

丛 书 序

交通科学在近 70 年来发展突飞猛进,不断拓展其外延并丰富其内涵;尤其是近 20 年来,智能交通、车联网、车路协同、自动驾驶等概念成为学者研究的热点问题的同时,也已成为媒体关注的热点;应用领域的专家及实践者则更加关注交通规划、土地利用、出行行为、交通控制和管理、交通政策和交通流仿真等问题的最近研究进展及对实践的潜在推动力。信息科学和大数据技术的飞速发展更以磅礴之势推动着交通科学和工程实践的发展。可以预见在不远的将来,车路协同、车联网和自动驾驶等技术的应用将根本改变人类的出行方式和对交通概念的认知。

多方式交通及支撑其运行的设施及运行管理构成了城市交通巨系统,并与时空分布极广的出行者之间形成了极其复杂的供需网络/链条。城市间的公路、航空、铁路和地铁等日益网络化、智能化,让出行日益快捷。有关城市或城市群的规划则呈现“住”从属于“行”的趋势。如此庞杂的交通系统激发了人们的想象力,使交通问题涉及面极广,吸引了来自不同学科和应用领域的学者和工程技术专家。

因此,为顺应学科发展需求,由科学出版社推出的这套《交通与数据科学丛书》将首先是“兼收并蓄”的,以反映交通科学的强交叉性及其各分支和方向的强相关性。其次,“数‘理’结合”,我们推动将数据科学与传统针对交通机理性的研究有机结合。此外,该丛书更是“面向未来”的,将与日新月异的科学和技术同步发展。“兼收并蓄”“数‘理’结合”和“面向未来”,将使该丛书顺应当代交通科学的发展趋势,促进立足于实际需求和工程应用的实际问题开展科研攻关与创新,进而持续推动交通科学研究成果的“顶天立地”。

该丛书内容将首先是对交通科学理论和工程实践的经典总结,同时强调经典理论和实践与大数据和现代信息技术的结合,更期待据此提出的新理论、新模型和新方法;研究对象可为道路交通、行人流、轨道交通和水运交通等,可涵盖车车和车路联网技术、自动驾驶技术、交通视频技术、交通物联网和交通规划及管理。书稿形式可为专著、编著、译著和专题等,中英文不限。该丛书主要面向从事交通科学研究和工程应用的学者、技术专家和在读研究生等。

该丛书编委会聚集了我国一批优秀的交通科学学者和工程应用专家,基于他

们的治学态度和敬业精神，相信能够实现丛书的目标并保证书稿质量。最后，上海麓通信息科技有限公司长期以来为丛书的策划和宣传做了大量工作，在此表示由衷的感谢！

张 鹏

2019 年 3 月

前言

预测地点间人、物、信息的流动是交通科学、经济地理学、区域经济学等诸多领域长期以来的一个重要研究主题,在城市规划、交通工程、疾病传播防控、紧急事件管理等方面也具有重要的应用价值.一百多年来,研究者们陆续提出了多种预测地点间流动量的模型——在地理和经济学中称为空间交互模型,在交通科学中称为出行分布预测模型.其中最有影响力的是引力模型,这是因为,在许多地点间流动量分布现象中都存在类似牛顿万有引力定律的规律,即两地间的某种流动量正比于两地的规模乘积,反比于两地间距离的幂函数.而引力模型就是类比这样的引力定律而提出的,并在许多领域获得了广泛而成功的应用.但为什么会有这样的引力定律存在?这是一个非常有趣和有价值的问题.一些研究用最大熵原理、确定与随机效用理论、博弈论等来为引力模型提供理论依据,更有若干比引力模型更准确、更普适的新模型被陆续提出.本书将对这些空间交互和出行分布预测的理论与方法进行系统的介绍.

全书共分8章.第1章介绍在各种社会系统中发现的引力定律实例,并介绍几类典型的引力模型及其计算与参数估计方法.

第2章主要介绍从宏观角度导出引力模型的最大熵原理,这是作者所知的最早的引力模型推导方法;还介绍最大熵原理在解释群体出行距离分布特征等方面的应用,以及随机分组模型这种更一般化的熵模型.

第3章介绍从微观角度导出引力模型的效用理论,包括确定效用理论和随机效用理论两类,其中,基于随机效用理论的离散选择模型是经济学、社会学等领域描述个体选择行为的重要模型,本章对其进行重点介绍.

第4章主要介绍大数据时代提出的新型无参数空间交互模型,包括辐射模型和人口权重机会模型.在此之前,先对启发新型无参数模型建模的传统介入机会模型进行介绍.最后讨论传统引力模型、介入机会模型与新型无参数模型的内在联系.

第5章介绍作者与合作者提出的目的地选择博弈(DCG)模型,包括DCG模型的建模框架与求解算法、DCG模型和引力模型、介入机会模型、辐射模型、人口权重机会模型预测精度的对比,以及简化DCG模型导出引力模型的过程.

第6章介绍个体连续移动行为建模,与前五章介绍的群体空间交互模型有所差别.作者从早期的随机游走、列维飞行等模型开始,结合进入大数据时代后多种

个体移动模式的实证分析结果, 重点介绍一些有代表性的个体移动行为模型.

第 7 章介绍作者与合作者提出的个体群体移动模式预测统一模型, 该模型是个体移动行为模型与群体空间交互模型的组合. 本章重点介绍统一模型的建模思路、模型规则和预测结果, 并详细介绍从不同角度简化的统一模型的数学解析过程.

第 8 章对全书内容进行总结, 通过概念地图给出各章介绍的核心模型之间的逻辑关系, 并讨论空间交互与人类移动建模研究方向的挑战性问题.

随着人、物、信息流动大数据的获取日益丰富, 数据内容日益精细, 基于大数据的空间交互与出行行为研究进展十分迅速, 新的研究成果不断涌现, 因此, 从大量成果素材中选取合适的理论与方法是一件困难的事情. 作者尽量把该方向最基本、最重要的理论和方法 (包括作者与合作者最近几年的研究成果) 撰写出来, 希望本书能起到抛砖引玉的作用.

在此感谢狄增如、樊瑛、高自友、韩晓、韩筱璞、Henrik Jeldtoft Jensen、贾斌、姜锐、来颖诚、Petter Minnhagen、孙晓燕、汪秉宏、王文旭、杨紫陌、曾安、赵琛、赵毅鸣、赵志丹、周涛等合作者对本书中相关研究成果所做出的贡献. 感谢吴金闪和刘二见对本书撰写所给予的帮助. 特别要感谢家人的支持与鼓励.

本书的研究和写作得到了国家自然科学基金委员会 (项目编号: 71822102, 71621001, 71671015, 61304177) 的经费支持, 出版则得到北京交通大学基本科研业务费 (项目编号: 2016JBZ007) 资助, 在此表示感谢.

由于作者的水平有限, 不妥之处在所难免, 敬请读者批评指正.

闫小勇

2018 年 10 月

目 录

丛书序

前言

第 1 章 引力万有 —— 从物理到社会	1
1.1 引力定律的发现	1
1.2 人类社会中的引力定律实例	2
1.2.1 铁路网络中的引力定律	2
1.2.2 公路网络中的引力定律	3
1.2.3 通勤出行中的引力定律	4
1.2.4 城市内出行中的引力定律	5
1.2.5 人口迁移中的引力定律	5
1.2.6 国际贸易中的引力定律	6
1.2.7 货船海运中的引力定律	7
1.2.8 手机通信中的引力定律	8
1.2.9 社交网络中的引力定律	8
1.2.10 科研网络中的引力定律	9
1.3 引力模型简介	11
1.3.1 无约束引力模型	11
1.3.2 单约束引力模型	11
1.3.3 双约束引力模型	12
1.3.4 参数估计方法	12
1.4 小结	13
第 2 章 睹微知著 —— 最大熵原理	14
2.1 熵的概念	14
2.1.1 统计物理熵	14
2.1.2 信息熵	15
2.1.3 最大熵	15
2.2 由最大熵原理导出引力模型	16
2.2.1 构建最大熵模型	16
2.2.2 导出引力模型	17
2.3 出行距离分布的根源	18

2.3.1	出行距离幂律分布现象	18
2.3.2	由最大熵原理导出出行成本分布	18
2.3.3	出行成本与距离的关系	19
2.3.4	导出出行距离分布	20
2.4	更一般化的熵模型 —— 随机分组模型	21
2.4.1	幂律分布的解释	21
2.4.2	随机分组模型	22
2.4.3	模型参数确定方法	25
2.5	小结	25
第 3 章	择善而行 —— 确定与随机效用理论	27
3.1	基本概念	27
3.1.1	效用与效用最大化	27
3.1.2	离散选择模型	27
3.1.3	效用最大化准则	28
3.1.4	效用水平与尺度	28
3.1.5	随机效用理论	29
3.1.6	二项 Logit 模型	29
3.2	由确定效用理论导出引力模型	31
3.2.1	推导过程	31
3.2.2	存在问题	33
3.3	由随机效用理论导出引力模型	34
3.3.1	多项 Logit 模型	34
3.3.2	导出引力模型	35
3.3.3	红蓝巴士问题	35
3.4	小结	37
第 4 章	以简驭繁 —— 大数据时代的无参数模型	38
4.1	介入机会模型回顾	38
4.1.1	基本思想	38
4.1.2	推导过程	38
4.1.3	IO 模型特点	40
4.2	辐射模型	41
4.2.1	模型假设与推导过程	41
4.2.2	辐射模型与 IO 模型的关系	42
4.2.3	辐射模型特点	43
4.3	人口权重机会模型	44

4.3.1	模型假设	44
4.3.2	预测精度	46
4.3.3	PWO 模型与 IO 模型、辐射模型的关系	50
4.4	几种模型的内在联系	52
4.4.1	人口均匀分布下模型的对比	52
4.4.2	序贯选择和全局选择	53
4.5	小结	55
第 5 章	势均力敌 —— 目的地选择博弈	57
5.1	目的地选择博弈模型	57
5.1.1	模型基本框架	57
5.1.2	模型求解算法	59
5.2	DCG 模型预测结果	60
5.2.1	模型预测精度	60
5.2.2	扩展到更多问题	62
5.3	导出引力模型	65
5.3.1	简化 DCG 模型	65
5.3.2	势博弈与拥挤博弈	65
5.3.3	推导过程	66
5.3.4	与其他引力模型理论基础的比较	68
5.4	目的地选择与路径选择组合模型	69
5.4.1	DCG 模型与 UE 分配模型的组合	69
5.4.2	PWO 模型与 UE 分配模型的组合	71
5.5	小结	73
第 6 章	时空漫步 —— 个体连续移动行为建模	75
6.1	随机游走类模型简介	75
6.1.1	普通随机游走	75
6.1.2	列维飞行与列维游走	77
6.1.3	连续时间随机游走	77
6.2	个体移动模式实证分析	78
6.2.1	货币流通数据	78
6.2.2	手机通信数据	79
6.2.3	出行日志数据	82
6.3	个体连续移动模型	83
6.3.1	探索与偏好返回模型	83
6.3.2	周期性随机游走模型	85

6.3.3 层次性交通系统游走模型	88
6.4 小结	90
第 7 章 浑然一体 —— 个体群体移动模式预测统一模型	92
7.1 统一模型	93
7.1.1 建模思路	93
7.1.2 模型规则	95
7.2 预测结果	95
7.2.1 城市间出行预测	95
7.2.2 城市内出行预测	99
7.2.3 无记忆与无竞争模型	100
7.3 模型解析	101
7.3.1 有限空间简化个体模型解析	101
7.3.2 无限空间简化个体模型解析	105
7.3.3 简化群体模型解析	108
7.4 小结	111
第 8 章 方兴未艾 —— 总结与展望	113
参考文献	115
索引	124

第1章 引力万有 —— 从物理到社会

1.1 引力定律的发现

大多数人在中学物理课就学过牛顿的万有引力定律^①：两个物体之间的引力正比于两者质量的乘积，反比于两者距离的二次方。如果两物体质量分别为 m_1, m_2 ，质心间距离为 r ，则两者之间的引力为

$$F = G \frac{m_1 m_2}{r^2}, \quad (1.1)$$

其中， G 是引力常量。这就是牛顿发现的物体间相互吸引力所遵循的规律^[1]。引力相互作用是自然界的四大基本相互作用之一（另外三种分别是电磁相互作用、弱相互作用及强相互作用）。

有趣的是，引力定律不仅在物理中存在，在人类社会中人、物和信息在地点间的流动现象中也广泛存在。早在 1846 年，比利时的一位土木工程师 Desart 就在比利时铁路客运数据中发现，两个车站间的客运量存在类似牛顿万有引力定律的规律^[2]：两个车站间的客运量正比于两个车站所在地的人口数乘积，反比于两车站间距离的幂函数，见图 1.1。图中直线的幂指数为 2.25，非常接近于牛顿万有引力定律中的距离幂指数 2。换句话说，铁路车站间的客流量就可以用以下模型来计算：

$$T_{ij} = \alpha \frac{m_i m_j}{d_{ij}^\beta}, \quad (1.2)$$

其中， T_{ij} 是车站 i 到车站 j 的客流量； m_i 是车站 i 所在地的人口数； d 是两车站间的铁路里程； α 和 β 是两个参数。这几乎就是最早的引力模型的形式。

Desart 在比利时铁路客运中发现的类似引力定律，是在社会系统中发现引力定律的最早记载^[3]，其发现时间比美国经济学家 Carey 在人口迁移现象中发现类似引力定律^[4]的时间（1858 年）还要早 12 年。后来，英国地理学家和人口统计学家 Ravenstein 从 19 世纪 70~80 年代英国人口普查数据中，也发现两地人口迁移量与两地人口乘积成正比、与两地距离成反比的引力定律^[5]。而 Reilly 在 1929 年的著作^[6]中研究零售市场问题时也发现，零售中心从其周围某个城镇吸引到的顾客数量与该零售中心的规模成正比，与两地间距离的二次方成反比。此后，在社会、经

^① 牛顿在 1687 年的《自然哲学的数学原理》一书^[1]中写到：宇宙中每个质点都以一种力吸引其他各个质点。这种力与各质点的质量的乘积成正比，与它们之间距离的二次方成反比。

济、地理、交通等许多领域，陆续发现了许多类似这些引力定律的现象。下面我们将介绍人类社会中观察到的一系列引力定律的典型实例。

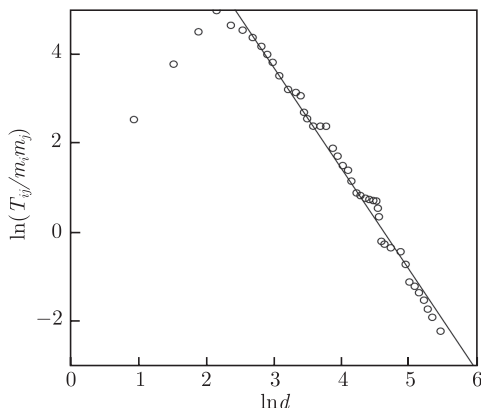


图 1.1 比利时铁路客运量中的引力定律^[3]

横轴表示车站间距离的对数，纵轴表示客运量与人口关系 $T_{ij}/(m_i m_j)$ 的对数，图中直线为 $y \sim x^{-2.25}$

1.2 人类社会中的引力定律实例

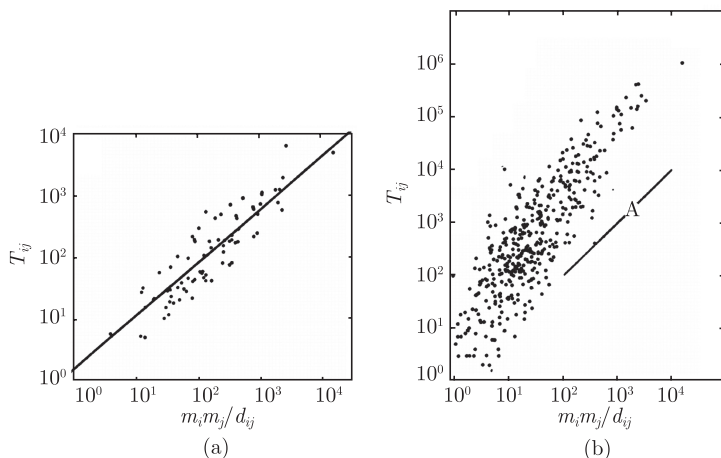
1.2.1 铁路网络中的引力定律

1846 年，比利时工程师 Desart 发现铁路网中存在引力定律，恰好在 100 年后的 1946 年，美国语言学家 Zipf^①又在美国的铁路网及公路网中发现了引力定律^[7]。Zipf 在美国 13 个城市之间的铁路快运网络上，发现两城市间铁路货运量与两地人口乘积比两地距离 $m_i m_j / d_{ij}$ 具有幂律关系，幂指数为 0.85，见图 1.2。不仅如此，在美国 29 个城市间的铁路客运量方面，与 $m_i m_j / d_{ij}$ 也具有幂律关系，幂指数也接近于 1。上述实证结果就引出了一种简化的引力模型：

$$T_{ij} \sim \frac{m_i m_j}{d_{ij}}. \quad (1.3)$$

当然，Zipf 的引力模型中的距离只取一次幂（出行量直接反比于距离），这在很多实际应用中并不恰当。实际中还是要像式 (1.2) 那样为距离函数增加可调参数。

① Zipf 更著名的发现是英文中单词的出现频率与其频率排序成反比，即 Zipf 律^[8]。它表明只有少数词被经常使用，而大多数词很少被使用。有趣的是，Zipf 律在人的移动行为模式中也普遍存在，将在第 6 章介绍。

图 1.2 美国铁路网络中的引力定律^[7]

(a) 美国 13 个城市间的铁路货运量 (纵轴) 和人口乘积与距离比值 (横轴) 之间的关系, 图中直线为 $y \sim x^{0.85}$. (b) 美国 29 个城市间的铁路客运量 (纵轴) 和人口乘积与距离比值 (横轴) 之间的关系,

图中直线为 $y \sim x^1$

1.2.2 公路网络中的引力定律

Zipf 的工作中不仅研究了铁路网络运量与城市人口、距离之间的关系, 还研究了高速公路网络客运量与城市人口、距离之间的关系, 发现同样存在引力定律, 结果见图 1.3(a).

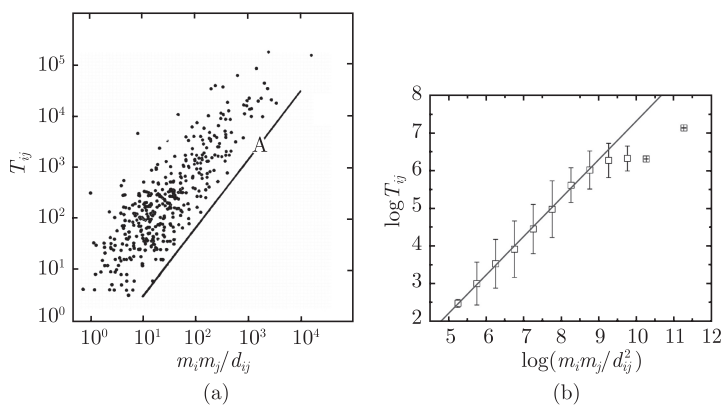


图 1.3 公路网络中的引力定律

(a) 美国 29 个城市间的高速公路客运量 (纵轴) 和人口乘积与距离比值 (横轴) 之间的关系, 图中直线为 $y \sim x^{1.25}$ ^[7]. (b) 韩国 30 个城市间的铁路客运量 (纵轴) 和人口乘积与距离二次比值 (横轴)

之间的关系, 图中直线为 $y \sim x^1$ ^[9]

目前,大多数高速公路网络的进出口都能自动记录车辆数据.基于这种数据,Jung 等研究了 2005 年韩国高速公路网络中人口超过 20 万的 30 个城市间交通量与人口、距离之间的关系.他们发现每个城市的进出流量与人口紧密相关,而且城市间交通量正比于两城市人口乘积,反比于两城市间高速公路距离的二次方,见图 1.3(b).这一结果与物理中的牛顿万有引力定律非常类似,结合 Zipf 早期的研究结果,说明在城市间的交通量中,引力定律是普遍存在的.

1.2.3 通勤出行中的引力定律

Viboud 等从美国 2000 年的人口普查数据^①中,统计出了 3109 个郡(县)之间的居民通勤出行数据^[11].他们发现,两郡之间的通勤流量与两郡人口乘积正相关,与两郡之间的距离负相关,非常类似于引力定律,见图 1.4.因此,他们用一个多参数的引力模型:

$$T_{ij} = \alpha \frac{m_i^{\tau_1} m_j^{\tau_2}}{d_{ij}^{\beta}} \quad (1.4)$$

来描述郡之间的通勤流量.引力模拟的参数估计结果显示,在距离 $d < 119$ km 的范围内,参数为 $(\tau_1, \tau_2, \beta) = (0.30, 0.64, 3.05)$,而在距离 $d \geq 119$ km 的范围内,参数为 $(\tau_1, \tau_2, \beta) = (0.24, 0.14, 0.29)$.

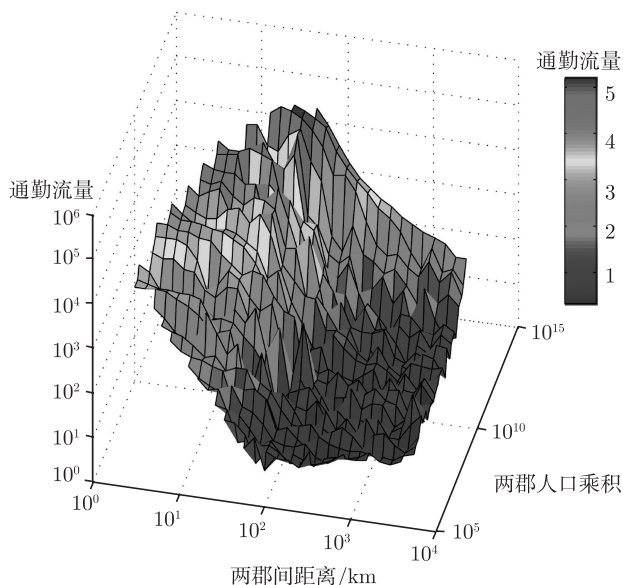


图 1.4 美国两郡间通勤量与两郡人口乘积、两郡间距离关系的三维视图^[11]

^① 美国人口普查要求每个有工作的居民都填报其居住地和work地,因此可从人口普查数据中统计出居民通勤出行数据,参见 <https://www.census.gov/population>.

1.2.4 城市内出行中的引力定律

前述介绍的三类案例都是城市间的人群移动行为,在城市内的居民日常出行行为中也普遍存在引力定律.随着公交 IC 卡的普及,研究者可以基于大规模的 IC 卡刷卡数据,开展地铁与公交客流分布规律的研究. Goh 等分析了韩国首尔地铁网络中的客流分布规律^[12],发现地铁车站间客流量与站间距之间具有显著的相关性,可以用引力模型很好地描述,见图 1.5. Goh 等还分析了首尔常规公交网络的客流分布规律^[13],发现同样存在类似的引力定律. Hong 等则分析了韩国 5 个中等城市公交网络上的客流分布规律^[14],发现这些城市内小区间的公交客流量与距离之间具有显著的相关性,可以用引力模型很好地描述. Smith 等在伦敦地铁客流数据中也发现了类似规律^[15]. 这些结果说明,引力定律在城市内出行模式中也是广泛存在的.

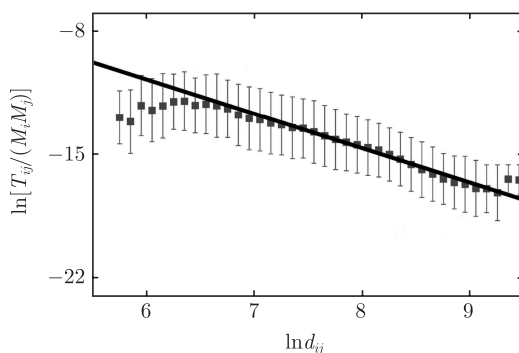


图 1.5 韩国首尔地铁车站间客流量与站间距的关系^[12]

纵轴是车站 i 与 j 之间的客流量 T_{ij} 与两站周边雇员数量乘积 $M_i M_j$ 比值的对数,横轴是两站之间的距离 d_{ij} 的对数

1.2.5 人口迁移中的引力定律

人的移动行为不仅包括短期进行的日常出行行为,还包括相对长期才进行的居住地迁移行为.我们在 1.1 节已经介绍过, Carey 和 Ravenstein 在 19 世纪都发现过人口迁移中存在的引力定律.而 Tobler 用美国 1975~1980 年 48 个州之间的人口迁移数据也验证了这一点^[16]①,见图 1.6.

前述介绍的各种不同空间尺度、时间尺度下的人类移动行为中都存在引力定律.这些移动行为可以从时、空两个维度划分为四种典型类型,见表 1.1. 本书后续

① 我们在后面第 5 章提出的模型也使用了美国人口迁移数据进行验证. 美国人口迁移数据的获取方法非常有意思:它是基于国民每年向国税局提交的个人所得税申报表上的地址变化,从这些表格中可以提取出州之间甚至郡之间的详细人口迁移数据. 参见 <https://www.irs.gov/statistics/soi-tax-stats-migration-data>.

章节将以人类移动行为作为重点研究对象. 但读者应该知道, 本书介绍的各种模型不仅可用于分析预测人的空间移动, 也同样可用于分析货物、信息的空间流动. 下面将介绍这两方面的一些案例.

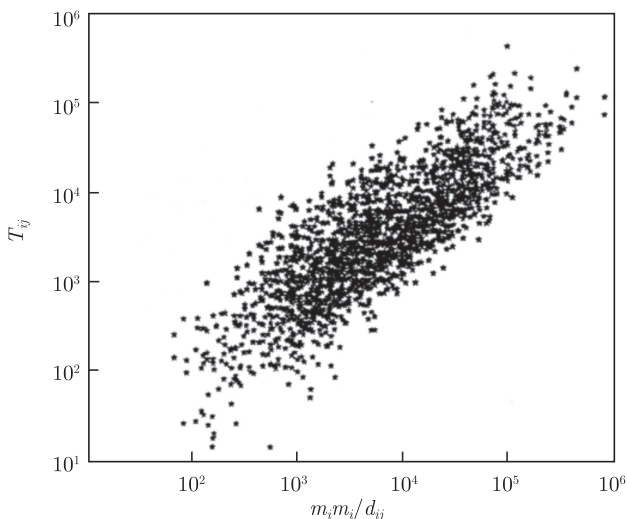


图 1.6 美国 48 个州之间的人口迁移量 (纵轴) 和两州人口乘积与州间距离比值 (横轴) 的关系^[16]

表 1.1 人类移动行为时空分类

空间	时间	
	短期	长期
城市内	出行	搬迁
城市间	旅行	迁移

1.2.6 国际贸易中的引力定律

引力定律不仅在人的空间移动中存在, 在货物的空间流动中也普遍存在. 特别是国家间的货物流动——国际贸易中, 双边贸易规模符合引力定律: 经济越发达、相互距离越近的国家之间, 贸易规模就越大. 图 1.7 是 Fagiolo 用全球 159 个国家之间 2000 年的双边贸易数据 (贸易规模折算为美元) 绘制的国际贸易量与双边 GDP 乘积比距离的关系图^[17]. 从中可以看到, 国际贸易量符合引力定律. 预测国际贸易量的基本引力模型就是式 (1.4), 只不过其中的变量 m_i 不再是使用地点人口数量, 而是使用反映国家经济水平的 GDP. Fagiolo 利用这组双边贸易数据估计出的引力模型主要参数为 $(\tau_1, \tau_2, \beta) = (1.471, 1.338, 0.727)$.

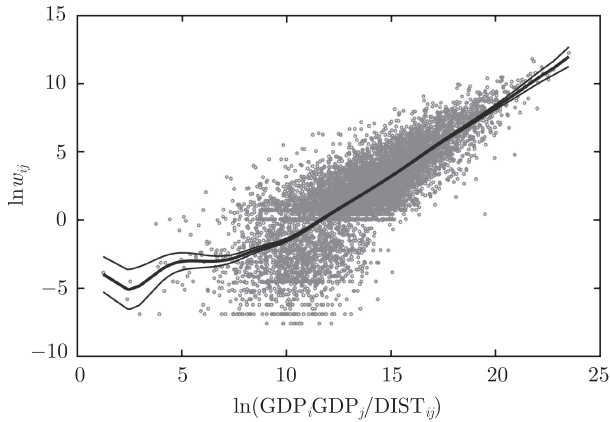


图 1.7 贸易量 w_{ij} 与双边 GDP 乘积比距离 $DIST_{ij}$ 的关系图^[17]

1.2.7 货船海运中的引力定律

货物船只的海上运输量中也存在引力定律. Kaluza 等利用 www.seaweb.com 网站提供的船舶港口自动识别系统报告数据, 分析了全球 16363 艘货船在 951 个港口间的海运量与港口间距离的关系^[18], 如图 1.8 所示. 他们发现, 港口间货船海运量可以用如下双约束引力模型^[10]来描述:

$$T_{ij} = a_i O_i b_j D_j f(d_{ij}), \quad (1.5)$$

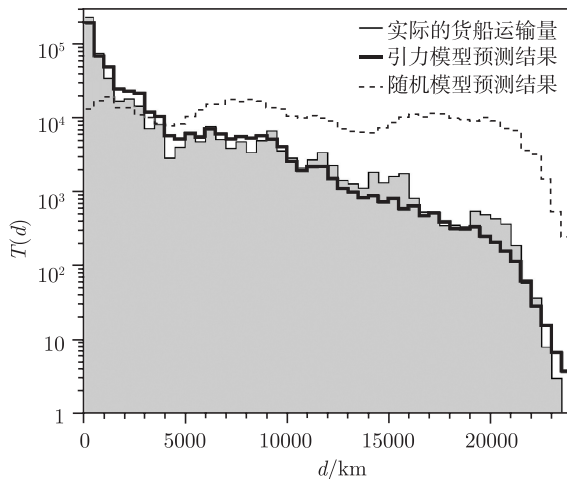


图 1.8 货船海运中的引力定律^[18]

港口间货船运输量与港口距离的统计关系. 贴近直方图的线为双约束引力模型的预测结果, 偏离直方图的线为随机模型预测结果

其中, O_i 是港口 i 驶出的货船总量; D_j 是驶入到港口 j 的货船总量; $f(d_{ij})$ 是一个距离函数; a_i 和 b_j 是两个迭代运算因子 (详见 1.3.3 小节). 他们使用的距离函数是带有指数截断的幂律函数 $f(d) = d^{-\beta}e^{-d/\kappa}$, 其中, 幂指数为 $\beta = 0.59$, 而指数函数中的参数 $\kappa = 4900$ km. 也就是说, 在近 5000 km 的范围内, 货船海运量都服从引力定律.

1.2.8 手机通信中的引力定律

除了人的移动和货物的流动, 在信息的流通中也存在引力定律. Krings 等分析了比利时 250 万名匿名手机用户 2006 年在 571 个城镇之间的手机通信量^[19], 发现地点间通信量正比于两地人口乘积, 负比于两地距离的二次方, 符合标准的引力定律, 见图 1.9.

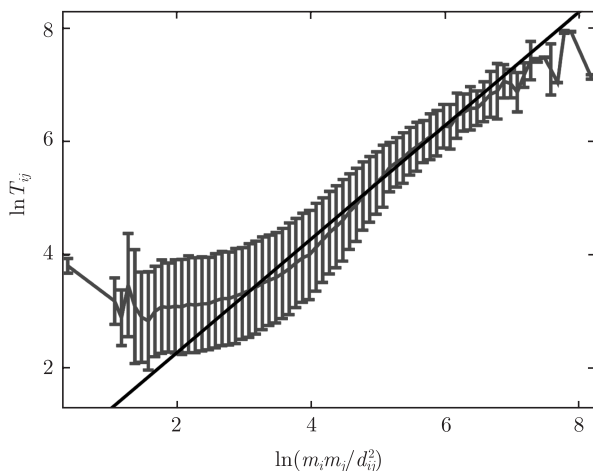


图 1.9 手机通信中的引力定律^[19]

横轴是两地人口乘积与两地距离二次方比值的对数, 纵轴是实际观测到的两地间手机通信量的对数

事实上, 信息流通本身对距离并不像人或物的移动对距离那么敏感, 故在信息流通中也会存在引力定律就不太容易被理解. 我们认为, 信息流通的背后反映的主要是社交网络 —— 手机相互通信的双方很可能是亲人、朋友、工作或商务关系等, 整体上都是反映了社会交往中的关系. 而社交网络中存在引力定律相对就好理解了: 人口越多、距离越近的城市之间, 相互交往的人就会更多.

1.2.9 社交网络中的引力定律

前述社交人数随距离变化的规律已在很多社交数据中被发现. Liben Nowell 等在 LiveJournal^①这一在线社交网络上发现^[20], 两人的交友概率 P 与两人的地理

① 美国的一个博客网站, <https://www.livejournal.com>.

距离 d 之间具有近似幂律关系, 即 $P(d) \sim d^{-1.2}$. 类似规律在社交媒体 Facebook 上也存在^[21, 22], 只是幂指数约为 -1 . 不仅在线上社交媒体, 在线下社会交往中也存在类似规律. Illenberger 等用瑞士休闲社交调查数据研究了交往概率与距离之间的关系^[23], 发现二者之间具有幂指数为 -1.4 的幂律关系. 这些结果都说明社交概率会随地理距离增加而降低, 即越近的人才越容易开展社交. 不仅如此, Garcia-Gavilanes 等还发现国家之间的用户社交强度符合引力定律^[24], 即社交强度不仅与地理距离负相关, 还与两地规模乘积正相关. 他们用社交媒体 Twitter 上 1300 万用户的社交数据, 分析了上百个国家之间的用户社交强度 (用平均提及次数表示) 与国家规模 (用 Twitter 用户数表示)、地理距离之间的关系, 发现这一关系能用引力模型很好地描述, 见图 1.10.

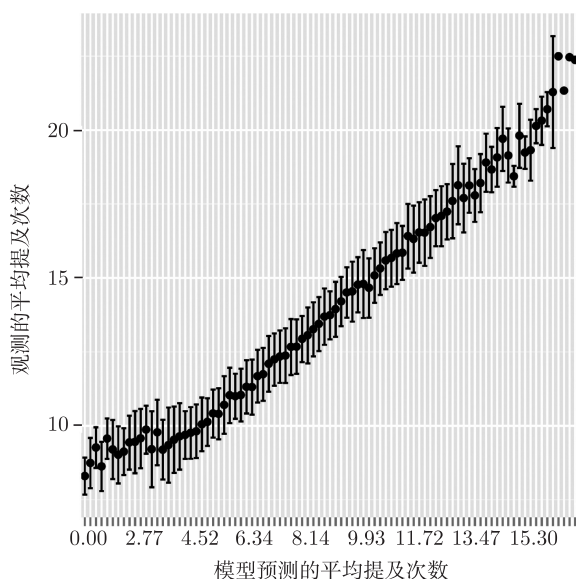


图 1.10 在线社交网络中的引力定律^[24]

横轴是引力模型预测得到的国家间 Twitter 用户平均提及次数 (反映用户社交强度), 纵轴是实际观测到的对应国家间 Twitter 用户平均提及次数, 两者的皮尔森相关系数为 0.919

1.2.10 科研网络中的引力定律

科学研究中学者之间的合作关系也是社会交往关系的一种. 这种关系可以从储存在文献数据库中的论文共同作者信息中获取, 即如果有多个人是同一篇论文的作者, 那么他们两两之间都具有合作关系. 进一步地, 可以将这种学者合作关系扩展到城市之间的科研合作关系——如果城市 i 与城市 j 的学者在一篇论文中有合作, 那么就给这两城市间的合作强度 w_{ij}^{Col} 加一, 而城市 i 的科研强度则可以用

$s_i = \sum_j w_{ij}^{\text{Col}}$ 来体现.

Pan 等利用科学信息研究所 (Institute for Scientific Information, ISI) 数据库中 2003~2010 年的论文数据, 从中提取了 18199 个城市之间的合作强度矩阵^[25]. 他们发现, 城市间的科研合作网络服从引力定律, 即 $w_{ij}^{\text{Col}} \sim s_i s_j / d_{ij}^\alpha$, 见图 1.11(a)、(b). 这反映了科研水平越高、距离越近的两个城市之间, 科研合作的强度就越强. 不仅如此, 他们在论文引用网络中也发现了类似的规律, 见图 1.11(c)、(d), 这说明引力定律在科研网络中也是普遍存在的.

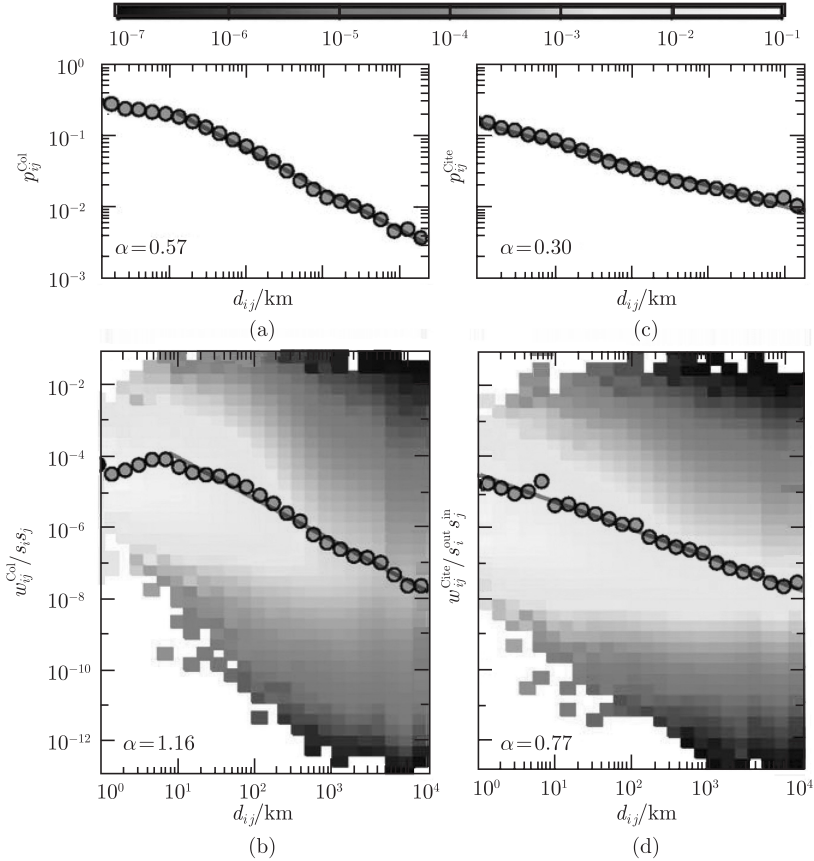


图 1.11 科研网络中的引力定律^[25]

- (a) 科研合作网络中, 两城市之间存在合作的概率与距离之间的关系. 图中直线为 $y \sim x^{-0.57}$. (b) 科研合作网络中, 两城市之间合作强度与城市科研水平乘积的比值 (纵轴) 与距离 (横轴) 的关系. 图中直线为 $y \sim x^{-1.16}$. (c) 论文引用网络中, 两城市之间存在论文引用的概率与距离之间的关系. 图中直线为 $y \sim x^{-0.30}$. (d) 论文引用网络中, 两城市之间引用强度与城市引用总量 s_i^{out} 被引用总量 s_j^{in} 乘积的比值 (纵轴) 与距离 (横轴) 的关系. 图中直线为 $y \sim x^{-0.77}$.