

# 人类个体出行行为的统计实证

闫小勇

(石家庄铁道大学交通运输学院 石家庄 050043; 中国科学技术大学近代物理系 合肥 230026)

**【摘要】**统计了欧洲某城市230位居民在6周内的出行活动中的停留时间分布和出行距离分布。首先假设人类出行的停留时间和距离在群体和个体水平上都符合幂律分布,然后用最大似然估计法对这些分布的幂指数进行估计,最后用Kolmogorov-Smirnov检验法对假设的真伪性进行检验。结果发现:在停留时间分布方面,无论在群体水平还是个体水平上都不具有幂律分布特征,而是以相对幂律分布更高的概率进行长时间停留;在出行距离分布方面,群体的出行距离服从带有指数截断的幂律分布,但有198个个体的出行距离分布并不具有幂律特征,而是在某个特征性距离上出现峰值。对这些统计特征的形成机制进行了解释。

**关键词** 实证统计; 人类动力学; 人类移动性; 出行模式

**中图分类号** O414.2

**文献标识码** A

doi:10.3969/j.issn.1001-0548.2011.02.002

## Empirical Statistics on Individual Human Travel Behavior

YAN Xiao-yong

(Department of Transportation Engineering, Shijiazhuang Tiedao University Shijiazhuang 050043;

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

**Abstract** Empirical statistics of the waiting time distributions and the travel distance distributions of 230 residents in an European city for a six-week period is presented. On the hypothesis that the waiting time distributions and the travel distance distributions are power-law at both population and individual levels, the maximum likelihood method is used to estimate the parameters of the distributions. And then the Kolmogorov-Smirnov method is used to test the hypotheses. The results show that the waiting time distributions, neither at population level nor at individual level, are power-law. The probability of that an individual staying at a location for a long time is larger than the prediction of the power-law distribution. The results also show that the travel distance distribution of human can be fitted with a truncated power-law at population level. Yet, at individual level, 198 users have peaked travel distance distributions, peaking at the distance between his or her two most visited locations. The mechanisms of these statistical characteristics are explained.

**Key words** empirical statistics; human dynamics; human mobility; travel patterns

人类个体行为是隐藏在许多复杂社会经济现象背后的驱动力,定量理解人类行为是现代科学的一个重要研究课题<sup>[1-3]</sup>。在人类动力学研究领域,科学家已通过大量的实证统计发现一些人类行为,如邮件通讯<sup>[1]</sup>、网页浏览<sup>[4]</sup>及在线电影点播<sup>[5]</sup>等的时间间隔近似服从幂律分布,这种幂律分布特性无论是在群体水平上还是在个体水平上都可以得到证实<sup>[5]</sup>。一些解释人类行为时间间隔幂律分布形成机制的模型<sup>[1,6-8]</sup>也被陆续提出,这些研究结果不断深化着人类对自身行为规律的理解。

除了发现人类行为的时间间隔分布中广泛存在的幂律现象之外,近年来的实证研究还发现在人类

的空间运动行为中也存在着幂律分布特性。在刻画人类空间运动模式方面,停留时间分布和出行距离分布是两个关键的统计特征。其中,停留时间分布对人类个体的空间扩散速度有决定性影响<sup>[9-11]</sup>,而出行距离分布则直接反映城市的经济性和效率<sup>[12]</sup>,并对传染病的传播范围和群体感染率等有显著影响<sup>[13]</sup>。通过对美元流通数据<sup>[9]</sup>、手机通信记录<sup>[10-11]</sup>以及GPS数据<sup>[14]</sup>等进行实证统计,人们已经发现人类出行过程中的停留时间和出行距离都近似服从幂律分布。但是,这些统计都是在群体水平上进行的,目前仍缺乏在个体水平上的统计分析。

文献[11]认为,根据在群体水平上统计得到的人

收稿日期: 2011-01-28

基金项目: 国家自然科学基金(10975126); 高校博士点专项基金(20093402110032); 河北省教育厅科研计划(Z2009139)

作者简介: 闫小勇(1980-),男,硕士,主要从事交通系统复杂性与人类动力学方面的研究。

类出行距离和停留时间服从幂律分布，可以推断出群体中的每个个体的出行距离和停留时间也同样服从幂律分布。但文献[15]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幂律分布不能推断出个体也服从幂律分布。尽管文献[11]已经将个体的出行距离和停留时间服从幂律分布作为基本假设构建人类个体的空间运动模型，但这些假设的合理性既未得到理论证明，也缺乏实证统计结果的支持。为了更清晰地理解人类个体的出行模式，有必要在个体水平上对人类出行行为的统计特征进行实证统计。

本文对欧洲某城市230位居民在6周内的出行日志所记录的数据进行统计分析，结果发现：无论在群体水平还是个体水平上，人类出行的停留时间分布都不具有幂律特征，而是以相对幂律分布更高的概率进行长时间停留；在出行距离分布方面，群体的出行距离服从带有指数截断的幂律分布，但在个体层面上，有198位居民的出行距离分布并不具有幂律特征，而是在某个特征性距离上出现峰值——该特征性距离一般是个体的家到工作单位(或学校)之间的距离。这些结果说明，对于人类个体的停留时间与出行距离服从幂律分布的假设是值得商榷的。

## 1 数据说明

本文所分析的数据来源于在瑞士弗劳恩费尔德市所进行的出行日志调查数据<sup>[16]</sup>。出行日志被认为是目前技术水平下大量采集城市居民出行行为数据最为有效的方法之一<sup>[17]</sup>，它主要通过问卷调查表的形式，收集居民在一段时期内连续的出行信息。本文中，出行的概念是指人在两个不同的活动地点(如家庭、单位、学校、商场等)之间的一次移动，在活动地点之内的小范围移动并不被认为是出行。

以往对人类空间运动模式所进行的实证研究所采用的数据多属于媒介数据<sup>[18]</sup>，这些数据不能精确反映人类，特别是个体层面的出行行为。例如，美元流通数据反映了钞票在不同人手中的传递过程，而手机通信数据中记录的用户位置只是用户通信时所处的位置，并不一定是用户的出行目的地(比如用户在路途中接打电话或收发短信)。与这些媒介数据相比，出行日志数据可以更精确地反映人类个体的出行行为。

本文所分析的数据集中记录了230位城市居民在6周内的出行情况，所记录的总出行次数为36 761

次，平均每位居民的出行次数为159.83次，人均日出行次数为3.81次。每条出行记录包含了该次出行的起讫点位置、出发和到达时刻以及出行目的等。其中，起讫点位置由经纬度坐标定义，单位为秒，出发与到达时刻的单位为分。通过起讫点坐标可计算出每次出行的空间直线距离，根据前一次出行的到达时刻与下一次出行的出发时刻的间隔时间，则可以计算出个体在不同活动地点上的停留时间。此外，该数据集中还记录了每位居民的年龄、性别、职业等个人信息。

## 2 停留时间分布的统计结果

### 2.1 群体水平上的统计结果

图1a给出了数据集中所有个体在活动地点的停留时间分布，图中横坐标为停留时间，纵坐标为概率密度。从图中可以看出，尽管群体进行短时间( $< 5$  h)停留的概率远高于长时间( $> 5$  h)停留的概率，但是相对于幂律分布而言，真实的停留时间分布中长时间停留的概率显然更高。事实上，如果假设一个分布是幂律的，总可以通过最大似然估计得到一个幂指数，该假设的真伪性可以通过Kolmogorov-Smirnov检验<sup>[19]</sup>(以下简称K-S检验)进行判断。本文在群体水平上统计的停留时间间隔分布为幂律分布的假设无法通过阈值为0.9的K-S检验，说明群体的停留时间分布并不符合幂律分布。

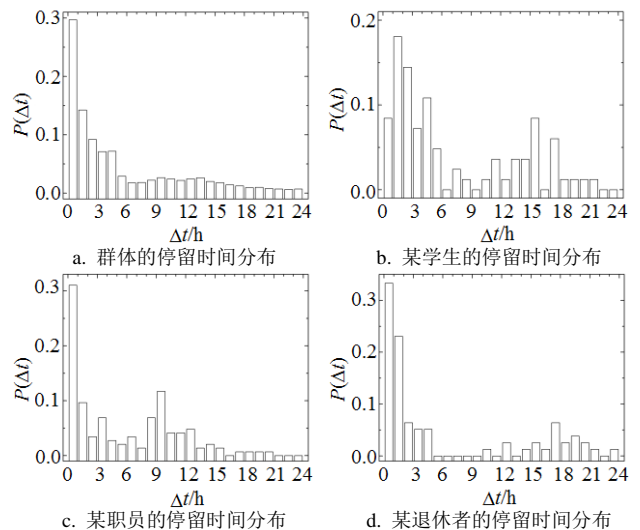


图1 停留时间分布

### 2.2 个体水平上的统计结果

本文对群体中每个个体的停留时间进行统计分析，发现所有个体的停留时间分布为幂律分布的假设都无法通过阈值为0.9的K-S检验。由于个体数量过多，此处不列出每个个体的停留时间分布统计结

果, 只挑选3个典型的个体(分别从学生、职员和退休者中选取, 为保持分析的一致性和可比性, 后续图中列举的个体与此处相同), 并给出他们的停留时间分布统计结果, 分别见图1b~图1d。从图中可以看出, 个体同样具有以相较幂律分布更高的概率进行长时间停留的特征(在后段出现明显的峰值), 而且该特征与群体水平上的统计结果相比更为显著。显然, 在个体水平上统计的停留时间是不符合幂律分布的。

### 3 对停留时间分布统计结果的解释

为解释人类个体为何具有以高概率进行长时间停留的特征, 本文对每个个体在统计时段内的日常活动时间序列进行了分析。图2是一个典型个体在一周内的活动情况, 图中的空白区域表示个体在某地点的停留, 黑色竖线表示在不同地点间的出行。横轴的长度与时间成正比。从图中可以看出, 该个体每天夜间都会长时间地停留在某一地点而不进行出行活动。显然, 这是由于人的生理周期性所决定的——绝大多数的人每天晚上都会在家里休息。虽然某一个体每天晚上在家里休息的时间可能不一样, 但这一时间会基本保持稳定, 而且一般都会长于在其他地点停留的时间。另外, 个体每天的出行次数是有限的(平均为3.81次), 平均意义上个体每天的停留地点不会超过4个。换句话说, 在个体的停留时间样本中, 在家的长时间停留所占的比例超过了1/4。以上这些原因使得个体的停留时间分布曲线在后段出现明显的峰值效应, 表现出强烈的非幂律分布特征。



图2 某个体在一周内的活动情况

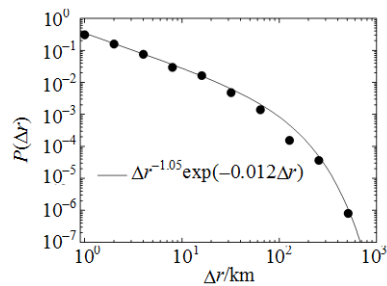
由于不同个体在家休息的时间并不相同(例如小学生可能会很长, 而夜生活丰富的人可能会短一些), 当把不同个体聚集在一起, 在群体水平上统计停留时间分布时, 分布曲线后段的峰值效应会变得不显著。当个体数量达到一定程度时, 在群体上统计的停留时间分布可能能够用幂函数近似, 这或许是之前在群体水平上<sup>[9-11]</sup>统计的停留时间分布具有幂律特性的原因之一。但无论如何, 在个体层面上假设人类出行的停留时间服从幂律分布是不合理的。

## 4 出行距离分布的统计结果

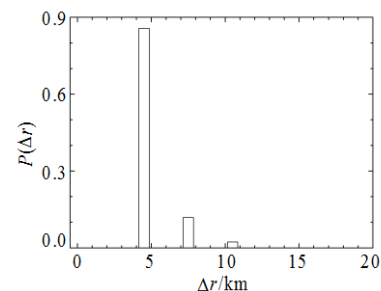
### 4.1 群体水平上的统计结果

图3中横坐标为出行距离, 纵坐标为概率密度。图3a给出了数据集中所有个体的出行距离分布, 该

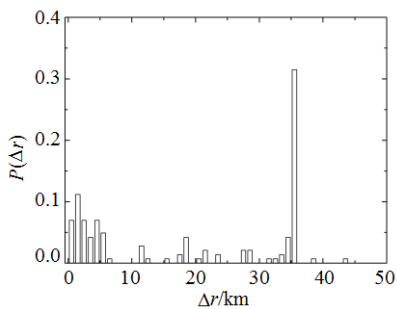
分布近似服从带有指数截断的幂律分布  $P(\Delta r) \sim \Delta r^{-1.05} \exp(-0.012\Delta r)$ , 能够通过阈值为0.9的K-S检验。本文的统计结果与文献[10-11]中得到的结果类似, 暗示着群体水平上出行距离分布的幂律特征反映了人类群体运动行为中的某种共同机制。



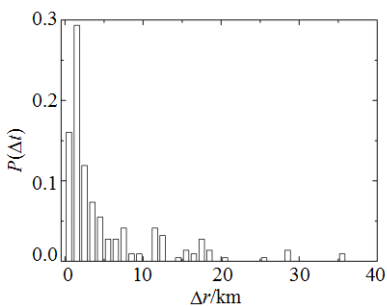
a. 群体的出行距离分布



b. 某学生的出行距离分布



c. 某职员的出行距离分布



d. 某退休者的出行距离分布

图3 出行距离分布

### 4.2 个体水平上的统计结果

从个体水平上对出行距离分布进行统计发现, 有198个个体(占个体总数的86%)的出行距离分布为幂律分布的假设无法通过阈值为0.9的K-S检验。该结果说明, 在群体水平上统计得到的出行距离分布的幂律特性在个体水平上并不是普遍存在的。

进一步观察个体的出行距离分布(图3b~图3d

是3个典型个体的出行距离分布图), 可发现具有非幂律出行距离分布特征的个体, 其出行距离分布图具有明显的峰值, 但峰值出现的位置并没有明显的规律性。下一节将对个体出行距离呈现有峰值分布现象的原因进行解释。

## 5 对出行距离分布统计结果的解释

### 5.1 对个体出行距离分布统计结果的解释

为解释个体出行距离分布非幂律的原因, 本文为每个个体建立了出行网络<sup>[20]</sup>, 并对出行网络的结构特征进行分析。出行网络的构建方法是: 用个体的访问地点作为节点, 如果个体在地点*i*、*j*之间出行了*w*次, 则在*i*、*j*之间连接一条无向边, 并令该边的权为*w*。图4a~图4c是3个典型个体的出行网络示意图。

通过计算个体出行网络边权的基尼系数<sup>[21]</sup>, 可以衡量网络权重分布的不平均程度。基尼系数是衡量一组数值不平均程度的指标, 取值介于0~1之间, 基尼系数越大说明数值的不平均程度越高。本文中出行网络边权的基尼系数*G*的计算公式<sup>[21]</sup>为:

$$G = \frac{N+1}{N-1} - \frac{2}{N(N-1)u} \sum_{i=1}^n p_i w_i \quad (1)$$

式中, *N*为网络中的边数; *w<sub>i</sub>*是第*i*条边的权; *p<sub>i</sub>*是权为*w<sub>i</sub>*的边在所有边中的排序, 权最大的边排序为1, 权最小的边排序为*N*; *u*为边权的平均值。

图4d是所有用户的基尼系数分布, 从图中可以看出, 大部分个体出行网络的边权分布是非常不平均的(近80%的个体的基尼系数大于0.4)。这意味着, 对于这些个体而言, 大部分出行发生只在少数节点之间, 而多数节点之间的出行次数很少。

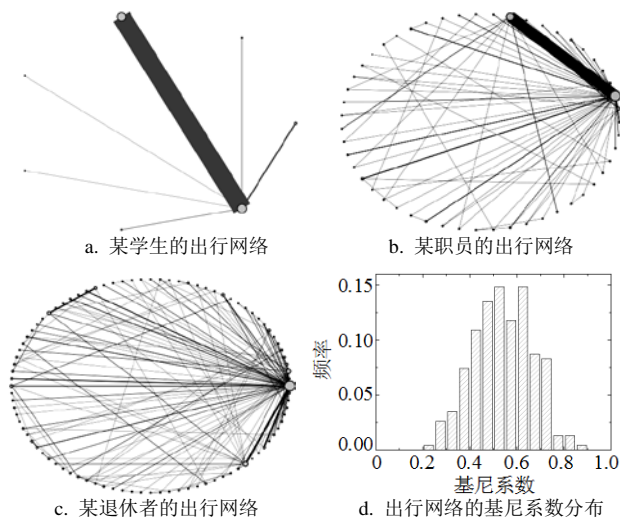


图4 个体的出行网络及出行网络的基尼系数分布

图中, 出行网络图的线条粗细与边权的大小成正比。

观察所有个体的出行网络可发现, 大部分个体出行网络的一个显著特征是: 在某两个地点之间出行的次数明显高于其他地点之间的出行次数(例如图4a和图4b的出行网络)。本文将该两点之间的出行称为“支配性出行”。用支配性出行的次数与个体总出行次数的比值*d*, 可以衡量支配性出行的支配程度。通过统计群体的*d*值分布(见图5a)可以发现, 超过85%的个体的*d*值大于0.1, 说明大部分个体有超过10%的出行发生在两个固定地点之间。

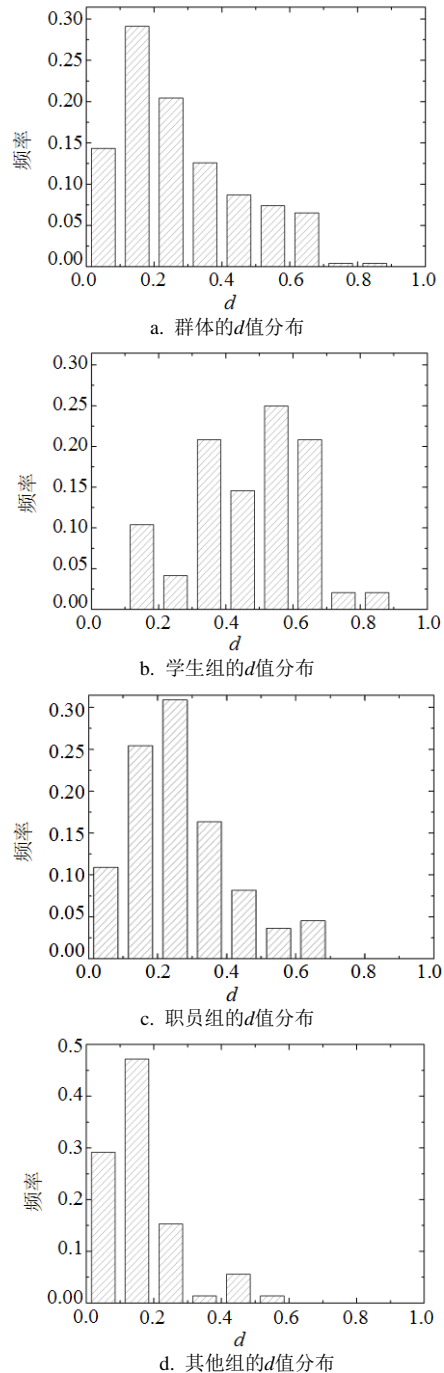


图5 支配程度*d*值的分布

进一步地, 本文将数据集中的所有个体分为学生、职员和其他(包括退休者、失业者和家庭主妇等)3组, 并分别统计了该3组群体的 $d$ 值分布, 结果见图5b~图5c。从图中可以看出, 学生组的 $d$ 值最大, 其平均 $d$ 值为0.474, 且没有人的 $d$ 值小于0.1; 职员组的 $d$ 值整体上也较大, 平均 $d$ 值为0.265, 只有10%左右的个体 $d$ 值小于0.1; 而对于其他组来说,  $d$ 值整体上较小, 平均 $d$ 值为0.167, 近1/3的个体 $d$ 值小于0.1, 超过2/3的个体 $d$ 值小于0.2。

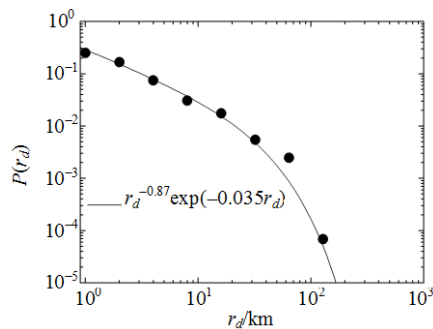
对于学生和职员的出行网络具有高支配度的原因很容易解释: 多数职员和学生每个工作日都会频繁地在家与单位(或学校)之间出行, 该两点之间的出行次数会多于其他地点之间的出行次数, 形成高支配性的出行。而对于退休者、家庭主妇等, 他们可以更自由地选择出行目的地, 因此形成高支配性出行的可能性降低。

高支配性出行的存在使得个体的出行距离分布在其最常去的两个地点(例如家和单位)之间的距离上出现峰值。当个体出行的支配程度较高时( $d > 0.1$ ), 除非个体的支配性出行距离很短, 否则个体的出行距离分布很难出现幂律特性。在之前的实证研究<sup>[10-11]</sup>中, 由于对出行距离分布的统计是在群体水平上进行的, 掩盖了个体出行模式之间的差异, 因此无法揭示个体出行距离分布的真实特征。

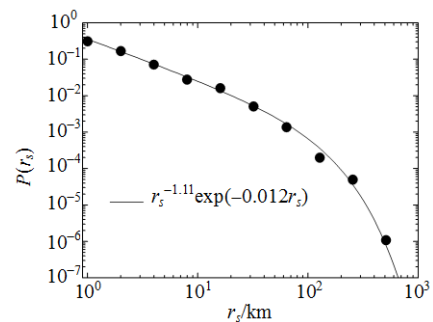
## 5.2 对群体出行距离分布统计结果的进一步分析

通过前述统计结果可知, 本文数据集中大部分个体的出行距离分布并非幂律分布, 那么为何群体水平上出行距离具有幂律分布特征呢? 为更深入地理解该现象, 本文分别统计了群体中个体支配性出行距离 $r_d$ 的分布和群体的非支配性出行距离 $r_s$ 的分布。其中, 非支配性出行是在群体的所有出行中去除了支配性出行后的剩余部分, 约占总出行次数的75%, 二者的统计结果分别见图6a和图6b。

经采用最大似然法估计, 二者都近似服从带有指数截断的幂律分布(均能通过阈值为0.9的K-S检验)。其中, 支配性出行距离的分布函数为 $P(r_d) \sim r_d^{-0.87} \exp(-0.035r_d)$ , 非支配性出行距离的分布函数为 $P(r_s) \sim r_s^{-1.11} \exp(-0.012r_s)$ 。当两个同样具有截尾幂律特性的分布混合在一起后, 产生了群体水平上出行距离分布的幂律效应, 且幂指数介于二者之间。但是, 对于支配性与非支配性出行距离幂律分布的形成机制尚不清楚, 仍有待于进行更深入的研究。



a. 群体的支配性出行距离分布



b. 群体的非支配性出行距离分布

图6 群体支配性出行和非支配性出行的距离分布

## 6 总结

本文分别在群体和个体水平上统计了230位城市居民在6周内的出行活动的停留时间分布和出行距离分布。统计结果显示, 无论在群体水平还是个体水平上, 人类出行的停留时间分布都不具有幂律特征, 而是以相对幂律分布更高的概率进行长时间停留; 而在出行距离分布方面, 群体的出行距离服从带有指数截断的幂律分布, 但198个个体的出行距离分布并非幂律的, 而是在某个特征性距离上出现峰值。本文对这些统计特性形成的可能机制进行了解释, 并讨论了造成群体与个体水平上统计结果差异的原因。

本文的研究结果说明, 在文献[11]提出的人类个体空间运动模型中, 对于人类个体的停留时间与出行距离服从幂律分布的假设是不合理的。因此, 尽管文献[11]等的模型引入了探索和返回两个新颖的动力学机制, 但由于其基本假设的缺陷, 该模型可能并不能真实地反映人类个体的出行模式。

需要指出的是, 人类个体出行距离与停留时间的非幂律分布并不意味着人类的出行行为不具有复杂性特征。恰恰相反, 人类独特的出行模型源于人类高度的社会性, 使人类具有比随机扩散的物理粒子更为复杂的空间运动模式。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<sup>[22]</sup>。过于追求普适标度律的存在而忽视人类活动的社会性特性, 可能会使人们陷入认识的误

区。当然，科学的发展总是在探索中前进的，之前在群体水平上进行的实证研究为人们定量理解人类出行行为提供了一个起点。但是，由于人类社会活动和自身的复杂性，距离真正理解人的出行行为并预测人的出行活动还很远。希望本文的研究结果能够深化对人类个体出行模式的理解，并为今后建立更符合实际的人类出行动力学模型提供借鉴。

### 参 考 文 献

- [1] BARABÁSI A-L. The origin of bursts and heavy tails in human dynamics[J]. *Nature*, 2005, 435: 207-211.
- [2] 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.
- [3] 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 Systems and Complexity Science*, 2010, 7(2-3): 132-144.
- [4] DEZSÖ Z, ALMAAS E, LUKÁCS A, et al. Dynamics of information access on the web[J]. *Physical Review E*, 2006, 73(6): 066132.
- [5] ZHOU T, KIET H A T, KIM B J, et al. Role of activity in human dynamics[J]. *Europhysics Letters*, 2008, 82: 28002.
- [6] VÁZQUEZ A, OLIVEIRA J G, DEZSÖ Z, et al. Modeling bursts and heavy tails in human dynamics[J]. *Physical Review E*, 2006, 73: 036127.
- [7] MALMGREN R D, STOUFFER D B, MOTTER A E, et al. A poissonian explanation for heavy tails in e-mail communication[J]. *Proceedings of the National Academy of Sciences*, 2008, 105: 18153.
- [8] HAN X P, ZHOU T, WANG B H. Modeling human dynamics with adaptive interest[J]. *New Journal of Physics*, 2008, 10: 073010.
- [9] BROCKMANN D, HUFNAGEL L, GEISEL T. The scaling laws of human travel[J]. *Nature*, 2006, 439: 462-465.
- [10] GONZÁLEZ M C, HIDALGO C A, BARABÁSI A L. Understanding individual human mobility patterns[J]. *Nature*, 2008, 453: 779-782.
- [11] SONG C M, KOREN T, WANG P, et al. Modelling the scaling properties of human mobility[J]. *Nature Physics*, 2010, 6: 818-823.
- [12] BARTHÉLEMY M. Spatial networks[J]. *Physics Reports*, 2011, 499: 1-101.
- [13] NI S J, WENG W G. Impact of travel patterns on epidemic dynamics in heterogeneous spatial metapopulation networks[J]. *Physical Review E*, 2009, 79: 016111.
- [14] JIANG B, YIN J J, ZHAO S J. Characterizing the human mobility pattern in a large street network[J]. *Physical Review E*, 2009, 80: 021136.
- [15] HIDALGO C A. Conditions for the emergence of scaling in the inter-event time of uncorrelated and seasonal systems[J]. *Physica A*, 2006, 369: 877-883.
- [16] SCHLICH R, AXHAUSEN K W. Habitual travel behaviour: evidence from a six-week diary[J]. *Transportation*, 2003, 30: 13-36.
- [17] STOPHER P R, GREAVES S P. Household travel surveys: where are we going[J]. *Transportation Research Part A*, 2007, 41: 367-381.
- [18] VESPIGNANI A. Predicting the behavior of techno-social systems[J]. *Science*, 2009, 325: 435-428.
- [19] CLAUSET A, SHALIZI C R, NEWMAN M E J. Power-law distributions in empirical data[J]. *SIAM Review*, 2009, 51(4): 661-703.
- [20] SONG C M, QU Z H, BLUMM N, et al. Limits of predictability in human mobility[J]. *Science*, 2010, 327: 1018-1021.
- [21] COWELL F A. *Measuring Inequality*[M]. Oxford: Oxford University Press, 2009: 65-70.
- [22] MAHMASSANI H S. *In perpetual motion: travel behavior research opportunities and application challenges*[M]. Amsterdam: Elsevier Science Press, 2002: 381-414.

编辑 蒋晓