→工作→休闲→家”的出行序列,可能抓住了城市中大部分居民日常出行的基本模式。这种模式根源于人类出行行为所固有的社会性和目的性,也是人与物理粒子在空间运动行为上的本质差别^[149]。认识这一点对于建立实用化的人类空间运动行为预测模型有重要意义。

4.5 麦克斯韦·玻尔兹曼统计解释

实证研究^[146-147]已经揭示,一般而言,幂律的行程分布在个体层面上是不存在的,那么群体层面上的出行距离幂律分布^[2,131]就不能看作是展现标度关系的若干个体的简单叠加。但这种群体上的标度律从何而来仍不清楚,需要给出合理的解释。城市人群的运动与气体分子的运动有很多相似之处:在个体层面上的移动没有任何规则,但在群体层面上却

涌现出了规则的模式。受此启发,用统计力学中经典的麦克斯韦·玻尔兹曼(Maxwell-Boltzmann)模型来解释城市人群出行距离幂律分布的现象。

将城市中出行的人类个体看做一个近独立粒子系统中运动的粒子。人在城市中出行需要耗费成本,类似于气体分子的移动需要消耗能量。假设群体中包含的个体数量为 N ,总成本耗费值为 C ,且个体的出行行为是相互独立的,则群体成本分布的微观状态数为:

$$\Omega = \frac{N!}{\prod_i n_i!} \quad (14)$$

满足约束条件 $\sum_i n_i = N$, $\sum_i n_i c_i = C$ 。其中 n_i 是出行成本为 c_i 的个体数量,且成本相同的出行被认为是“不可分辨的全同粒子”。

根据最大熵原理,系统最可能出现的成本分布是微观状态数 Ω 最多的分布(最可几分布),用拉格朗日乘法求解式(14),可得到该系统的最可几成本分布为:

$$P(c) \propto \exp(-c/\bar{c}) \quad (15)$$

式中, $\bar{c} = C/N$ 是个体耗费的平均成本。

一般认为出行成本是由出行时间和货币费用两部分组成的,可以表示为时间 t 和费用 m 的加权和形式^[156]: $c \approx \eta t + \mu m$ 。其中,货币费用通常与出行距离 r 具有线性关系^[156]: $m \approx \nu r$ 。而对于混合交通方式的出行者,出行时间与距离则具有对数关系^[157]: $t \approx \phi \ln r + \psi$ 。这种对数关系源于不同出行距离采用交通方式的速度差异,如人们在进行几百米的出行时往往是步行或骑自行车,而几十公里时就要使用公交、小汽车等交通方式,当几千公里时就要乘坐更快速的火车或飞机了,这使得出行时间和距离之间呈现非线性关系。将式(15)中的变量 c 替换为出行距离 r 后,可以得到一个带有指数截断的幂律分布为:

$$P(r) \propto (\beta/r + 1/\kappa) r^{-\beta} \exp(-r/\kappa) \quad (16)$$

式中, $\beta = \eta\phi/\bar{c}$; $\kappa = \bar{c}/\mu\nu$ 。这一结果与文献^[131,147]的实证结果很好地吻合。

该模型不仅能够解释混合交通方式使用者群体出行距离的幂律分布,还可以解释单一交通方式使用者群体出行距离的指数分布现象。这是因为当群体采用单一交通方式时,可以认为出行速度是稳定的,故出行时间和距离之间具有近似线性的关系,由此可以推出群体的出行距离分布为 $P(r) \propto \exp(-ar)$ 。这一结果覆盖了4.4节中大多数针对单一交通方式使用者群体的出行距离分布的实证

研究结果^[137-144]。

4.6 小结与讨论

除前述介绍的几种典型模型之外, 在人类空间运动建模方面的工作还包括文献[158]提出的自相似最小作用力游走模型, 以及文献[159]提出的考虑记忆效应和距离影响的空间运动模型等。限于篇幅, 本文不再对这些模型的细节进行介绍。

总体看来, 现有模型从不同角度对人类空间运动行为中多种统计特性的产生机制进行了有意义的探讨, 深化了人们对于自身空间运动行为的理解。但是, 人类行为本身是高度复杂的, 目前还没有一种模型能够解释人类空间运动的所有统计特性, 在人类独特的空间运动模式背后或许还有更为本质的机制尚未被触及, 仍有待于研究者进行更为深入的探索。

另外, 现有模型大多是通过唯象机制来解释各种标度特性的可能起源, 模型往往过于简化, 很难直接运用于出行行为预测; 而交通领域传统上使用的出行预测4阶段模型^[129]又过于宏观, 对人类空间运动行为特征的影响考虑不足, 所需的基础调查数据量大且模型预测精度和时效性都较差。如何吸取人类动力学的丰富研究成果来发展传统的交通预测理论与方法, 是人类动力学和交通领域的研究者都应该关注的一个重要问题。最近, 文献[160]报道根据人类空间运动行为特征构造的一个无参数的“辐射”模型, 该模型预测区域间人口流动性的精度比交通领域中传统的引力模型提高了很多, 而且不需要进行大规模的交通调查即可实施预测, 这是非常有吸引力的。总体来说, 这方面的研究工作才刚刚起步, 研究空白特别多, 尚有大量的工作需要深入开展。

5 应用研究概述

如前所述, 人类行为无论从时间方面还是空间方面都展现出和以前泊松假设所不同的特性, 包括间隔时间和行程距离的幂律分布。这些对人类行为的新认识可能对很多应用方向产生深远的影响。首先, 对人类行为时间特性的研究发现, 人类行为时间间隔的幂律分布会明显减缓传播速度。群体层面上时间的异质性对传播速度的影响非常大, 相比之下, 个体层面上的异质性对传播影响较小。其次, 对人类行为空间特性的研究发现人类出行中的长程连边可以加快病毒的传播, 短程连边对局域传播起作用, 个体的出行率、返回率的改变会诱导传播过

程中的相变。最后, 对人类行为的分析可以更好地帮助对用户行为进行预测, 并提高信息推荐的精确性。

5.1 人类行为时间特性对于传播的影响

本文所叙及的研究, 不仅帮助人们更好地了解自己的行为特征, 而且对很多应用场景都有启发作用, 其中讨论最多的应属人类动力学特征对于疾病或信息在人群中传播的影响。

虽然从1840年起, 人们就开始尝试运用简单的数学模型来刻画和预测流行病的扩散, 然而直到二十世纪的四、五十年代, 以微分方程为主的模型才开始受到重视, 微分方程也进而成为研究流行病的主流方法。其中, 最经典、最具影响力的是SIS模型和SIR模型。SIR模型是1920年由Reed和Frost提出的, 该模型将人群分为3组: 未患病但易感的人(susceptible)、患病且具备传播能力的人(infectious)、治愈并终身免疫的人(recovered)。麻疹、腮腺炎就是典型的这类流行病。如果患者必将走向死亡, 如艾滋病, 则也可以用SIR模型刻画, 只是此时的R就不再是recovered, 而是removed。假设一个患者总以几率 β 感染易感者, 且以几率 γ 康复, 那么这3类人的演化机制满足:

$$\frac{ds}{dt} = -\beta is, \quad \frac{di}{dt} = \beta is - \gamma i, \quad \frac{dr}{dt} = \gamma i \quad (17)$$

式中, i 、 s 和 r 分别代表患者、易感者和康复者所占的比例。

但对于大多数疾病来说, 没有终身免疫的特点, 康复后又回到可以被感染的状态。对于这种性质的疾病, 可用类似的微分方程刻画:

$$\frac{ds}{dt} = -\beta is + \gamma i, \quad \frac{di}{dt} = \beta is - \gamma i \quad (18)$$

这就是经典的SIS模型。

此后, 针对不同的疾病的特点, 人们相对于经典模型提出了丰富的改进, 如对于短期免疫的疾病, 人们更倾向于用SIRS模型去描述, 针对存在潜伏期的疾病, 文献[161-162]提出了SIER模型和SIES模型。

疾病与流言的传播都是以人为载体的, 仅考虑疾病性质而剥离了人的行为特征去研究传播过程会与实际有较大的偏差。经典的疾病传播模型中存在一些和人类真实行为时间上的阵发性、记忆性和异质活跃性不一致的假设: 1) 人活动的时间间隔相同, 即所有人都均匀地在每个时间步活动一次, 这里的“活动”是指接触其他人或自己康复; 2) 所有人活动的频率在群体中是无差别的, 即每个个体的

活跃性是一样。

为揭示人类活动的阵发性对病毒传播的影响,文献[11]用人们查收E-mail的两组数据(3 188个用户之间发送的129 135封邮件和1 729 165个用户之间发送的39 046 030封邮件),根据真实时间间隔,模拟了邮件通信网络上计算机病毒的传播情况。以天和小时为单位的统计结果都表明人类行为的时间间隔满足幂律分布这一事实,会极大程度上减慢病毒的传播。

图49是在实证数据的网络上根据真实时间间隔,模拟网络上计算机病毒的传播过程,表现的是每天新被感染的人数随天数的变化情况。其中主图是以天为单位的统计结果,而子图是以小时为单位的统计结果。实线为模拟结果,虚线为基于泊松过程假设所得到的理论预测值,点划线表现的是指数衰减的过程。可以看出,相比于原先基于泊松过程假设的理论预测,传播的真实过程更接近于指数衰减。即人类实际行为的阵发性会导致传播过程的明显变慢。

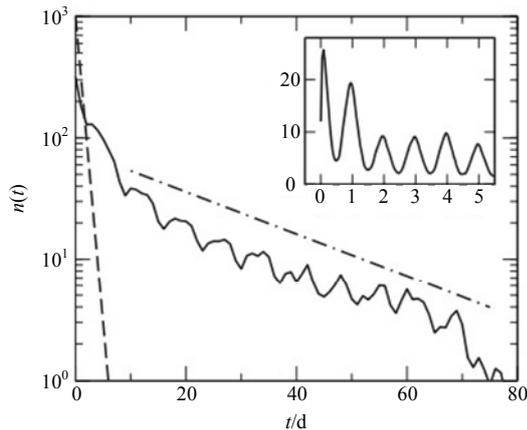
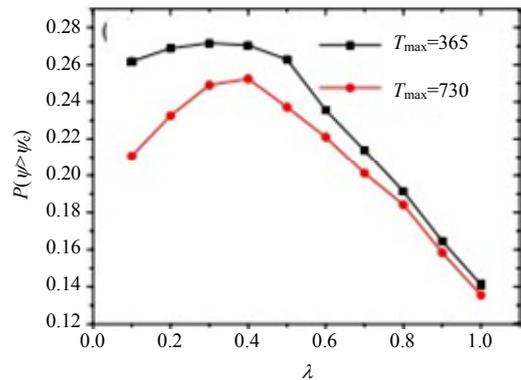


图49 新被感染人数随天数变化情况^[11]

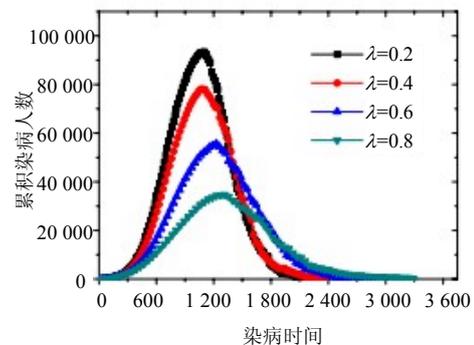
图50a反映了病毒的存活率随时间异质性的变化情况, λ 越大,时间序列的异质性就越小。从图中可以看出,当 $\lambda > 0.4$ 时,病毒的存活概率明显随着 λ 的增大而减小。图50b是染病人数随着时间的变化,不同曲线对应不同 λ 的值,说明 λ 的增大使疾病存在时间延长,但是峰值降低。

有了实证的启发,近年来很多工作围绕阵发性对传播的影响给出了一些定量地理论分析。文献[163]运用SIR模型通过模拟,讨论了时间分布的阵发性对于病毒存活概率和存活时间的影响。考虑了等待时间对传播的影响,并假设等待时间是满足幂律分布的,即 $p(t_w) \propto t_w^{-(1+\lambda)}$, 其中幂指数在1~2之间,即 $\lambda \in (0,1)$ 。图50为仿真结果,反映了时间分

布的幂指数对于存活概率和存活时间的影响,其中 T_{\max} 为模型生成等待时间序列最大值,模拟结果显示:时间间隔分布的异质性越小,病毒存活概率越小,存活时间却越长。



a. 病毒的存活率随时间异质性变化



b. 染病人数随着时间的变化

图50 时间分布的阵发性对病毒存活概率和存活时间的影响^[163]

关注人类行为的阵发性对传播速度的影响,文献[164]在树状网络上用SI模型给出了模拟及相应的理论分析结果。如前面第2节所介绍,大量的实证表明,人类个体活动的时间间隔或回复时间间隔也可以用幂律函数刻画。在该工作中假设系统中的每个个体的活动时间间隔都满足幂律分布,且相互独立,形式如下:

$$P(\tau) \propto \tau^{-\alpha}, \quad \alpha \in (2,3) \quad (19)$$

由文献[165]知,个体从收到信息后的 Δ 时间长度内无法回复的概率为:

$$g(\Delta) = \frac{1}{\langle \tau \rangle} \int_{\Delta}^{\infty} P(\tau) d\tau \quad (20)$$

可知回复时间的间隔分布为:

$$g(\Delta) \propto \Delta^{-(\alpha-1)} \quad (21)$$

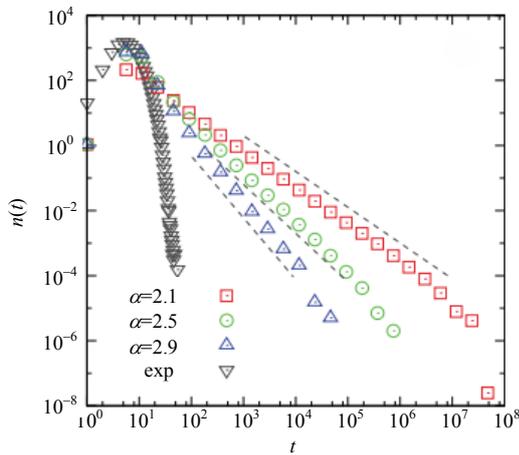
且每个时间步新感染的人数为:

$$n(t) \propto t^{-(\alpha-1)} \quad (22)$$

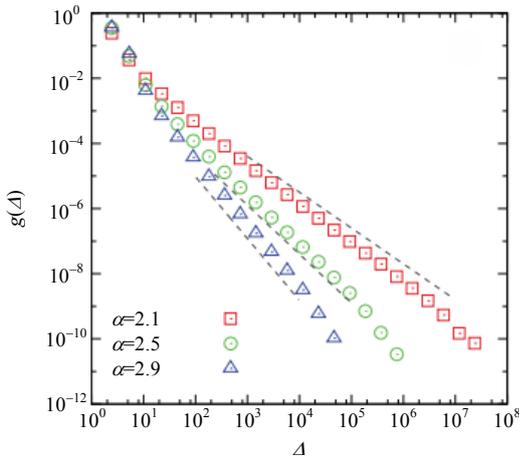
图51为 α 分别为2.1、2.5和2.9时的模拟结果对比,并且将其结果与假设人类行为时间间隔满足指

数分布的结果进行了对比。图51a为每个时间步新感染的人数，从侧面反映出人类活动时间间隔的异质性会放缓传播的步伐——放缓步伐并不代表降低了传播的危险性，事实上，传播持续的时间更长了。而幂指数越大，时间间隔异质性越小，新感染的人越多，但是传播来得快，去得也快。图51b则反映了人类活动时间间隔和回复时间间隔的关系，与式(21)结果相符。

除了直接分析真实数据，文献[164]探究了排队论模型中时间间隔的记忆效应对于传播速度的影响。图52对比了根据排队论生成的时间间隔序列和随机生成的满足幂律分布的时间间隔序列对于传播速度和回复时间间隔分布的影响，发现根据排队论形成的时间序列，相对于打乱了顺序的时间序列，对于传播速度有明显的延迟作用，且排队论所生成的个体活动的时间间隔分布的幂指数与回复时间间隔的幂指数相等。



a. 每个时间步新感染的人数随着时间的变化趋势

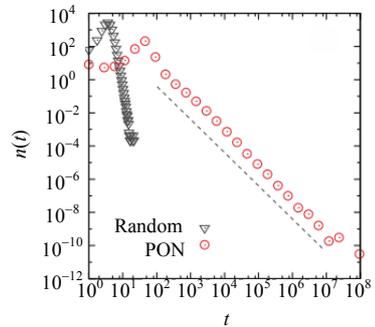


b. 人类活动时间间隔和回复时间间隔的关系

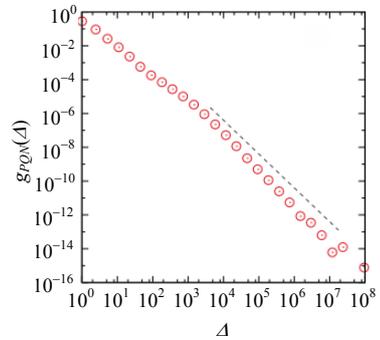
图51 个体时间间隔分布的幂指数 α 对于传播速度和回复时间间隔的影响^[164]

鉴于时间的阵发性可以体现在个体与群体两个

层面上，文献[166]进一步比较了二者对于疾病传播速度的影响。所谓群体层面的时间异质性，是指每个人活动的时间序列都是规则的，即时间间隔保持不变，但人与人之间的活跃性(时间间隔的倒数)有较大的差别，满足幂律分布。个体层面的时间异质性，表现为个体自身活动的时间间隔满足幂律分布，但不同个体的活跃性及时间间隔所满足的分布形式都是相同的。真实系统中可能两种异质性都存在。从图53中可以看出，当系统感染人数超过一半的时间 t^* ，从侧面度量了传播的速度。群体层面上时间的异质性对传播速度的影响非常大，相比之下，个体层面上时间的异质性对传播影响较小。而且与个体层面上时间间隔异质性越强，传播越慢的结论相反，从群体层面来说，时间的异质性越强，传播越快。

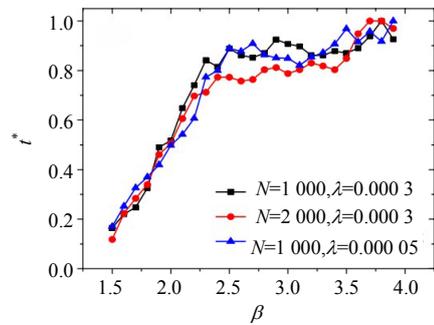


a. 每个时间步新感染的人数随着时间的变化趋势

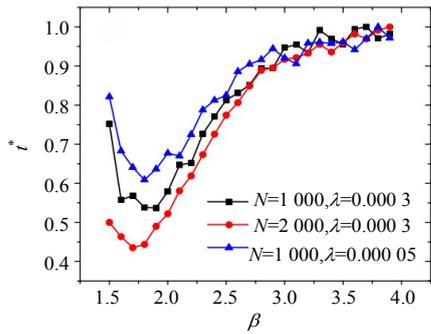


b. 根据排队论生成的时间间隔分布所对应的回应时间间隔分布

图52 时间异质性对于疾病传播速度的影响^[166]



a. 群体层面上



b. 个体层面上
图53 对比了群体层面上和个体层面上
的时间异质性对于传播速度的影响^[166]

如第2节所述,很多学者对于幂律分布的说法是否准确一直争论不下。除了幂律分布这种主流说法,文献[30]认为人类活动的间隔时间分布是对数正态的或是介于幂律和指数之间的。文献[37]分析了31 183个用户之间E-mail通信数据,认为人类行为时间间隔若满足对数正态分布,更符合人类的真实传播行为。图54为在真实网络上,不同的时间间隔分布的模拟结果比较。图中黑色的空心圆为实际的传播数据,对比于3组模拟结果,其中黑色线为人类行为时间间隔满足对数正态分布所得到的传播结果,红色线为指数分布,蓝色线为幂律分布。通过对比可知,假设人类行为时间间隔分布满足对数正态分布可以更好地预测真实的传播行为。

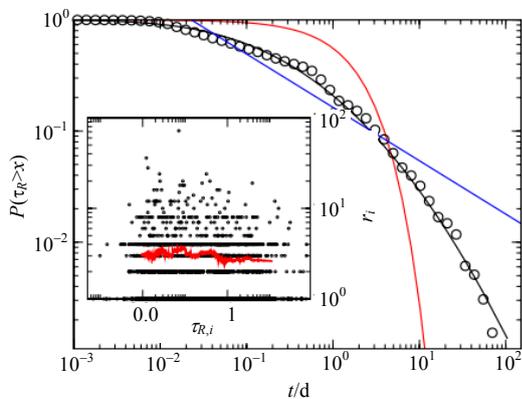
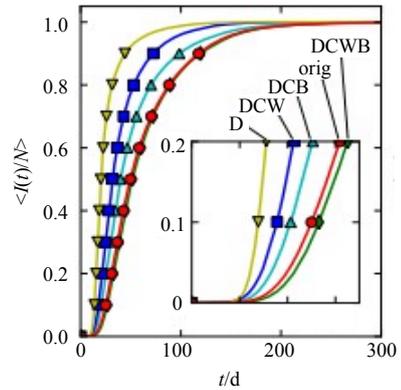


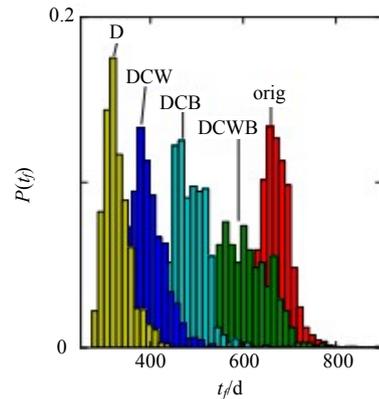
图54 真实网络上,不同时间间隔分布的模拟结果^[167]

在人类活动时间间隔满足阵发性的基础上,为探讨时间序列的记忆性对传播过程的影响,文献[168]基于时间跨度为9个月的电话网络的实证数据,在规模为 $N=4.6 \times 10^6$ 的网络上模拟传播。通过对比网络是否含权重、网络是否有向、通话时间分布是否有记忆性等各类情况,模拟结果表明,人类行为时间上的记忆性会在一定程度上减慢传播过程。图55a展示了五组SI模型的传播过程,感染率都为1,且都是基于实证数据,其中有:原始的具有记忆性的时

间序列在有向含权网络上(\circ),有记忆性时间序列在无向不含权网络上(\diamond),有记忆性时间序列在无向含权网络上(\triangle),无记忆性时间序列在有向含权网络上(\square),无记忆性时间序列在无向不含权网络上(∇)。图55b展示了5组传播过程完成完全传播的时间分布。对比表明,人类行为时间上的记忆性会一定程度地减慢传播过程。



a. 5组SI模型的传播过程



b. 5组传播过程完成完全传播的时间分布

图55 5组SI模型的传播过程和传播时间分布^[168]

5.2 人类行为空间特性对于传播的影响

5.2.1 种群模型

研究人类种群模型对传播过程的影响,要追溯到文献[169]提出的层次种群模型。在该模型中,总人群数目 N 被平均分配为 m 份各含有 n 个个体的子种群, $N=nm$ 。个体可以通过子种群之间的连边进行移动,疾病只能在子种群内部进行直接感染,染病个体需通过子种群间连边旅行到其他种群内部后才能感染其他种群的个体。图56是一个层次种群模型示意图,其中共有3层,每个节点有两个分支节点,每个底层子种群节点内部有8个个体^[169]。每个子种群(大圆圈)内部有8个个体(黑实心圈)。可以把层次结构想象为子种群的组织结构,大到一个国家下面

的不同省市, 小到一个学校的不同院系等。疾病或者信息刚开始只会在一个地方、一个部门开始, 然后逐渐地扩散至全校、全市乃至全国。

随后, 基于种群的传染病模型研究便如雨后春笋般涌现出来, 特别是考虑个体移动特性的种群模型。图57是文献[170]研究的种群入侵动力学示意图, 该模型不仅考虑了用户的种群结构, 而且考虑了用户的出行率 σ 以及返回率 τ 对传播过程的影响, 其中出行率 σ_{ij} 表示用户从子群 i 迁移到子群 j 的概率, 大小与初始和目标子种群中包含的个体数目相关, 返回率 τ 表示用户由于记忆效应, 返回其初始子种群的概率。对于微观层面两个子种群间人群迁移和疾病传播的过程来说, 一般每个子种群都有一部分用户留在种群内部, 同时有部分种群用户旅行到其他的种群。

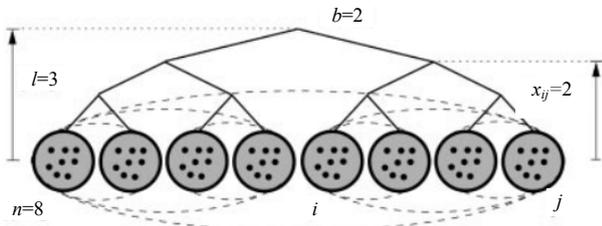
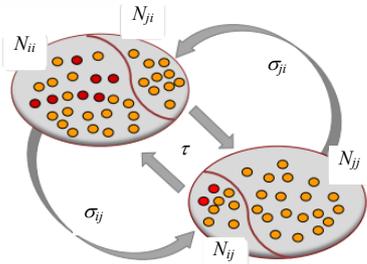
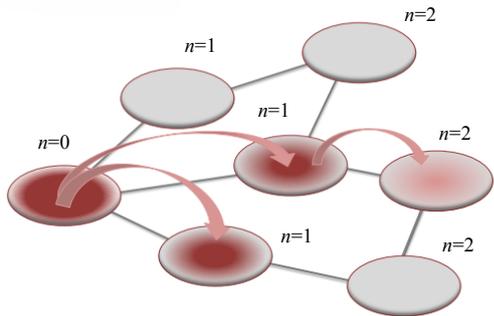


图56 层次种群模型示意图^[169]



a. 微观层面子种群内部个体移动示意图(箭头表示个体在不同子种群间来回, 黑色节点表示已感染节点)

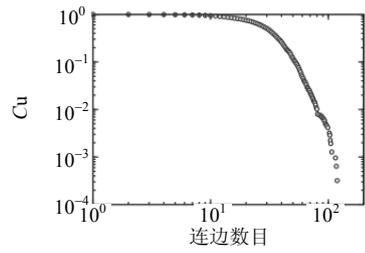


b. 宏观层面下疾病在子种群之间传播示意图

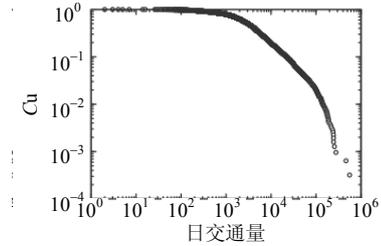
图57 子种群传染病入侵示意图^[170]

为了更贴近人类实际出行的特征, 文献[170]首先实证研究了美国和法国交通网络的统计特征, 发现用户的出行规律有着胖尾特性, 如图58所示。可以发现交通网络在连边数目和交通流量上都

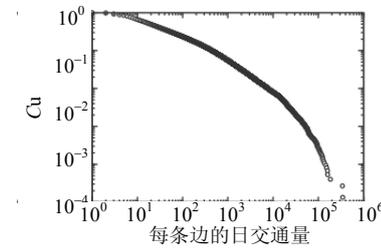
有着很强的异质性^[170]。



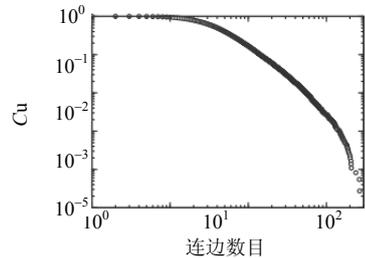
a. 美国交通网连边数示意图



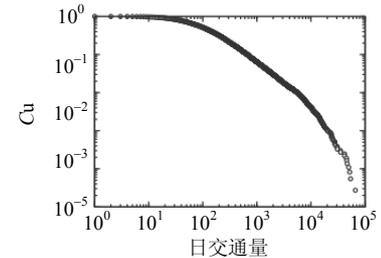
b. 美国每天的交通量示意图



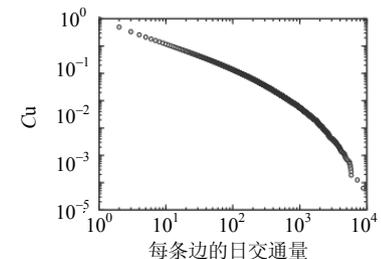
c. 美国每条边日常交通量的累积分布示意图



d. 法国交通网连边数示意图



e. 法国每天的交通量示意图

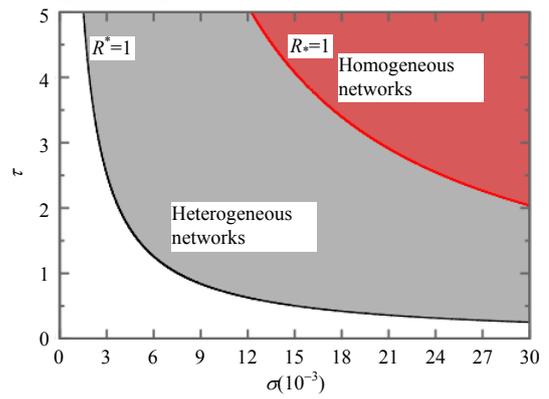


f. 法国每条边日常的交通量的累积分布示意图

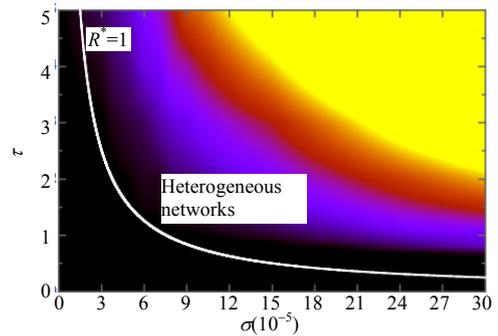
图58 美国和法国的交通网络统计特性示意图^[170]

为了研究用户出行率 σ 和返回率 τ 对传播过程的影响,文献[170]还讨论了 $\sigma-\tau$ 空间上疾病全局爆发和衰减消失的相图,如图59所示。从图可以发现,用户的出行率 σ 和返回率 τ 对传播的影响是一样的。当出行率和返回率增大时,有利于传播,当两者都大于其各自的阈值时,疾病就能全局传播开来。同时,该图另外一个有趣的地方是反映了基于度分布对传播过程的影响:度分布的异质性能降低传播的阈值。用户出行率高或者旅行时间长都能使疾病更容易爆发。其中黑色和红色的线分别表示在异质结构的网络和同质结构的网络中疾病全局爆发的阈值线。说明异质网络比同质网络更容易受出行率和旅行时间影响。

为了更充分地验证理论分析结果,文献[170]使用了美国大陆所有州的交通网络数据(美国2000年人口普查数据),在这个数据集中,每个子种群代表一个州,两个州之间的连接由两个州之间的交通流量表示。作者使用SIR模型研究了出行率和感染比例的关系,在模拟网络和真实网络中观察到了相似的相变现象,如图60所示。

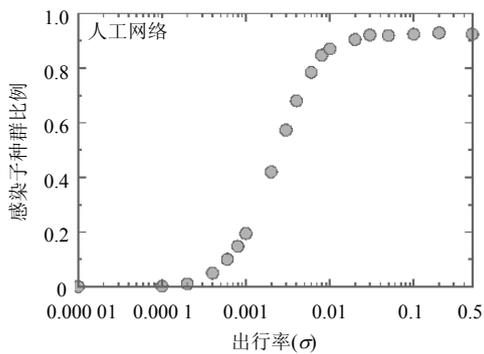


a. 数值解析结果

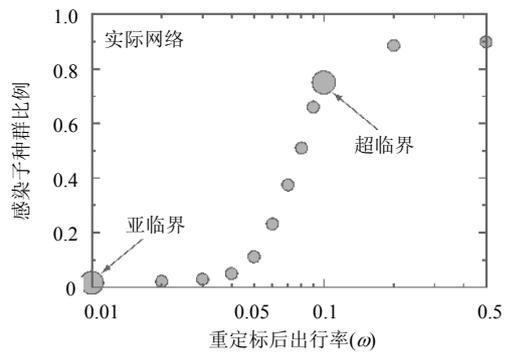


b. 实验模拟结果

图59 疾病全局爆发和衰减消失相图^[170]



a. 模拟情况下平均感染子种群比例和出行率之间的关系



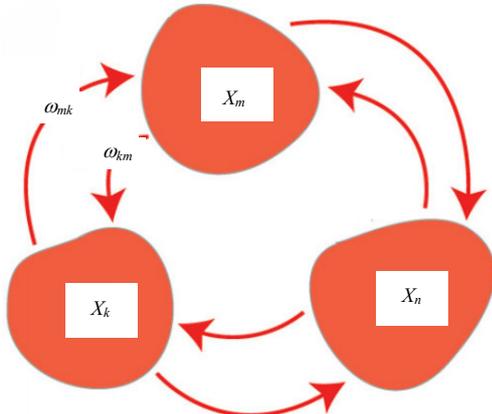
b. 真实网络中感染子种群比例和出行率之间的关系



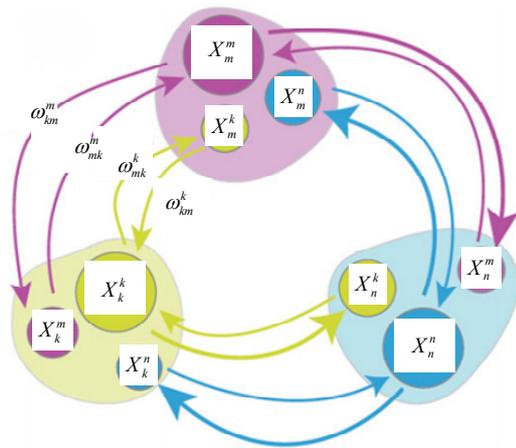
c. 处于超临界态(上)和亚临界态(下)情况下,美国疾病传播情况的瞬时斑图

图60 基于美国实际交通流数据的SIR传播动力性行为^[170]

文献[171]研究了基于个体双向移动的疾病传播模型, 如图61所示。在该模型中, 不但考虑了用户在不同子种群之间的移动, 而且考虑同一个体在目的节点和初始节点往返移动速率 ω 对传播过程的影响。图61a表示没有区分一个子种群内不同个体差异性的移动示意图, 图61b表示区分种群中不同个体移动的示意图。



a. 不同子群之间个体随机的进行移动, 任意两个子群之间的个体按照一定出行率随机移动, 同一子群内部个体没有区分



b. 不同子群内部个体按照不同的有区别的出行率进行移动
图61 种群移动示意图^[171]

研究发现, 普通的反应扩散模型和交互模型之间有着显著的区别。在SI传播模型情况下, 普通的扩散模型的前向速度会随着个体的转移率而不断增长, 解析结果为:

$$c = 2\sqrt{\alpha D(1 - \beta/\alpha)} \sim \sqrt{\omega} \quad (23)$$

但是在用户个体存在差异的模型中, 前向速度存在一个饱和值, 解析结果为:

$$c = \frac{2\sqrt{6D\alpha\omega}}{(\alpha + 3\omega)} \quad (24)$$

图62是传播速度和个体移动率关系图。图中, 蓝色菱形表示不考虑个体双向移动情况下, 按照SI模型的传播示意图; 红色菱形、红色圆圈以及红色方块分别表示考虑个体双向移动情况下, 使用标准

SI模型、SIR模型以及易感者S固定不移动情况下的模拟曲线; 黑色虚线对应于个体移动有区别的模型, 点划线表示个体移动率没有区别的传播示意图; 所有结果在网络包含102个站点, 每个站点包含104个体^[171]。

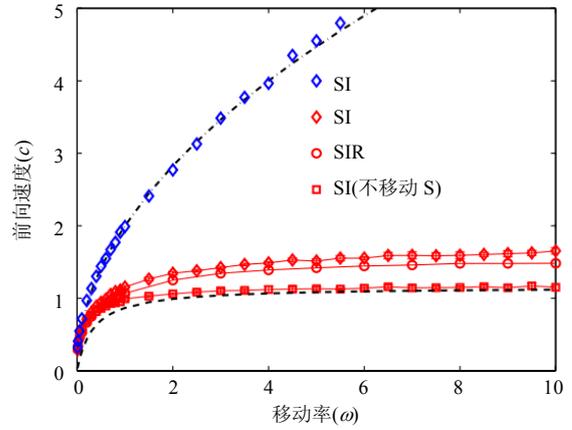


图62 前向速度与移动率的关系^[171]

文献[171]进一步将个体移动率 ω 拆分成 ω^+ 和 ω^- 两部分, 前者表示个体离开子种群的移动率, 后者表示个体返回子种群的移动率。在此基础上发现模型存在另外一种传播阈值, 该阈值由用户的返回率 ω^- 决定。如图63所示, 传播模型为SIR, 包含102个节点, 每个节点有N个个体, $\alpha=1, \beta=0.1$ 。其中无标度网络中幂指数 $\gamma=1.5$, 最大度为50, 最小度为5。ER网络中有500个节点, 平均度为10。实验结果为模拟50次取平均^[171]。随着返回率的增加, 疾病几乎不可能在全局范围爆发。这主要是因为用户的返回率较高, 导致用户在目标区域内呆的时间过少, 从而不能接触并感染目标区域易感节点。而且这种相变和网络的拓扑结构无关。

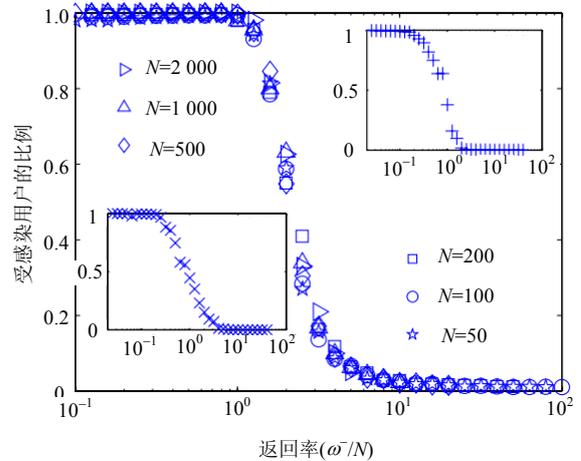
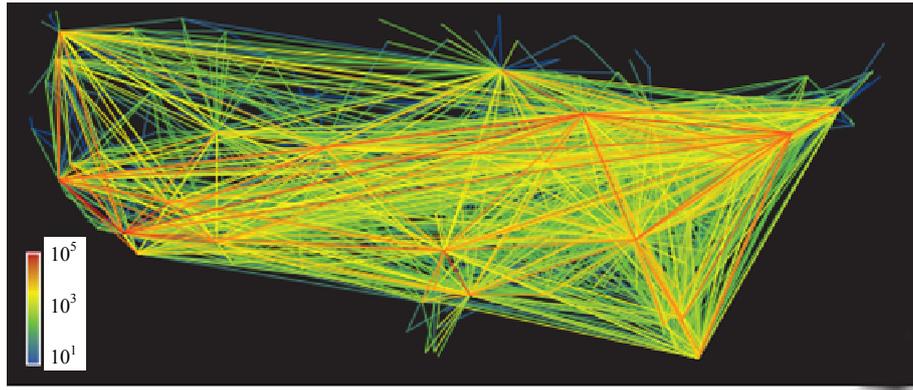


图63 随机模拟结果示意图, 表示受感染用户的比例 ρ 和用户的返回率 ω^- 相变图^[171]



a. 美国大陆航空网络示意图



b. 美国大陆陆地交通网络示意图

图64 多尺度移动网络示意图^[172]

总的来说,通过引入人类种群的概念,可以使用更精细、更接近人类空间运动实际情况的模型来研究疾病传播情况。

5.2.2 交通模式对疾病传播的影响

这方面的研究主要集中于航空网络和地面交通网络对疾病传播过程的影响,一般也运用了6.2.1节中种群模型的概念。文献[172]首先分析了来自29个国家的移动数据,发现可以使用引力模型在300 km的范围内来描述交通流特征。其中基于引力模型的交通流量为:

$$w_{ij} = C \frac{N_i^\alpha N_j^\gamma}{f(d_{ij})} \quad (25)$$

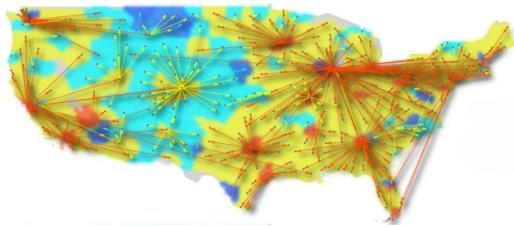
式中, N_i 和 N_j 分别表示子种群 i 和 j 中的个体数目; C 是一个比例常数; α 和 γ 是基于子种群大小的可调参数; $f(d_{ij})$ 是一个距离相关的函数。该文献又整合了一个全球种群结构的传播模型,该模型针对航空网络和普通网络使用不同的移动返回时间,从而评估多尺度移动过程对传播动力学的影响。图64是航空交通网络和陆地交通网络示意图。边的颜色和宽度表示的是对数刻度下的流量和密度^[172]。

分析发现,尽管陆地交通流网络的流量要比航

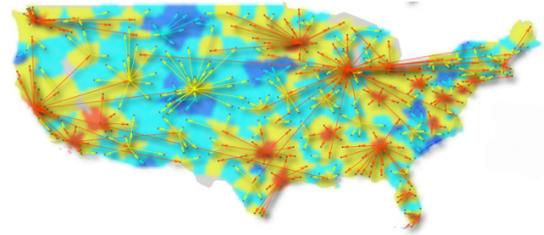
空网络的流量大一个数量级,但是疾病传播的全局时空斑图主要还是由航空网络决定。这也说明了短程的交通流量在很大程度上只对子区域的传播起作用。图65展示了传染病入侵树示意图。图65a是只有航空流量的入侵传播示意图,可以看出航空枢纽通过直接的航线在疾病传播到其他地方时起着决定性的作用;图65b是考虑了航空流量和陆地交通流量的入侵传播示意图,可以看出普通交通流量可以降低航空枢纽的作用,并增加局部传播的重要性。其中红色节点表示的第一批被海外感染的城市,黄色表示第二级被感染城市,青色表示第三级被感染城市,蓝色表示第四级被感染城市^[172]。

2009年的H1N1流行之初,人们控制了离/达墨西哥的国际航线,使其流量降低40%,但即使这样,也没能阻止H1N1的全球流行。文献[173]提出了一个理论模型研究旅行控制对阻止和延迟疾病传染的作用,该模型结合人类航空旅行和短程移动的数据,同时包含高精度的世界各地人口统计数据。定量地讨论了限制不同移动模式对传播延迟的影响,以及疾病在不同情况下可能的爆发情况,见图66。该文献指出全球流行疾病无法使用限制旅行方式进行控

制的根本原因是由于人类移动模式的异质性。



a. 只有航空流量的入侵树



b. 考虑了航空流量和陆地交通流量的入侵树

图65 疾病入侵树^[172]

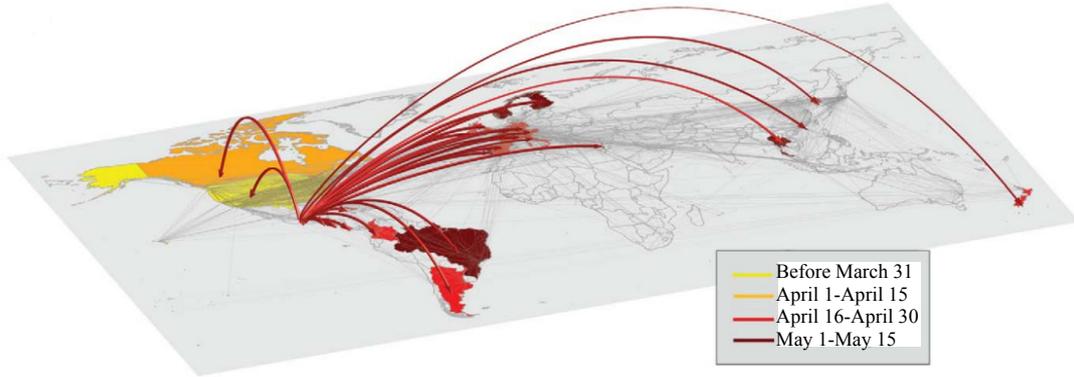


图66 2009年H1N1全球大流行前期传播过程示意图(箭头指向表示已感染的个体旅行到未感染国家, 地图颜色表示当地开始传播日^[173])

人类的出行行为受到所处的交通系统的影响和制约, 而人的活动同时也影响着交通系统的发展, 两者之间呈现出一种共演化的关系^[134,136]。文献[121]提出了层次交通网络模型来描述人类出行位移的胖尾分布。显然, 个体间的空间距离将不可避免地阻碍或延滞接触活动。因此, 探索人类行为的时空相关特性对于动力学过程的本质影响具有极为重要的理论与现实意义。基于此, 文献[174]研究了空间层次网络中3种接触模式对于信息/病毒传播的影响。这3种接触模式分别是: 1) HETD, 网络中任意两个相连节点之间的距离由节点之间的欧拉距离表示, 例如晶格网络中节点坐标的距离, 该网络既表征了节点之间的距离, 又表征了层次和距离之间的相关性。2) HOTD, 网络中任意两个相连节点之间的距离相等, 该网络消除了节点之间的距离和层次距离相关性。3) RNTD, 将HETD网络中节点之间的欧拉距离随机重匹配, 该网络保持了节点之间的距离分布, 但层次和距离之间的相关性被打破。实验结果表明, 与同质/随机的接触过程相比, 因层次结构和距离相关联而导致的时延能够显著地降低传播速度, 并呈现逐级递增的波峰, 具体结果如图67所示。图中显示了新增节点 $n(t)$ 与可变性随时间变化的趋势, 其中, 黑色实线表示节点之间不存在随机重连, 而黑色方框、红色圆圈、绿色三角及蓝色三角分别表示存在以2%、5%、10%及20%概率随机重匹配的结果^[174]。可变性定义如下:

$$\Delta[i(t)] = \frac{\sqrt{\langle i(t)^2 \rangle - \langle i(t) \rangle^2}}{\langle i(t) \rangle} \quad (26)$$

式中, $i(t)$ 表示在 t 时刻已感节点数目; $\langle i(t) \rangle$ 表示对 t 时刻已感染节点数目求平均。

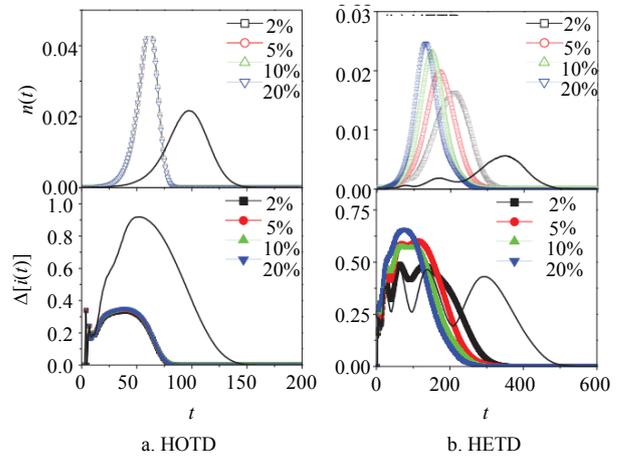


图67 空间层次网络对病毒传播的影响^[174]

该结果初看起来非常有利于监控工作的开展, 但是, 更为深入的研究表明, 高层的初始感染节点将导致极大的可变性, 而低层初始节点则导致了长时间的多峰可变性, 从而造成了空间层次网络中传播过程的不可预测性, 给舆情/疫情的监测与防控带来巨大的挑战。具体模拟结果如图68所示, 在两种不同参数的网络中, 发现节点最大波峰值会随着初始感染节点层数的增长而逐渐变小; 到达最大波峰的时间随着层数的增大而逐渐增大; 最大可变性在

HOTD网络和随机网络中,会随着层级增大而增大,但在异质性网络中,会随着初始节点数目层级的数目而减小^[174]。

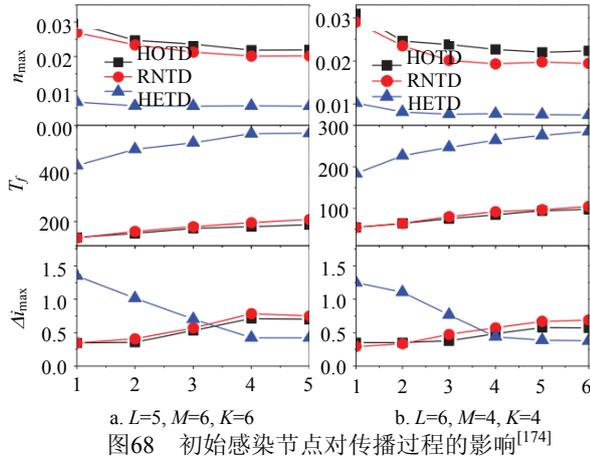


图68 初始感染节点对传播过程的影响^[174]

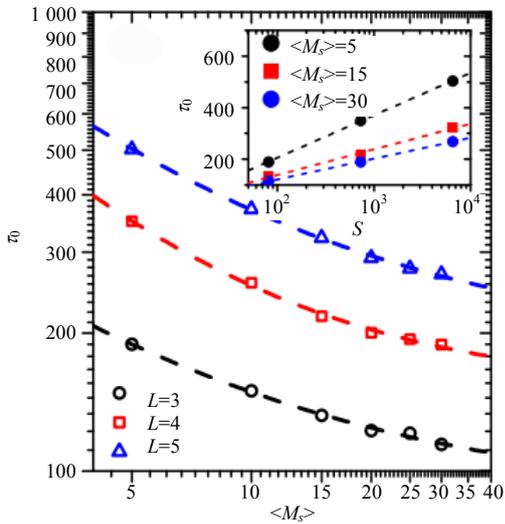


图69 网络规模L和感染节点峰值时刻 τ_0 关系(其中内嵌图是人口密度和感染节点峰值时刻 τ_0 的关系^[175])

文献[175]使用种群模型研究了该层次网络中传播峰值与人口密度和网络规模的关系,发现峰值出现时刻 τ_0 与人口密度存在负相关的关系;同时 τ_0 与网络规模存在对数的正相关关系,如图69所示。以上研究揭示了网络拓扑结构及距离对传播过程的影响,同时可为疾病的免疫和控制策略提供借鉴。

5.2.3 行为模式对手机病毒传播的影响

文献[176]研究了手机病毒的传播模式,发现蓝牙病毒可以在足够长时间内感染所有易感用户,但是由于受人类移动模式影响,传播速度很慢,使开发者文有很多时间开发出反病毒软件来防御该类型病毒。与之相反,彩信病毒可以在几小时内感染所有的手机,但是由于受易感数目的影响(手机操作系统份额),传播存在一个受手机市场份额影响的阈值。当市场份额小于该阈值时,该手机操作系统类型的用户被分割成许多小的“孤岛”;当份额大于该阈值时,手机网络存在一个大的集团,这样一个大的连接集团整体都存在被病毒感染的可能性。图70和图71分别为手机病毒传播机制示意图和手机病毒传播模式示意图。从图中可以看出,感染BT(蓝牙)病毒的手机可以感染周围蓝牙侦测范围内的所有带蓝牙功能的手机用户,该传播过程由用户移动模式决定。同时, MMS(彩信)病毒可以感染一个用户电话本中出现过的所有用户,该过程可以导致独立于用户地理地点的长程传播模式^[176]。

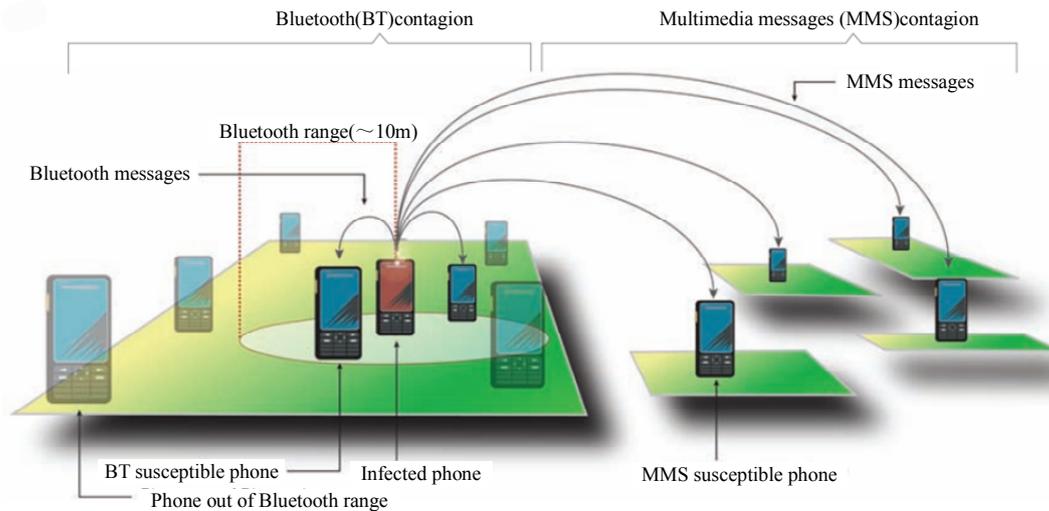
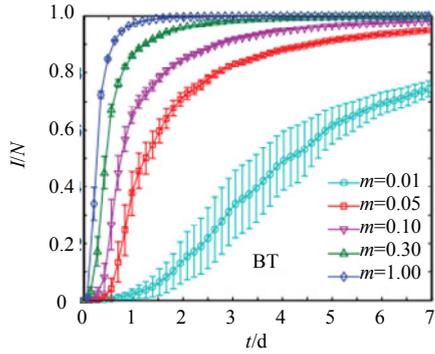
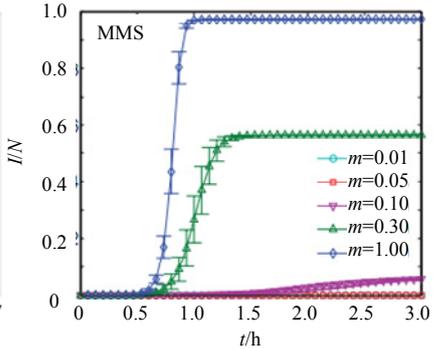


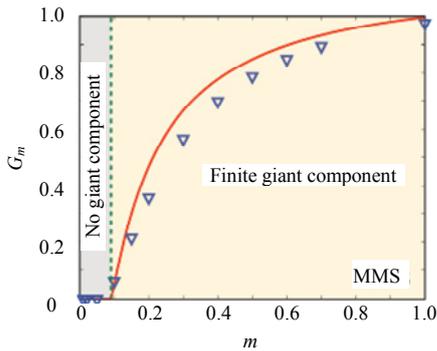
图70 手机病毒传播机制示意图^[176]



a. BT病毒感染节点比例随时间变化示意图



b. MMS病毒随手机操作系统市场份额变化示意图



c. 最大连通集团规模和手机操作系统市场份额关系示意图(其中三角形是实际饱和值, 线条表示通过渗流理论的预测值)
图71 手机病毒传播模式示意图^[176]

5.3 空间位置预测

人类行为的预测对通讯服务、安全事务、公共健康、资源分配及信息服务等都有着重大的意义。文献[133]依据熵值分析得出了理论上人类空间访问行为有93%可以精确预测的(指预测下一次访问的地点, 精确到手机基站覆盖的范围), 并定义了人类行为的3种熵:

1) 用户访问不同地点的熵, 即:

$$S_i^{rd} \equiv \log_2 N_i \tag{27}$$

式中, N_i 表示历史中个体 i 所到达过的不同地点数目。

2) 用户在不同地点被观察到的比例熵, 即:

$$S_i^{unc} \equiv -\sum_{j=1}^{N_i} p_i(j) \log_2 p_i(j) \tag{28}$$

式中, $p_i(j)$ 表示个体 i 访问地点 j 的次数占总访问

次数的比例。

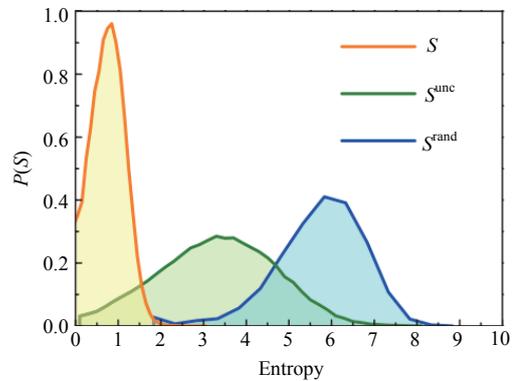
3) 用户的真实行为序列熵, 即:

$$S \equiv \left(\frac{1}{n} \sum_i A_i\right)^{-1} \ln n \tag{29}$$

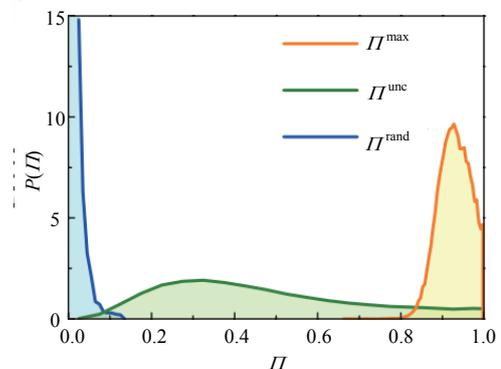
式中, A_i 表示从 i 位置开始, 前面从未出现过的子序列的最短长度。例如对序列(1,2,1,2,2), 从位置1开始, 依照定义, $A_1=1, A_2=1, A_3=3$ 。图72a表示是用户3种熵的变化趋势。从图中可以看出, 用户的真实行为序列熵分布范围比其他两种都小, 预示用户真实行为序列具有更大的可预测性, 依据费诺不等式^[177], 可以得到熵和可预测性关系为:

$$S = S_F(\Pi^{\max}) = -[\Pi^{\max} \log_2 \Pi^{\max} + (1 - \Pi^{\max}) \log_2 (1 - \Pi^{\max})] + (1 - \Pi^{\max}) \log_2 (N - 1) \leq S_F(\Pi) \tag{30}$$

式中, Π^{\max} 表示可预测性的最大值; N 表示访问不同地点数目; S_F 是费诺函数——单调下降的凹函数。通过解该方程, 可得到用户最大可预测性值 Π^{\max} , 得到可预测性分布如图72b所示, 该图显示用户可预测性与熵值相对应, 其中共抽取了50 000名用户, 每个用户打电话时所记录基站数目大于两个, 平均每小时电话通话0.5次^[133]。注意, 用户真实行为序列具有平均值为93%的极大可预测性, 且不同用户可预测性相差很少。



a. 用户3种熵的分布图



b. 用户的3种熵所对应的可预测性最大值分布
图72 手机用户熵与可预测性分布^[133]

费诺不等式既可表示在输出序列 Y 已知情况下计算输入序列 X 的平均不确定性,也可表示输入序列 X 经信道传输后信息量的损失情况^[177]。具体公式为:

$$H(X|Y) \leq H(p_E) + p_E \log(r-1) \quad (31)$$

式中, $H(X|Y)$ 为条件熵; p_E 为平均错误率; X 、 Y 为输入和输出序列; r 为输入符号的数目。费诺不等式的含义可以理解为知道输出序列 Y 时,输入序列 X 的平均不确定性可以由两方面消除:1) 确定输出是否有错,消除这种不确定性所需信息量为 $H(p_E)$;2) 确定输出出错后,找出具体是哪一个符号错误,消除该不确定性所需最大信息量为 $\log(r-1)$ 。

费诺不等式只能给出可预测性的理论上限,并不是具体的预测算法,也不代表就一定能够找到合适的算法达到这个极限。因此,知道极限之后,对预测算法的研究依然是一个具有重要实践价值的问题。文献[178]研究了通过用户当前位置和移动记录的数据来判断该用户要去的下一地点。对于该任务,使用了两种不同的方法。1) 考虑时间的动态贝叶斯网络(DBN),它可以将时间和其他信息都融入判断;2) 转换为分类问题,使用神经网络和决策树做分类。使用这两种方法,都得到了平均准确率高于60%的预测结果。文献[179]设计的算法获得了2011~2012年诺基亚移动数据挖掘竞赛“下一地点预测”任务的冠军,其精确度超过了含时的马尔科夫模型。文献[180]利用用户的GPS数据,向用户推荐他们可能感兴趣的地点。如一个用户以前经常去咖啡馆,那么当这个用户到达一个广场之后,就可以向他推荐广场附近有名的咖啡馆。该文一共研究了3个算法:1) 协同矩阵分解;2) 协同张量和矩阵分解;3) 一个基于排序优化的新型改进算法,更能够帮助用户找到合适的地点。

有趣的是,空间位置预测的思路可以自然推广到对人类很多在线行为的预测——只要我们把一类的行为看成一个地点。如淘宝上的每一个商品品类都可以看成一个地点,也可以得到一个个体在不同品类间跳转的网络,类似于图40d~图40f,如果把所有个体的跳转聚合起来,还可以得到针对用户整体的品类跳转网络,如图73所示。也可以用费诺不等式分析用户访问某个品类的可预测性,如图73所示的3个系统,其可预测性分别为豆瓣0.93、淘宝0.84、手机阅读0.83。

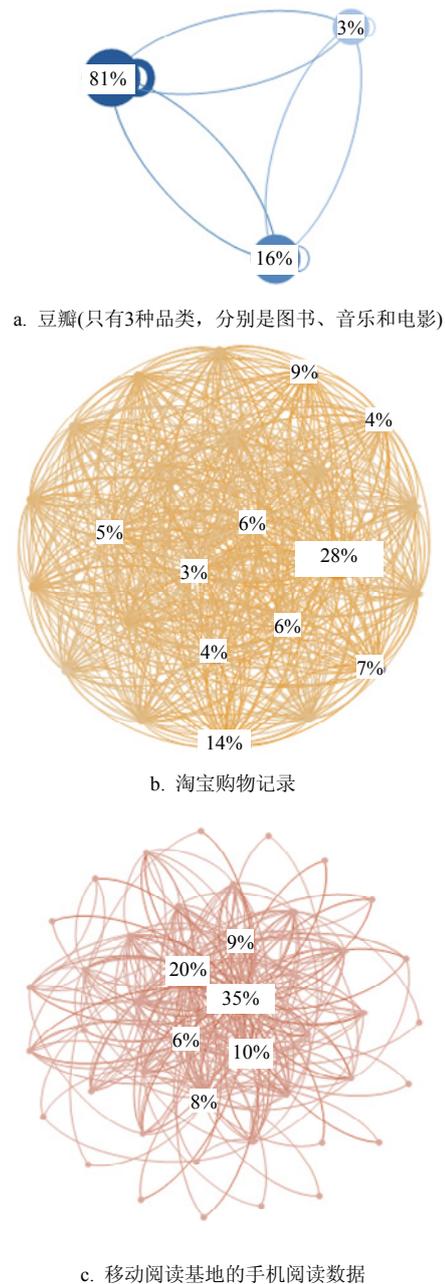


图73 3种真实系统中真实用户群体形成的品类访问跳转网络

5.4 信息推荐

互联网时代最大的问题之一就是信息过载问题:我们可以获取的信息总量爆炸性地增长,与此同时,我们处理甄别信息的能力却没有相应地提高,因此越来越难以从海量信息中找到我们需要或喜欢的内容。解决信息过载问题最有效的方式就是通过推荐系统,利用自动化算法向用户推荐关联的内容^[181]。最为简单的推荐系统的模型是一个“用户-商品”的二部分网络,如果用户选择过某件商品,就连一条边。推荐系统的功能就是从现在已经知道的选择关系中分析出用户的喜好,在所有用户没有选

择过的产品中找到用户最喜欢的, 并且推荐给用户。

深入挖掘用户的行为模式有望提高推荐的效果或在更复杂的场景下进行推荐。如新用户和老用户具有很不一样的选择模式: 一般而言, 新用户倾向于选择热门的商品, 而老用户对于小众商品关注更多^[182], 新用户所选择的商品相似度更高, 老用户所选择的商品多样性较高^[18]。有些混合推荐算法允许通过调节一个单参数来调节推荐结果的多样性和热门程度^[183], 在这种情况下就可以考虑给不同用户赋予不同的参数, 甚至允许用户自己移动一个滑钮调节这个参数——当用户想看热门时, 算法提供热门推荐; 当用户想找点很酷的产品时, 算法也可以提供冷门推荐。

贯通本文的是人类行为时间和空间上的统计特性, 对这些性质的深入挖掘, 也可以用于提高推荐效果或者设计针对特定场景的应用。最简单的例子是在推荐前考虑用户从事相关活动随时间变化的活跃性。如在进行手机个性化阅读推荐时, 如果数据显示某个用户只在7~8点之间有一个小时左右手机阅读行为(可能是上班时在地铁或者公交车上), 那么9点发送一个电子书阅读的短信广告就是很不明智的选择。从含时数据中还可以分析出影响用户选择的长期和短期的兴趣, 通过将这两种效应分离出来, 可以明显地提高推荐的精确度^[184-186]。事实上, 简单假设用户兴趣随时间按照指数递减, 也能够得到改进的推荐效果^[187-188]。

随着移动互联网的飞速发展以及GPS及其他手机定位技术的发展和普及, 基于位置信息的推荐已经成为推荐系统的研究热点和重要应用场景, 而这个问题的解决需要能够对用户的移动模式有深入的

理解, 包括预测用户的移动轨迹和判断用户在当前位置是否有可能进行餐饮购物活动等^[133,189], 同时还要有定量的办法去定义用户之间以及地点之间的相似性^[190-191]。事实上, 即便简单把位置信息作为一个单独属性加以考虑, 也可以明显提高广告推荐^[192]和朋友推荐^[193]的精确度。

除此之外, 不同用户打分的模式也不一样^[194], 用户针对不同商品的行为模式也不一样^[195-196]——想象你在网上下载一首歌和团购房子时的区别, 这些模式都可以挖掘刻画并利用来提高推荐的效果。总而言之, 推荐引擎针对合适的对象在合适的时间和合适的地方推荐合适的内容(4S标准)。通过分析不同用户在选择、评价、时间、空间等等行为模式上的不同, 推荐系统最终的目的是猜测目标用户当前的意图, 并且针对不同的意图匹配或组合不同的算法结果, 将其推荐给用户。这不仅需要更高级的数据分析能力, 还需要有丰富经验了解业务逻辑的工作人员配合完成。这种称为“情境计算”的思路, 有可能较大幅度地提高推荐及其他信息服务(譬如搜索)的质量^[197]。

本文举一个具体的例子来说明如何通过深入分析人类行为的统计特性来实现更好的推荐。文献^[198]分析了MovieLens和WikiLens这两个典型的推荐系统, 其中用户用1~5分给对象评分。首先观察一些特殊的对象——平均得分高于4.5分或者低于2.0分的对象(这样的对象在所有对象中占比不到1%), 发现在高分对象后面的打分普遍高, 而在低分对象后面的打分普遍也低。这个差异非常明显, 在MovieLens中两个均值分别是4.16和2.72, 而在WikiLens中两个均值分别是4.13和2.63。

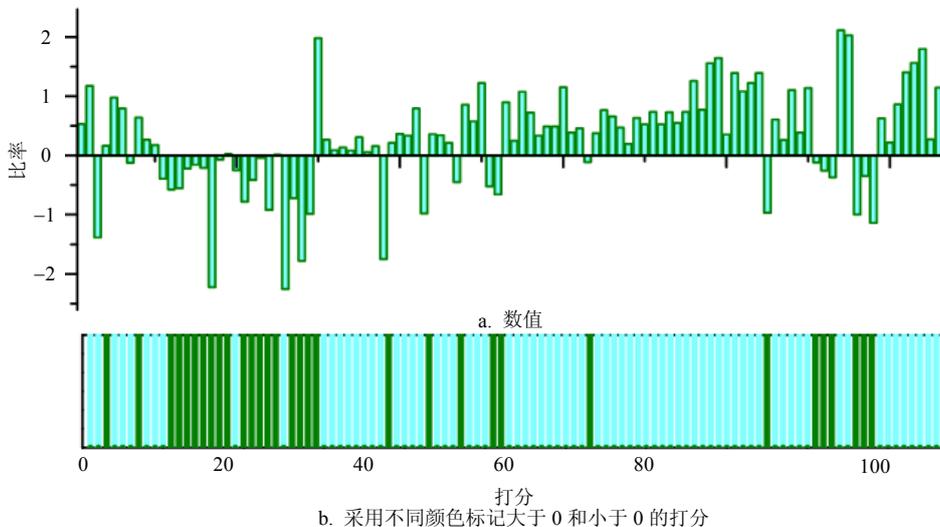


图74 MovieLens中一个典型用户打份去除掉自身打分倾向和对象质量不同的影响后, 所得到的打分偏差的序列^[198]

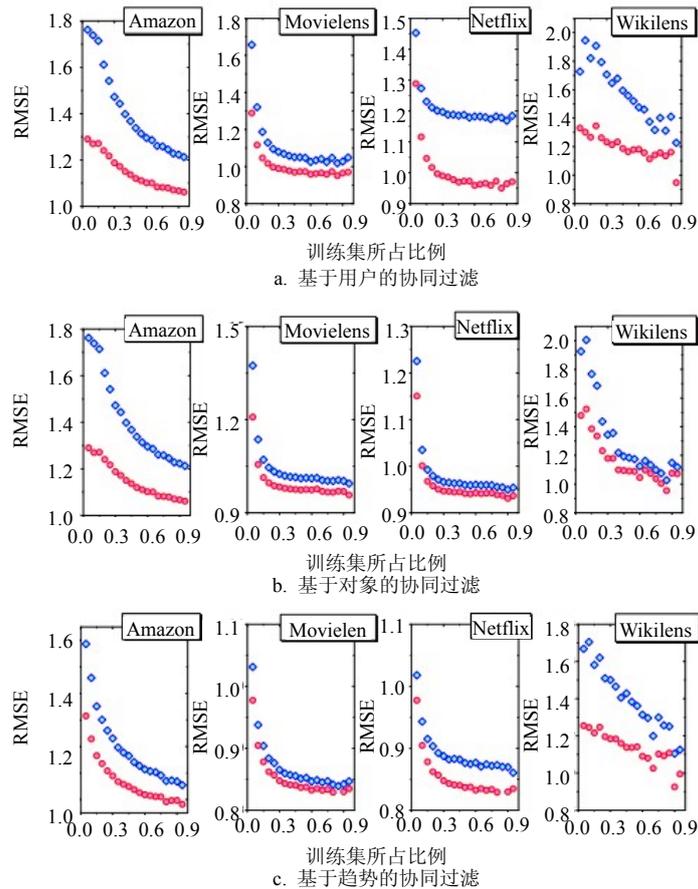


图75 3种经典算法(蓝色数据点)和利用锚定效应优化后的对应算法(红色数据点)所得到的用户评分预测均方根误差^[198]

考虑到有的用户倾向于打高分,有的用户倾向于打低分,有的对象质量好,得分普遍高;有的对象恰恰相反。在去除掉用户打分和对象质量导致的影响后(具体细节请参考文献[198]),将所有的打分分作两类^[198]:偏高的打分(>0)和偏低的打分(<0)。如图74所示,所有大于0的打分,都是偏高的打分;小于0的打分则是偏低的。偏高的打分和偏低的打分出现具有明显的记忆性,也就是说偏高分和偏低分倾向于集中一起出现。通过和随机化后的参照模型进行对比,可以证明这种记忆效应具有显著性!进一步分析显示,这种倾向于和最近的打分保持相近的偏差随着打分间隔次数呈对数下降的趋势。

20世纪70年代,文献[199]指出,人们在进行判断时常过分看重那些显著的、难忘的证据,甚至从中产生歪曲的认识。如医生在估计病人因极度失望而导致自杀可能性时,常常容易想起病人自杀的偶然性事件,从而夸大极度失望病人将自杀的概率,这就是人们在判断中存在的锚定效应。更严谨地说,锚定效应是指当人们对某个事件做定量估测时,会将某些特定数值作为起始值,起始值像锚一样制约着估测值。在做决策的时候,会不自觉地给

予最初获得的信息过多的重视。锚定效应在众多领域判断与决策问题的研究中得到验证^[200],从日常生活中的现象,如促销广告用词对购买数量决策的影响到风险预测性问题,如估计股市指数的变化;从一般知识性问题、博彩估计问题、法律判断问题、协商谈判问题、价格估计问题,到自我效能评估、软件评估问题等。许多研究从不同角度证明锚定效应是一种普遍存在的、十分活跃又难以消除的判断偏差。尽管文献[199]所讨论的锚定偏差和原来在社会学中熟悉的锚定效应有所不同,但是两者的心理机制或有共通之处。特别地通过对非受控实验数据统计分析得到的结果,研究思路不同于传统社会学控制实验的思路,对于传统社会学家,或有可兹借鉴之处,这也是开展人类动力学研究所一直企望达到的目标。

进一步地,文献[198]利用所发现的锚定效应,提出了具有普适性的提高用户评分预测精确性的优化方法。针对Amazon、MovieLens、Netflix、WikiLens共4种有代表性的推荐系统和基于用户的协同过滤^[201]、基于对象的协同过滤^[202]、基于趋势的协同过滤^[203]这3种有代表性的方法,如图75所示,优化算

法都能够明显降低评分预测的均方根误差, 亦即提高预测的精度。

6 结论与讨论

谢谢读者有耐心读到这篇长文的最后, 这也是对本文6位作者工作的最好回报。本文有很多特色, 它首次尝试在一篇综述中同时介绍人类行为中的时间和空间特性, 而且强调了这两种迥异统计对象的共性——它们都具有阵发的特性, 而且可以用很多类似的方法进行分析, 它还注意了理论和实际应用的结合。特别希望读者关注的是, 本文仔细分析了很多工作的精妙之处和不足的地方, 也包括一些学界尚存的激烈争论, 但相信这些深入的和深刻的洞见才是一篇综述的精髓。当然, 限于水平和文章的篇幅, 本文还有很多可以进一步改进的地方, 如含时网络^[204]和空间网络^[135]的研究进展, 跟本文讨论的问题息息相关, 但是我们吝嗇笔墨; 一些重要的概念和方法, 如Lévy游走、最大熵原理、空间位置预测方法等, 都应该给出更仔细的介绍; 若干关键的解析过程, 如Barabási模型的解析、如何从泊松个体到群体标度律的涌现等都没有给出。

本文在现象层面的结论是比较明晰的。就人类行为的时间特性而言, 尽管不同系统的人类行为时间间隔分布形式有所不同, 但是都有比泊松过程所预测的分布更胖的尾部, 人类的时间行为还具有“强阵发弱记忆”的特性, 并且表现出非常明显的波动性和周期性。就人类行为的空间特性而言, 人类个体的空间运动范围具有局域化的特点, 移动步长分布呈现不规则的多样化特征, 并不服从某种特定的分布形式, 但在群体层面上, 位移分布具有明显的统计规律, 其中混合交通方式下的位移分布服从幂律或截尾幂律分布, 而单一交通方式下的位移分布则服从指数或近似指数分布。与现象相对应的理论模型研究方面还有所欠缺, 任务优先机制、记忆和兴趣、周期与波动、社会交互作用都被认为可以产生行为时间上的阵发性, 但对于真实系统, 到底什么系统中哪个机制起主要作用还不清楚。空间模型方面就更加混乱, 还没有被广泛认可的机制。应用研究更弱, 主要的研究都还集中在行为时空特性对于传播动力学的影响, 其他方面的探索都才刚刚起步。

2000多年前, 古希腊德尔菲神庙上就刻上了“认识你自己”的神谕——几千年来认识自我一直是整个人类的梦想。如今, 生物基因工程的发展和无所不在的传感器, 使得从生物层面和行为层面定量化

认识自我变得可能。人类动力学的蓬勃发展也受益于大数据的采集、存储和处理技术的发展。简而言之, 人类动力学研究承载着人类科学和哲学的千年梦想, 技术和理念上恰在其机, 又存在重大的社会经济价值, 正处于科研成果大规模井喷的边缘。人类动力学是一门典型的交叉学科, 从理念、方法到应用, 都融合了计算机科学、统计物理学、社会科学、数学等多个学科的菁华。青年学者和研究生可以很容易地进入这个方向, 在学习研究的过程中体会不同学科的研究精神和方法论, 得到系统的科研训练, 而优秀的科研成果需要深刻的分析和综合的视野, 对于研究人员有很高的要求。

科学是一把双刃剑, 既可以为我们带来福祉, 也可能被别有用心的人利用而产生伤害。人类动力学的研究因为直接涉及到人的行为数据, 尤其有可能被不恰当地使用而侵犯我们的隐私。文献[205]针对150万手机用户轨迹数据的分析显示, 即便时空分辨率很低(空间到基站, 时间到小时), 如果能够在4个时间地点看到一个人, 就可以以超过95%的概率唯一确定这个人——实际上, 很多企业和科研团队都有机会拿到这样的加密数据, 而从网络新闻中获知某些知名人物在哪些时间地点出席什么会议也不是一件难事, 这些人的私密通话很可能被识别、泄露。文献[206]发现通过对Facebook上点击Like按钮的分析, 可以比较好的精确度预测出一个人的性取向——是异性恋还是同性恋。实际上, 如果知道手机轨迹信息, 这个预测可能还更精确。对于绝大多数一般用户而言, 这绝对不是一个好消息, 因为你的社交网络行为和手机应用, 可能会泄露你个人隐私的内核。作为本文的结束, 我们特别呼吁相关的科研人员, 关注科学研究可能带来的社会负面影响, 同时探索用科学的手段保护个人隐私的方法。

参 考 文 献

- [1] BARABÁSI A L. The origin of bursts and heavy tails in human dynamics[J]. *Nature*, 435(2005): 207-211.
- [2] BROCKMANN D, HUFNAGEL L, GEISEL T. The scaling laws of human travel[J]. *Nature*, 2006(439): 462-465.
- [3] BARABÁSI A L. The architecture of complexity[J]. *Control Systems Magazine, IEEE*, 2007, 27(4): 33-42.
- [4] ZHOU Tao, HAN Xiao-pu, WANG Bing-hong. Towards the understanding of human dynamics[M]. *Science matters: humanities as complex systems*. Singapore: World Scientific Publishing, 2008.
- [5] 郭进利, 周涛, 李季明, 等. 人类动力学模型[M]. 香港: 上海系统科学出版社, 2008.

- GUO Jin-li, ZHOU Tao, LI Ji-ming, et al. Model of human dynamics[M]. Hong Kong: Shanghai Systems Science Press, 2008.
- [6] 李楠楠, 周涛, 张宁. 人类动力学基本概念与实证分析[J]. 复杂系统与复杂性科学, 2008, 5(2):15-24.
LI Nan-nan, ZHOU Tao, ZHANG Ning. Concept and empirical results of human dynamics[J]. Complex System and Complexity Science, 2008, 5(2): 15-24.
- [7] 汪秉宏, 韩筱璞. 人类行为的动力学与统计力学研究[J]. 物理, 2010, 39(1): 28-37.
WANG Bing-hong, HAN Xiao-pu. The dynamics and statistical mechanics of human behaviors[J]. Physics, 2010, 39(1): 28-37.
- [8] 韩筱璞, 汪秉宏, 周涛. 人类行为动力学研究[J]. 复杂系统与复杂性科学, 2010, 7(2-3): 132-144.
HAN Xiao-pu, WANG Bing-hong, ZHOU Tao. Researches of human dynamics[J]. Complex System and Complexity Science, 2010, 7(2-3): 132-144.
- [9] BARABÁSI A L, Bursts: the hidden patterns behind Everything we do, from your E-mail to bloody crusades[M]. New York: Plume Books, 2010.
- [10] WATSON J B. Psychology as the behaviorist views it[J]. Psychological Review, 1913, 20(2): 158-177.
- [11] VAZQUEZ A, RACZ B, LUKACS A, et al. Impact of non-Poissonian activity patterns on spreading processes[J]. Phys Rev Lett, 2007, 98(15): 158702.
- [12] PASTOR-SATORRAS R, VESPIGNANI A. Epidemic spreading in scale-free networks[J]. Phys Rev Lett, 2001, 86(14): 3200-3203.
- [13] 史定华. 复杂网络的随机刻画和演化规律[J]. 力学进展, 2008(38): 679-691.
SHI Ding-hua. Stochastic features of complex networks and their patterns in dynamic evolution[J]. Advances in Mechanics, 2008(38): 679-691.
- [14] HUFNAGEL L, BROCKMANN D, GEISEL T. Forecast and control of epidemics in a globalized world[J]. Proc Natl Acad Sci, 2004(101): 15124-15129.
- [15] HAN Xiao-pu, WANG Bing-hong, ZHOU Chang-song, et al. Scaling in the global spreading patterns of pandemic Influenza A (H1N1) and the role of control: empirical statistics and modeling[EB/OL]. <http://arxiv.org/pdf/0912.1390>.
- [16] 周涛. 网络“水军”不可放任[N]. 人民日报, 2009-12-27 (7).
ZHOU Tao. Do not indulge spammer[N]. People's Daily, 2009-12-27(7).
- [17] MOKHTARIAN P L, SALOMON I. In perpetual motion: travel behavior research opportunities and application challenges[M]. Amsterdam: Elsevier Science Press, 2002.
- [18] SHANG Ming-sheng, LÜ Lin-yuan, ZHANG Yi-cheng, et al. Empirical analysis of web-based user-object bipartite networks[J]. Europhys Lett, 2010, 90(4): 48006.
- [19] POISSON S D. Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile, Précédées des Règles générales du Calcul des Probabilités, Paris, Imprimeur-Libraire pour les Mathematiques[M]. Paris: Bachelier, 1837.
- [20] HAIGHT F A. Handbook of the Poisson distribution[M]. New York: John Wiley & Sons, 1967.
- [21] REYNOLDS P. Call center staffing: The complete, practical guide to workforce management[M]. Lebanon: The Call Centre School Press, 2003.
- [22] GREENE J H. Production and inventory control handbook[M]. New York: McGraw-Hill, 1987.
- [23] ANDERSON H R. Fixed broadband wireless system Design[M]. New York: Wiley, 2003.
- [24] «运筹学»教材编写组. 运筹学[M]. 北京: 清华大学出版社, 2005.
Edit team of operations research. Operations research[M]. Beijing: Tsinghua University Press, 2005.
- [25] 汪荣鑫. 随机过程[M]. 西安: 西安交通大学出版社, 2006.
WANG Rong-xin. Stochastic process[M]. Xi'an: Xi'an Jiaotong University Press, 2006
- [26] ECKMANN J P, MOSES E, SERGI D. Entropy of dialogues creates coherent structures in e-mail traffic[J]. Proc Natl Acad Sci, 2004, 101(40): 14333-14337.
- [27] JOHANSEN A. Probing human response times[J]. Physica A 2004, 338(1): 286-291.
- [28] MALMGREN R D, STOUFFER D B, MOTTER A E, et al. A poissonian explanation for heavy tails in e-mail communication[J]. Proc Natl Acad Sci, 2008, 105(47): 18153-18158.
- [29] ANTENEODO C, MALMGREN R D, CHIALVO D. Poissonian bursts in e-mail correspondence[J]. The European Physical Journal B, 2010, 75(3): 389-394.
- [30] STOUFFER D B, MALMGREN R D, AMARAL L A N. Log-normal statistics in e-mail communication patterns [EB/OL]. [2013-05-11]. <http://arxiv.org/pdf/physics/0605027>
- [31] WUCHTY S, UZZI B. Human communication dynamics in digital footsteps: a study of the agreement between self-reported ties and email networks[J]. PloS one, 2011, 6(11): e26972.1- e26972.8.
- [32] OLIVEIRA J G, BARABÁSI A L, Human dynamics: darwin and einstein correspondence patterns[J]. Nature, 2005, 437(27): 1251-1251.
- [33] VÁZQUEZ A, OLIVEIRA J G, DEZS Z, et al. Modeling bursts and heavy tails in human dynamics[J]. Phys Rev E, 2006, 73(3): 036127.
- [34] 李楠楠, 张宁, 周涛. 人类通信模式中基于时间统计的实证研究[J]. 复杂系统与复杂性科学, 2008, 5(3): 43-47.
LI Nan-nan, ZHANG Ning, ZHOU Tao. Empirical analysis on temporal statistics of human correspondence patterns[J]. Complex System and Complexity Science, 2008, 5(3): 43-47.
- [35] LI Nan-nan, ZHANG Ning, ZHOU Tao. Empirical analysis on temporal statistics of human correspondence patterns[J]. Physica A, 2008, 387(25): 6391-6394.
- [36] MALMGREN R D, STOUFFER D B, CAMPANHARO A

- S L O, et al. On universality in human correspondence activity[J]. *Science*, 2009, 325(5948): 1696-1700.
- [37] HONG Wei, HAN Xiao-pu, ZHOU Tao, et al. Heavy-tailed statistics in short-message communication[J]. *Chin Phys Lett*, 2009, 26(2): 4.
- [38] ZHAO Zhi-dan, XIA Hu, SHANG Ming-sheng, et al. Empirical analysis on the human dynamics of a large-scale short message communication system[J]. *Chin Phys Lett*, 2011, 28(6): 068901.
- [39] WU Ye, ZHOU Chang-song, XIAO Jing-hua, et al. Evidence for a bimodal distribution in human communication[J]. *Proc Natl Acad Sci*, 2010, 107(44): 18803-18808.
- [40] CANDIA J, GONZ LEZ M C, WANG Pu, et al. Uncovering individual and collective human dynamics from mobile phone records[J]. *J Phys A*, 2008, 41(22): 224015.
- [41] 樊超, 郭进利, 纪雅莉, 等. 基于图书借阅的人类行为标度律分析[J]. *图书情报工作*, 2010, 54(15): 35-39.
FAN Chao, GUO Jin-li, JI Ya-li, et al. Analysis of human behavior scaling law based on library loans[J]. *Library and Information Service*, 2010, 54(15): 35-39.
- [42] SCALAS E, KAIZOJI T, KIRCHLER M, et al. Waiting times between orders and trades in double-auction markets[J]. *Physica A*, 2006(366): 463-471.
- [43] WANG S C, TSENG J J, TAI C C, et al. Network topology of an experimental futures exchange[J]. *Eur Phys J B*, 2008, 62(1): 105-111.
- [44] DEZSÖ Z, ALMAAS E, LUKÁCS A, et al. Dynamics of information access on the web[J]. *Phys Rev E*, 2006, 73(6): 066132.
- [45] GONCALVES B, RAMASCO J J. Human dynamics revealed through Web analytics[J]. *Phys Rev E*, 2008, 78(2): 026123.
- [46] 赵庚升, 张宁, 周涛. 网页浏览中的标度行为研究[J]. *统计与决策*, 2009, 277(1): 18-19.
ZHAO Geng-sheng, ZHANG Ning, ZHOU Tao. Scaling behavior in web browsing[J]. *Statistics and Decision*, 2009, 277(1): 18-19.
- [47] RADICCHI F. Human activity in the web[J]. *Phys Rev E*, 2009, 80(2): 026118.
- [48] LESKOVEC J, HORVITZ E. Planetary-scale views on a large instant-messaging network[C]//In Proceedings of the 17th international Conference on World Wide Web. [S.l.]: ACM, 2008: 915-924.
- [49] CHEN Guan-xiong, HAN Xiao-pu, WANG Bing-hong. Multi-level scaling properties of instant-message communications[J]. *Physics Procedia*, 2010, 3(5): 1897-1905.
- [50] VOJNOVIC M. On mobile user behaviour patterns[C]//2008 IEEE International Zurich Seminar on Communications. [S.l.]: IEEE, 2008: 26-29.
- [51] GOETZ M, LESKOVEC J, MCGLOHON M, et al. Modeling blog dynamics[C]//Proceedings of the International Conference on Weblogs and Social Media. [S.l.]: AAAI Press, 2009.
- [52] WANG Peng, ZHOU Tao, HAN Xiao-pu, et al. Modeling correlated human dynamics[EB/OL]. [2012-12-10]. <http://arxiv.org/pdf/1007.4440>.
- [53] 任学藻, 杨紫陌, 汪秉宏. 演化网络的Mandelbrot律[J]. *电子科技大学学报*, 2011, 40(2): 164-167.
REN Xue-zao, YANG Zi-mo, WANG Bing-hong. Mandelbrot law of evolution networks[J]. *Journal of University of Electronic Science and Technology of China*, 2011, 40(2): 164-167.
- [54] WANG Peng, LEI Ting, YEUNG C H, et al. Heterogenous human dynamics in intra-and inter-day time scales[J]. *Europhys Lett*, 2011, 94(1): 18005.
- [55] GUO Jin-li, FAN Chao, GUO Z H. Weblog patterns and human dynamics with decreasing interest[J]. *Eur Phys J B*, 2011, 81(3): 341-344.
- [56] CHOI S, WETTE R. Maximum likelihood estimation of the parameters of the gamma distribution and their bias[J]. *Technometrics*, 1969, 11(4): 683-690.
- [57] SONG Ya-ding, ZHANG Chuang, WU Ming. The study of human behavior dynamics based on blogosphere[C]//Proceedings of the Web Information Systems and Mining (WISM), 2010 International Conference on. [S.l.]: IEEE, 2010: 87-91.
- [58] KAN A, CHAN J, HAYES C, et al. A time decoupling approach for studying forum dynamics[J]. *World Wide Web*, 2012, 1(1): 1-26.
- [59] BAO Yuan-yuan, XIN Zhan-hong. Human activity pattern on microblogging interaction[C]//Proceedings of the Information Management, Innovation Management and Industrial Engineering (ICIII). [S.l.]: IEEE, 2011: 303-306.
- [60] CHUN H, KWAK H, EOM Y H, et al. Comparison of online social relations in volume vs interaction: a case study of cyworld[C]//Proceedings of the Proceedings of the 8th ACM SIGCOMM Conference on Internet measurement [S.l.]: ACM, 2008: 57-70.
- [61] IPEIROTIS P G. Analyzing the amazon mechanical turk marketplace[J]. *XRDS: Crossroads, the ACM Magazine for Students*, 2010, 17(2): 16-21.
- [62] SLIJPER H, RICHTER J, SMEETS J, et al. The effects of pause software on the temporal characteristics of computer use[J]. *Ergonomics*, 2007, 50(2): 178-191.
- [63] BAEK S K, KIM T Y, KIM B J. Testing a priority-based queue model with Linux command histories[J]. *Physica A*, 2008, 387(14): 3660-3668.
- [64] PICKARD G, PAN W, RAHWAN I, et al. Time-critical social mobilization[J]. *Science*, 2011, 334(6055): 509-512.
- [65] ZHOU T, KIET H A T, KIM B J, et al. Role of activity in human dynamics[J]. *Europhys Lett*, 2008, 82(2): 28002.
- [66] HU Hai-bo, HAN Ding-yi. Empirical analysis of individual popularity and activity on an online music service system[J]. *Physica A*, 2008, 387(23): 5916-5921.
- [67] WANG Peng, XIE Xiao-yi, YEUNG C H, et al. Heterogenous scaling in the inter-event time of on-line bookmarking[J]. *Physica A*, 2011, 390(12): 2395-2400.

- [68] WANG Jin-long, GAO Ke, LI Gang. Empirical analysis of customer behaviors in Chinese e-commerce[J]. *J Networks*, 2010, 5(10): 1177-1184.
- [69] BOHORQUEZ J C, GOURLEY S, DIXON A R, et al. Common ecology quantifies human insurgency[J]. *Nature*, 2009, 462(7275): 911-914.
- [70] TANG Da-hai, HAN Xiao-pu, WANG Bing-hong. Stretched exponential distribution of recurrent time of wars in China[J]. *Physica A*, 2010, 389(13): 2637-2641.
- [71] ZHU Jun-fang, HAN Xiao-pu, WANG Bing-hong. Statistical property and model for the inter-event time of terrorism attacks[J]. *Chin Phys Lett*, 2010, 27(6): 068902.
- [72] LAMBIOTTE R, AUSLOOS M, THELWALL M. Word statistics in Blogs and RSS feeds: Towards empirical universal evidence[J]. *J Informetrics*, 2007, 1(4): 277-286.
- [73] LI Na, YAN Han, ZHANG Wen-yao, et al. Temporal statistical analysis on human article creation patterns[EB/OL]. [2013-01-15]. <http://arxiv.org/pdf/1108.0833>.
- [74] HUI P, CHAINTREAU A, SCOTT J, et al. Pocket switched networks and human mobility in conference environments [C]//Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-Tolerant Networking. [S.l.]: ACM, 2005: 244-251.
- [75] SCHERRER A, BORGNAT P, FLEURY E, et al. Description and simulation of dynamic mobility networks [J]. *Computer Networks*, 2008, 52(15): 2842-2858.
- [76] EAGLE N, PENTLAND A. Reality mining: Sensing complex social systems[J]. *Personal and Ubiquitous Computing*, 2005, 10(4): 255-268.
- [77] BARRAT A, CATTUTO C, COLIZZA V, et al. High resolution dynamical mapping of social interactions with active RFID[EB/OL]. [2013-01-15]. <http://arxiv.org/pdf/0811.4170>.
- [78] NEYLON C, CATTUTO C, VAN DEN BROECK W, et al. Dynamics of person-to-person interactions from distributed RFID sensor networks[J]. *PLoS One*, 2010, 5(7): e11596.
- [79] ZHAO K, STEHL J, BIANCONI G, et al. Social network dynamics of face-to-face interactions[J]. *Phys Rev E*, 2011, 83(5): 056109.
- [80] TAKAGUCHI T, NAKAMURA M, SATO N, et al. Predictability of conversation partners[J]. *Phys Rev X*, 2011, 1(1): 011008.
- [81] RYBSKI D, BULDYREV SV, HAVLIN S, et al. Scaling laws of human interaction activity[J]. *Proc Natl Acad Sci*, 2009, 106(31): 12640-12645.
- [82] EISLER Z, BARTOS I, KERT SZ J. Fluctuation scaling in complex systems: Taylor's law and beyond[J]. *Adv Phys*, 2008, 57(1): 89-142.
- [83] JIANG Zhi-qiang, CHEN Wei, ZHOU Wei-xing. Detrended fluctuation analysis of intertrade durations[J]. *Physica A*, 2009, 388(4): 433-440.
- [84] CRANE R, SORNETTE D. Robust dynamic classes revealed by measuring the response function of a social system[J]. *Proc Natl Acad Sci USA*, 2008, 105(41): 15649-15653.
- [85] 赵飞, 刘金虎, 查一龙, 等. 在线协同写作的人类动力学分析[J]. *物理学报*, 2011, 60(11): 118902.
- ZHAO Fei, LIU Jin-hu, ZHA Yi-long, et al. Human dynamics analysis in online collaborative writing[J]. *Acta Phys Sin*, 2011, 60(11): 118902.
- [86] SUMI R, YASSERI T, RUNG A, et al. Edit wars in wikipedia[C]In Proceedings of 2011 IEEE Third International Conference on and 2011 IEEE third International Conference on Social Computing (SocialCom). [S.l.]: IEEE, 2011: 724-727.
- [87] KUJAWSKI B, HOLYST J, RODGERS G J. Growing trees in internet news groups and forums[J]. *Phys Rev E*, 2007, 76(3): 036103.
- [88] RATKIEWICZ J, FORTUNATO S, FLAMMINI A, et al. Characterizing and modeling the dynamics of online popularity[J]. *Phys Rev Lett*, 2010, 105(15): 158701.
- [89] LI Hui. Workload dynamics on clusters and grids[J]. *The Journal of Supercomputing*, 2009, 47(1): 1-20.
- [90] MRYGLOD O, HOLOVATCH Y. Towards journalometrical analysis of a scientific periodical: a case study[J]. *Condensed Matter Phys*, 2007(10):129.
- [91] MRYGLOD O, HOLOVATCH Y, MRYGLOD I. Editorial process in scientific journals: analysis and modeling[J]. *Scientometrics*, 2012, 91(1): 101-112.
- [92] WU Ye, ZHOU Chang-song, CHEN Mao-ying, et al. Human comment dynamics in on-line social systems[J]. *Physica A*, 2010, 389(24): 5832-5837.
- [93] GOH K I, BARABÁSI A L. Burstiness and memory in complex systems[J]. *Europhys Lett*, 2008, 81(4): 48002.
- [94] BUNDE A, EICHNER J, KANTELHARDT J, et al. Long-Term Memory: a natural mechanism for the clustering of extreme events and anomalous residual times in climate records[J]. *Phys Rev Lett*, 2005, 94(4): 048701.
- [95] LIVINA VN, HAVLIN S, BUNDE A. Memory in the occurrence of earthquakes[J]. *Phys Rev Lett*, 2005, 95(20): 208501.
- [96] CAI Shi-min, FU Zhong-qian, ZHOU Tao, et al. Scaling and memory in recurrence intervals of Internet traffic[J]. *Europhys Lett*, 2009, 87(6): 68001.
- [97] YU Jie-fei, HU Yan-qing, YU Min, et al. Analyzing netizens' view and reply behaviors on the forum[J]. *Physica A*, 2010, 389(16): 3267-3273.
- [98] NEWMAN MEJ. Assortative mixing in networks[J]. *Phys Rev Lett*, 2002, 89(20): 208701.
- [99] 史定华. 无标度网络: 基础理论和应用研究[J]. *电子科技大学学报*, 2010, 39(5): 644-650.
- SHI Ding-hua. Scale-free networks: Basic theory and applied research[J]. *Journal of University of Electronic Science and Technology of China*, 2010, 39(5): 644-650.
- [100] ZHOU Tao, ZHAO Zhi-dan, YANG Zi-mo, et al. Relative clock verifies endogenous bursts of human dynamics[J]. *Europhys Lett*, 2012, 97(1): 18006.
- [101] YASSERI T, SUMI R, KERT S Z J. Circadian patterns of wikipedia editorial activity: a demographic analysis[J].

- PLoS ONE, 2012, 7(1): e30091.
- [102]JIANG Zhi-qiang, REN Fei, GU Gao-feng, et al. Statistical properties of online avatar numbers in a massive multiplayer online role-playing game[J]. *Physica A*, 2010, 389(4): 807-814.
- [103]HIDALGO C A. Conditions for the emergence of scaling in the inter-event time of uncorrelated and seasonal systems[J]. *Physica A*, 2006, 369(2): 877-883.
- [104]ZHAO Zhi-dan, ZHOU Tao. Empirical analysis of online human dynamics[J]. *Physica A*, 2012, 391(11): 3308-3315.
- [105]GRABOWSKI A, KRUSZEWSKA N, KOSIŃSKI R. Dynamic phenomena and human activity in an artificial society[J]. *Phys Rev E*, 2008, 78(6): 066110.
- [106]GRABOWSKI A, KRUSZEWSKA N, KOSIŃSKI R. Properties of on-line social systems[J]. *Eur Phys J B*, 2008, 66(1): 107-113.
- [107]CHMIEL A, SOBKOWICZ P, SIENKIEWICZ J, et al. Negative emotions boost user activity at BBC forum[J]. *Physica A*, 2011, 390(16): 2936-2944.
- [108]GROSS D. Fundamentals of queueing theory[M]. [S.l.]: John Wiley & Sons, 2008.
- [109]VÁZQUEZ A. Exact results for the barabási model of human dynamics[J]. *Phys Rev Lett*, 2005, 95(24): 248701.
- [110]GABRIELLI A, CALDARELLI G. Invasion percolation and critical transient in the barabási model of human dynamics[J]. *Phys Rev Lett*, 2007, 98(20): 208701.
- [111]ANTENEDO C. Exact results for the Barabási queueing model[J]. *Phys Rev E*, 2009, 80(4): 041131.
- [112]MASUDA N, KIM J, KAHNG B. Priority queues with bursty arrivals of incoming tasks[J]. *Phys Rev E*, 2009, 79(3): 036106.
- [113]GUO Jin-li. A model of human behavior dynamics and exact results[J]. *Acta Phys Sinica*, 2010, 59(6): 3851-3855.
- [114]CAJUEIRO D O, MALDONADO W L. Role of optimization in the human dynamics of task execution[J]. *Phys Rev E*, 2008, 77(3): 035101.
- [115]OLIVEIRA J, VAZQUEZ A. Impact of interactions on human dynamics[J]. *Physica A*, 2009, 388(2): 187-192.
- [116]MIN B, GOH K I, KIM I M. Waiting time dynamics of priority-queue networks[J]. *Phys Rev E*, 2009, 79(5): 056110.
- [117]VAZQUEZ A. Impact of memory on human dynamics[J]. *Physica A*, 2007(373): 747-752.
- [118]HAN Xiao-pu, ZHOU Tao, WANG Bing-hong. Modeling human dynamics with adaptive interest[J]. *New J Phys*, 2008, 10(7): 073010.
- [119]SHANG M S, CHEN G X, DAI S X, et al. Interest-driven model for human dynamics[J]. *Chin Phys Lett*, 2010(27): 048701.
- [120]CHMIEL A, KOWALSKA K, HOLYST J A. Scaling of human behavior during portal browsing[J]. *Phys Rev E*, 2009, 80(6): 066122.
- [121]HAN Xiao-pu, HAO Qiang, WANG Bing-hong, et al. Origin of the scaling law in human mobility: Hierarchy of traffic systems[J]. *Phys Rev E*, 2011, 83(3): 036117.
- [122]JO H H, MOON E, KASKI K. Optimized reduction of uncertainty in bursty human dynamics[J]. *Phys Rev E*, 2012, 85(1): 016102.
- [123]KIM S. Self-organized pinning and interface growth in a random medium[J]. *Phys Rev Lett*, 1992, 69(24): 3539-3542.
- [124]BAK P, SNEPPEN K. Punctuated equilibrium and criticality in a simple model of evolution[J]. *Phys Rev Lett*, 1993, 71(24): 4083-4086.
- [125]STEHL J, BARRAT A, BIANCONI G. Dynamical and bursty interactions in social networks[J]. *Phys Rev E*, 2010, 81(3): 035101.
- [126]JO H H, PAN R K, KASKI K. Emergence of bursts and communities in evolving weighted networks[J]. *PloS One*, 2011, 6(8): e22687.
- [127]KUMPULA J M, ONNELA J-P, SARAM K J, et al. Emergence of communities in weighted networks[J]. *Phys Rev Lett*, 2007, 99(22): 228701.
- [128]SORRIBES A, ARMENDARIZ BG, LOPEZ-PIGOZZI D, et al. The origin of behavioral bursts in decision-making circuitry[J]. *PLoS Comp Bio*, 2011, 7(6): e1002075.
- [129]MEYER M D, MILLER E J. Urban transportation planning: a decision-oriented approach[M]. New York: McGraw-Hill, 2001.
- [130]MONTROLL E W, WEISS G H. Random walks on lattices II[J]. *J Math Phys*, 1965(6): 167-182.
- [131]GONZÁLEZ M C, HIDALGO C A, BARABÁSI A L. Understanding individual human mobility patterns[J]. *Nature*, 2008, 453(7196): 779-782.
- [132]SONG C, KOREN T, WANG P, et al. Modelling the scaling properties of human mobility[J]. *Nat Phys*, 2010(6): 818-823.
- [133]SONG C, QU Z, BLUMM N, et al. Limits of predictability in human mobility[J]. *Science*, 2010, 327(5968): 1018-1021.
- [134]RHEE I, SHIN M, HONG S, et al. On the levy-walk nature of human mobility[C]//Proceedings of the INFOCOM'08. [S.l.]: IEEE, 2008: 924-932.
- [135]BARTH L M. Spatial networks[J]. *Phys Rep*, 2011, 499(1): 1-101.
- [136]JIANG Bin, YIN Jun-jun, ZHAO Si-jian. Characterizing the human mobility pattern in a large street network[J]. *Phys Rev E*, 2009, 80(2): 021136.
- [137]LIANG Xiao, ZHENG Xu-dong, LV Wei-feng, et al. The scaling of human mobility by taxis is exponential[J]. *Physica A*, 2012, 391(5): 2135-2144.
- [138]PENG Cheng-bin, JIN Xiao-gang, WONG K-C, et al. Collective human mobility pattern from taxi trips in urban area[J]. *PloS One*, 2012, 7(4): e34487.
- [139]RAMBALDI S, BAZZANI A, GIORGINI B, et al. Mobility in modern cities: looking for physical laws[C]//Proceedings of the ECCS. [S.l.]: ECCS, 2007: 132-141.
- [140]BAZZANI A, GIORGINI B, RAMBALDI S, et al. Statistical laws in urban mobility from microscopic GPS

- data in the area of Florence[J]. *J Stat Mech*, 2010(5): P05001.
- [141] RICCARDO G, ARMANDO B, SANDRO R. Towards a statistical physics of human mobility[J]. *Int J Mod Phys C*, 2012, 23(09): 1250061.
- [142] JIANG Bin, JIA Tao. Exploring human mobility patterns based on location information of US flights[EB/OL]. [2013-01-15]. <http://arxiv.org/pdf/1104.4578>.
- [143] ROTH C, KANG S M, BATTY M, et al. Structure of urban movements: polycentric activity and entangled hierarchical flows[J]. *PLoS One*, 2011, 6(1): e15923.
- [144] 王明生, 黄琳, 闫小勇. 探索城市公交客流移动模式[J]. *电子科技大学学报*, 2012, 41(1): 2-7.
WANG Ming-sheng, HUANG Lin, YAN Xiao-yong. Exploring the mobility patterns of public transport passengers[J]. *Journal of University of Electronic Science and Technology of China*, 2012, 41(1): 2-7.
- [145] PETROVSKII S, MASHANOVA A, JANSEN V A. Variation in individual walking behavior creates the impression of a Lévy flight[J]. *Proc Natl Acad Sci USA*, 2011, 108(21): 8704-8707.
- [146] 闫小勇. 人类个体出行行为的统计实证[J]. *电子科技大学学报*, 2011, 40(2): 168-173.
YAN Xiao-yong. Empirical statistics on individual human travel behavior[J]. *Journal of University of Electronic Science and Technology of China*, 2011, 40(2): 168-173.
- [147] YAN Xiao-yong, HAN Xiao-pu, WANG Bing-hong, et al. Diversity of individual mobility patterns[EB/OL]. [2013-01-15]. <http://arxiv.org/pdf/1211.2874>.
- [148] CHALASANI V, ENGEBRETSEN O, DENSTADLI J, et al. Precision of geocoded locations and network distance estimates[J]. *J Transport Stat*, 2005, 8(2): 1-15.
- [149] MAHMASSANI H S. In perpetual motion: travel behavior research opportunities and application challenges[M]. Amsterdam: Elsevier Science Press, 2002.
- [150] BROCKMANN D. Statistical mechanics: The physics of where to go[J]. *Nat Phys*, 2010, 6(10): 720-721.
- [151] VISWANATHAN G, BULDYREV SV, HAVLIN S, et al. Optimizing the success of random searches[J]. *Nature*, 1999, 401(6756): 911-914.
- [152] HU Yan-qing, ZHANG Jiang, HUAN Di, et al. Toward a general understanding of the scaling laws in human and animal mobility[J]. *Europhys Lett*, 2011, 96(3): 38006.
- [153] SCHLICH R, AXHAUSEN K W. Habitual travel behaviour: Evidence from a six-week travel diary[J]. *Transportation*, 2003(30): 13-36.
- [154] EAGLE N, PENTLAND A S. Eigenbehaviors: identifying structure in routine[J]. *Behav Ecol Sociobiol*, 2009, 63(7): 1057-1066.
- [155] YAN Xiao-yong, HAN Xiao-pu, ZHOU Tao, et al. Exact solution of the gyration radius of an individual's trajectory for a simplified human regular mobility model[J]. *Chin Phys Lett*, 2011, 28(12): 120506.
- [156] DE DIOS ORT ZAR J, WILLUMSEN L G. Modelling transport[M]. New York: Wiley, 2001.
- [157] RIETVELD P, ZWART B, VAN WEE B, et al. On the relationship between travel time and travel distance of commuters[J]. *Ann Reg Sci*, 1999, 33(3): 269-287.
- [158] LEE K, HONG S, KIM S J, et al. Slaw: a new mobility model for human walks[C]//Proceedings of the INFOCOM 2009. [S.l.]: IEEE, 2009: 855-863.
- [159] HAN Xiao-pu, WANG Bing-hong. Impacts of distance and memory in the emergence of scaling mobility pattern of human[J]. *Physics Procedia*, 2010, 3(5): 1907-1911.
- [160] SIMINI F, GONZ LEZ M C, MARITAN A, et al. A universal model for mobility and migration patterns[J]. *Nature*, 2012, 484(7392): 96-100.
- [161] MA Zhi-en, ZHOU Yi-cang, WU Jian-hong. Modeling and dynamics of infectious diseases[M]. Beijing: Higher Education Press, 2009.
- [162] WENDI W, ZHIEN M. Global dynamics of an epidemic model with time delay[J]. *Nonlinear Analysis: Real World Applications*, 2002, 3(3): 365-373.
- [163] NI Shun-jiang, WENG Wen-guo. Impact of travel patterns on epidemic dynamics in heterogeneous spatial metapopulation networks[J]. *Phys Rev E*, 2009, 79(1): 016111.
- [164] MIN B, GOH KI, VAZQUEZ A. Spreading dynamics following bursty human activity patterns[J]. *Phys Rev E*, 2011, 83(3): 036102.
- [165] FELLER W. An introduction to probability theory and its applications[M]. New York: John Wiley & Sons, 2008.
- [166] YANG Z, CUI A-X, ZHOU T. Impact of heterogeneous human activities on epidemic spreading[J]. *Physica A*, 2011, 390(23-24): 4543-4548.
- [167] IRIBARREN J L, MORO E. Impact of human activity patterns on the dynamics of information diffusion[J]. *Phys Rev Lett*, 2009, 103(3): 038702.
- [168] KARSAI M, KIVEL M, PAN R, et al. Small but slow world: How network topology and burstiness slow down spreading[J]. *Phys Rev E*, 2011, 83(2): 025102.
- [169] WATTS D J. Multiscale, resurgent epidemics in a hierarchical metapopulation model[J]. *Proc Natl Acad Sci USA*, 2005, 102(32): 11157-11162.
- [170] BALCAN D, VESPIGNANI A. Phase transitions in contagion processes mediated by recurrent mobility patterns[J]. *Nature physics*, 2011, 7(7): 581-586.
- [171] BELIK V, GEISEL T, BROCKMANN D. Natural human mobility patterns and spatial spread of infectious diseases[J]. *Phys Rev X*, 2011, 1(1): 011001.
- [172] BALCAN D, COLIZZA V, GONCALVES B, et al. Multiscale mobility networks and the spatial spreading of infectious diseases[J]. *Proc Natl Acad Sci U.S.A.*, 2009, 106(51): 21484-21489.
- [173] AJARDI P, POLETTI C, RAMASCO J J, et al. Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic[J]. *PLoS ONE*, 2011(6): e16591.
- [174] ZHAO Zhi-dan, LIU Ying, TANG Ming. Epidemic variability in hierarchical geographical networks with

- human activity patterns[J]. *Chaos*, 2012, 22(2): 023150.
- [175] HAN Xiao-pu, ZHAO Zhi-dan, HADZIBEGANOVIC T, et al. Epidemic patterns and long-range mobility control on hierarchical geographical networks[EB/OL]. <http://arxiv.org/pdf/1302.2508>.
- [176] WANG P, GONZALEZ M C, HIDALGO C A, et al. Understanding the spreading patterns of mobile phone viruses[J]. *Science*, 2009, 324(5930): 1071-1076.
- [177] 田宝玉, 周炯磐, 吴伟陵. 工程信息论[M]. 北京: 北京邮电大学出版社, 2004.
- TIAN Bao-yu, ZHOU Jiong-pan, WU Wei-ling. *Engineering information theory*[M]. Beijing: Beijing University of Posts and Telecommunications Press, 2004
- [178] ETTER V, KAFSI M, KAZEMI E. Been there, done that: what your mobility traces reveal about your behavior[C]// *Proceedings of the Nokia Mobile Data Challenge 2012 Workshop*. Nokia. [S.l.]: [s.n.], 2012: 1-6.
- [179] 朱寅, 杨强. 诺基亚移动数据挖掘竞赛[J]. *计算机学会通讯*, 2012, 8(8): 67-70.
- ZHU Yin, YANG Qiang. Nokia mobile data challenge[J]. *Communications of China Computer Federation*, 2012, 8(8): 67-70.
- [180] ZHENG V W, ZHENG Y, XIE X, et al. Towards mobile intelligence: Learning from GPS history data for collaborative recommendation[J]. *Artif Intell*, 2012(184): 17-37.
- [181] LU L, MEDO M, YEUNG C H, et al. Recommender Systems[J]. *Phys Rep*, 2012, 519(1): 1-49.
- [182] ZHANG Cheng-jun, Zeng An. Behavior patterns of online users and the effect on information filtering[J]. *Physica A*, 2012, 391(4): 1822-1830.
- [183] ZHOU T, KUSCSIK Z, LIU J-G, et al. Solving the apparent diversity-accuracy dilemma of recommender systems[J]. *Proc Natl Acad Sci USA*, 2010, 107(10): 4511-4515.
- [184] MIN S H, HAN I. Detection of the customer time-variant pattern for improving recommender systems[J]. *Expert Syst Appl*, 2005, 28(2): 189-199.
- [185] XIANG Liang, YUAN Quan, ZHAO Shi-wan, et al. Temporal recommendation on graphs via long-and short-term preference fusion[C]// *Proceedings of the Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.]: ACM, 2010: 723-732.
- [186] LIU NN, ZHAO M, XIANG E, et al. Online evolutionary collaborative filtering[C]// *Proceedings of the Proceedings of the fourth ACM Conference on Recommender Systems*. [S.l.]: ACM, 2010: 95-102.
- [187] LIU Ji, DENG Gui-shi. Link prediction in a user-object network based on time-weighted resource allocation[J]. *Physica A: Statistical Mechanics and its Applications*, 2009, 388(17): 3643-3650.
- [188] KOREN Y. Collaborative filtering with temporal dynamics [C]// *Proceedings of the proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and data mining*. [S.l.]: ACM, 2009: 447-456.
- [189] CHO E, MYERS S A, LESKOVEC J. Friendship and mobility: user movement in location-based social networks[C]// *Proceedings of the Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.]: ACM, 2011: 1082-1090.
- [190] ZHENG V W, ZHENG Y, XIE X, et al. Collaborative location and activity recommendations with GPS history data[C]// *Proceedings of the Proceedings of the 19th International Conference on World Wide web*. [S.l.]: ACM, 2010: 1029-1038.
- [191] CLEMENTS M, SERDYUKOV P, DE VRIES A P, et al. Personalised travel recommendation based on location co-occurrence[EB/OL]. [2013-01-15]. <http://arxiv.org/pdf/1106.5213>.
- [192] DAO TH, JEONG SR, AHN H. A novel recommendation model of location-based advertising: Context-Aware Collaborative Filtering using GA approach[J]. *Expert Syst Appl*, 2012, 39(3): 3731-3739.
- [193] SCELLATO S, NOULAS A, MASCOLO C. Exploiting place features in link prediction on location-based social networks[C]// *Proceedings of the Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.]: ACM, 2011: 1046-1054.
- [194] KOREN Y, SILL J. OrdRec: an ordinal model for predicting personalized item rating distributions[C]// *Proceedings of the Proceedings of the fifth ACM Conference on Recommender Systems*. [S.l.]: ACM, 2011: 117-124.
- [195] VIG J, SEN S, RIEDL J. Navigating the tag genome[C]// *Proceedings of the Proceedings of the 16th International Conference on Intelligent User Interfaces*. [S.l.]: ACM, 2011: 93-102.
- [196] CHEN L, PU P. Critiquing-based recommenders: survey and emerging trends[J]. *User Model User-adap*, 2012, 22(1-2): 125-150.
- [197] 刘译璟, 张韶峰, 周涛. 推荐引擎: 信息暗海的领航员[J]. *中国计算机学会通讯*, 2012, 8(6): 22-25.
- LIU Yi-jing, ZHANG Shao-feng, Zhou Tao. Recommender: pilot for the dark sea of information[J]. *Communications of China Computer Federation*, 2012, 8(6): 22-25.
- [198] YANG Zi-mo, ZHANG Zi-ke, ZHOU Tao. Anchoring bias in online voting[J]. *Europhys Lett*, 2012, 100(6): 68002.1-68002.6.
- [199] TVERSKY A, KAHNEMAN D. Judgment under uncertainty: Heuristics and biases[J]. *Science*, 1974, 185(4157): 1124-1131.
- [200] FURNHAM A, BOO H C. A literature review of the anchoring effect[J]. *The Journal of Socio-Economics*, 2011, 40(1): 35-42.
- [201] GOLDBERG D, NICHOLS D, OKI BM, et al. Using collaborative filtering to weave an information tapestry[J]. *Commun Acn*, 1992, 35(12): 61-70.

- [202] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms[C]// Proceedings of the Proceedings of the 10th international conference on World Wide Web. [S.l.]: ACM, 2001: 285-295.
- [203] CACHEDA F, CARNEIRO V, FERN N D, et al. Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems[J]. ACM Transactions on the Web (TWEB), 2011, 5(1): 1-33.
- [204] HOLME P, SARAM KI J. Temporal networks[J]. Physics reports, 2012, 519(3): 97-125.
- [205] DE MONTJOYE Y A, HIDALGO C A, VERLEYSSEN M, et al. Unique in the crowd: the privacy bounds of human mobility[J]. Scientific Reports, 2013(3): 1376.1-1376.4.
- [206] KOSINSKI M, STILLWELL D, GRAEPEL T. Private traits and attributes are predictable from digital records of human behavior[J]. Proc Natl Acad Sci USA, 2013, 110(15): 5802-5805.

编辑 蒋晓