



人类出行距离分布的统计实证

Empirical analysis of human travel
length distribution

闫小勇 kaiseryxy@163.com

中国科学技术大学近代物理系

电子科技大学互联网科学中心

石家庄铁道大学交通运输学院



提纲

- 1 背景
- 2 数据描述
- 3 统计结果
- 4 对个体出行距离非幂律分布的解释
- 5 对群体出行距离幂律分布的解释
- 6 总结与讨论

1 背景

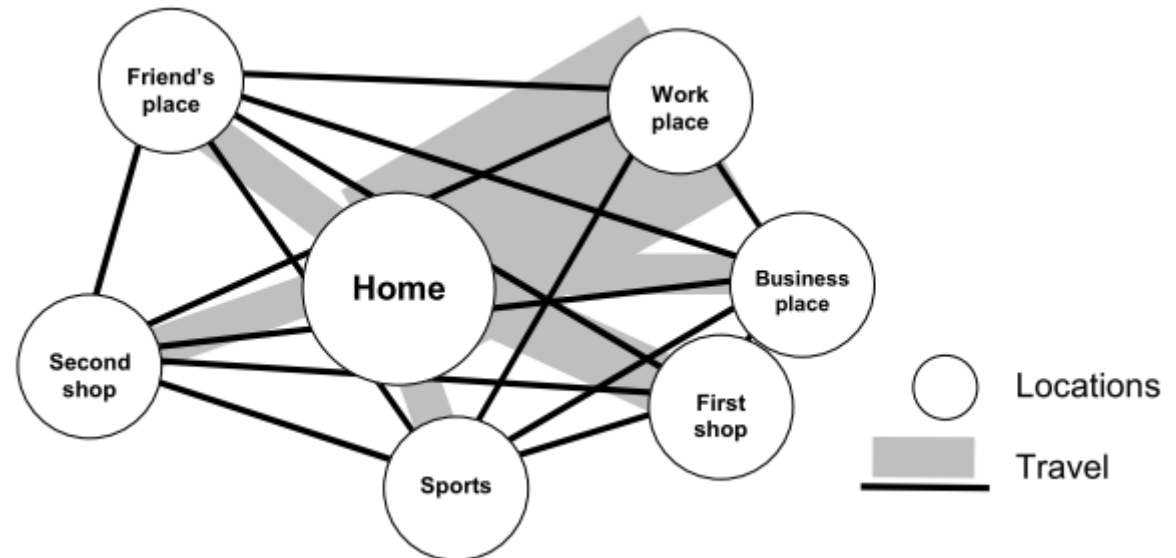
- 理解人类的出行行为对于研究受人类空间运动驱使的各种复杂物理现象具有重要意义。
 - 例如，一些传染病正是由于人的空间运动和相互接触才发生传播，研究人类的出行规律有助于深化对疾病传播机制的理解，进而提出有效的预防和控制措施。
 - 另一个更直接的例子是交通流研究，人在不同地点之间的出行直接导致了交通网络上的各种复杂流动现象。只有在了解人类出行规律的基础上，才能合理规划交通设施，预防和控制交通拥堵。



出行距离分布

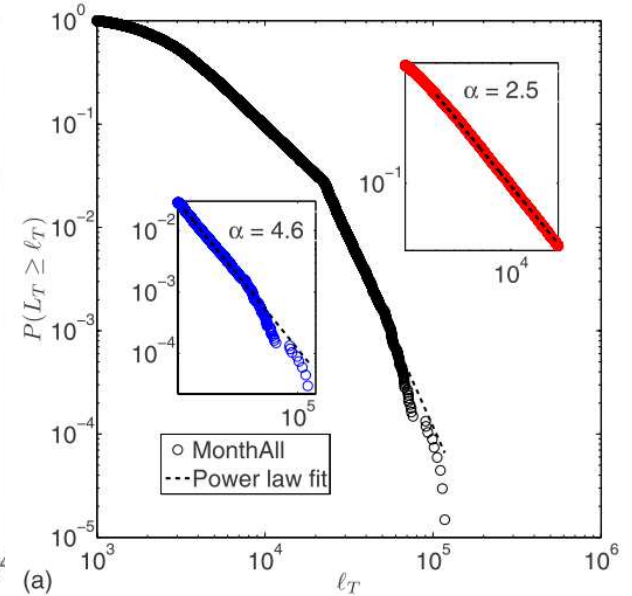
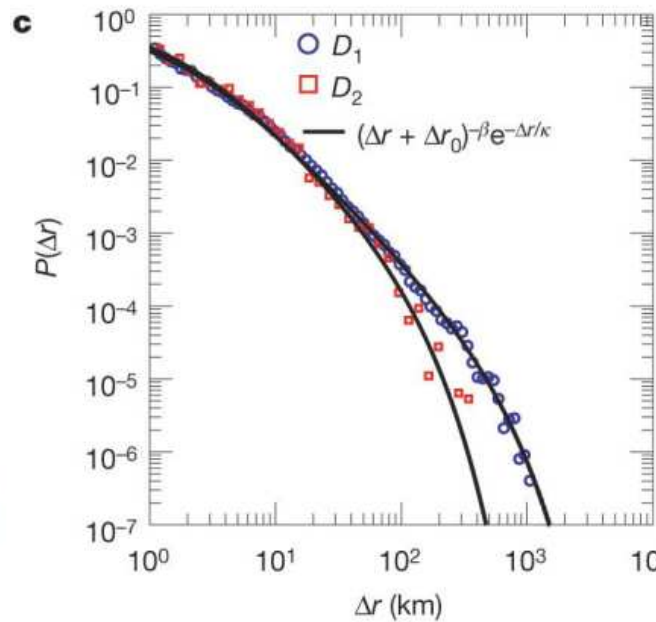
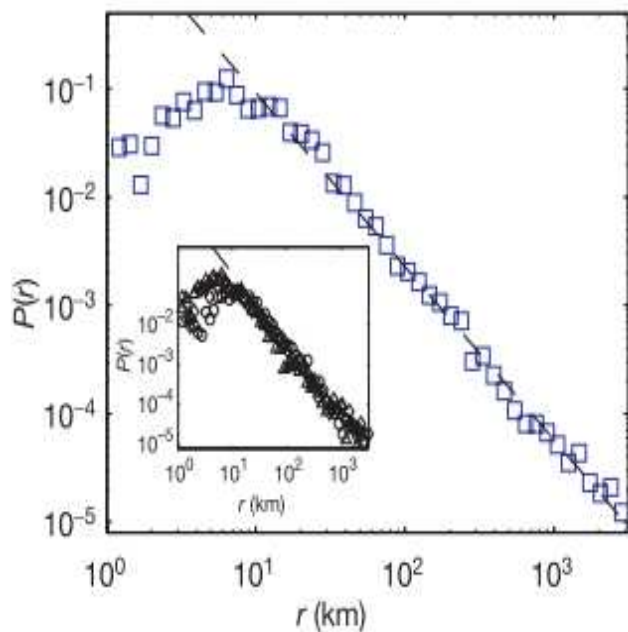
- 出行距离分布是刻画人类空间运动模式的关键统计特征，它直接反映了城市的经济性和效率 [Phys. Rep., **499** 1 2011]，同时对传染病的传播范围和速度等有显著影响 [PRE **79**, 016111 2009]。

在这里，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。



相关文献回顾（实证）

- 对美元流通数据 [Nature 439, 462 2006]、手机通讯记录 [Nature 453, 779 2008] 以及GPS数据 [PRE 80, 021136 2009] 的实证研究都发现，在群体水平上人类的出行距离近似服从幂律分布。



相关文献回顾（模型）

- 韩筱璞等 [PRE **83**, 036117 2011] 分析了交通系统的层次性对人类长程旅行模式的影响，揭示出交通系统自身所具有的层次性可以直接导致人类旅行距离分布的标度特性。
- Song等 [*Nat. Phys.*, **6**, 818 2010] 将个体出行距离和停留时间服从幂律分布作为两个基本假设构建了一个人类个体的空间运动模型。通过引入探索和返回两种动力学机制，再现了人类出行模式中的几种异常标度现象。
- 在我们之前的研究工作中 [arXiv:1011.5111]，根据人类日常活动序列的规则性特征，构建了一个双吸引点周期性随机游走模型，再现了实证研究中发现的人类个体活动区域的椭圆律和活动轨迹回转半径的稳定性特征。但该模型产生的出行距离分布是**有峰值的分布**（峰值出现在个体的家与工作地之间的距离上）。



问题1：个体的出行距离分布是否是幂律的？

- 现有的关于人类出行距离分布的实证研究都是在群体水平上进行的，**缺乏个体水平上的实证统计**。仅根据群体水平上的幂律分布并不能推断出个体也服从幂律分布。

问题2：如何解释群体出行距离幂律分布这一现象？

- 如果幂律的行程分布不是个体层面上的特征，那么如何解释群体出行距离幂律分布这一现象？



主要工作

- 通过对瑞士弗劳恩菲尔德市230位居民在6周内的出行日志数据进行了统计分析，结果发现：
 - 在群体水平上，出行距离服从截尾幂律分布。
 - 在个体水平上，幂律分布并不是个体出行距离分布的普遍特征——多数个体的出行距离呈现有峰值的分布。
- 对个体出行距离非幂律和群体出行距离幂律的原因进行了解释。

2 数据描述

数据来源：瑞士弗劳恩费尔德市所进行的出行日志调查数据。

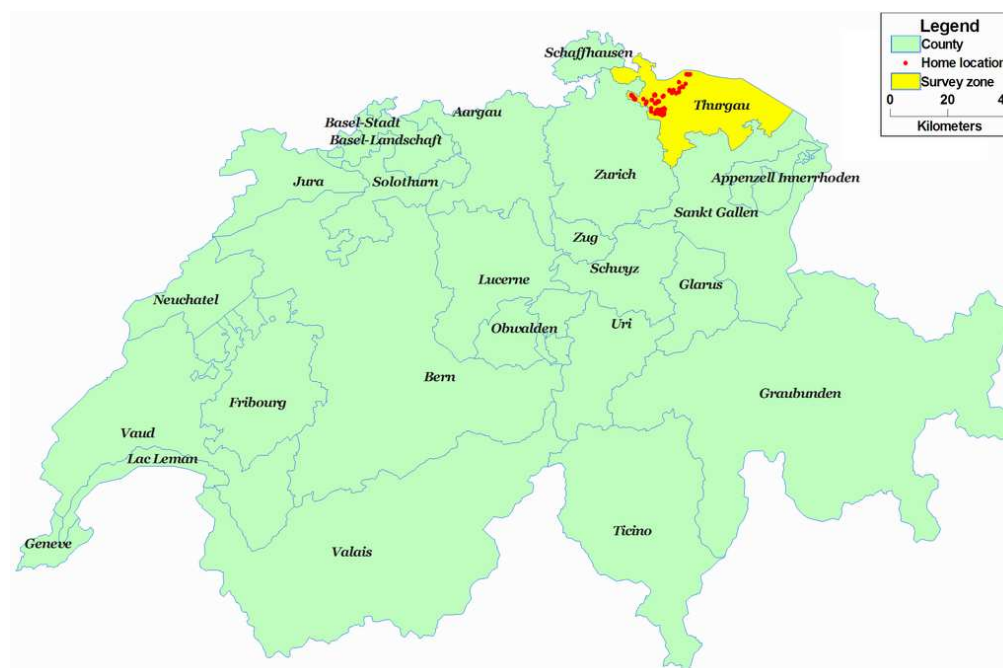
个体数量：230 人

调查持续时间：42 天

总出行次数：36761 次

人均出行次数：160次

人均日出行次数：3.81 次/人日

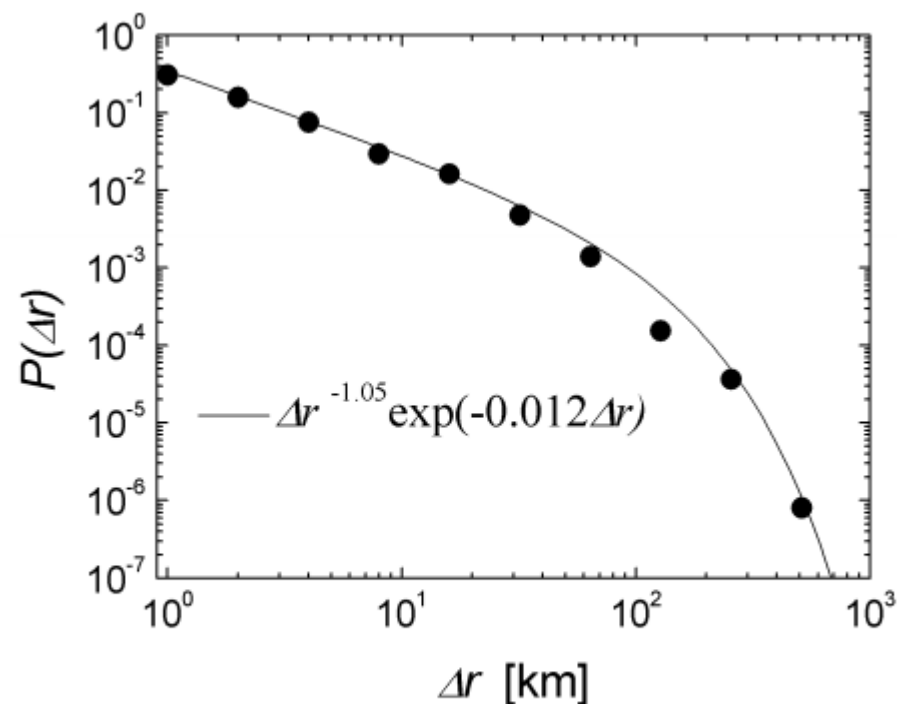


数据集中的每条记录包含了一次出行的起讫点位置、出发和到达时刻以及出行目的等。其中，起讫点位置由经纬度坐标定义（精度为秒），通过起讫点坐标可计算出每次出行的空间直线距离。

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。

3 统计结果

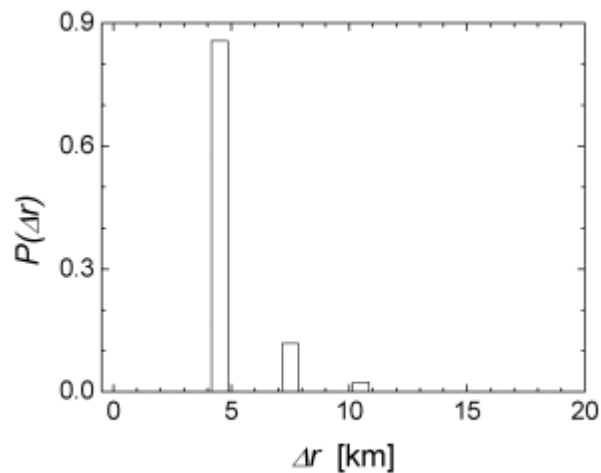
- 在**群体水平**上对出行距离分布的统计结果与之前的研究结果类似：群体的出行距离近似服从带有指数截断的幂律分布。



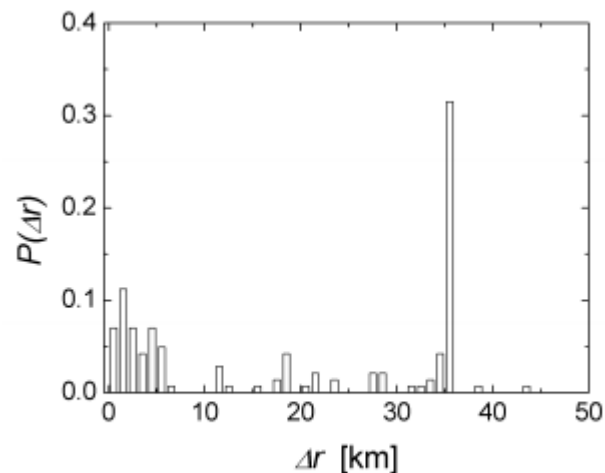
a. 群体的出行距离分布

个体的出行距离分布

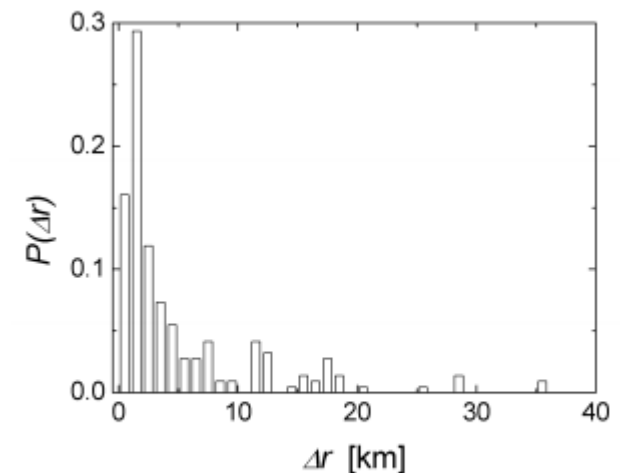
- 从个体水平上进行的统计发现，多数个体的出行距离分布并不是幂律的，而是有一个明显的峰值。
- 对于不同个体，这个峰值可能出现在 0-100 km 的范围内，并没有明显的规律性。



b. 某学生的出行距离分布



c. 某职员出行距离分布



d. 某退休者的出行距离分布



个体出行距离分布的假设检验

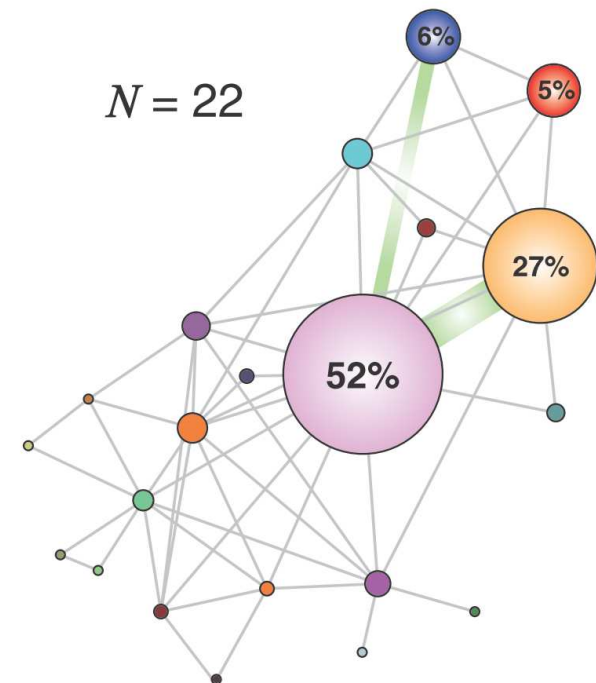
- 首先对每个个体出行距离分布做出服从幂律分布的假设，然后用最大似然估计法估计相应的参数，最后再用K-S检验方法进行检验。
- K-S检验的结果显示，在全部230个个体中，有198个个体（占样本数的86%）的 p 值小于0.1。
- 这一结果说明，尽管在群体水平上统计得到人类的出行距离分布具有幂律特征，但这一特征在个体水平上并不是普遍存在的。

4 对个体出行距离非幂律分布的解释

- 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，并对出行网络的结构特征进行分析。

- 出行网络的构建方法：

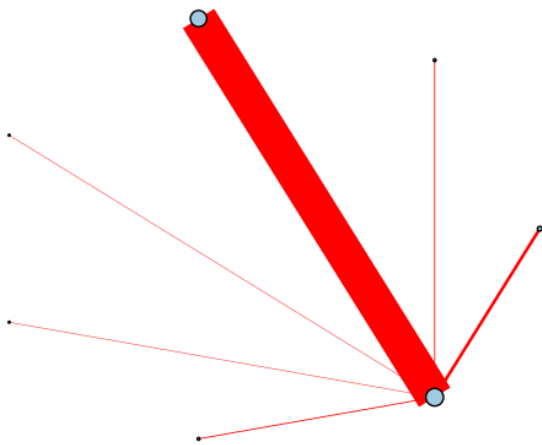
- 用个体的访问地点作为节点，如果个体在地点 i 、 j 之间出行了 w 次，则在 i 、 j 之间连接一条边，并令该边的权值为 w ，这样就可以根据个体访问地点的序列构建一个含权的出行网络（Mobility Networks）[*Science* 327 1018 2010]。



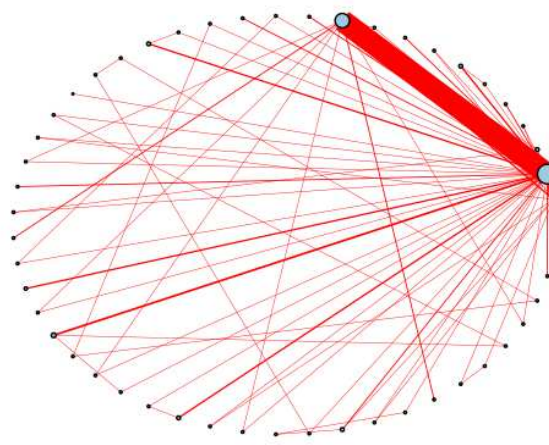
支配性出行

- 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在其他地点之间的出行次数。
- 我们将这两点之间的出行称为“支配性出行”（Dominant Trip）。

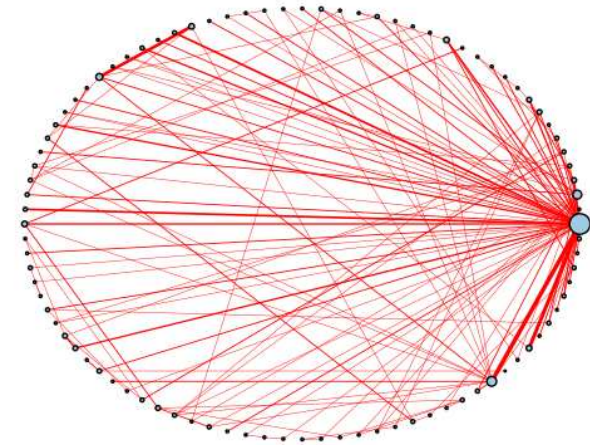
Pupil, Male, 15 years old



Employed, Male, 43 years old

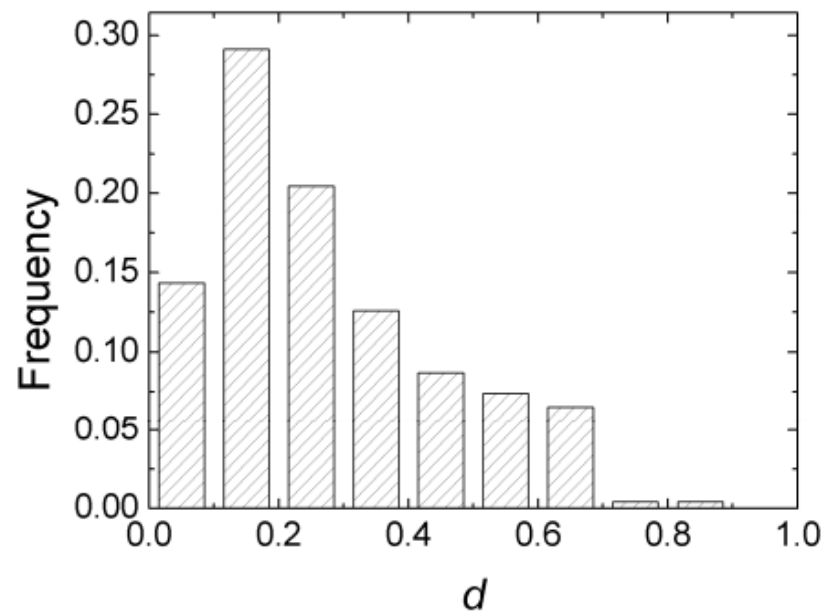


Retired, Male, 70 years old



群体的支配程度分布

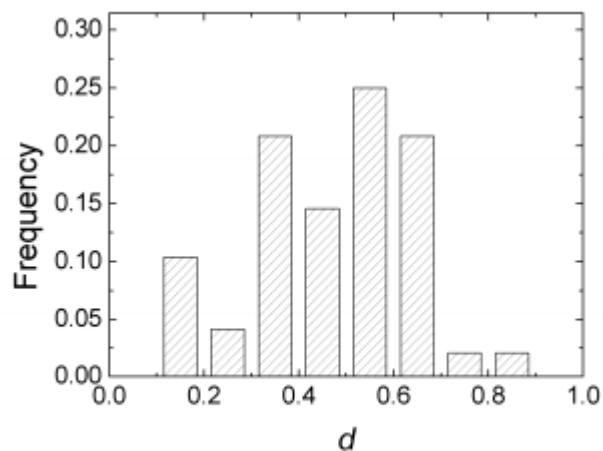
- 用支配性出行的次数 m 与个体总出行次数 M 的比值 d ，可以衡量支配性出行的支配程度。
- 通过统计群体的 d 值分布，可以发现，超过85%的个体的 d 值大于 0.1，说明大部分个体有超过10%的出行发生在两个固定地点之间。



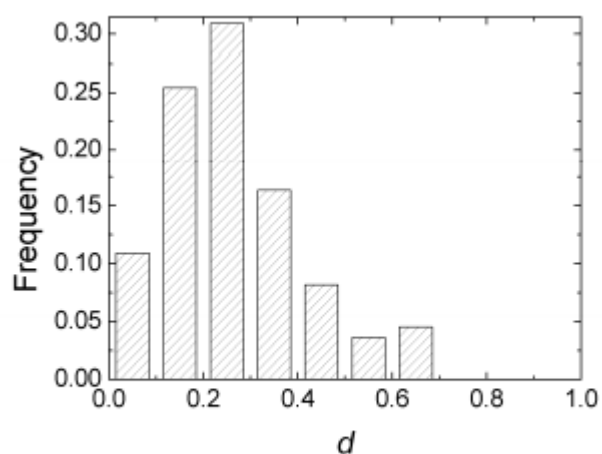
a. 群体的 d 值分布

不同类型群体的支配程度分布

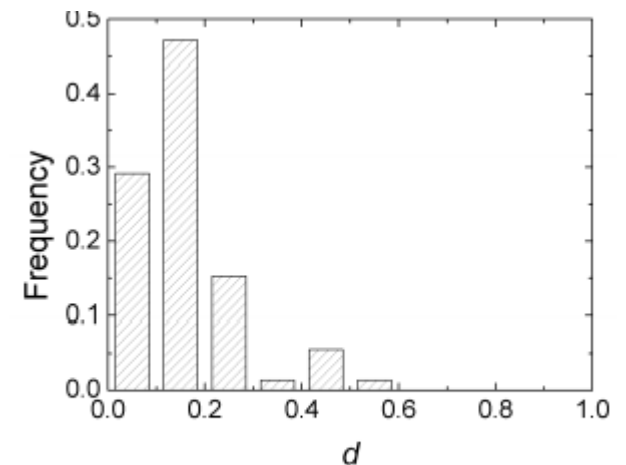
- 进一步地，我将居民分为学生、职员和其他（包括退休者、家庭主妇、无业者等）三类，并分别统计了这三类群体的 d 值分布。
- 结果显示，学生组的 d 值最大，其平均 d 值为 0.474，且没有人的 d 值小于 0.1；职员组的 d 值整体上也较大，平均 d 值为 0.265，只有 10% 左右的个体 d 值小于 0.1；而对于其他组来说， d 值整体上较小，平均 d 值为 0.167，近 1/3 的个体 d 值小于 0.1，超过 2/3 的个体 d 值小于 0.2。



b. 学生组的 d 值分布



c. 职员组的 d 值分布



d. 其他组的 d 值分布

支配性出行对个体出行距离分布的影响

- 对于学生和职员出行网络具有高支配度的原因很容易解释：多数职员和学生在工作日都会频繁地在家与单位（或学校）之间出行，这两点之间的出行次数会多于其他地点之间的出行次数，形成高支配性出行。

individual's motion is dominated by his/her most visited location.

- 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征。
- 在之前的实证研究中，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 对群体出行距离幂律分布的解释

- 前述实证统计结果显示，幂律分布并不是个体出行距离分布的普遍特征。那么，为何群体水平上出行距离分布具有幂律特征呢？
- 我们可以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。这样，统计物理中的能量分布律能够帮助我们理解群体出行距离幂律分布这一现象。

模型

- 人在空间中移动需要消耗能量，这里假设个体移动时消耗的能量与移动距离之间具有关系 $\varepsilon_i = f(r_i)$
- 假设群体中包含的个体数量为 N ，总能量消耗值为 E ，且个体的空间运动是相互独立的，则群体能量分布的微观状态数为：

$$\Omega = \frac{N!}{\prod_i n_i!}$$

- 系统满足约束条件：
$$\sum_i n_i = N \quad \sum_i n_i \varepsilon_i = E$$



模型求解

- 根据波尔兹曼能量分布律，系统最可能出现的能量分布是微观状态数最多的分布（最可几分布）。
- 用拉格朗日乘数法（求解过程略）可求出系统能量的最可几分布为：

$$p(\varepsilon_i) \sim e^{-\alpha\varepsilon_i}$$

能量与距离的关系

- 1、旅行时间与消耗的能量成正比 [NJP 5 48.1 2003]
- 2、时间与距离的关系 [Ann. Reg. Sci. 33 269 1999]

$$t = b_0 + b_1 \frac{r^a - 1}{a}$$

- 用拉格朗日函数插值方法，可将上式近似表示为：

$$t = b_0 + b_1 [c(r-1) + (1-c)\ln(r)]$$

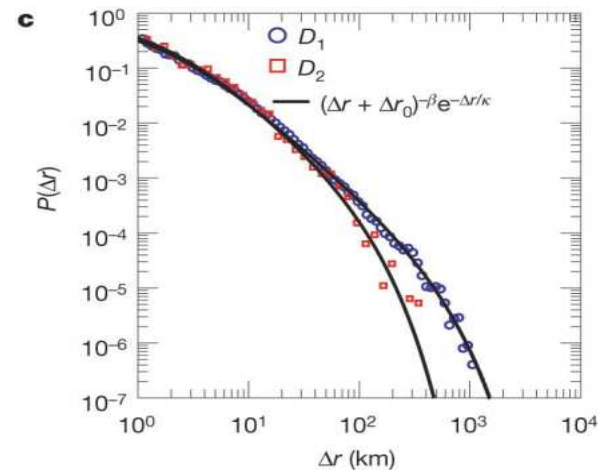
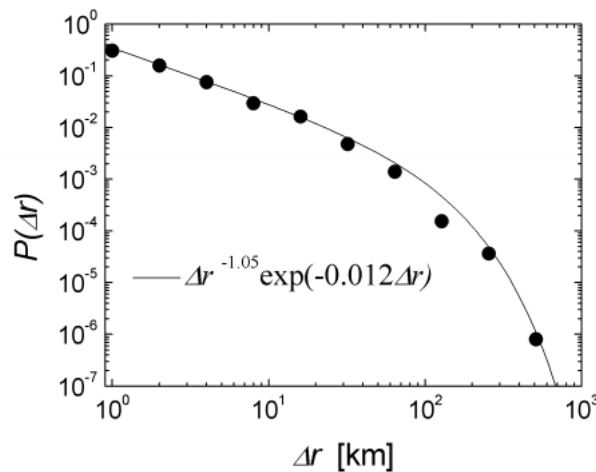
- 根据1、2，可得到能量与距离的关系：

$$\varepsilon = \beta r + \gamma \ln(r) + \delta$$

距离幂律分布的导出

- 根据前述获得的能量-距离关系 $\varepsilon_i = \beta r_i + \gamma \ln(r_i) + \delta$ 以及群体的能量分布 $p(\varepsilon_i) \sim e^{-\alpha \varepsilon_i}$, 可导出群体出行距离的最可几分布为：

$$p(r_i) \sim r_i^{-\alpha\gamma} \cdot e^{-\alpha\beta r_i}$$



6 总结与讨论

- 对230位居民在6周内的出行行为进行了统计，结果发现：
 - 在群体水平上，人类的出行距离服从截尾幂律分布，与对手机通讯记录所进行的实证研究结果一致；
 - 在个体水平上，幂律并不是出行距离分布的普遍特征——多数个体的出行距离呈现有峰值的分布。
- 尽管个体的出行行为受到各种社会活动因素的影响，具有复杂多样的特征；但在群体层面上，城市人群的移动和气体分子的运动却呈现出了惊人相似的模式——在这里，简单的能量率对于描述复杂的人类群体出行模式同样适用。
- **“ 当我们找到可应用于完全不同的复杂系统的定律时，我们可期望对其深入地理解。 ”**

——哈肯，*《信息与自组织：复杂系统中的宏观方法》*，1987
- 我们希望本文的结果能够深化人类对自身空间运动行为的理解，并有助于我们在疾病传播、交通预测等领域建立更符合实际的人动模型。



汇报结束，谢谢！

请各位老师同学批评指正！

闫小勇 kaiseryxy@163.com