

基于大数据的流空间研究进展

杨延杰^{1,2}, 尹丹^{1,2}, 刘紫玟^{1,2}, 黄庆旭^{1,2*}, 何春阳^{1,2}, 吴康³

1. 北京师范大学地表过程与资源生态国家重点实验室, 人与环境系统可持续研究中心, 北京 100875;
2. 北京师范大学地理科学学部自然资源学院, 土地资源与区域发展研究中心, 北京 100875;
3. 首都经济贸易大学城市群系统演化与可持续发展的决策模拟研究北京市重点实验室, 北京 100070)

摘要:流空间是认识城市网络结构和演化的重要手段。近年来大数据的快速发展为流空间研究提供了新的机遇和挑战。论文系统综述了基于大数据的流空间研究进展。首先, 论文梳理了基于大数据流空间研究的背景和历史, 然后总结了基于大数据的流空间研究的主题、数据类型、方法和主要发现, 最后展望了未来的研究挑战。2011年以后, 基于大数据的流空间研究呈指数增长趋势, 中英文论文年均发表量从2010年的11篇增长到2018年的106篇。大数据主要从提供新的数据源、激发新的分析方法和提供新的研究视角三方面推进了流空间研究。常用于流空间研究的大数据主要包括手机信令数据、社交媒体签到数据、公共交通刷卡数据和出租车轨迹数据, 它们比传统统计数据更能直接提供人流、物流和信息流的时空动态信息。研究方法也从传统的基于距离的重力模型发展为网络分析方法。未来在交叉学科研究、大数据和传统数据的耦合、大数据与深度学习和云计算等新方法的结合方面仍需进一步探索, 从理论、数据和方法上全面深化流空间研究。

关键词:流空间; 大数据; 城市网络; 人员移动; 城市可持续性

流空间是以信息技术为基础的人流、物流和信息流等组成的双向或者多向流动的网络化的动态空间^[1-2]。它使城市地理学的研究视野由基于位空间(space of places)理论的城市等级静态结构转向由人流、物流和信息流构建的城市网络结构和关系^[3]。早期的研究多使用统计和问卷调查数据, 但统计数据获取难度大、数据种类少, 对小城市的记录有限, 难以有效表征真实的人流、物流和信息流, 而问卷数据获取难度较大, 流空间的研究遇到了数据瓶颈。大数据为流空间研究带来了新的机遇^[4]。一方面, 大数据记录了丰富和实时的人流、物流和信息流, 为流空间研究提供了广泛的数据来源, 有助于开展精细化和多尺度的流空间研究^[5-6]。另一方面, 基于大数据的流空间研究也涌现出一些新的研究方法, 比如复杂网络^[7]和人员移动建模^[8]等。因此,

及时了解基于大数据的流空间研究的数据特点、分析方法和存在的挑战, 对深化流空间研究具有重要意义。

已有学者从多个方面对流空间研究进行了综述。在理论上, 孙中伟等^[3]将流空间的形成机制总结为“技术—经济—空间”模式。沈丽珍等^[9]认为流空间具有虚体空间和实体空间的二元属性。在方法上, 陆锋等^[6]认为流空间的人员移动研究方法主要包括统计物理学方法、数据挖掘方法、复杂网络分析方法和空间交互理论方法。刘瑜等^[5]将人员移动模式的基本度量方法归纳为移动步长、角度分布和回旋半径等。在主题上, 秦萧等^[10]综述了国内外应用大数据进行城市时空间行为研究的最新进展, 从数据获取方式、数据结构和内容, 以及分析手段与方法角度构建了基于大数据应用的城市时空间

收稿日期: 2019-07-15; 修订日期: 2019-10-31。

基金项目: 北京市科技新星项目(Z181100006218049)。[Foundation: Beijing Nova Program, No. Z181100006218049.]

第一作者简介: 杨延杰(1992—), 男, 山西太原人, 硕士生, 主要从事大数据和城市景观过程模拟研究。

E-mail: yjyang@mail.bnu.edu.cn

*通信作者简介: 黄庆旭(1984—), 男, 四川成都人, 副教授, 主要从事城市景观过程和城市可持续性研究。

E-mail: qxhuang@bnu.edu.cn

引用格式: 杨延杰, 尹丹, 刘紫玟, 等. 基于大数据的流空间研究进展 [J]. 地理科学进展, 2020, 39(8): 1397-1411. [Yang Yanjie, Yin Dan, Liu Ziwen, et al. Research progress on the space of flow using big data. Progress in Geography, 2020, 39(8): 1397-1411.]
DOI: 10.18306/dlkxjz.2020.08.013

行为研究方法框架。但是,目前对基于大数据开展的流空间研究缺乏系统的综述,对用于流空间研究的大数据类型、主题和方法仍没有一个清晰的认识。

本文目的是综述基于大数据的流空间研究进展。首先,辨析流空间和大数据的基本概念;然后,梳理基于大数据流空间研究的背景和历史,并利用系统综述法检索和筛选相关文献,了解基于大数据流空间研究的趋势和特点;在此基础上,总结基于大数据流空间研究的数据类型和研究方法;最后归纳已有研究存在的问题和不足,并对未来的研究提出建议。

1 流空间和大数据的基本定义

1.1 流空间的基本定义

1989年,Manuel Castells在《*The Informational City: Information Technology, Economic Restructuring, and the Urban-Regional Progress*》一书中首次提出流空间^[1]。之后,他在《*The Rise of the Network Society*》^[11]一书中进一步解释了流空间的定义,他认为流空间是通过流动而运作的共享时间之社会实践的物质组织。这也是目前最广泛采用的流空间定义,其在谷歌学术和百度学术中的被引频次分别达到39292次和8532次。之后,Appadurai^[12]提出了技术景观、金融景观和媒体景观等5个主要的流空间类型。中国学者在此基础上补充和本地化流空间概念。如甄峰^[13]提出的灰空间作为虚空间和实空间之间的过渡性空间,与流空间十分相似;孙中伟等^[3]从地理学视角分析了流空间的形成机制和特征,认为流空间是在网络空间引导下位空间的新表现形式;高鑫等^[14]认为流空间是信息流引导下的一系列流以及通过各种流的不断运动所形成的空间场域;目前,沈丽珍等^[15]对流空间的定义最为常用,他们认为流空间是围绕人流、物流、资金流、技术流和信息流等要素流动而建立起来的空间,以信息技术为基础的网络流线和快速交通流线为支撑,创造一种有目的的、反复的、程式化的动态运动。综合以上定义,本文中使用的流空间的定义是围绕人流、物流和信息流等核心要素,基于流动而建立起来的具有多尺度特征的动态空间。基于大数据的流空间研究既包括城市内部人员移动特征和城市间的要素流动,也包括区域和全球等大尺度的相互联系。

近年来,随着信息技术的快速发展,以及城市化和全球化的崛起,城市内和城市之间的交流除了传统的人流和物流之外,信息和技术流也愈发丰富。因此城市成为流空间研究的重要载体。

1.2 大数据的基本定义

1980年,著名未来学家、社会思想家 Alvin Toffler^[16]在《*The Third Wave*》一书中,首次提及大数据,但这一概念并没有引起人们的关注。进入21世纪,随着网络技术的发展,数据开始呈现爆炸式增长,大数据的概念和应用开始受到关注。2001年,Laney^[17]使用3Vs模型(volume、variety和velocity)来描述数据的特征,模型被Gartner、IBM和微软等公司广泛用于描述大数据的特征。之后,Gartner公司将大数据定义为具有高容量、高增长率和类型多样化的信息资产^[18]。Nature杂志在“Big data: Science in the petabyte era”专题文章中,将大数据定义为数据规模大,在有限时间内无法用目前的方法来获取、管理和处理的数据集^[19]。一些学者也对大数据的定义进行了探讨^[20-21]。但总体上,目前还没有对大数据形成一致的定义。参考已有研究^[22-23],本文中使用的大数据的定义是基于传感器和网络等手段,以非目的性的方式得到的,具有数据规模大、更新快、价值密度低和类型丰富等特征,以用户个体为信息记录单元、能够有效揭示人类行为特征的数据。因此,本文所使用的大数据不涉及遥感大数据。作为区别于传统数据的新数据类型,大数据已经在互联网、金融业和交通业等行业得到广泛应用。

近年来,越来越多的学者关注大数据在地理学中的应用和发展。如宋长青^[24]归纳了地理大数据的研究范式;程昌秀等^[25]探讨了如何利用大数据开展地理复杂性研究;裴韬等^[23]从地理大数据的特征、地理大数据挖掘的核心问题和挖掘方法等方面进一步总结了大数据在地理学研究中的应用。此外,也有学者总结了特定类型大数据在地理学中的应用。如龙瀛等^[26]总结了公共交通智能卡数据在定量城市研究中的应用;Ilieva等^[27]总结了社交媒体数据在城市可持续研究中的应用;Martí等^[28]梳理了社交媒体数据在城市研究中的挑战、机会和局限性。但目前研究对大数据在流空间研究中的应用缺乏系统的认识,及时对基于大数据的流空间研究进行综述显得十分必要。

2 流空间的研究历程

流空间的提出可以追溯到位空间(图1)。位空间主要是在区域尺度研究以距离为核心的,基于区域经济一体化的空间相互作用^[1,29]。国内外科学家发展了一系列理论和量化方法研究位空间。其中,代表性的理论有农业区位论、工业区位论和中心地理论等^[30]。这些理论模型主要关注区域和城市由于距离和空间相互作用产生的分布规律。代表性的量化方法包括借鉴物理学万有引力定律的引力模型和潜力模型。这些方法一般以城市人口、GDP等指标表示城市的质量,以2个城市的实际距离或运输成本等作为距离因子。

随着全球化和信息技术的进步,全球在近30 a经历了快速的时空压缩过程,流空间研究应运而生^[1]。流空间研究突破对距离和实体流的限制,关注区域间的网络结构和虚拟流。早期的流空间实证研究将世界城市作为全球流空间的关键节点,数据主要来源于统计数据和调查问卷。较为著名的研究团队包括全球化与世界城市研究网络(Globalization and World Cities Research Network, GaWC)团队和欧洲多中心巨型城市区域可持续发展管理(Sustainable Management of European Polycentric Mega-City Regions, Polynet)团队。如GaWC开发了连锁网络模型来衡量城市与世界城市网络的连通性^[31]。Beaverstock等^[32]通过商业银行办事处数据研究了20世纪90年代中期以来银行人员的全球流动。Taylor^[33]利用经济、文化、政治和社会统计数据,从城市网络视角综合评估了世界主要城市的重

要程度。国内也有一些学者基于传统统计数据开展了流空间研究^[34-37]。在此阶段,由于传统数据无法反映个体的流动路径,也难以有效表征信息和文化等虚拟流,流空间的研究主要集中于区域尺度的城市联系,较少关注个体层面的流动^[38]。

基于大数据的流空间研究萌芽于21世纪初期。随着通信技术及计算机技术的飞速发展和广泛应用,以及大量人类行为数据的积累,流空间研究获得大数据助益。多基于位置服务和社交媒体服务的大规模开放数据提供了丰富、详细和实时的信息,降低了流空间研究的数据获取成本,有助于研究人员更加全面地开展流空间研究^[22]。通过提供新的数据源,激发新的分析方法,流空间的研究内容和研究尺度得到了深化^[39-41]。

3 流空间主要研究进展

3.1 论文检索和筛选

基于文献主题(含题目、关键词和摘要)检索,在CNKI中文核心数据库和Web of Science核心数据库集中检索2018年9月28日前发表的中英文论文。本文参考已有研究得到可用于流空间研究的常用大数据类型^[27-28,42],并选择用户量较多的代表性网站和应用以及大数据的一些专有名词作为关键词。同时,参考沈丽珍等^[15]的研究,选择不同流要素类型作为流空间关键词。具体地,在CNKI核心数据库中,大数据检索词为大数据、互联网、手机信令、固定电话、百度地图、百度迁徙、高德地图、滴滴、公交刷卡、签到数据、大众点评、携程、百度指数、贴

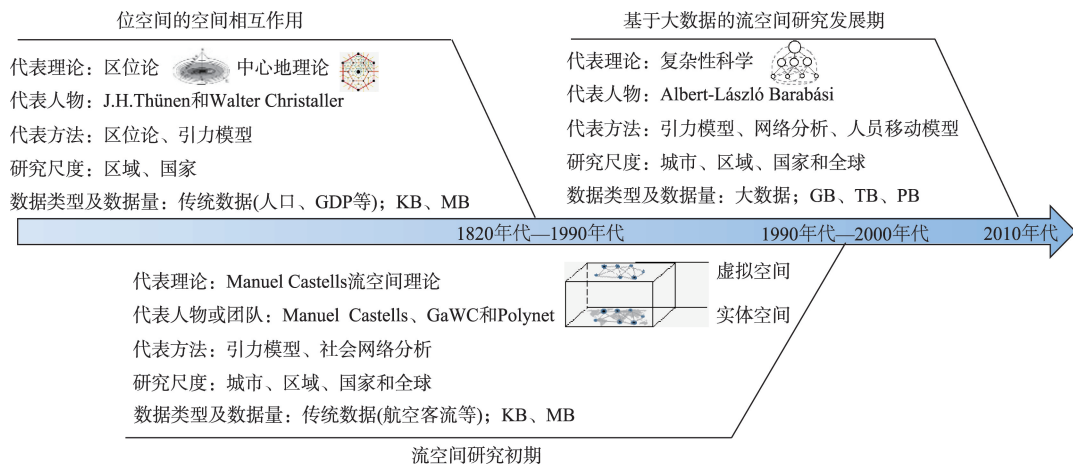


图1 流空间研究的发展历程

Fig.1 History and progress of the studies on the space of flow

吧、微博、网络发帖、网络关注度、网络游记、腾讯、豆瓣同城、数据挖掘、数据爬取;流空间相关检索词为流空间、城市网络、人流、人口流动、人口迁徙、旅游流、信息流、通信流、交通流。在英文数据库中,大数据相关检索词为 big data、mobile phone、telephone、taxi、smart card、Amap、social media、Four-square、Pinterest、Flicker、facebook、Baidu、Google、Tmall、Jingdong、Weibo、twitter、tencent、check-in data;流空间相关检索词为 population flow、travel flow、tourism flow、human mobility、passenger flow、traffic flow、information flow、telecommunication flow、material flow、flow data、urban network、city network、space of flow。

最终共检索到中文论文 1019 篇,英文论文 670 篇。在此基础上,挑选基于大数据的流空间研究论文,即研究内容为基于大数据的流空间研究,最终共筛选到中文论文 268 篇,英文论文 260 篇。由于以上关键词无法涵盖所有基于大数据的流空间研究,可能会造成一些遗漏。但是,由于已经尽可能多地考虑了相关关键词,这些遗漏不会影响本文的主要结论。

3.2 研究趋势和特点

2011 年以后,中英文论文数量呈指数增长趋势(图 2)。2001—2010 年,中文和英文论文年均发表量分别在 2 篇左右,增长趋势较慢。2011 年以后,中英文论文数量呈指数增长趋势,中英文论文年均发表量从 2010 年的 11 篇增长到 2018 年的 106 篇。被引频次前 5 名的中文论文的合计被引频次为 570 次,占论文总被引频次的 20.87%。被引频次前 5 名的英文论文的合计被引频次为 3919 次,占论文总被

引频次的 54.37%。共有 7 篇论文被引频次超过 100 次(表 1)。

3.3 主要研究主题

通过对中英文文献的题目、摘要和关键词进行单词云分析表明,城市网络和人员移动是基于大数据的流空间研究的 2 大主题(图 3)。中文文献的现有研究以城市网络为研究主题,研究尺度集中在全国、区域和城市群,主要研究内容包括揭示城市联系、等级结构和网络演化等方面。中文文献中,前 10 位的高频词分别是城市网络、大数据、人口流动、中国、空间结构、旅游流、东北地区、社会网络分析、流空间和复杂网络。从时间动态上看,基于大数据的流空间研究关注的流要素类型呈现出从信息流到人口流动和旅游流的变化特征。产生这种变化的一个主要原因是大数据的获取难度降低。

英文文献的现有研究以人员移动为研究主题,研究尺度集中在城市内部,主要研究内容是揭示人员移动的特征和预测人员移动规律。英文文献中,前 10 位的高频词分别是系统、人员移动、行为、轨迹、预测、出租车、大数据、城市规划、旅游和交通。在聚类结果上,基于大数据的流空间研究主要关注移动行为和格局,以及人员移动预测。

3.4 大数据在流空间研究中的优势

(1) 大数据丰富了流空间研究内容。首先,大数据提供了丰富的基于个体记录的数据,为测量虚拟流的强度和方向提供了数据支撑,有效促进了虚拟流视角下的流空间研究。如陈映雪等^[52]的研究表明,中国城市网络信息关系呈现出“一超多强”的空间极化不对称格局;王波等^[46]的研究发现,信息技术的发展降低了距离对交往的限制,增加了人们

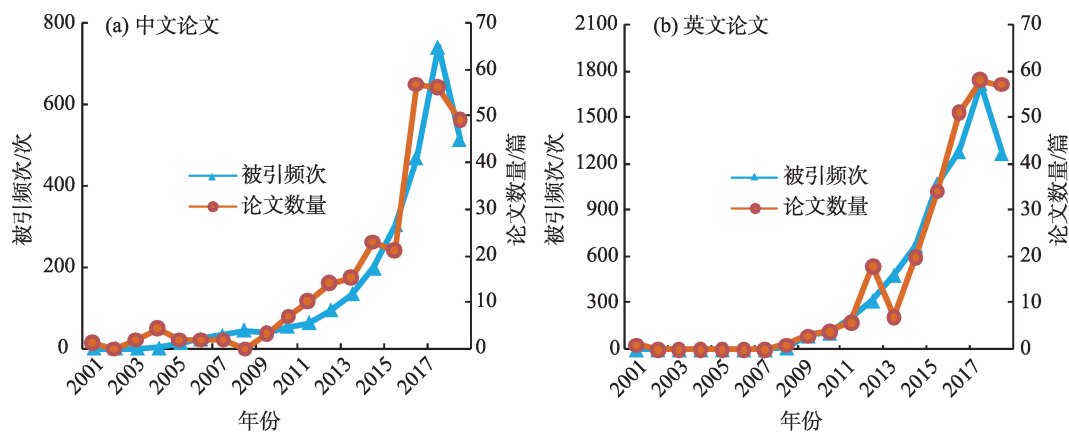


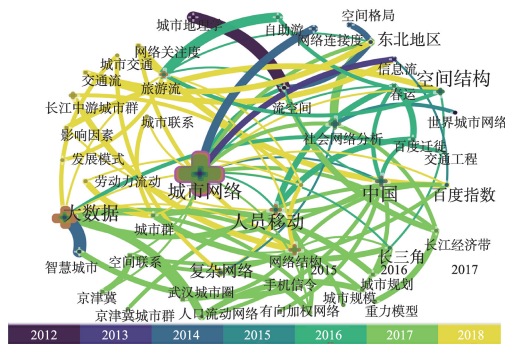
图 2 2001—2018 年论文数量和被引频次的年际变化

Fig.2 Changes in the number of papers and their citations, 2001–2018

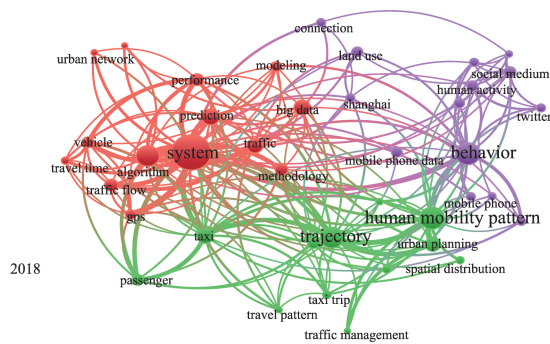
表1 高被引的中文和英文论文

Tab.1 Top five papers published by Chinese and English journals with the highest citations

文献类型	文献来源	发表年份	论文题目	期刊	引用量/次
中文	李山等 ^[43]	2008	基于百度指数的旅游景区络空间关注度: 时间分布及其前兆效应	地理与地理信息科学	203
	甄峰等 ^[44]	2012	基于网络社会空间的中国城市网络特征——以新浪微博为例	地理学报	304
	熊丽芳等 ^[45]	2013	基于百度指数的长三角核心区城市网络特征研究	经济地理	132
	王波等 ^[46]	2013	基于微博用户关系的网络信息地理研究——以新浪微博为例	地理研究	81
	董超等 ^[47]	2014	基于通信流的吉林省流空间网络格局	地理学报	55
英文	Gonzalez 等 ^[48]	2008	Understanding individual human mobility patterns	<i>Nature</i>	2042
	Song 等 ^[49]	2010	Limits of predictability in human mobility	<i>Science</i>	1015
	Lu 等 ^[50]	2012	Predictability of population displacement after the 2010 Haiti earthquake	<i>PNAS</i>	144
	Liu 等 ^[40]	2014	Uncovering Patterns of Inter-Urban Trip and Spatial Interaction from social media Check-In data	<i>PLoS One</i>	122
	Long 等 ^[51]	2015	Combining smart card data and household travel survey to analyze jobs-housing relationships in Beijing	<i>Environment and Urban Systems</i>	55



(a) 中文论文



(b) 英文论文

注:对选取的268篇中文文献和260篇英文文献的题目、摘要和关键词进行单词云分析;单词云中的字号大小表征单词出现的频率,频率越高字号越大。图a基于CiteSpace制作,依据单词共现关系和出现时间,将其用不同颜色聚类;图b基于VOSviewer制作,依据单词的共现关系将其用不同颜色聚类。

图3 基于大数据的流空间研究单词云

Fig.3 Word clouds of the studies on the space of flow using big data

的社会交往距离;熊丽芳等^[53]的研究表明,基于百度信息流的城市网络具有明显的层级特征。其次,大数据可以实时记录人流、物流和信息流的信息,弥补了传统数据在揭示城市网络的演化特征研究中数据连续性较差的不足,丰富了流空间的动态研究。如熊丽芳等^[45]的研究表明,长三角地区各城市信息联系总量增加,年际变化减小。最后,大数据有效揭示了人员移动的时空规律。如Xia等^[54]的研究发现,工作日和周末的人口流动规律存在明显差异。

(2) 大数据丰富了流空间研究尺度。基于大数据的研究尺度横跨全球、国家、区域和城市内部,而

基于传统数据的研究由于数据的可比性和完整性,大多集中在国家和区域尺度。在全球尺度,大数据有助于计算全球城市联系和划分城市等级。如Devriendt等^[55]基于谷歌搜索数据,从信息流的角度衡量城市的重要性,并得到全球100个大城市的信息流排名。在国家及区域尺度,大数据有助于揭示城市联系的时空规律。如赵梓渝等^[56]、刘望保等^[57]分别基于百度迁徙数据研究了中国城市人口流动特征;陈映雪等^[52]和甄峰等^[44]基于新浪微博数据研究了中国信息流的时空格局;Liu等^[40]基于社交媒体数据分析了中国城际间的旅行和空间互动模

式。在城市尺度,大数据有助于识别人口流动的热点区域,并揭示人口流动规律。如龙瀛等^[58]根据公交刷卡数据识别和可视化了北京地区的通勤出行;钟炜菁等^[59]利用手机信令数据从城市内部和居民个体的角度分析了上海市人口的空间动态分布特征;黄洁等^[60]基于公共交通刷卡数据分析了北京市地铁客流的时空分布特征。此外,已有学者基于带有地理信息的照片^[61]和旅游攻略^[62]等数字足迹数据,分析了游客在城市热点区域的流动特征。

除上述的优势外,大数据在流空间研究中还有一些应用潜力。首先,大数据可以表征人流、物流和信息流等的相互作用。如春节期间的大规模迁徙改变了人口的空间格局,进而引起信息流格局产生较大变化。同样,信息流也可以通过影响人的决策来引起人流强度和方向的变化。比如反映交通实时拥堵的信息流可以影响用户对通勤线路的选择,进而改变交通流。其次,大数据有利于开展多种流共同作用下的流空间研究。通过多源大数据综合分析区域人流、物流和信息流,可以全面揭示区域流空间特征。此外,大数据有利于从微观角度探究流向和流强度。大数据具有精细化程度高的特点,可以通过聚焦于小区域来探究流的强度和流向等特征。最后,大数据更有利于刻画流空间的复杂性。城市的交通流和人流规模庞大,具有较高的时空异质性。基于车辆的定位信息和手机的信息可以连续记录交通流和人流的变化。海量的时空信息为开展流空间的复杂性研究提供了数据基础。

3.5 主要研究数据

与传统数据相比,大数据在研究流空间方面具有独特的优势。与基于航空、铁路和公路的客运班次数据相比,大数据可以反映实际产生的人员移

动。与问卷调查等方法获取的人口流动数据相比,大数据可以通过用户通话和刷卡等行为附带产生的地理位置来记录用户的移动信息,避免了主观性强、数据获取成本高的问题。目前,用于流空间研究的大数据主要包括手机数据、社交媒体数据、公共交通刷卡数据、出租车轨迹数据和其他数据^[25,42]。这些大数据均以个体为信息记录单元,具有较为连续的信息,可以有效揭示人流和信息流等要素的时空规律。同时,这些大数据也存在各自的优势和劣势(表2)。

手机信令数据具有覆盖范围广、来源丰富、实时性强的特点。该数据可以记录个体移动的时空信息和个体之间的信息交流,时空精度较高。大量的研究表明,手机信令数据在城市网络特征和人员移动规律研究中具有重要的作用。如Trasarti等^[63]基于手机信令数据分析了城市不同区域之间的相互联系;Çolak等^[66]通过结合手机信令数据和城市公路网数据,对城市早高峰交通流的特征进行了综合分析;王垚等^[63]基于手机信令数据分析了江西省北部地区的城市联系的特征和差异。但手机信令数据的质量受到基站分布的影响,城市地区的数据质量明显高于农村地区。

社交媒体数据来源于用户发布的具有地理位置信息的照片、文本和视频等,具有时间跨度长、覆盖范围广和语义丰富等特点。随着移动互联网的快速发展,社交媒体数据呈现爆发式增长,为流空间研究提供了重要的数据源。目前,社交媒体数据已经被广泛应用于交通网络、消费联系和信息交流等研究。如Zhang等^[67]基于微博数据分析了长江三角洲地区城市间联系特征;Li等^[68]基于微博数据分析了城市内部人员移动模式,有效揭示了城市内部

表2 基于大数据的流空间研究中主要涉及的大数据

Tab.2 Major types of data involved in research on the space of flow using big data

大数据	数据概述	特点	流空间类型	文献来源
手机信令数据	记录手机用户在发生通话、收发短信、位置更新等事件时手机连接的基站位置信息	覆盖人群广,时间连续性好;存在精度受基站分布影响、数据质量取决于用户通话的频率和数据跳站等问题	人流、信息流	王垚等 ^[63]
社交媒体签到数据	来源于社交平台,具有地理位置信息	容易获取,具有语义特征;存在数据不能覆盖所有年龄段的人群、签到数据不完整和签到位置间不合理的快速移动等问题	人流、信息流	Zhen等 ^[64]
公共交通刷卡数据	记录乘客的卡号、上下车位置和时间等信息	数据获取成本低,覆盖人群广;数据连续性差,空间分辨率低,只能反映用户在不同刷卡位置间的移动	人流、交通流	陆锋等 ^[65]
出租车轨迹数据	记录地理位置、时间和是否载客等信息	数据精度高,可以详细记录车辆的行驶轨迹;存在覆盖范围集中于城市内部、移动轨迹受路网限制、覆盖人群较小等问题	人流、交通流	Barbosa等 ^[42]

的流动规律;秦静等^[61]利用Flickr数据探究了北京市入境游客热点区域的空间分布格局和流动特征。但社交媒体数据存在数据冗余、有效数据的密度较低等问题,数据处理较为复杂。

公共交通刷卡数据记录了用户乘坐公交或地铁出行的信息,内容包括上下车的位置、时间和费用等。公共交通刷卡数据可以反映用户的出行轨迹,从而有效揭示城市居民出行的时空特征。目前,公共交通刷卡数据已被广泛应用于探究城市内部区域间人口流动特征和人员移动时空规律等。如Zhou等^[69]基于刷卡数据分析了北京交通的通勤模式;Wang等^[70]基于地铁刷卡数据,分析了城市旅游流与城市功能结构之间的关系;Zhong等^[71]基于公交刷卡数据比较了伦敦、新加坡和北京市人员移动的日变化规律。但公共交通刷卡数据受到城市路网限制,覆盖范围也集中于城市内部。

出租车轨迹数据记录了出租车每隔一定时间的位置信息,主要包括时间、空间位置和是否载客等^[72]。基于出租车轨迹数据分析流空间,可以有效揭示城市内部出租车乘客的流动特征。另外,基于出租车轨迹数据,可以深入挖掘用户的出行行为,如不同时刻的动态分布、周末和周内的出行规律等。如Zhang等^[73]基于出租车轨迹数据估计了不同时间内出租车的服务范围,进而揭示了城市内出租车的流动规律;Liu等^[39]基于出租车轨迹数据分析了上海市人员移动的时空特征;Tang等^[74]基于出租车轨迹数据探究了哈尔滨市人员移动的规律。与公共交通刷卡数据相比,出租车轨迹数据具有地理信息丰富和空间分布范围广的优势。但由于出租车乘客主要集中在特定人群,基于出租车轨迹数据的流空间研究需要注意数据代表性问题。

来自迁徙、物流和新闻等的其他大数据也引起学者们的广泛关注,并已经应用于流空间研究。这些数据主要来自于互联网企业,记录了海量的用户行为信息,可以有效揭示人流、物流和信息流,极大地丰富了用于流空间研究的大数据。如迁徙数据记录了城市间人流量,可以表征城市间人流的热度。如赖建波等^[75]运用复杂网络方法揭示了腾讯迁徙数据下中国春运期间城市间人口流动的空间格局;魏冶等^[76]利用百度迁徙数据分析了中国春运人口流动网络的富人俱乐部现象与不平衡性。同时,物流大数据可以通过刻画城市间物流信息来揭示基于物流的城市联系特征。如Xi等^[77]基于天猫

购物网站的在线商品揭示中国城市网络的结构特征;李鲁奇等^[78]基于“宅急送”网站的物流数据分析了“双十一”期间中国快递流通的时空结构和效率。此外,新闻大数据也成为衡量信息流的有效手段。如Lin等^[79]利用新浪新闻数据研究城市联系特征,并将其应用于城市景观过程模拟研究;秦昆等^[80]基于GDELT新闻数据分析了国家间的交互关系。将不同来源、不同类型的大数据应用于流空间研究,有利于从不同视角揭示流空间特征,进而不断丰富流空间研究。

3.6 主要研究内容和方法

在基于大数据流空间研究中,研究内容和方法随学科的交叉和大数据的信息挖掘应运而生。地理学研究者主要关注流空间的时空动态研究。依据研究对象的不同,现有研究内容主要分为2类。

第一类是从流的数量、方向和强度等方面开展基于几何空间的流空间特征研究,常用的研究方法包括流量、流向和流强度分析等。大数据为分析流量大小、流向和流强度的时空特征提供了数据基础。具体地,在流量分析方面,大数据记录了海量的用户行为信息,可以直观地表征人流和信息流等数量。如潘碧麟等^[81]基于微博数据估计人口流动总量,并通过构建人口流动矩阵揭示了成渝城市群人口流动规律;Fan等^[82]基于手机数据分析了城市人口通量的时间变化规律。在流向分析方面,大数据记录了用户随时间变化的移动轨迹,有利于分析流向变化特征。如Yao等^[83]基于逐步时空聚类法识别了流的时空变化趋势;von Landesberger等^[84]基于移动图聚类方法识别了伦敦市常住人口的日常流动规律;彭卉等^[85]利用停留点提取和高斯核密度方法揭示了基于手机信令数据的北京市周末城乡交互模式。由于大数据存在的大量交叉点和轨迹重叠等问题,开展流向分析前需要进行复杂的数据聚类处理。在流强度分析方面,熊丽芳等^[45]基于百度指数衡量信息流强度,并分析了中国长三角核心区城市网络的时空演变规律。这些流分析方法可以有效衡量流的时空规律,从而直观地表现流特征。

第二类是基于流空间视角探究拓扑空间的网络结构,主要的研究方法是网络分析。网络分析是将社会个体视为节点,将个体之间的关系视为节点之间的联系,并描述节点之间的联系所构成的网络。在流空间研究中,网络分析一般应用于城市间和城市内部的流空间研究。在城市间,已有研究一

般以城市作为分析单位,从关系角度研究城市间人流、物流和信息流等的网络特征^[86]。将网络分析方法应用于基于大数据的流空间研究,可以有效揭示城市间的网络结构。如赵映慧等^[87]运用优势流分析法、C-value和D-value层次分析法从信息流角度分析了东北三省城市网络联系格局;Zhang等^[88]运用度中心性构建长三角地区城市网络,进而分析了城市间的人口流动特征;王姣娥等^[89]从城市节点、流量和子网络视角比较了中国城市网络的结构特征。在城市内部,已有研究主要以市辖区、街道等行政单元或者以景点为代表的热点区域为目标区域来揭示城市内部不同区域间的流动规律。如Liu等^[90]基于出租车数据构建空间嵌入式网络,采用复杂网络方法揭示了上海市的旅行格局;靳诚等^[91]基于流入度、流出度和网络密度等方法探讨了南京市旅游流的空间特征。但需要注意的是,网络分析方法在有效利用大数据中的语义信息方面存在明显不足。

此外,越来越多的研究者利用机器学习和多智能体建模等方法开展不同流的模拟和预测研究。代表性的研究有人员移动建模。人员移动建模是指通过统计模型分析人员移动特征,以及基于机器学习和多智能体建模等方法对个体或者群体在地理空间移动过程进行模拟和预测^[5-6,92]。大数据可以弥补传统数据样本量少和数据覆盖范围小等问题,为构建人员移动模型提供了重要的数据来源^[5,93]。如Huang等^[94]基于时空语义模型,从时间、空间和语义上定量表达了人类活动模式;Ling等^[95]利用多个模型预测了地铁网络中的动态客流;Barchiesi等^[8]基于机器学习算法,利用Flickr数据探究了人员移动的规律。用于人员移动模拟与预测的模型还包括多智能体^[96]和神经网络^[97]等模型。但人员移动行为的不确定性给模型的有效性增加了难度,且现有

模型在模型普适性和机理解释方面都存在不足。未来应该通过融入大数据中蕴含的语义信息,进一步提高模型的准确性。主要方法、原理及特点归纳见表3。

4 讨论

4.1 现有研究不足

大数据也是一把“双刃剑”,我们需要客观地思考大数据,避免数据“陷阱”^[20,101]。目前,基于大数据的流空间研究面临着跨学科研究挑战大、数据自身存在缺陷和数据分析方法滞后等问题。

在理论上,基于大数据的流空间研究还没有形成完整的研究框架。大数据作为一种新的数据源,本身并不是科学知识^[22]。目前,基于大数据的流空间研究涉及物理学、地理学和计算机科学等领域。不同学科的研究性质不同,进而导致不同领域基于大数据的流空间研究的方法和主题各有侧重。

在内容上,基于大数据的流空间研究在理解不同流的作用方式存在不足。现有研究重点关注由人流和信息流等构成的网络的结构和演化规律,以及基于用户视角的人员移动特征。但不同流之间并非独立存在,而是相互影响,共同作用于流空间。利用大数据探究不同流之间的关系和相互作用规律,有助于深入认识流空间。

在数据上,大数据在流空间研究中的有效性有待商榷。首先,不同类型的大数据所代表的用户群体存在差异,并且很难覆盖所有的人群^[26]。如微博数据主要来自于年轻人,出租车轨迹数据主要来自于城市内部以出租车为主要出行方式的居民。其次,大数据的获取时间不同,数据特征可能存在较大差异。如何根据研究内容选择合适的数据类型和数据范围,从而使数据具有代表性,是目前研究

表3 基于大数据的流空间研究中主要涉及的方法

Tab.3 Representative methods used to analyze the space of flow using big data

方法	原理	特点	典型应用
流分析	通过探究流量、方向和强度等特征来揭示流的时空特征,主要包括流量分析、流向分析和流强度分析等	流分析方法可以有效衡量流的时空规律,从而直观地表现流特征	人口通量变化 ^[82] 高性能空间聚类 ^[98]
网络分析	通过社会网络分析和复杂网络分析等方法探究网络中各节点的联系程度,并揭示人流、物流和信息流等在不同节点间的流动特征	可以有效揭示流空间的节点特征和网络特征;难以利用大数据中的语义信息	信息流强度 ^[99] 网络密度和中心度 ^[100]
人员移动建模	通过建立模型挖掘海量数据中所隐藏的规律,揭示人员移动特征和预测人员移动行为	人员移动模型有助于揭示流空间规律的深层机制,深化人们对流空间的理解;现有模型在模型普适性和机理解释方面存在不足	总体模型 ^[8] 时空语义模型 ^[94]

需要重点关注的问题。此外,大数据的地理位置信息可能存在误差。如手机信令数据的空间精度受基站分布影响,基于手机提取的轨迹与用户实际移动轨迹可能存在差异^[92,102]。最后,如何有效结合大数据和传统数据开展流空间研究也是值得思考的问题。

在方法上,目前还没有形成成熟的方法体系。网络分析和人员移动建模等方法被广泛用于基于大数据的流空间研究中,极大地丰富了流空间研究方法。但无论是基于城市网络揭示城市联系、等级和演化特征,还是探究人员移动特征和模拟人员移动规律方面,目前的研究方法主要针对特定的情景和数据,存在普适性低、迁移性差等问题。此外,目前的聚类方法在海量数据计算方面不能很好地满足分析要求,需要开发适合大数据的高性能聚类算法^[98]。最后,大数据价值密度低、数据结构复杂,如何通过海量的数据进行处理和分析,并发现新规律和价值,也是目前研究需要解决的重要难题。

4.2 构建基于大数据的流空间研究框架

大数据不仅有利于从不同尺度、不同视角验证已有的流空间理论和模型,而且可以从微观角度详细刻画人流、物流和信息流的时空规律,为充分认识流空间带来了新的方向和机遇。未来可以通过加深理论研究、丰富数据源和完善分析方法来构建基于大数据的流空间研究框架,丰富基于大数据的流空间研究(图4)。

在理论上,基于大数据的流空间研究需要通过开展交叉学科研究进一步深化理论基础。一方面,已有学者基于大数据分别从不同视角探索了流空间特征。未来研究需要通过融合地理学、物理学和计算机科学等理论,发展跨学科平台,进一步提高

大数据的应用潜力,丰富流空间研究的理论体系和实践应用。另一方面,在流空间研究中充分利用数据科学理论,为大数据的管理、预处理和计算提供理论依据。

在内容上,基于大数据的流空间研究需要从广度和深度上继续丰富。首先,构建“多视角、多尺度和多方法”的研究框架来创新流空间研究。如在宏观尺度上揭示城市间人口流动和信息传递,在微观尺度开展人员移动模式的预测;又如从静态角度揭示流空间的空间格局,从动态角度探究流空间的时间序列特征,最终从宏观和微观层次、静态和动态视角,完善基于大数据的流空间研究体系。其次,结合大数据探究不同流之间的关系,从而认识不同流的影响方式,以及如何共同作用于流空间。最后,探究流空间对社会发展的影响方式,充分发挥流空间在城市规划、交通优化等方面的优势。

在数据上,基于大数据的流空间研究需要丰富大数据的类型和提高数据利用效率。首先,在使用大数据之前,需要对大数据的质量、完整性和空间精度等方面进行评估。未来的研究可以将大数据与传统数据相结合开展流空间研究,充分结合大数据的相关关系分析和传统数据因果关系研究^[101]。同时,大数据存在有偏性问题,不同类型的数据代表不同人群活动的特点,综合不同类型的大数据开展研究,有利于纠正单一类型的大数据带来的有偏性影响,提高大数据的代表性和可靠性。其次,将云计算和高性能服务器应用到流空间研究中,可以极大地提高数据处理效率。云计算作为一种新的计算模式,具有较强的可靠性和可扩展性。如Gao等^[103]通过构建基于Hadoop (Hadoop Distributed File System)的可扩展的地理处理平台,明显减少了处理Flicker数据的时间。未来研究需要结合这些技术和设备提高数据处理效率。最后,大数据的存储和管理是保证数据安全和高效使用的重要手段。目前,传统的集中式数据库已经无法满足存储大数据的要求,而分布式数据库具有扩展能力强、数据存储和分析能力强等特点,有利于用户高效地查询和处理数据。未来可以通过利用分布式数据库提高多源异构数据的存储和管理效率。

在方法上,基于大数据的流空间研究需要结合复杂性科学和深度学习等构建新的模型。当前研究主要是从节点和网络两方面揭示流空间的网络特征。但流空间是一个复杂系统,传统的网络分析方法难以揭示流空间的复杂性。目前,已有学者开

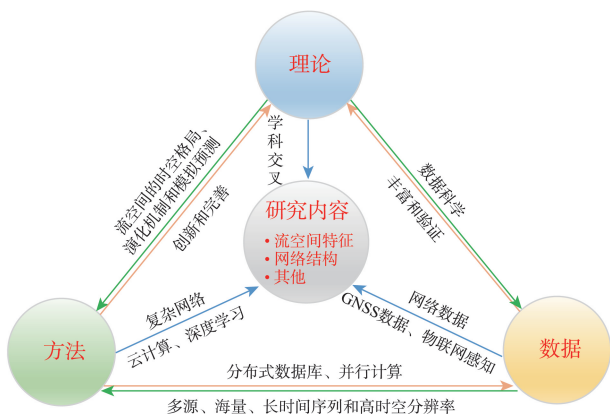


图4 基于大数据的流空间研究框架

Fig.4 Research framework of the space of flow using big data

始基于复杂性科学构建流空间研究方法。如 Yan 等^[104]提出了一种基于复杂网络理论的新方法来研究交通流时间序列变化特征。未来可以将复杂性科学的理论方法广泛应用于流空间研究中,通过建立大数据模型开展定量分析,进而充分挖掘大数据的潜在信息。此外,深度学习可以从大数据中自动提取特征,已被广泛应用于计算机视觉和语音识别等各领域。已有学者开始将深度学习应用于交通流分析和预测。如 Song 等^[105]利用深度学习方法揭示了工作日和周末城市交通流的特征;Liu 等^[106]利用深度学习方法从复杂的交通模式中自动提取有效信息,并实现了交通流预测。未来可以将深度学习应用于地理标记图片的信息识别和社交媒体数据的语义挖掘等方面,提取代表性的特征信息。

5 结论

大数据有力地促进了流空间研究。大数据记录了详细和实时的人流、物流和信息流动态,克服了传统统计数据 and 问卷数据无法直接反映流动信息和获取难度高的问题,为流空间研究提供了丰富的数据源。自2011年以后,基于大数据的流空间研究呈指数增长趋势,中英文论文年均发表量从2010年的11篇增长到2018年的106篇。

基于大数据的流空间研究集中在2个主题,分别是以城市为节点的城市网络结构和以个体为对象的人员移动规律。主要研究方法包括网络分析、流分析和人员移动建模等。但是,目前基于大数据的流空间研究在理论框架系统性、数据有效性和数据挖掘方法等方面仍面临挑战。未来在交叉学科研究、大数据与传统数据的耦合、大数据与深度学习和云计算等新方法的结合方面仍需进一步探索,从而进一步深化流空间研究。

参考文献(References)

- [1] Castells M. The informational city: Information technology, economic restructuring and the urban-regional progress [M]. New York, USA: Blackwell, 1989.
- [2] 魏冶. 流空间视角的沈阳市空间结构研究 [D]. 长春: 东北师范大学, 2013. [Wei Ye. Spatial structure of Shenyang in the perspective of space of flows. Changchun, China: Northeast Normal University, 2013.]
- [3] 孙中伟, 路紫. 流空间基本性质的地理学透视 [J]. 地理与地理信息科学, 2005, 21(1): 109-112. [Sun Zhongwei, Lu Zi. A geographical perspective to the elementary nature of space of flows. Geography and Geo-information Science, 2005, 21(1): 109-112.]
- [4] 甄峰, 秦萧, 席广亮. 信息时代的地理学与人文地理学创新 [J]. 地理科学, 2015, 35(1): 11-18. [Zhen Feng, Qin Xiao, Xi Guangliang. The innovation of geography and human geography in the information era. Scientia Geographica Sinica, 2015, 35(1): 11-18.]
- [5] 刘瑜, 康朝贵, 王法辉. 大数据驱动的人类移动模式和模型研究 [J]. 武汉大学学报信息科学版, 2014, 39(6): 660-666. [Liu Yu, Kang Chaogui, Wang Fahui. Towards big data-driven human mobility patterns and models. Geomatics and Information Science of Wuhan University, 2014, 39(6): 660-666.]
- [6] 陆锋, 刘康, 陈洁. 大数据时代的人类移动性研究 [J]. 地球信息科学学报, 2014, 16(5): 665-672. [Lu Feng, Liu Kang, Chen Jie. Research on human mobility in big data era. Journal of Geo-information Science, 2014, 16(5): 665-672.]
- [7] 蒋小荣, 汪胜兰. 中国地级以上城市人口流动网络研究: 基于百度迁徙大数据的分析 [J]. 中国人口科学, 2017(2): 35-46, 127. [Jiang Xiaorong, Wang Shenglan. Research on China's urban population mobility network: Based on Baidu Migration big data. Chinese Journal of Population Science, 2017(2): 35-46, 127.]
- [8] Barchiesi D, Preis T, Bishop S, et al. Modelling human mobility patterns using photographic data shared online [J]. Royal Society Open Science, 2015, 2(8): 150046. doi: 10.1098/rsos.150046.
- [9] 沈丽珍, 甄峰, 席广亮. 解析信息社会流动空间的概念、属性与特征 [J]. 人文地理, 2012, 27(4): 14-18. [Shen Lizhen, Zhen Feng, Xi Guangliang. Analyzing the concept, attributes and characteristics of the attributes of space of flow in the information society. Human Geography, 2012, 27(4): 14-18.]
- [10] 秦萧, 甄峰, 熊丽芳, 等. 大数据时代城市时空行为研究方法 [J]. 地理科学进展, 2013, 32(9): 1352-1361. [Qing Xiao, Zhen Feng, Xiong Lifang, et al. Methods in urban temporal and spatial behavior research in the big data era. Progress in Geography, 2013, 32(9): 1352-1361.]
- [11] Castells M. The rise of the network society [M]. The information age: Economy, society, and culture Vol 1. New York, USA: Blackwell, 1996: 407-459.
- [12] Appadurai A. Cultural dimensions of globalization [M]. Minnesota, USA: the University of Minnesota Press, 1996: 27-85.
- [13] 甄峰. 信息时代新空间形态研究 [J]. 地理科学进展, 2004, 23(3): 16-26. [Zhen Feng. Researches on new spatial forms in information era. Progress in Geography, 2004, 23(3): 16-26.]

- [14] 高鑫, 修春亮, 魏冶. 城市地理学的“流空间”视角及其中国化研究 [J]. 人文地理, 2012, 27(4): 32-36, 160. [Gao Xin, Xiu Chunliang, Wei Ye. Study on the sinicization of space of flows basing on the visual angle of urban geography. Human Geography, 2012, 27(4): 32-36, 160.]
- [15] 沈丽珍, 顾朝林. 区域流动空间整合与全球城市网络构建 [J]. 地理科学, 2009, 29(6): 787-793. [Shen Lizhen, Gu Chaolin. Integration of regional space of flows and construction of global urban network. Scientia Geographica Sinica, 2009, 29(6): 787-793.]
- [16] Alvin T. The third wave [M]. New York, USA: Bantam Books, 1980.
- [17] Laney D. 3D data management: Controlling data volume, velocity, and variety [R]. META Group Research Note, 6. Stamford, USA: META Group, 2001.
- [18] Ji C Q, Li Y, Qiu W M, et al. Big data processing in cloud computing environments [R]. 12th International Symposium on Pervasive Systems, Algorithms and Networks. San Marcos, USA, 2012: 17-23.
- [19] Graham-Rowe D, Goldston D, Doctorow C, et al. Big data: Science in the petabyte era [J]. Nature, 2008, 455: 8-9.
- [20] Sagioglu S, Sinanc D. Big data: A review [R]. 2013 International Conference on Collaboration Technologies and Systems (CTS). San Diego, USA, 2013: 42-47.
- [21] Hashem I A T, Yaqoob I, Anuarx N B, et al. The rise of "big data" on cloud computing: Review and open research issues [J]. Information Systems, 2015, 47: 98-115.
- [22] 甄峰, 王波. “大数据”热潮下人文地理学研究的再思考 [J]. 地理研究, 2015, 34(5): 803-811. [Zhen Feng, Wang Bo. Rethinking human geography in the age of big data. Geographical Research, 2015, 34(5): 803-811.]
- [23] 裴韬, 刘亚溪, 郭思慧, 等. 地理大数据挖掘的本质 [J]. 地理学报, 2019, 74(3): 586-598. [Pei Tao, Liu Yaxi, Guo Sihui, et al. Principle of big geodata mining. Acta Geographica Sinica, 2019, 74(3): 586-598.]
- [24] 宋长青. 地理学研究范式的思考 [J]. 地理科学进展, 2016, 35(1): 1-3. [Song Changqing. On paradigms of geographical research. Progress in Geography, 2016, 35(1): 1-3.]
- [25] 程昌秀, 史培军, 宋长青, 等. 地理大数据为地理复杂性研究提供新机遇 [J]. 地理学报, 2018, 73(8): 1397-1406. [Cheng Changxiu, Shi Peijun, Song Changqing, et al. Geographic big-data: A new opportunity for geography complexity study. Acta Geographica Sinica, 2018, 73(8): 1397-1406.]
- [26] 龙瀛, 孙立君, 陶遂. 基于公共交通智能卡数据的城市研究综述 [J]. 城市规划学刊, 2015(3): 70-77. [Long Ying, Sun Lijun, Tao Sui. A review of urban studies based on transit smart card data. Urban Planning Forum, 2015(3): 70-77.]
- [27] Ilieva R T, McPhearson T. Social-media data for urban sustainability [J]. Nature Sustainability, 2018, 1: 553-565.
- [28] Martí P, Serrano-Estrada L, Nolasco-Cirugeda A. Social media data: challenges, opportunities and limitations in urban studies [J]. Computers, Environment and Urban Systems, 2019, 74: 161-174.
- [29] 艾少伟, 苗长虹. 从“地方空间”、“流动空间”到“行动者网络空间”: ANT 视角 [J]. 人文地理, 2010, 25(2): 43-49. [Ai Shaowei, Miao Changhong. "Space of places", "space of flows" and "'space of actor-networks": From the perspective of ANT. Human Geography, 2010, 25(2): 43-49.]
- [30] 许学强, 周一星, 宁越敏. 城市地理学 [M]. 北京: 高等教育出版社, 2009: 191-240. [Xu Xueqiang, Zhou Yixing, Ning Yuemin. Urban geography. Beijing, China: Higher Education Press, 2009: 191-240.]
- [31] Taylor P J. Specification of the world city network [J]. Geographical Analysis, 2001, 33(2): 181-194.
- [32] Beaverstock J V, Boardwell J T. Negotiating globalization, transnational corporations and global city financial centres in transient migration studies [J]. Applied Geography, 2000, 20(3): 277-304.
- [33] Taylor P J. Leading world cities: Empirical evaluations of urban nodes in multiple networks [J]. Urban Studies, 2005, 42(9): 1593-1608.
- [34] 赵渺希, 刘铮. 基于生产性服务业的中国城市网络研究 [J]. 城市规划, 2012, 36(9): 23-28, 38. [Zhao Miaoxi, Liu Zheng. Research on China's city network based on production service industry. City Planning Review, 2012, 36(9): 23-28, 38.]
- [35] 李仙德. 基于上市公司网络的长三角城市网络空间结构研究 [J]. 地理科学进展, 2014, 33(12): 1587-1600. [Li Xiande. Spatial structure of the Yangtze River Delta urban network based on the pattern of listed companies network. Progress in Geography, 2014, 33(12): 1587-1600.]
- [36] 吴康, 方创琳, 赵渺希. 中国城市网络的空间组织及其复杂性结构特征 [J]. 地理研究, 2015, 34(4): 711-728. [Wu Kang, Fang Chuanglin, Zhao Miaoxi. The spatial organization and structure complexity of Chinese intercity networks. Geographical Research, 2015, 34(4): 711-728.]
- [37] 赵渺希, 钟焯, 徐高峰. 中国三大城市群多中心网络的时空演化 [J]. 经济地理, 2015, 35(3): 52-59. [Zhao Miaoxi, Zhong Ye, Xu Gaofeng. Polycentric progress of the three major city regions in China. Economic Geography, 2015, 35(3): 52-59.]
- [38] 王垚, 钮心毅, 宋小冬. “流空间”视角下区域空间结构研究进展 [J]. 国际城市规划, 2017, 32(6): 27-33. [Wang Yao, Niu Xinyi, Song Xiaodong. Research progress of re-

- gional spatial structure under the perspective of space of flow. *Urban Planning International*, 2017, 32(6): 27-33.]
- [39] Liu Y, Kang C G, Gao S, et al. Understanding intra-urban trip patterns from taxi trajectory data [J]. *Journal of Geographical Systems*, 2012, 14(4): 463-483.
- [40] Liu Y, Sui Z W, Kang C G, et al. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data [J]. *PLoS One*, 2014, 9(1): e86026. doi: 10.1371/journal.pone.0086026.
- [41] Long, Y, Thill J C. Combining smart card data and household travel survey to analyze jobs-housing relationships in Beijing [J]. *Computers, Environment and Urban Systems* 2015, 53: 19-35.
- [42] Barbosa H, Barthelemy M, Ghoshal G, et al. Human mobility: Models and applications [J]. *Physics Reports*, 2018, 734: 1-74.
- [43] 李山, 邱荣旭, 陈玲. 基于百度指数的旅游景区络空间关注度: 时间分布及其前兆效应 [J]. *地理与地理信息科学*, 2008, 24(6): 102-107. [Li Shan, Qiu Rongxu, Chen Ling. Cyberspace attention of tourist attractions based on Baidu Index: Temporal distribution and precursor effect. *Geography and Geo-information Science*, 2008, 24(6): 102-107.]
- [44] 甄峰, 王波, 陈映雪. 基于网络社会空间的中国城市网络特征: 以新浪微博为例 [J]. *地理学报*, 2012, 67(8): 1031-1043. [Zhen Feng, Wang Bo, Chen Yingxue. China's city network characteristics based on social network space: An empirical analysis of Sina micro-blog. *Acta Geographica Sinica*, 2012, 67(8): 1031-1043.]
- [45] 熊丽芳, 甄峰, 王波, 等. 基于百度指数的长三角核心区城市网络特征研究 [J]. *经济地理*, 2013, 33(7): 67-73. [Xiong Lifang, Zhen Feng, Wang Bo, et al. The research of the Yangtze River Delta core area's city network characteristics based on Baidu Index. *Economic Geography*, 2013, 33(7): 67-73.]
- [46] 王波, 甄峰, 席广亮, 等. 基于微博用户关系的网络信息地理研究: 以新浪微博为例 [J]. *地理研究*, 2013, 32(2): 380-391. [Wang Bo, Zhen Feng, Xi Guangliang, et al. A study of cybergeography based on micro-blog users' relationship: With a case of Sina micro-blog. *Geographical Research*, 2013, 32(2): 380-391.]
- [47] 董超, 修春亮, 魏冶. 基于通信流的吉林省流空间网络格局 [J]. *地理学报*, 2014, 69(4): 510-519. [Dong Chao, Xiu Chunliang, Wei Ye. Network structure of 'space of flows' in Jilin Province based on telecommunication flows. *Acta Geographica Sinica*, 2014, 69(4): 510-519.]
- [48] González M C, Hidalgo C A, Barabási A L. Understanding individual human mobility patterns [J]. *Nature*, 2008, 453: 779-782.
- [49] Song C M, Qu Z H, Blumm N, et al. Limits of predictability in human mobility [J]. *Science*, 2010, 327: 1018-1021.
- [50] Lu X, Bengtsson L, Holme P. Predictability of population displacement after the 2010 Haiti earthquake [J]. *PNAS*, 2012, 109(29): 11576-11581.
- [51] Long Y, Thill J C. Combining smart card data and household travel survey to analyze jobs-housing relationships in Beijing [J]. *Environment and Urban Systems*, 2015, 53: 19-35.
- [52] 陈映雪, 甄峰, 王波, 等. 基于微博平台的中国城市网络信息不对称关系研究 [J]. *地球科学进展*, 2012, 27(12): 1353-1362. [Chen Yingxue, Zhen Feng, Wang Bo, et al. A study of internet information asymmetry relations among Chinese cities based on the micro-blog platform. *Advances in Earth Science*, 2012, 27(12): 1353-1362.]
- [53] 熊丽芳, 甄峰, 席广亮, 等. 我国三大经济区城市网络变化特征: 基于百度信息流的实证研究 [J]. *热带地理*, 2014, 34(1): 34-43. [Xiong Lifang, Zhen Feng, Xi Guangliang, et al. Characteristics of the city network in the three major economic zones of China: A study based on Baidu information flow. *Tropical Geography*, 2014, 34(1): 34-43.]
- [54] Xia F, Wang J Z, Kong X J, et al. Exploring human mobility patterns in urban scenarios: A trajectory data perspective [J]. *IEEE Communications Magazine*, 2018, 56(3): 142-149.
- [55] Devriendt L, Boulton A, Brunn S, et al. Searching for cyberspace: The position of major cities in the information age [J]. *Journal of Urban Technology*, 2011, 18(1): 73-92.
- [56] 赵梓渝, 魏冶, 庞瑞秋, 等. 基于人口省际流动的中国城市网络转变中心性与控制力研究: 兼论递归理论用于城市网络研究的条件性 [J]. *地理学报*, 2017, 72(6): 1032-1048. [Zhao Ziyu, Wei Ye, Pang Ruiqiu, et al. Alter-based centrality and power of Chinese city network using inter-provincial population flow. *Acta Geographica Sinica*, 2017, 72(6): 1032-1048.]
- [57] 刘望保, 石恩名. 基于ICT的中国城市间人口日常流动空间格局: 以百度迁徙为例 [J]. *地理学报*, 2016, 71(10): 1667-1679. [Liu Wangbao, Shi Enming. Spatial pattern of population daily flow among cities based on ICT: A case study of "Baidu Migration". *Acta Geographica Sinica*, 2016, 71(10): 1667-1679.]
- [58] 龙瀛, 张宇, 崔承印. 利用公交刷卡数据分析北京职住关系和通勤出行 [J]. *地理学报*, 2012, 67(10): 1339-1352. [Long Ying, Zhang Yu, Cui Chengyin. Identifying commuting pattern of Beijing using bus smart card data. *Acta Geographica Sinica*, 2012, 67(10): 1339-1352.]
- [59] 钟炜菁, 王德, 谢栋灿, 等. 上海市人口分布与空间活动

- 的动态特征研究: 基于手机信令数据的探索 [J]. 地理研究, 2017, 36(5): 972-984. [Zhong Weijing, Wang De, Xie Dongcan, et al. Dynamic characteristics of Shanghai's population distribution using cell phone signaling data. *Geographical Research*, 2017, 36(5): 972-984.]
- [60] 黄洁, 王姣娥, 靳海涛, 等. 北京市地铁客流的时空分布格局及特征: 基于智能交通卡数据 [J]. 地理科学进展, 2018, 37(3): 397-406. [Huang Jie, Wang Jiaoe, Jin Haitao, et al. Investigating spatiotemporal patterns of passenger flows in the Beijing metro system from smart card data. *Progress in Geography*, 2018, 37(3): 397-406.]
- [61] 秦静, 李郎平, 唐鸣镛, 等. 基于地理标记照片的北京市入境旅游流空间特征 [J]. 地理学报, 2018, 73(8): 1556-1570. [Qin Jing, Li Langping, Tang Mingdi, et al. Exploring the spatial characteristics of Beijing inbound tourist flow based on geotagged photos. *Acta Geographica Sinica*, 2018, 73(8): 1556-1570.]
- [62] 靳诚, 徐菁, 黄震方, 等. 南京城市内部景点间游客流动特征分析 [J]. 地理学报, 2014, 69(12): 1858-1870. [Jin Cheng, Xu Jing, Huang Zhenfang, et al. Analyzing the characteristics of tourist flows between the scenic spots in inner city based on tourism strategies: A case study in Nanjing. *Acta Geographica Sinica*, 2014, 69(12): 1858-1870.]
- [63] 王垚, 钮心毅, 宋小冬, 等. 人流联系和经济联系视角下区域城市关联比较: 基于手机信令数据和企业关联数据的研究 [J]. 人文地理, 2018, 33(2): 84-91, 146. [Wang Yao, Niu Xinyi, Song Xiaodong, et al. The comparison of regional urban relations between people flow and capital flow: A study based on mobile phone signaling data and firm interlock data. *Human Geography*, 2018, 33(2): 84-91, 146.]
- [64] Zhen F, Cao Y, Qin X, et al. Delineation of an urban agglomeration boundary based on Sina Weibo microblog 'check-in' data: A case study of the Yangtze River Delta [J]. *Cities*, 2017, 60: 180-191.
- [65] Trasarti R, Olteanu-Raimond A M, Nanni M, et al. Discovering urban and country dynamics from mobile phone data with spatial correlation patterns [J]. *Telecommunications Policy*, 2015, 39(3-4): 347-362.
- [66] Çolak S, Lima A, González M C. Understanding congested travel in urban areas [J]. *Nature Communications*, 2016, 7: 10793. doi: 10.1038/ncomms10793.
- [67] Zhang W Y, Derudder B, Wang J H, et al. Using location-based social media to chart the patterns of people moving between cities: The case of Weibo-users in the Yangtze River Delta [J]. *Journal of Urban Technology*, 2016, 23(3): 91-111.
- [68] Li L, Yang L, Zhu H H, et al. Explorative analysis of Wuhan intra-urban human mobility using social media check-in data [J]. *PLoS One*, 2015, 10(8): e0135286. doi: 10.1371/journal.pone.0135286.
- [69] Zhou J P, Murphy E, Long Y. Commuting efficiency in the Beijing metropolitan area: An exploration combining smartcard and travel survey data [J]. *Journal of Transport Geography*, 2014, 41: 175-183.
- [70] Wang Z Y, Hu Y X, Zhu P, et al. Ring aggregation pattern of metro passenger trips: A study using smart card data [J]. *Physica A: Statistical Mechanics and Its Applications*, 2018, 491: 471-479.
- [71] Zhong C, Batty M, Manley E, et al. Variability in regularity: Mining temporal mobility patterns in London, Singapore and Beijing using smart-card data [J]. *PLoS One*, 2016, 11(2): e0149222. doi: 10.1371/journal.pone.0149222.
- [72] 杨喜平, 方志祥. 移动定位大数据视角下的人群移动模式及城市空间结构研究进展 [J]. 地理科学进展, 2018, 37(7): 880-889. [Yang Xiping, Fang Zhixiang. Recent progress in studying human mobility and urban spatial structure based on mobile location big data. *Progress in Geography*, 2018, 37(7): 880-889.]
- [73] Zhang S, Tang J J, Wang H X, et al. Revealing intra-urban travel patterns and service ranges from taxi trajectories [J]. *Journal of Transport Geography*, 2017, 61: 72-86.
- [74] Tang J J, Liu F, Wang Y H, et al. Uncovering urban human mobility from large scale taxi GPS data [J]. *Physica A: Statistical Mechanics and Its Applications*, 2015, 438: 140-153.
- [75] 赖建波, 潘竟虎. 基于腾讯迁徙数据的中国“春运”城市间人口流动空间格局 [J]. 人文地理, 2019, 34(3): 108-117. [Lai Jianbo, Pan Jinghu. Spatial pattern of population flow among cities in China during the Spring festival travel rush based on "Tencent Migration" data. *Human Geography*, 2019, 34(3): 108-117.]
- [76] 魏冶, 修春亮, 王绮, 等. 中国春运人口流动网络的富人俱乐部现象与不平衡性分析 [J]. 人文地理, 2018, 33(2): 124-129. [Wei Ye, Xiu Chunliang, Wang Qi, et al. Rich-club phenomenon and disequilibrium of China's population flow network during spring festival travel period. *Human Geography*, 2018, 33(2): 124-129.]
- [77] Xi G L, Zhen F, He J L, et al. City networks of online commodity services in China: Empirical analysis of Tmall clothing and electronic retailers [J]. *Chinese Geographical Science*, 2018, 28(2): 231-246.
- [78] 李鲁奇, 孔翔. “双十一”期间中国快递流通的时空结构与效率: 基于时间地理学视角 [J]. 地理研究, 2019, 38(8): 1891-1904. [Li Luqi, Kong Xiang. The tempo-spatial structure and efficiency of China's express service during the "Double Eleven Shopping Carnival": A time-

- geographic approach. *Geographical Research*, 2019, 38(8): 1891-1904.]
- [79] Lin J Y, Li X. Simulating urban growth in a metropolitan area based on weighted urban flows by using web search engine [J]. *International Journal of Geographical Information Systems*, 2015, 29(10): 1721-1736.
- [80] 秦昆, 罗萍, 姚博睿. GDELT数据网络化挖掘与国际关系分析 [J]. *地球信息科学学报*, 2019, 21(1): 14-24. [Qin Kun, Luo Ping, Yao Borui. Networked mining of GDELT and international relations analysis. *Journal of Geo-information Science*, 2019, 21(1): 14-24.]
- [81] 潘碧麟, 王江浩, 葛咏, 等. 基于微博签到数据的成渝城市群空间结构及其城际人口流动研究 [J]. *地球信息科学学报*, 2019, 21(1): 68-76. [Pan Bilin, Wang Jianghao, Ge Yong, et al. Spatial structure and population flow analysis in Chengdu-Chongqing urban agglomeration based on Weibo Check-in big data. *Journal of Geo-information Science*, 2019, 21(1): 68-76.]
- [82] Fan Z D, Pei T, Ma T, et al. Estimation of urban crowd flux based on mobile phone location data: A case study of Beijing, China [J]. *Computers, Environment and Urban Systems*, 2018, 69: 114-123.
- [83] Yao X, Zhu D, Gao Y, et al. A stepwise spatio-temporal flow clustering method for discovering mobility trends [J]. *IEEE Access*, 2018, 6: 44666-44675.
- [84] von Landesberger T, Brodtkorb F, Roskosch P, et al. Mobility graphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering [J]. *IEEE Transactions on Visualization and Computer Graphics*, 2016, 22(1): 11-20.
- [85] 彭卉, 杜云艳, 易嘉伟, 等. 基于手机数据的北京市城市与近郊交互模式挖掘 [J]. *地球信息科学学报*, 2019, 21(1): 101-110. [Peng Hui, Du Yunyan, Yi Jiawei, et al. Mining urban-rural spatial interaction pattern from mobile data of Beijing. *Journal of Geo-information Science*, 2019, 21(1): 101-110.]
- [86] 宋冬林, 齐文浩. 东北区域经济一体化演变的社会网络分析 [J]. *吉林大学社会科学学报*, 2018, 58(4): 97-107, 206. [Song Donglin, Qi Wenhao. The social network in the evolution of regional economic integration of Northeast China. *Jilin University Journal Social Sciences Edition*, 2018, 58(4): 97-107, 206.]
- [87] 赵映慧, 高鑫, 姜博. 东北三省城市百度指数的网络联系层级结构 [J]. *经济地理*, 2015, 35(5): 32-37. [Zhao Yinghui, Gao Xin, Jiang Bo. The urban network connection of three provinces in Northeast China based on Baidu Index. *Economic Geography*, 2015, 35(5): 32-37.]
- [88] Zhang X, Kloosterman R C. Connecting the 'workshop of the world': Intra- and extra- service networks of the Pearl River Delta city-region [J]. *Regional Studies*, 2016, 50(6): 1069-1081.
- [89] 王姣娥, 景悦. 中国城市网络等级结构特征及组织模式: 基于铁路和航空流的比较 [J]. *地理学报*, 2017, 72(8): 1508-1519. [Wang Jiao'e, Jing Yue. Comparison of spatial structure and organization mode of inter-city networks from the perspective of railway and air passenger flow. *Acta Geographica Sinica*, 2017, 72(8): 1508-1519.]
- [90] Liu X, Gong L, Gong Y X, et al. Revealing travel patterns and city structure with taxi trip data [J]. *Journal of Transport Geography*, 2015, 43: 78-90.
- [91] 靳诚, 徐菁. 南京市对外交通节点与酒店之间游客流动空间特征分析 [J]. *人文地理*, 2016, 31(5): 55-62. [Jin Cheng, Xu Jing. Study on the tourists flow among external transport nodes and hotels in Nanjing. *Human Geography*, 2016, 31(5): 55-62.]
- [92] 刘瑜, 肖昱, 高松, 等. 基于位置感知设备的人类移动研究综述 [J]. *地理与地理信息科学*, 2011, 27(4): 8-13. [Liu Yu, Xiao Yu, Gao Song, et al. A review of human mobility research based on location aware devices. *Geography and Geo-information Science*, 2011, 27(4): 8-13.]
- [93] 李婷, 裴韬, 袁焯城, 等. 人类活动轨迹的分类、模式和应用研究综述 [J]. *地理科学进展*, 2014, 33(7): 938-948. [Li Ting, Pei Tao, Yuan Yecheng, et al. A review on the classification, patterns and applied research of human mobility trajectory. *Progress in Geography*, 2014, 33(7): 938-948.]
- [94] Huang W, Li S N. Understanding human activity patterns based on space-time- semantics [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2016, 121: 1-10.
- [95] Ling X M, Huang Z R, Wang C C, et al. Predicting subway passenger flows under different traffic conditions [J]. *PLoS One*, 2018, 13(8): e0202707. doi: 10.1371/journal.pone.0202707.
- [96] Wu L, Zhi Y, Sui Z W, et al. Intra-urban human mobility and activity transition: Evidence from social media check-in data [J]. *PLoS One*, 2014, 9(5): e97010. doi: 10.1371/journal.pone.0097010.
- [97] Duan Z T, Yang Y, Zhang K, et al. Improved deep hybrid networks for urban traffic flow prediction using trajectory data [J]. *IEEE Access*, 2018, 6: 31820-31827.
- [98] Chen Y R, Huang Z, Pei T, et al. HiSpatialCluster: A novel high-performance software tool for clustering massive spatial points [J]. *Transactions in GIS*, 2018, 22(5): 1275-1298.
- [99] 蒋大亮, 孙焯, 任航, 等. 基于百度指数的长江中游城市群城市网络特征研究 [J]. *长江流域资源与环境*, 2015, 24(10): 1654-1664. [Jiang Daliang, Sun Ye, Ren Hang, et al. Analyses on the city network characteristics of middle

- Yangtze urban agglomeration based on Baidu Index. Resources and Environment in the Yangtze Basin, 2015, 24(10): 1654-1664.]
- [100] 孙阳, 姚士谋, 张落成. 中国沿海三大城市群城市空间网络拓展分析: 以综合交通信息网络为例 [J]. 地理科学, 2018, 38(6): 827-837. [Sun Yang, Yao Shimou, Zhang Luocheng. Spatial expansion of urban network for the three coastal agglomerations of China: A study based on integrated traffic information network. Scientia Geographica Sinica, 2018, 38(6): 827-837.]
- [101] 吴志峰, 柴彦威, 党安荣, 等. 地理学碰上“大数据”: 热反应与冷思考 [J]. 地理研究, 2015, 34(12): 2207-2221. [Wu Zhifeng, Chai Yanwei, Dang Anrong, et al. Geography interact with big data: Dialogue and reflection. Geographical Research, 2015, 34(12): 2207-2221.]
- [102] 康朝贵, 刘瑜, 邬伦. 城市手机用户移动轨迹时空熵特征分析 [J]. 武汉大学学报(信息科学版), 2017, 42(1): 63-69, 129. [Kang Chaogui, Liu Yu, Wu Lun. An analysis of entropy of human mobility from mobile phone data. Geomatics and Information Science of Wuhan University, 2017, 42(1): 63-69, 129.]
- [103] Gao S, Li L N, Li W W, et al. Constructing gazetteers from volunteered big geo-data based on Hadoop [J]. Computers, Environment and Urban Systems, 2017, 61: 172-186.
- [104] Yan Y, Zhang S, Tang J J, et al. Understanding characteristics in multivariate traffic flow time series from complex network structure [J]. Physica A: Statistical Mechanics and Its Applications, 2017, 477: 149-160.
- [105] Song H Y, You D. Modeling urban mobility with machine learning analysis of public taxi transportation data [J]. International Journal of Pervasive Computing and Communications, 2018, 14(1): 73-87.
- [106] Liu Z D, Li Z J, Wu K S, et al. Urban traffic prediction from mobility data using deep learning [J]. IEEE Network, 2018, 32(4): 40-46.

Research progress on the space of flow using big data

YANG Yanjie^{1,2}, YIN Dan^{1,2}, LIU Ziwen^{1,2}, HUANG Qingxu^{1,2*}, HE Chunyang^{1,2}, WU Kang³

(1. Center for Human-Environment System Sustainability (CHESS), State Key Laboratory of Earth Surface Processes and Resource Ecology (ESPRE), Beijing Normal University, Beijing 100875, China;

2. School of Natural Resources, Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China;

3. Beijing Key Laboratory of Megaregions Sustainable Development Simulation, Capital University of Economics and Business, Beijing 100070, China)

Abstract: The studies on the space of flow play an important role in understanding the structure and change of urban networks. The rapid development of big data provides new opportunities and challenges for the studies on the space of flow in recent years. This article systematically reviewed the research progress on the space of flow based on big data. First, we retrospect the background and history of the studies on the space of flow, then summarized the themes, the types of big data, the methods used for the studies and the major findings, as well as discussed the research challenges. We found an exponential growth of studies on the space of flow using big data after 2011. The annual number of published papers increased from 11 to 106 during 2011–2018. Big data deepen the research on the space of flow by providing new data sources, inspiring new analytical methods, and new research perspectives. Four types of big data—mobile phone, social media, smart card, and taxi trajectory data are commonly used in the studies on the space of flow, which can provide information on spatiotemporal flows (such as population flow, material flow, and information flow) directly. Research methods have also evolved from distance-based gravity models to network analysis. In the future, the research on the space of flow using big data can be further improved by validating the effectiveness and representativeness of the big data, the integration of big data and traditional data, and the information mining from big data using new methods such as deep learning and cloud computing.

Keywords: space of flow; big data; urban network; human mobility; urban sustainability